# Estimation of 3D Shape from Shading and Binocular Disparity

Dicle Nahide Dövencioğlu

**A thesis submitted in fulfilment of the requirements of the degree of Doctor of Philosophy.**

**January 3, 2013**


School of Psychology

College of Life and Environmental Sciences

UNIVERSITY OF BIRMINGHAM

# ABSTRACT

How does the visual system make use of various sources of information from the three-dimensional (3D) geometry world? To infer distances in a 3D scene, the brain uses multiple cues such as binocular disparity, which provides metric estimates of depth; or shading, which is inherently ambiguous and requires additional interpretation. In this thesis, I use psychophysical and functional magnetic resonance imaging (fMRI) techniques to address the following questions: (i) how does the visual system resolve ambiguities in a luminance signal, in particular, separating shading cues to shape from luminance variations caused by the changes in the surface material, (ii) when both shading and binocular disparity are available, how do these cues interact to produce a coherent 3D shape estimate, (iii) what is the neural substrate to this cue integration between shading and disparity?

First, in Chapter 3, I examine how first- and second-order luminance signals in a luminance pattern are perceived, and ask if observers can benefit from the phase relationship of these signals as a cue to shape. I show that observers learn to exploit the phase relationship after training, and that the changes through training can be explained as associative learning. Next, in Chapter 4, I ask whether decomposing shading and reflectance cues to infer shape can be done in very short presentation times. Through training, observers learn to use the phase relationship of first- and second-order luminance signals, where the change in their performance seems to be at a perceptual level. In Chapter 5, I challenge the dorsal visual cortical area V3B/KO which was previously indicated as a crucial locus when integrating disparity and motion parallax cues to infer 3D shape. I present evidence that the involvement of V3B/KO in 3D shape processing can be extended to disparity and shading

signals. Moreover, I find a distinct relation between neural activity in this cortical area and perceptual judgements of individual observers. Finally, in Chapter 6, I carry on investigating cue integration to gain further insight into the individual variations. I report systematic differences between observers, where about half of the population benefit from having shading and disparity together; the other half fail to establish cue integration. Nevertheless, after training, I show that these participants learn to exploit shading and disparity information.

**TABLE OF CONTENTS**

## LIST OF FIGURES

## LIST OF TABLES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| 2D | Two-dimensional (space) |
| 3D | Three-dimensional (space) |
| AM | Amplitude modulation |
| arcmin | Minute of arc |
| BOLD | Blood oxygenation level dependent |
| c/deg | Cycle per degree |
| cm | Centimetre |
| CRT | Cathode ray tube |
| fMRI | Functional magnetic resonance imaging |
| Hz | Hertz |
| IPD | Inter-pupillary distance |
| LM | Luminance amplitude |
| m | Metre |
| mm | Millimetre |
| ms | Millisecond |
| MVPA | Multi-voxel pattern analysis |
| pc | Per cent correct |
| r.m.s. | Root mean square |
| RDS | Random dot stereogram |
| s | Second |
| SVM | Support vector machine |
| $cd/m^2$ | Candela per metre squared |
| ROI | Region of interest |

# CHAPTER 1:

# General Introduction

## 1.1 Introduction

Our internal representation of the 3D world is so remarkably accurate that we can interact with the environment and direct our actions without hesitation (e.g. playing tennis, driving). Understanding the 3D geometry of the world predominantly relies on the perception of the distances of objects with respect to the observer. The visual system's ability to estimate depth might seem effortless, but it surely requires many computations exploiting multiple sources of information to depth. The computations will be especially complex if the sources vary in the nature of information provided. Above all, the light entering each eye is converted to electrical signals by a film of photoreceptors before being transmitted to the visual cortex via the optic nerve. Therefore the visual system must acquire the 3D percept from 2D input signals.

The brain uses multiple depth cues to infer distances in a 3D scene. Of these cues, some, such as binocular disparity or motion parallax, provide metric (absolute) estimates of distance, and can easily be mathematically defined. In the absence of these cues, the visual system is still capable of constructing the 3D geometry by recruiting pictorial cues: perspective, shading, cast shadows, and occlusion. Even though it is impossible to tell the

absolute distance of a point from pictorial cues alone, one can still judge depth order, and a rough geometry of the scene. Moreover, despite the computational discrepancy between metric and pictorial cues, depth judgements seem to benefit from having both of these present. In the next section I will give background information regarding the nature of the two particular cues to depth, disparity and shading, which form the basis of this thesis.

### 1.1.1 Shape from Shading

While artists have exploited shading for centuries, and earlier researchers were intrigued by the phenomenon (Brewster, 1826; Rittenhouse, 1786), the term *shape-from-shading* was coined fairly recently (Horn, 1975).

Shading information is inherently ambiguous, in the sense that a shading pattern might be interpreted in different ways by different people and when using different measurement methods (Koenderink, van Doorn, Kappers, & Todd, 2001). Dutch graphic artist M. C. Escher demonstrates this beautifully in his lithograph (Figure 1.1, Convex and Concave, 1955). The scene is coherent when viewed locally; it demonstrates strong and meaningful local shading cues to depth despite the fact that the global architecture is impossible.

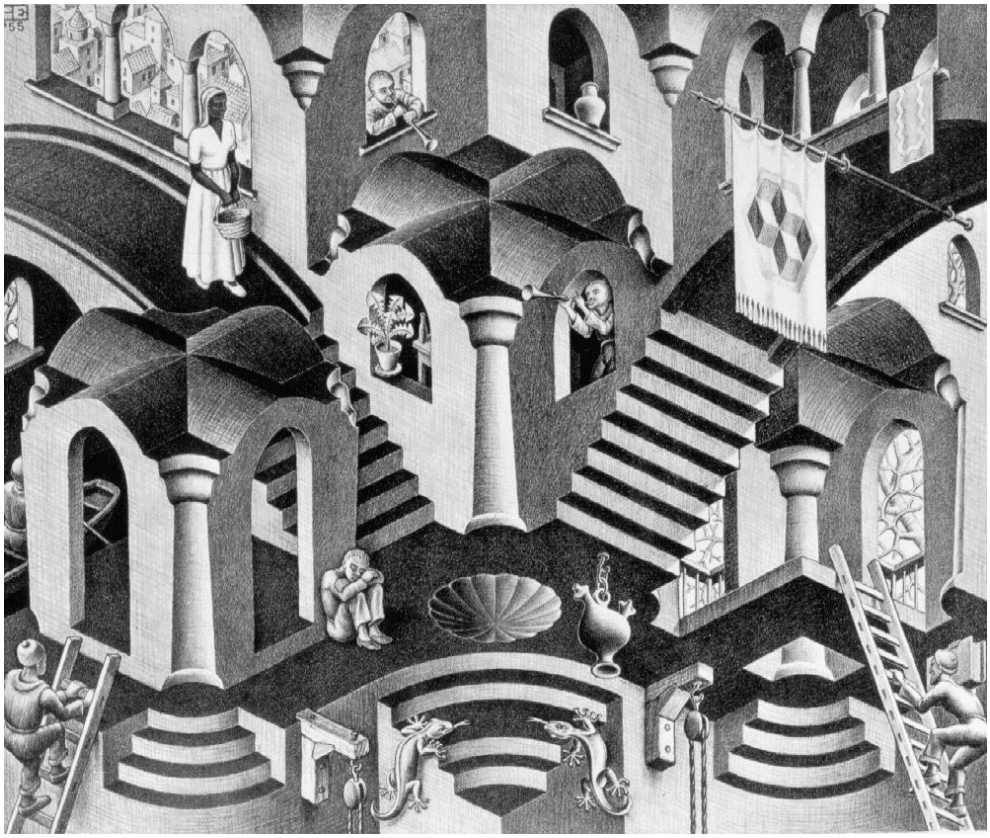


Figure 1.1: Ambiguity of shading information in M. C. Escher's lithograph *Convex and Concave* (1955). The artist exploits variable local luminance information and distorts the geometry locally to depict an impossible architecture introduced by the ambiguities in interpreting surface shape throughout the image. Initially the pillars seem to be going into the intersection of the arches, which can easily be reversed if the print is viewed upside down. It is almost impossible to maintain a coherent percept for the disc shape in the centre of this print, whether it is a dome on the ceiling, or a void on the floor.

Shape-from-shading is a mathematically ill-posed problem and is only tractable via the application of number of constraints. Typical constraints in machine vision include: uniform surface material and reflectance, Lambertian (matte) surface properties (Pentland, 1984), light source at infinity, and orthographic projection (Horn, 1975). Even when so constrained, shape-from-shading still presents as a single equation in two unknowns: surface orientation and light source direction. While it is often possible to infer the direction of the light source from the image (Koenderink, Pont, van Doorn, Kappers, & Todd, 2007; Koenderink, Van Doorn, & Pont, 2007), when this is not possible, observers will adopt a default or prior assumption for lighting direction. Research on light source priors suggests that human observers prefer to assume that light is coming from above-left, when there is no

other cue to lighting direction (Adams, 2007; Adams, Graf, & Ernst, 2004; Brewster, 1826; Mamassian & Goutcher, 2001; Sun & Perona, 1998). However, ambiguous shape-from-shading can also be resolved by assuming a default surface geometry. Convex surfaces are preferred more than concave (Liu & Todd, 2004).

Human observers use light priors to recover shape from shading in an automatic and pre-attentive process (Adams, 2007). Adams (2008) also reported that depending on the demands of the task, the light-from-above prior can aid a 'quick and dirty' recovery of the shape in a visual search task, but this process becomes more elaborate when the retinal frame of reference needs to be recalculated in a task, which requires fine estimation of the shape.

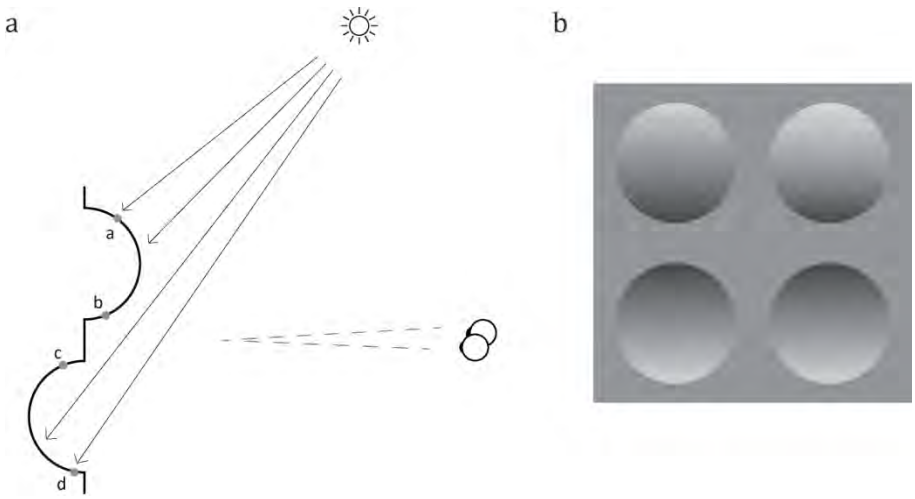



**Figure 1.2: (a)** Cross section of a surface viewed from the side: convex and a concave. Surface brightness is dependent on the surface curvature. Portions of the surface facing towards the light source (points a and d) look brighter than those facing away (points b and c). **(b)** Same phenomenon in (a) illustrated with simple linear luminance gradients. Top row discs have bright portions on top, suggesting that they are facing toward an above light, hence convex; while this is supported by bottom row, concave surfaces. The convexity/concavity matches can easily be reversed if the image is viewed upside down.

Yet another challenge for the visual system is the segregation of luminance variations caused by surface shape, from those caused by the changes in the reflectance properties of

the surface. To give an example, an appropriately painted flat surface and a curved but uniform reflectance surface under directional lighting can provide the same shading pattern, and this can be indistinguishable to human vision (Kingdom, 2008; Schofield, Hesse, Rock, & Georgeson, 2006; Schofield, Rock, Sun, Jiang, & Georgeson, 2010; Todd & Mingolla, 1983). Indeed, the lithograph in Figure 1.1 uses changes in reflectance to simulate curved surfaces. Such ambiguities can be resolved by the segregation of shading and reflectance cues. Kingdom (2008) describes this segregation as a layer decomposition process (also see 'intrinsic images', Barrow & Tanenbaum, 1978). Kingdom (2003) shows psychophysical evidence that addition of colour information to a luminance pattern aids layer decomposition. Similarly, visual texture can be a useful cue for layer decomposition. Adding a texture to the shading pattern has been shown to enhance the interpretation of shape from shading (Schofield *et al.*, 2006; Schofield *et al.*, 2010; Todd & Mingolla, 1983). Schofield and colleagues showed that when albedo textured (painted texture) and corrugated surfaces receive directional lighting, the resulting shading pattern conveys changes in local mean luminance (LM, a first-order signal) together with variations in the luminance difference between light and dark elements of the texture (AM, amplitude modulation, a second-order signal). The authors demonstrated that luminance variations caused by surface shape provide positively correlated first- and second-order signals. On the other hand, luminance variations caused by reflectance properties of the surface result in no such relationship, making the relationship between first- and second-order signals a cue for layer decomposition.

In some cases, LM-AM also produces a depth percept, especially when presented alone. However, the reported profiles of LM-AM are weaker than those reported for LM-only and LM+AM; and this is even more attenuated when LM-AM is presented together with LM+AM in a plaid.

In computer vision, occluding edges can be identified by in-phase changes in luminance and amplitude and may be treated as a material change. However, Schofield et al. (2010) refer to smooth changes in stimulus properties and to reflectance changes 'painted' onto a smooth surface rather than changes between objects. It may be then that the meaning of luminance and amplitude relationships themselves depends on image context and that the sharpness of each change is an important cue. Occluding edges also tend to produce changes in texture (e.g. dominant orientation) and these may be another cue to material changes even if luminance and amplitude are correlated.

Shape from shading algorithms often assume Lambertian surfaces with constant reflectance. This allows the interpretation of luminance changes in the image as shading and thus the estimation of surface orientation during shape recovery. However, in natural images, uniform reflectance is rare and changes in the luminance can also be caused by the changes in reflectance. Humans recover shape from shading in the natural environment without a problem. As might be expected, human observers overcome this problem by using other cues such as context or spatial arrangements (Gilchrist, 1977, 1988) to separate illumination and reflectance changes in addition to perceived brightness.

In an analogy with the intrinsic image representation in computer vision, the human visual system's ability to perceive lightness, brightness and transparency is explained as the image being represented in different layers separating the source of the change in luminance (Kingdom, 2008). When the human visual system disambiguates reflectance from shading (which is a direct input to recover shape), luminance changes that are aligned with changes in hue appear to be related to reflectance, and non-aligned hue and luminance are likely to be perceived as shading (Kingdom, 2003). The assumption of uniform albedo surfaces limits shape from shading algorithms applications to only a subset of natural images. Barron and Malik (2011) address this problem with a statistical formulation where they recover the

"most likely albedo and shape that explain a single image (p. 2521)". They incorporate low frequency priors on shape to show that recovering shape (and albedo) from shading is possible. In their following work (2012), the authors introduce further priors, such as flatness of the shape, surface orientation at the occluding contour, and second-order smoothness and this time their algorithm estimates illumination as well as shape and albedo. These examples show that shape from shading algorithms can succeed when the problem is addressed as an optimization problem."

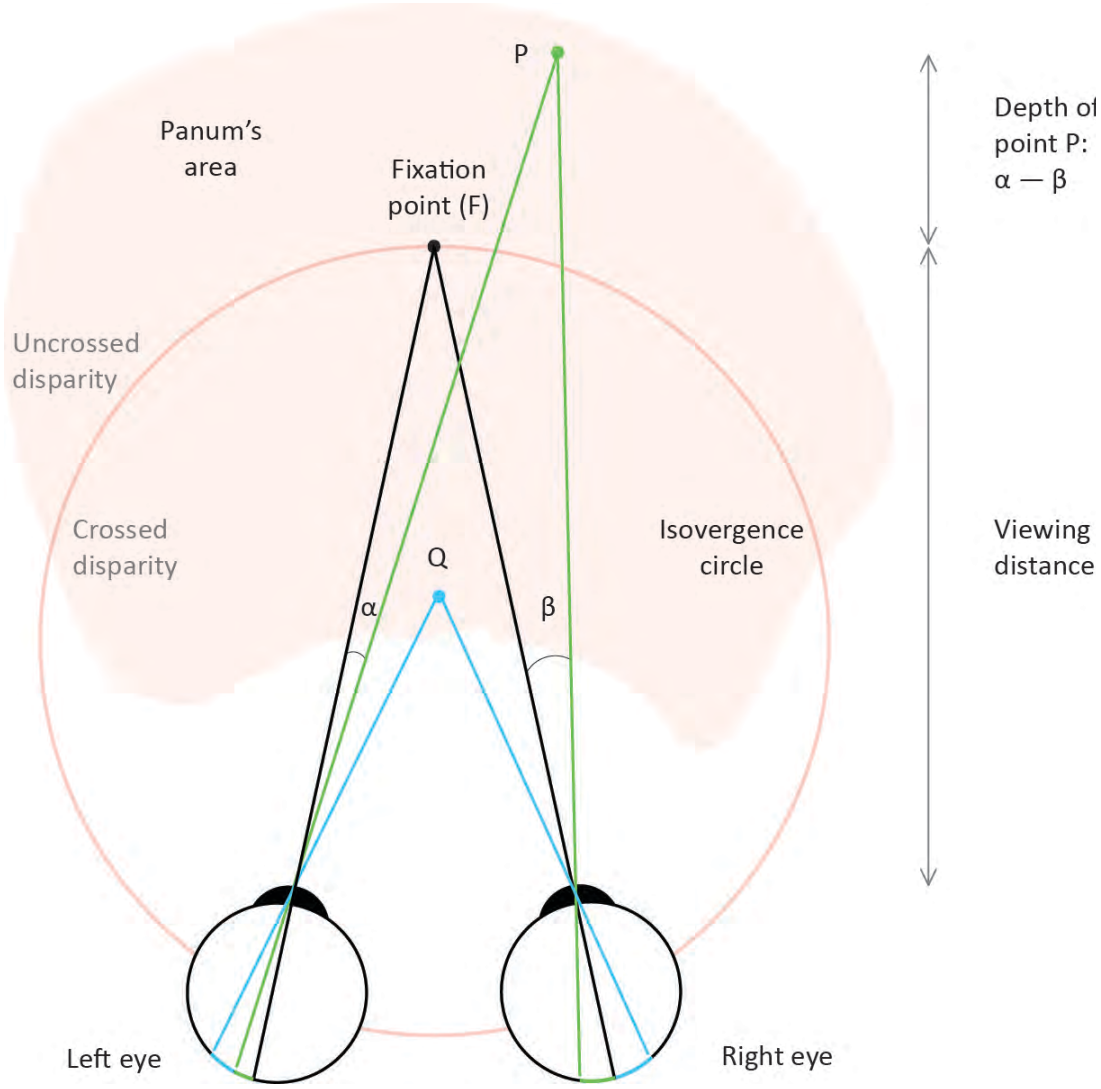
### 1.1.2 Binocular disparity

Not all information provided by our 3D environment is ambiguous. On the contrary, binocular disparity can provide a precise and metric cue to 3D depth. It is believed to be the most intricate ability of the human visual system, and is also considered a valuable one, since it was evolved at the expense of the wide field peripheral vision that might otherwise be provided by having two eyes.

The visual fields of the two eyes provide slightly different images, due to the separation of the eyes. When the overlapping portions of these visual fields are processed together, an ability to perceive distances results. Unlike shading, the extraction of binocular disparity can be understood with a relatively straightforward, tractable mathematical description.

The horizontal separation between two eyes is on average 6.5 cm (inter-pupillary difference, IPD), so the eyes have slightly different viewpoints. The brain uses the disparity between the two images registered on each retina to construct the 3D geometry of the world (Howard & Rogers, 2002; Julesz, 1971). The absolute disparity (or zero-order disparity) of a point around fixation is the angular separation of its projection on the retinae (for point P, this is illustrated by green arcs on each retina, Figure 1.3). For points in front of the fixation

plane (Q in Figure1.3), the left retinal image is to the left of the central fovea; where the right retinal image is more to the right, hence the eyes converge to fuse two images (crossed disparity).




igure 1.3: Illustration of geometry of binocular vision: two eyes fixating at a point F, where ach image falls in the centre (fovea) of the eyes. The absolute disparity of the fixation point ; zero (because projection of F into each retina is the centre of fovea), and so is any point alling on the isovergence circle. The absolute disparity of point P is the angular difference of :s projection to each eye's retina from each eye's fovea (α for left eye, β for right eye, ifferences also indicated by green arcs). The relative disparity between point P and point Q ; the difference between their absolute disparities, which can also be calculated by .

The relative disparity of two points, say P and Q, can simply be calculated by the difference between their absolute disparities, but to be able to infer the distance between these points, one has to take into account the viewing distance as follows. When the eyes are fixating at a point in space at a viewing distance of D, the relative distance (d) of another point (P) can be approximated by the following equation:

$$\text{relative disparity} \times \text{viewing distance}^2 = I \times dD^2$$

given that the inter-pupillary distance of the viewer (I) is known (Howard & Rogers, 2002). This approximation specifies an inverse relation between viewing distance and relative disparity, hence when two points are located at same distances from the fixation point, the one behind the fixation (at a larger viewing distance) has a smaller disparity when compared to the point located in front of the fixation.

The images in each retina also differ vertically, and when the visual stimulus has large eccentricity and the viewing distance is small, observers can benefit from vertical disparity (Rogers & Bradshaw, 1993).

When one eye's image is vertically magnified relative to the other eye's image, a fronto-parallel surface in the field of view appears to be rotated about a vertical axis. This phenomenon, known as the Ogle's induced effect (1938), is considered to be a strong evidence that the visual system exploits vertical disparities as well as horizontal disparities. Especially in perception of slant, vertical disparity signals can be combined with ambiguous horizontal disparities to provide an unambiguous estimate of slant (Backus, Banks, van Ee, & Crowell, 1999).

### 1.1.3 Neural correlates of 3D shape processing

Early studies in monkeys found evidence for neurons tuned specifically to binocular information (Hubel & Wiesel, 1970). More recently, human brain imaging studies have shown that binocular neurons are distributed throughout the occipital cortex (for reviews, see Cumming & DeAngelis, 2001; Parker, 2007). Dorsal and ventral visual pathways differ in the sense that binocular processing in the ventral regions (V3v, V4, LOC) represents depth in shape categories without discriminating fine differences in binocular disparity, where dorsal regions responsive to fine binocular disparity information (V3d, V3A, V3B/KO) represent the magnitude of disparity (Preston, Li, Kourtzi, & Welchman, 2008) and thus retain information about surface shape within object boundaries.

Neurophysiological and fMRI studies investigating shape from shading have mostly reported activity in the early visual areas. The orientation of shading gradient is often psychophysically shown to be crucial for shape from shading (Kleffner & Ramachandran, 1992; Ramachandran, 1988), and Humphrey et al. (1997) report one of the first examples of neurological evidence showing activity in V1 corresponding to the strong depth percept obtained when shading gradient is vertical, as compared to the weak and unstable percept for horizontal gradients.

Neural correlates of shading processing have also been studied in the context of object perception (Kourtzi, Erb, Grodd, & Bülthoff, 2003; Moore & Engel, 2001). Kourtzi et al. (2003) isolated shading and contour information in object presentations to demonstrate selective responses to shape discrimination (convex vs. concave) in the anterior sub-region of the LOC (lateral occipital cortex). When taken together, these studies, including Humphrey et al. (1997), report shape from shading related activity in a variety of regions in the occipital cortex, including the early visual cortex and both the dorsal and ventral pathways, but within each design they investigate limited regions in the occipital cortex, which make it difficult to demonstratedistributed activation related to shape processing.

Taira and colleagues (2001) provide fMRI evidence from the whole brain while observers discriminate convex and concave surfaces defined solely by shading gradients. Their results show that the intraparietal area is involved in shape from shading in humans, and because the parietal cortex is often reported to be involved in binocular processing, they suggest this region might be a locus for integration of shading and disparity cues. In another fMRI study using only monocular cues (shading and texture, Georgieva, Todd, Peeters, & Orban, 2008), the authors report caudal inferior temporal gyrus and additional lateral occipital cortex activity related to texture processing, but unlike Kourtzi et al. (2003) they report no evidence for shape from shading in the LOC.

Variations in the reported brain areas involved in the processing of shading might relate to the different stimulus sets used. Alternatively, they might reflect observer idiosyncrasies, as inferring shape from shading requires many assumptions, not least about light source direction. Data from electrophysiological recordings (Mamassian, Jentzsch, Bacon, & Schweinberger, 2003) demonstrate that light source direction is detected very early in the visual system (100 ms). In the same study, the authors also report individual observers' electrophysiological activity in the occipital and temporal cortex which is highly correlated to their perception, i.e. the variation in bias for left light preference. These findings not only suggest that shape from shading is achieved by bottom-up processes, but also that ambiguously shading scenes are encoded very quickly in the visual system. A more recent fMRI study to explore lighting prior in the brain (Gerardin, Kourtzi, & Mamassian, 2010) separates shape-from-shading into two stages and demonstrate that early and quick 'light processing' is followed by 'shape processing' at the higher stages of the visual system.

While it is more common to report neural basis of separate depth cues, other examples show neurons to be sensitive to multiple depth cues (Howard, 2003). In their study, Tsutsui and colleagues (2001) reported single cell recordings from monkey intraparietal

sulcus while they were shown disparity and texture gradient defined slated surfaces. They first specified neurons sensitive to disparity defined slant in the caudal intraparietal sulcus (CIP, a dorsal visual area bordering with and receiving direct fibre projections from V3A). Later, they discovered that these neurons are still selective to surface slant when it is defined by a linear texture gradient in the absence of disparity. In a following study (Tsutsui, Sakata, Naganuma, & Taira, 2002), the authors challenged CIP neurons to find that 77% of the neurons sensitive to texture showed selectivity for disparity defined slant. This first report of neurons selective to 3D depth invariant of the cue type raises an opportunity to further explore the phenomenon in humans.

More recently, Ban et al. (2012) investigated the neural activity specific to the integration of two computationally similar cues to depth: disparity and motion parallax. In addition to providing further evidence for cortical locations related to disparity and motion processing, they tested a region in the higher dorsal visual cortex (V3B/Kinetic Occipital, KO) and showed evidence for distinct neural patterns corresponding to cue integration. This finding presents an opportunity to test the cortical locus, V3B/KO, with other cue pairings such as disparity and shading; because despite their computational dissimilarity, psychophysical data already suggests that perception is enhanced when both shading and disparity signal the same shape, as described in the next section.

## 1.2 Cue integration to estimate depth

When multiple cues signal 3D geometry, the visual system merges all available information to estimate a coherent and/or more precise percept. Several models have been proposed to understand this process. One can assume that the visual system estimates a modular percept from each cue independently, and then combines these estimates linearly to obtain a unified percept (weak fusion, Clark & Yuille, 1990). Comparatively, in a strong fusion model, cues can

interact and a unified percept can be calculated without having access to the independent estimates per cue. Moreover, the integration is not necessarily linear: it can be modelled so as to maximise the likelihood of the percept (Nakayama & Shimojo, 1992). In a model that lies between these two ends of the spectrum, Landy and colleagues (1995) propose a Modified Weak Fusion for understanding depth cue combination. In this model, the resulting estimate is a weighted average of the individual estimates; hence it resembles weak fusion.

In an earlier example of cue integration, Ernst and Banks (2002) investigate visual and haptic cues while observers estimate height. The authors quantify visual cue dominance over the haptic cue by measuring variances for height estimates individually from visual and haptic cues, and then model behavioural data with a maximum likelihood integrator that minimises the variance (hence maximises reliability) in the combined estimate. In other words, they show that the human perceptual system weighs each cue according to its reliability to integrate an optimal result, where the resulting percept is most reliable when both visual and haptic cues are available.

When there are two cues available, the most efficient (optimal) way of combining the estimates from these cues would be to come up with the most reliable combined estimate (*S*). If the noise of each estimate is independent and Gaussian, the most reliable combined estimate, i.e. one with the lowest variance, is the Maximum Likelihood Estimate (MLE). This is calculated as a weighted sum of the individual estimates ($s_1$ and $s_2$), where weight for each cue, $\omega_1$ and $\omega_2$, is the reliability of the cue (Ernst, 2006):

$$S = s_1 \omega_1 + s_2 \omega_2$$

where $\omega_1 + \omega_2 = 1$.

$$S = s_1 \frac{1/\sigma_1^2}{1/\sigma_1^2 + 1/\sigma_2^2} + s_2 \frac{1/\sigma_2^2}{1/\sigma_1^2 + 1/\sigma_2^2}$$

In this manner, the reliability of the combined estimate is the sum of the reliabilities and it is increased, i.e. final estimate's variance ($\sigma_{12}$) is smaller than each individual estimate's variance:

$$\sigma_{12}=\sigma_1 \sigma_2\sigma_1+\sigma_2$$

This weighted sum is often referred to as 'optimal combination' of sensory estimates (Figure 1.4a). As can be seen from the example in Figure 1.4a, when two estimates $s_1$ and $s_2$ are combined to create $s_{12}$ the resulting estimate is closer to $s_2$, i.e. it is weighted more because it has lower variance (higher reliability).

The maximum likelihood estimation model was also used to show that humans combine texture and stereo vision cues (Hillis, Ernst, Banks, & Landy, 2002; Knill & Saunders, 2003) and; shading and stereo vision cues (Lovell, Bloj, & Harris, 2012) in a statistically optimal fashion.

When the discrepancy between reliability of two cues is relatively high, the more reliable cue might appear to capture the resulting percept, i.e. perceptual judgements might rely on the more reliable cue almost entirely. However, when higher reliability is compromised (e.g. with the addition of noise), resulting percept might rely on the previously less effective cue (Alais & Burr, 2004; Ernst & Banks, 2002).

In the light of optimal integration, Nandy and Tjan (2008) quantify the behaviour of the ideal observer with an index of integration. As illustrated in Figure 1.4b, two independent cues (red and blue) provide similar sensitivity discriminations (distributions along the sides of the square) in a two-fold classification (yellow and green, e.g. two shapes). When both cues are taken into account, the discrimination sensitivity improves as can be seen merely by the Euclidean distances (separation of distributions) along the sides of the square being shorter than those along the diagonal. They quantify further improvements in discrimination

sensitivity with an integration index ($\phi$) based on the quadratic summation of sensitivities ($S$) to individual cues:

$$\phi = \frac{S_{12}}{\sqrt{S_1^2 + S_2^2}}$$

According to this quantification, an index value of 1 indicates either independent processing of each cue or optimal integration of the cues, where anything less than 1 shows suboptimal integration. This value ($\phi$ = 1) has also been suggested to be a minimum bound for fusion, and for values greater than 1, the index would be an indicator of the fusion of two cues (also see Ban *et al.*, 2012). Details of this framework which are specific to disparity and shading cues are explained in section 5.1 of this thesis.

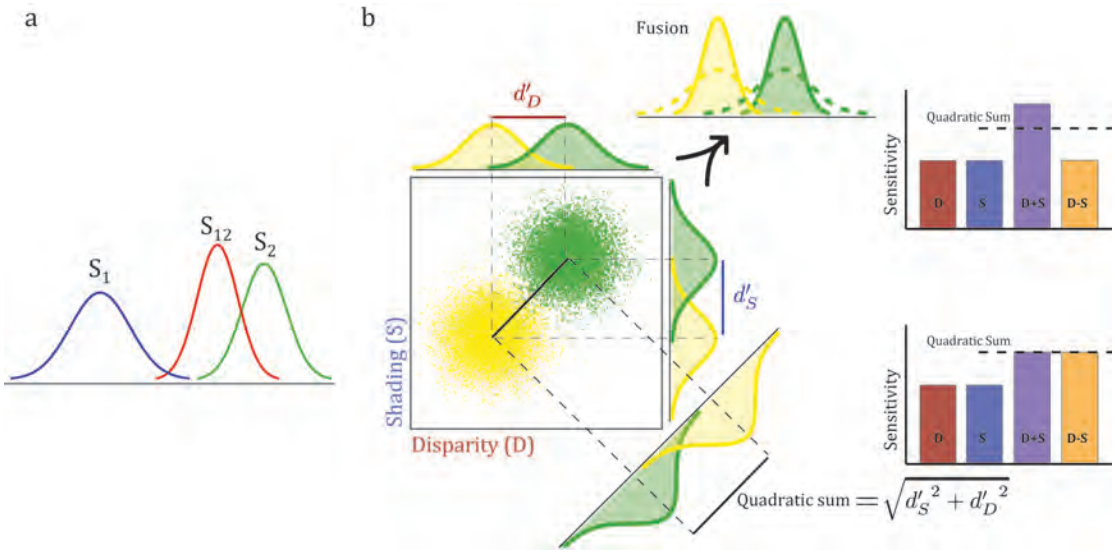



Figure 1.4: (a) Probability distributions of two estimates, $S_1$ and $S_2$, and the resulting estimate $S_{12}$ which is derived from the optimal combination of two estimates from two different cues. (b) Probability distributions to discriminate two shape classes (e.g. convex and concave) are illustrated by the dot density of the yellow and green clouds. Separation of distributions along the vertical side demonstrate estimation from shading cue, and along the horizontal side from the disparity cue. In this example, the x-axis indicates separation of convex and concave (left to right) when only disparity cue is available (as indicated by the S distributions projected on top of the square plot). A combined estimate from both disparity and shading provides a better discrimination. As the combination of independent cues is shown along the diagonal as the quadratic summation calculation; and fusion mechanism is shown as the multiplication of the probability distributions on top. Cue and stimulus specific discussion of the fusion mechanism is provided in Chapter 5.

The maximum likelihood estimation model was also used to show that humans combine texture and stereo vision cues (Hillis, Ernst, Banks, & Landy, 2002; Knill & Saunders, 2003), shading and stereo vision cues (Lovell, Bloj, & Harris, 2012) in a statistically optimal fashion.

When the discrepancy between the reliability of two cues is relatively high, the more reliable cue might capture the resulting percept, i.e. perceptual judgements might rely solely on the more reliable cue. However, when higher reliability is compromised (e.g. with the addition of noise), the resulting percept might rely on the previously less effective cue (Alais & Burr, 2004).

Bülthoff and Mallot (1988) demonstrate that shading information is completely overridden when edge information is available. However, their findings support the notion of shading as an additional cue which increases the perceived depth profile of a disparity defined surface. Similarly, Vuong et al. (2006) studied shading and disparity in combination; even though their results do not directly imply an increase in depth profiles, they show that perceptual reports become more precise when the two cues are both present. Furthermore, when a third cue, motion parallax, is added, both accuracy and reaction time have been reported to improve (Schiller, Slocum, Jao, & Weiner, 2011).

In a recent study by Lovell and colleagues (2012), when disparity cue was compromised by adding noise, shading information proved to be a reliable cue to shape. By matching the reliability of disparity to shading, the authors overcame the problem of shading cues often being overridden by disparity and showed comparable sensitivity for shape discrimination. Under these conditions, they found cue integration between disparity and shading which can be modelled by a maximum likelihood estimator.

Statistically optimal cue integration has been reported in adults, but infant studies suggest that the ability to integrate visual cues is acquired in later stages of development (Burr & Gori, 2012; Nardini, Bedford, & Mareschal, 2010). This later achievement supports the view that different modalities recalibrate each other during development, especially haptic information recalibrating visual information.

## 1.3 Overview of chapters

**Chapter 2.** The next chapter lays out a summary of the various methods and equipment used in this thesis while discussing their merits and shortcomings. Each experimental chapter describes the specific methods used in more detail.

**Chapter 3.** This first experimental chapter examines how first- and second-order luminance signals are perceived, and whether their phase relationship provides a cue to layer decomposition and hence shape for naive observers. We show that observers learn to benefit from the phase relationship information after training with binocular feedback, and perceive in-phase gratings as corrugated. However, the percept seems to be suddenly reversible with little reinforcement, suggesting categorical learning of stimulus labels rather than a perceptual improvement.

**Chapter 4.** The second experimental chapter shows that the phase relationship of first- and second-order signals cannot be discriminated at short presentation times, so processing phase information might be a high level function in the brain. Through training with intermittent feedback on performance, phase discrimination can be learned at a perceptual level. Here the results of training are specific to the trained stimulus properties. These findings suggest that training tunes the processes of an early and automatic mechanism that leads to layer decomposition of in- and anti-phase signals of first- and second-order stimuli.

**Chapter 5.** In this chapter, we challenge previously reported cortical locations for 3D shape processing with disparity and shading cues. We report multiple lines of evidence from behavioural and functional magnetic resonance (fMRI) data, suggesting that the dorsal area V3B/KO plays a crucial role in integration of these cues. Furthermore, we find a distinct relation between individual differences in perception and underlying neural activity for processing disparity and shading cues in combination.

**Chapter 6.** In this chapter individual differences in perceiving disparity and shading are explored more fully. Similar to the results from Chapter 5, we report systematic differences between observers: Although beyond optimal integration for the two cues is observable in half of the observers, it is far from optimal for the remaining observers; rather the results suggest that an independent processing mechanism is preferred. However, I show that fusion can be achieved for these participants after training.

**Chapter 7.** The final chapter summarises the findings in this thesis, and brings together the contributions of the substantive chapters to our understanding of estimating 3D shape from disparity and shading.

# CHAPTER 2:

# General Experimental Methods

The empirical chapters in this thesis employ a number of psychophysical methods and imaging techniques. This chapter serves as an overall summary of the methods used in each experiment, discussing each method's merits and limitations. I will start by describing the two general types of stimulus used in this thesis. Then I will describe the different psychophysical set-ups which I used to present stimuli. Finally, I will provide definitions for the psychophysics model that is used to interpret behavioural performance. This will be followed by a general description of the imaging techniques used in Chapter 5. Each empirical chapter in this thesis will also provide detailed stimulus specifications for the experiments under discussion.

## 2.1 Psychophysics

### 2.1.1 Stimulus generation

I used two sets of stimuli for the experiments in this thesis. The first set consisted of luminance gratings, employing first-order luminance gratings and second-order modulations of texture amplitude to simulate the effects of shading on a corrugated textured surface while retaining precise control over the stimulus properties, thus allowing fine manipulations to explore shape from shading. Next, I used random dot stereograms (RDS) superimposed with linear luminance gradients. RDS are popular for studying shape from disparity in isolation, but when there is additional luminance information, they also become advantageous tools for investigating disparity and shading cues in combination. In this thesis, RDS were used in both psychophysical and imaging experiments, while luminance gratings were used for psychophysical experiments only.

#### 2.1.1.1 Luminance gratings of first- and second-order information

In Chapters 3 and 4, I used sine wave luminance gratings added to noise texture to simulate the shading patterns found on albedo textured Lambertian surfaces. One sine wave grating has luminance modulations, where in the other grating, the texture elements are amplitude modulated. These sine waves are then added together, either spatially in-phase (LM+AM) or anti-phase (LM-AM), to simulate the luminance variations that would occur on an undulating surface (Figure 2.1). Stimuli were created following the original instructions reported in Schofield *et al.* (2006) except for the noise pattern in Chapter 3 (see below).

For single gratings, □□□,□ and □□□,□are created separately, and then imposed on a noise texture:

$$\square\square\square,\square=\square\times\cos 2\square\square\cos\square 1\square-\sin\square 1\square-\square 1 \qquad \text{(Eqn. 2.1)}$$

$$\square\square\square,\square=\square\times\cos 2\square\square\cos\square 2\square-\sin\square 2\square-\square 2 \qquad \text{(Eqn. 2.2)}$$

$$\square\square,\square=\square 01+\square\square\square,\square+\square\square\square,\square+\square\square\square,\square\times\square\square\square,\square \qquad \text{(Eqn. 2.3)}$$

where $\square$ is the spatial frequency of the modulation, $\square$ is the contrast of $\square\square\square,\square$, and $\square$ is the modulation depth of $\square\square\square,\square$. As for the angles, $\square 1$ and $\square 2$ are the orientations of LM and $\square\square$, respectively; and, $\square 1$ and $\square 2$ are the spatial phase angles of $\square\square$ and $\square\square$, respectively. The noise texture ($\square\square,\square$) has contrast *n* and is added to $\square\square$ and multiplied by $\square\square$. To plot an LM/AM grating (Figure 2.1, c), $\square\square$ and $\square\square$ modulated noise is added to 1 and this is scaled by the mean luminance of the monitor ($\square 0$) (Equation 2.3). In the case of an anti-phase grating, signals are spatially aligned anti-phase: $\square 1-\square 2=\square$ (Figure 2.1, c, left); while for the in-phase grating $\square 1-\square 2=0$ (Figure 2.1, c, right).

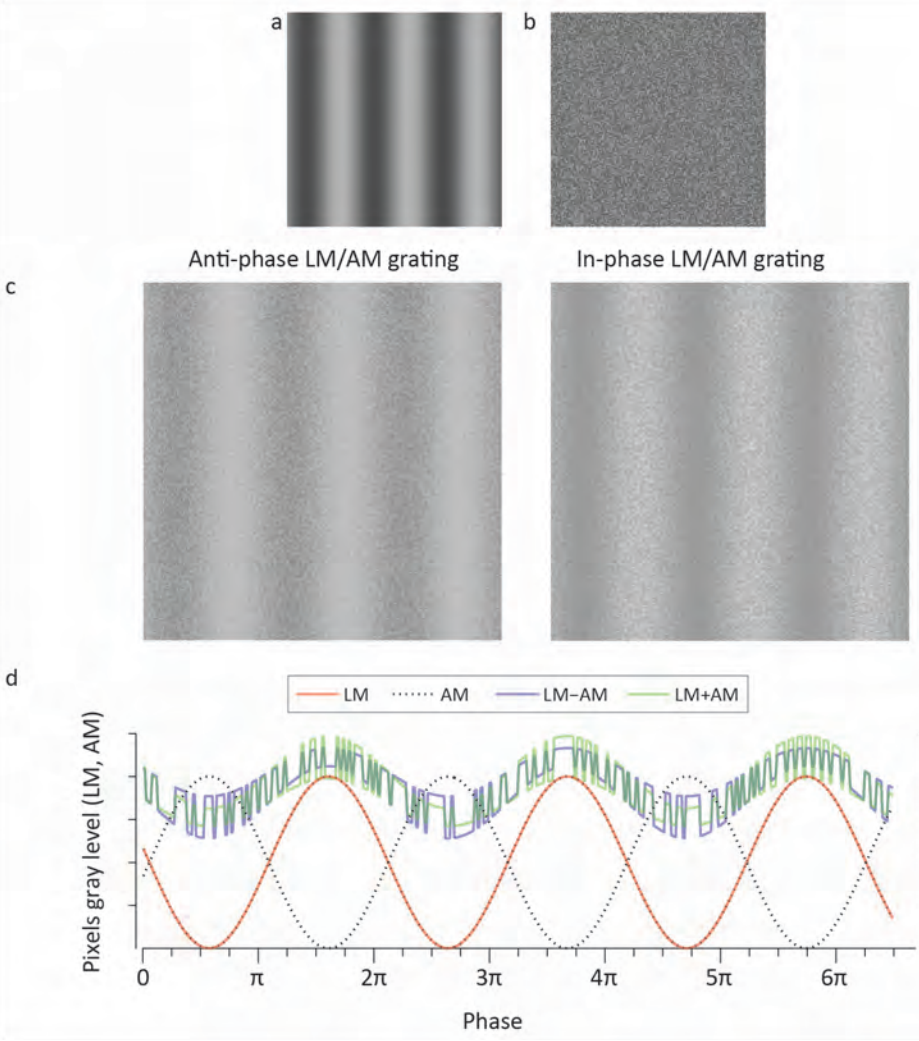



Figure 2.1: Stimulus creation for LM/AM mixes. All gratings in this example are oriented vertical: □=□°. (a)Luminance modulation (LM), and (b) binary noise texture were used to create stimuli. AM component is not visible unless there is a texture carrier present, so it is demonstrated in combination with LM and binary noise in (c). (c) On the left, an anti-phase grating resulting from addition of a, and modulated b; on the right an in-phase grating resulting from adding a, and modulated b. (d) The plot represents each components (LM, red solid line; AM, black dotted line) phase offset (x-axis). AM is plotted twice to indicate in- and anti-phase alignment with LM. When LM and AM are combined with 0° offset, an in-phase luminance and amplitude modulated noise texture results (green line), and when they are combined with an offset of □ an anti-phase texture results (purple line). The contrast ratio for AM in this plot (black dotted lines) depicts modulation, where AM's contrast ratio can only be seen in combination with a texture signal (blue and green lines). In these examples, higher than normal contrast ratios were used for LM (l = 0.4), AM (m = 0.4) and binary noise (n = 0.2) for printing purposes.

In a plaid stimulus, a pair of luminance gratings *(□□□)* and a plaid of amplitude modulations *(□□□)* are created first according to Equations 2.4 & 2.5.

□□□□,□=□□□,□+□□cos2□□cos□□□−sin□□□−□□+□□cos2□□cos□□□−sin□□□−□□

(Eqn. 2.4)

□□□□,□=□□□,□×□□cos2□□cos□□□−sin□□□−□□+□□cos2□□cos□□□−sin□□□−□□

(Eqn. 2.5)

where □ is the spatial frequency of the modulation, □□ and □□ are the contrasts of □□□□,□, and □□ and □□ are the modulation depths of □□□□,□. As for the angles, □□ and □□ are the orientations of □□□ components, □□ and □□ are the orientations of the □□□ components, where □□ and □□ stand for the spatial phase angle of □□□, and, □□ and □□ are the spatial phase of □□□. Essentially, Equation 2.4 gives a plaid consisting of two sine wave gratings added to noise, where the components of the plaid are separated by □□−□□° (which would always be equal to the separation angle for the components of □□□□,□, that is □□−□□°, in this thesis). Similarly, Equation 2.5 produces a plaid composed of two AM gratings of noise texture.

The plaid stimuli □□,□ used in Chapter 3 and Chapter 4 were created by combining the □□□,□ and □□□,□ components while scaling the image with the monitor's mean luminance (□*0*):

□□,□=□*01*+□□□□,□+□□□□,□ (Eqn. 2.6)

For instance, in the orthogonal plaid, as illustrated in Experiment 1 (Chapter 4, Figure 4.1E), the right diagonal's LM and AM components are orientated 45° clockwise from vertical □□=□□=*45*°, and the left diagonal's LM and AM components are orientated -45° clockwise from vertical □□=□□=*−45*°. In this example, the right diagonal is *in-phase*, so the difference between angles, □□ and □□ is equal to 0, whereas for the left diagonal, □□−□□=□, hence LM and AM are combined *anti-phase* with a phase offset of 180°.

□=□□□□−□□□□□□□□□+□□□□ (Eqn. 2.7)

Elements of the binary noise texture given by the component □□,□ had one of two intensity values: Highest luminance (□□□□), or lowest luminance (□□□□), and the overall noise contrast (□) was set to 0.1 in all experiments (Equation 2.7). The experimental designs manipulated AM signal level in the stimuli to probe performance, hence we used an interval of AM contrasts (□) that fell into a previously reported detectable signal range (Schofield & Georgeson, 1999).

In Chapter 3, instead of using binary noise texture, a texture consisting of Gabor patches was used to optimise for the requirements for stereoscopic presentation (see Methods section, Chapter 3). The single gratings were superimposed with disparity defined corrugated surfaces, and disparity modulation of the surfaces was achieved by showing two alternating image sequences through FE-1 goggles. A method similar to anti-aliasing was used to achieve disparities less than a pixel. Gabor micro-patterns were created offline, and each pixel was magnified into a 101×101 sub-pixel grid in which Gabor patches could be displaced with sub-pixel accuracy. The final micro-patterns were created by averaging the grey levels of the 101x101 sub-pixels in each display pixel. Thus small disparities were converted into subtle changes in grey level. Despite this conversion, such micro-patterns can convey small stereoscopic disparities well.

Grating and plaid stimuli were created and presented with custom software written in C++, using the frame store of a VSG graphics card (CRS Ltd, UK).

#### 2.1.1.2 Random Dot Stereograms with Shading

The experiments in Chapters 5 and 6 used stimuli composed of RDS patterns superimposed with a Blinn-Phong shading algorithm. In each quadrant of the stimulus, a hemisphere depicted a convex or a concave figure. Two convex figures on the left of fixation and two

concave figures on the right of fixation (or vice versa) depicted identical surfaces in a given interval, but were randomised for convex on the left and concave on the left during the trials.

Two parallel surfaces were created separately: the RDS (height field) and the shading surface. Then the elements of RDS were assigned an intensity level according to their corresponding projection point in the 2D image of the shading surface. The RDS subtended a square of □×□°, including four hemispheres of 1.5° radius, and each (dot) element of RDS had a radius of 0.03° at a viewing distance of 65 cm. Anti-aliasing on the RDS elements was used (oversampling ratio = 3) to achieve sub-pixel disparity. The dot density was set to 94 dots per degree square. The maximum depth amplitude used was 6 arcmin, i.e. the centre of the hemisphere would be -6 arcmin in the convex surface and 6 arcmin in the concave surface. These values were jittered ± 1 arcmin across trials to minimise adaptation.

Using smaller RDS elements with higher dot density facilitated the approximation of a continuous 3D surface and increasde the benefit from the shading cue while still allowing control of the disparity cue. This surface has no intensity variation; rather it serves as a height field map where elements are assigned a luminance value from the shading surface according to their □,□ position in the height field.

□=□+□□+□ Eqn. 2.8

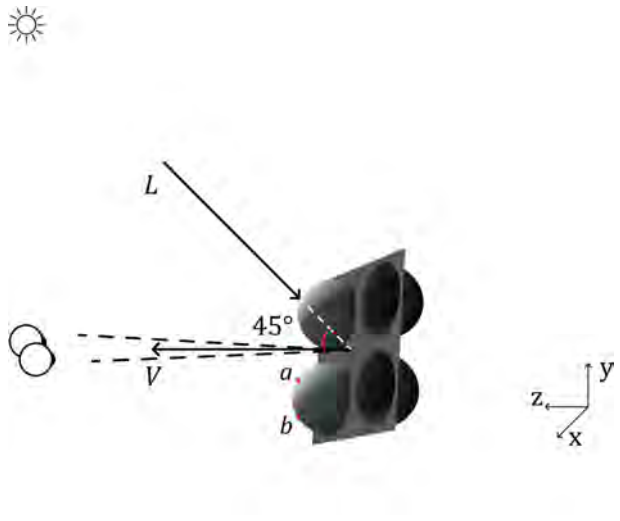



Figure 2.2: Cartoon of the settings for the shading model. Light source and the viewer are assumed to be at infinity. The normalised vector for the viewpoint (*V*) is orthogonal to the stimulus plane (xy-plane). Light source vector (*L*) is also normalised and it forms a 45° angle with the normal of the stimulus plane. Points a and b are at the same depth amplitude (hence have the same height field), but point a is brighter because it is facing towards the light source in this example of a convex hemisphere.

The shading surface was created monocularly using the Blinn-Phong algorithm implemented in Matlab (Lyon, 1993). This algorithm assumes that the light source and viewer are at infinity, and calculates the halfway vector (*H*) between the normalised vectors of a light source (*L*) and the viewpoint (*V*) (Equation 2.4). The viewpoint vector was set parallel to the normal of the stimulus plane central to the fixation: $\vec{V}(x,y,z)=0, 0, 1$, and a point light source was positioned over the observer's head, making a 45° angle with the surface normal: $\vec{L}(x,y,z)=0,-1,-1$ (Figure 2.2). The specularity factor was minimised to simulate a Lambertian surface so as to isolate the shading cue from any specular information.

For the disparity alone stimuli, the RDS elements had the same overall luminance distributions with the shading stimuli, but their position within the figure was scrambled so that the intensity variation did not indicate any shape from shading. For the shading alone stimuli, the RDS elements depicted a flat disparity surface, and the luminance variation was the cue to a curved 3D surface. For the binary luminance stimuli, the shading pattern was

binarised (two intensity levels) corresponding to the mean luminance of the top and bottom portions of the hemispheres. A peripheral grid of black and white squares that subtended □□×□□° surrounded the stimuli. This larger grid remained unchanged throughout a session and served as a stable background reference. The background of this reference grid was set to mid-level grey.

Stimuli were created with custom made scripts in Matlab (The MathWorks, MA, USA) and presented using the psychophysics toolbox (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997).

### 2.1.2 Stimulus presentation

#### 2.1.2.1 Monocular display

We used a dichotopic presentation for the study reported in Chapter 4. Luminance grating stimuli were presented on a ViewSonic P225f monitor at a refresh rate of 160 Hz.

The monitor's gamma nonlinearity was estimated using a ColourCal luminance meter (Cambridge Research Systems, CRS Ltd., UK) and corrected via the VSG's look-up tables (LUT) using a four parameter CRT display characterisation model (Brainard, Pelli, & Robson, 2002).

$$□ = □□□□□□ + □0□□□□□ + □0□$$

Eqn. 2.9

where $L$ (cd/m$^2$) is the luminance output read from the monitor, $j$ is the luminance value of the LUT that varies between 0 and □□□□ (for an pseudo 15-bit- graphics card as used here, □□□□=*32767*), and $L_{max}$, $j_0$, k and $\gamma$ are the parameters to be adjusted). A sample set of luminance values were first run from a linear LUT, and using the `fminsearch` function

in Matlab, four parameters were estimated for the lowest value of LUT (□0), the lowest luminance value that can be displayed by the monitor ($k$), $L_{max}$, and finally the gamma value ($\gamma$) to generate a new LUT to be used in the experiments.

#### 2.1.2.2 Stereoscopic displays

For the experiments in Chapter 3, I used a Clinton Monoray monitor (CRS Ltd, UK) that emits a yellow/green (DP104) phosphor light at a refresh rate of 150 Hz. FE-1 shutter goggles (CRS Ltd, UK) were fixed to a chin rest at a 0.6 m viewing distance. The goggles use alternating frame sequences to give a stereoscopic view. Minimising cross talk between the eyes is critical with this set-up. The FE-1 shutters have a very high transmittance ratio between their on and off states and a very fast switching function. When combined with the Monoray monitor with rapid decay phosphor, cross talk is minimised. This equipment set-up is superior to a combination of LCD shutters and standard RGB monitor. The Monoray monitor also has a high maximum luminance to counter attenuation from the goggles in their on state. The monitor's gamma nonlinearity was estimated using a ColourCal luminance meter (CRS Ltd, UK) with a neutral density filter to attenuate luminance to a level within the ColourCals operating range. This attenuation was corrected prior to deriving the monitor's gamma characteristic. Gamma was again corrected using the VSG's lookup tables. The method used was similar to that described above (**Equation 2.9**).

For the behavioural experiments in Chapter 5 and Chapter 6, psychophysical stimuli were presented on a mirror stereoscope in a Wheatstone configuration using a pair of ViewSonic P225f CRT monitors. Each monitor had a screen resolution of 1600 x 1200 pixels, and stimuli were displayed at a refresh rate of 100 Hz with an Nvidia Quadro 4400 graphics card. Linearisation of the graphics card outputs was achieved using photometric measurements. The haploscope setting allows us to show separate images to each eye through front-silvered mirrors; hence the experimenter can simulate binocular disparities

that specify a 3D object. Monitors and mirrors are mounted on adjustable arms, allowing us to position the arms with respect to an individual observer's inter-pupillary distance (IPD). The observer's head position was stabilised with a chin rest at a viewing distance of 0.5 m.

The stereoscopic presentation inside the scanner was achieved using two video projectors (JVC; D-ILA SX21): the optical images were combined using a beam-splitter and then projected into the scanner room (Preston *et al.*, 2008). Participants were lying down inside the bore, and they viewed images back projected on a translucent screen, with the aid of a front-surfaced mirror mounted on the head-coil at a viewing distance of 0.65 m. Stereoscopic presentation was achieved by using two projectors and separate spectral interference filters (INFITEC), which are narrowly bandpass, allowing the presentation of different spectra for each the eyes. The two projectors were matched and linearised using photometric measurements.

### 2.1.3 Psychophysics methods

The stimuli defined in the previous section were designed to simulate aspects of visual presentation in the physical world, while allowing us to quantify and manipulate their properties systematically. For the behavioural experiments in this thesis, I used classical measures of observers' responses to stimuli that were varied systematically. With this in mind, one stimulus dimension was manipulated at a time while everything else was kept constant, to control the measures of performance.

The psychometric function was used as a model to describe the relationship between the observer's performance and the physical characteristics of the stimulus (Wichmann & Hill, 2001). The independent variable was the physical quantity of the stimulus; that is stimulus intensity (plotted on the x-axis), while the observer's response, i.e. proportion correct, is a dependent variable (plotted on the y-axis). The psychometric function describes

the rate of change in performance in terms of the stimulus intensity. Ideally, it would be similar to a sigmoid function: for the highest absolute stimulus intensity (two extremes of the x-axis), one would get the highest proportion correct, but when stimulus intensity is too small to be discriminable, performance would be around chance level. I used a cumulative Gaussian shape to model the psychometric function in all experiments.

All the psychophysical experiments in this thesis employ a two-alternative forced choice method, where the observer is forced to choose between two possible responses, and usually instructed to make a guess when they are uncertain about the stimuli. Also, in these experiments, test questions are always designed to prompt the observer to discriminate between two types of stimulus, instead of detecting a signal; the reported thresholds refer to discrimination thresholds instead of detection thresholds.

In Chapters 3 and 4, usually, a single stimulus is presented and the participant is asked to choose between two responses, "Left" vs. "Right", to answer the question, "Which component of the plaid seemed more corrugated in depth?". For the training experiments in Chapter 3, two visual stimuli were presented simultaneously above and below the fixation at a single interval, and the observer responded to "Which grating seemed more corrugated in depth?" by choosing either "Above" or "Below".

A two-interval presentation method was used for the experiments in Chapter 5 and 6. One of the intervals always contained a standard stimulus, and the other interval contained the test stimulus that would vary in stimulus intensity; the order of standard and test stimuli was randomised. After the two intervals, observers were asked: "Which stimulus had greater depth?" and responded between "First" or "Second". The reason behind comparing various stimuli against a standard stimulus was to find the point of subjective equality. In other words, I measured the intensity at which observers report that they perceive the test stimulus equal to the standard stimulus. Observers' performance for discrimination is

expected to be around 50 per cent correct when standard and test stimuli are the same, and up to 100 per cent correct for those test stimuli that are separated most from the standard stimulus.

In most of the experiments, I used the method of constant stimuli: a number of stimulus intensities were chosen from an interval that contains both easily discriminable signals and lower contrast signals which are difficult (or impossible at times) to discriminate. Each stimulus level is presented multiple times (minimum 20 repetitions) and the mean of these repetitions is taken to represent the performance at that stimulus level. A psychometric function is then fitted to the data, where there is no limitation on the parameter slope and shift. These parameters would help us define precision (slope) and the point of subjective equality (PSE, shift) once the data has been fitted. When reporting group data, psychometric functions were fitted to individual observers' data, and then pooled to report mean and errors (fit-then-pool). This method is reported (Wallis, Baker, Meese, & Georgeson, 2013) to have an effect of maintaining the steepness of individual slopes when compared to fitting a psychometric function to pooled data points (pool-then-fit), but the difference between the two methods is suggested to be negligible. Chance level performance was at 50% correct, and threshold values were acquired at the inflection point of the function where available.

I have used the *discrimination thresholds* in psychophysics experiments in this thesis, and these were derived from the standard deviation (slope) of the fitted cumulative Gaussian functions. In experiments where I used the method of constant stimuli, I constrained the two parameters lapse rate and guess rate, and derived slope and point of subjective equality parameters from the fit. The standard deviation of the fitted function ($\sigma^2$) is very much related to the slope I use, 82% of the sigma squared is used to derive the slope. The Just Noticeable Difference (JND) can be calculated as the difference between this point and the 50% point. The variability of the probability density function for the combined estimate is

related to the JND defined at this level as JND corresponds to$2 \cdot \sigma$. Given that reliability is the inverse of the variance ($\sigma^2$), the reliabilities can be derived from the JND measurements.

In Chapter 6, a Bayesian adaptive staircase method was implemented to measure discrimination thresholds at an 82% correct level, which is similar to a conventional 3-up-1-down staircase procedure (QUEST, Watson & Pelli, 1983). This procedure assumes that participants' behaviour is explained by a canonical psychometric function, *i.e.* it has a fixed shape, while varying its position along the stimulus intensity axis (x-axis) to estimate a threshold. A cumulative Gaussian distribution around the initial guess for a threshold (maximum depth amplitude = 4.5 arcmin) is used as the prior probability density function (*pdf*) at the start of the experiment. Afterwards, data from each trial is used to update the estimated threshold (mean of *pdf*) while minimising the variance of the *pdf*. Participants completed at least 30 trials per experimental condition, and the convergence of the data from all trials was inspected before taking into account the final threshold estimate from QUEST.

The QUEST method is efficient in the sense that estimation for a threshold can be acquired much more quickly than for a constant stimulus method. In a constant stimulus method, on the other hand, many repetitions must be measured for each discrete stimulus level in order to be able to fit the psychometric model reliably. An advantage of the psychometric function is that if the constant levels of stimulus cover a sufficient range of stimulus intensities, and performance measurements have enough repetitions to provide a good fit of the function, then, by interpolation, one can infer the relation between stimulus and response in the full dataset. Compared to this, a single QUEST procedure for each stimulus condition is limited, because it gives a single estimation of the threshold, e.g. at an 82% correct level.

## 2.2 Functional Magnetic Resonance Imaging

In addition to measuring behavioural performance using the previously described methods, we also wanted to assess the neural activity underlying the behaviour (Huettel, 2009; Logothetis & Pfeuffer, 2004). This imaging technique relies on the haemodynamics (blood flow) of the body, and measures the levels of oxygen in the blood, *i.e.* blood oxygen level dependent (BOLD) signal.

In other words, instead of measuring a direct neurophysiological signal, MRI methods inherently assume that neural activity and blood flow are correlated. However, studies combining electrophysiological recordings with fMRI data confirm the speculation of BOLD signal reflecting the neural activity in the cortex (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001; Logothetis & Pfeuffer, 2004).

### 2.2.1 Data acquisition

All fMRI data was acquired at the Birmingham University Imaging Centre using a 3 Tesla Phillips Achieva MRI scanner with an 8-channel multi-phase array head coil. Two separate sessions were run for each participant: one for functional localisers to create an individual functional map of the participant's visual cortex, and another experimental session to identify neural activity related to the experimental condition.

An echo-planar (EPI) pulse sequence was used to acquire MR images quickly, where an entire image was acquired within 35 ms (*echo time, TE*), and this pulse sequence was repeated every 2s (*repetition time TR*). In total, 28 slices aligned orthogonal to the calcarine sulcus (near the coronal) were used to cover the occipital cortex. The voxel dimensions in the functional scans were *1.5×1.5×2* mm. In the functional localiser sessions, in order to be able to acquire a decent segregation between grey and white matter, a high-resolution T1-weighted structural (anatomical) scan was acquired for each participant with voxel size *1×1×1* mm.

This structural scan was later used to reconstruct the cortical surface and to register functional data acquired on different days in Talairach coordinates. In the functional scans, T2-weighted images were acquired in order to be able to segregate tissue fluid, i.e. blood.

### 2.2.2 Data analysis

#### 2.2.2.1 Pre-processing

High resolution anatomical scans were first translated by referencing the anterior and posterior commissures, and extreme points in the individual brain, and then resized to fit into a standard Talairach space. Inflated and flattened surface models were created for both hemispheres for each participant.

For functional runs, the data had to be corrected using pre-processing steps prior to analysis. First, a temporal interpolation was performed using TR and inter slice time value (*slice scan time correction*). This made sure that that 28 slices which were acquired over time could be interpreted as a whole volume at a given instance. Next, all volumes over time were aligned to the first volume as a reference, and displacements in the x, y, and z axes were later corrected by undoing the detected head motion (*3D motion correction*). After motion correction, each participant's functional data were aligned to their anatomical scan and transformed into a Talairach space. For temporal filtering, linear trends were detected and removed in each voxel's time series (*linear trend removal*). Time series were converted to the frequency domain first; after removing low frequency drifts (*high pass filtering*), series were converted back to time space.

In the multi-variate analysis, when it comes to the classification of functional image patterns (Section 2.2.2.4), the method relies highly on the spatial position of voxels over time. For this reason, we applied sensitive criteria for motion correction, and discarded data from participants who moved more than 4 mm in any direction (see Figure 2.3 for an example).

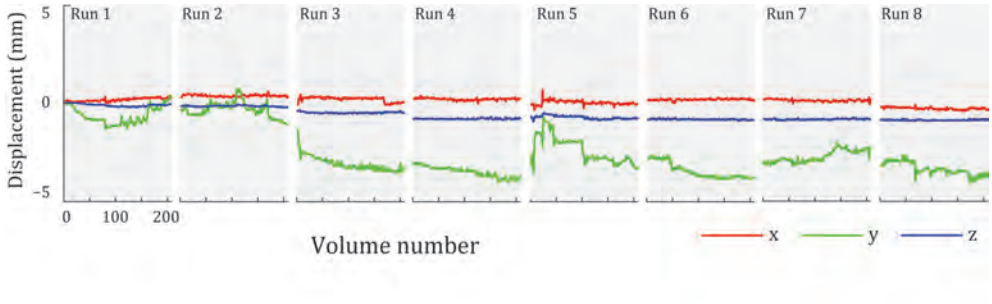



Figure 2.3: Example of head motion during experimental scans. Data plotted represent head movements from one participant (IM) who was discarded from further analysis. Displacement in three directions is plotted as a function of slice volume number (every 2 s, total of 208 volumes each run). Here, the criteria for data inclusion in analysis is breached by participant's >4mm head movements on y-axis.

### 2.2.2.2 Defining Regions of Interest

We acquired functional images from the whole visual cortex, but data was analysed only from separately defined regions of interest (ROIs). Defining ROIs was independent of the experimental stimuli, and relied on data from a separate imaging session. Specific stimuli (e.g. moving dots with kinetic borders) were presented and regions were delineated according to voxels' functional selectivity when compared to other regions of the visual cortex. This method has been criticised for disregarding anatomical specificity, and being context-sensitive (Friston, Rotshtein, Geng, Sterzer, & Henson, 2006). However, it can be advantageous when used with consideration of anatomical landmarks, and in an experimental design where conditions are contrasted within ROIs rather than across the cortex. Furthermore, using individual ROIs minimises anatomical variability across participants, and representing a group of voxels (ROI) rather than looking for effects in individual voxels simplifies data analysis.

Initially, in an individual localiser scanning session, retinotopic borders were identified for each participant. Retinotopic borders in the early (V1, V2), dorsal (V3d, V3A, V7), and ventral (V3v, V4) visual cortex were defined with the checker board stimuli (DeYoe *et al.*, 1996; Sereno *et al.*, 1995, see Figure 2.4). Rotating wedges (size of 24° of central angle)

and expanding concentric rings (from fovea to 16° eccentricity) of coloured checker boards were used as stimuli, and the temporal pattern of neural activity was correlated to the phase of the stimulus (Yamamoto *et al.*, 2009).

Retinotopic borders were used to define the area V3B/KO, together with a separate motion structure localiser. In the V3B/KO localiser, the stimuli were moving black and white random dot kinematograms which contrasted transparent motion and partially grouped motion (diagonal stripes) to define kinetic borders (also see Ban *et al.*, 2012). A separate localiser was also used to define the borders of hMT+/V5, where random static dots were contrasted with coherently moving dots (Huk, Dougherty, & Heeger, 2002). Finally, a localiser was run to define the lateral occipital cortex (LOC): grids showing black and white images of objects were compared with scrambled versions of the objects to find the cortical areas selective to intact object stimuli (Kourtzi & Kanwisher, 2000).

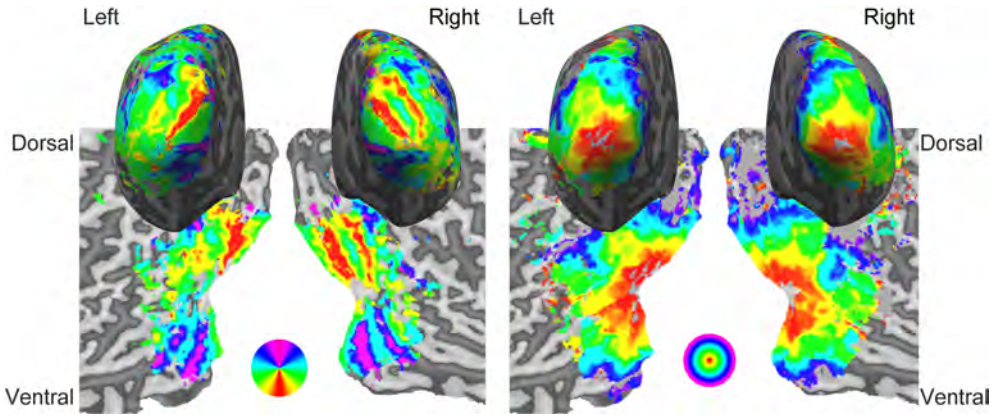



Figure 2.4: Examples of inflated and flattened cortex maps (for both left and right hemi-spheres) depicting retinotopic mapping for one participant. Retinotopic borders related to rotating wedge stimuli (flat maps on the left) and to expanding ring stimuli (flat maps on the right) are shown with relative look-up tables to indicate stimulus position in the visual field.

### 2.2.2.3 Multi-voxel Pattern Analysis (SVMlight)

FMRI data is traditionally represented as changes in the BOLD signal in a time course for each individual voxel (univariate analysis). This method restricts data representation to a point-

by-point system with temporal variation. In multi-voxel pattern analysis (MVPA), one can look for the differences in activity patterns from multiple voxels, hence adding a spatial element and making the analysis multivariate (Cox & Savoy, 2003; Haxby *et al.*, 2001; Kamitani & Tong, 2005). When looking at the fMRI data across the cortex, MVPA allows more subtle effects to be detected.

Initially, we selected voxels to be used by multivariate analysis, by sorting grey matter voxels according to their response (*t*-statistic) to all stimulus conditions versus the fixation baseline across all experimental runs (**Table 2.1**). For each ROI, 300 voxels were selected because the benefit from MVPA saturated for more than 300 voxels. We used a linear support vector machine (SVMlight toolbox) to deal with this large data set.

We calculated z-scores for the timecourse separately for each voxel and each experimental run to minimise baseline differences between runs. fMRI time series were shifted by 2 TRs (4 s) to account for the haemodynamic response lag when generating test patterns for the multivariate analysis. To remove potential univariate differences, we normalised by subtracting the mean of all voxels for a given volume (Serences & Boynton, 2007), with the result that each volume had the same mean value across voxels, and differed only in the pattern of activity.

We used an n-fold leave-one-out cross validation procedure: data from all the runs except one (typically 7 runs) were used to train the classifier (21 patterns, 3 per run) and the remaining data (3 patterns) were used to evaluate prediction accuracy. Each participant's data represents the mean prediction accuracy across cross-validation folds.

Table 2.1: Number of voxels representing each ROI (rows) averaged across 15 participants. The left column represents mean number of voxels in a region and the right column gives the mean number of significant voxels (*t*-statistic).

| | Mean number of voxels in the region | Mean number of significant voxels |
|---|---|---|
| V1 | 2011 | 1152 |
| V2 | 1981 | 1129 |
| V3v | 753 | 332 |
| V4 | 573 | 269 |
| LO | 809 | 304 |
| V3d | 804 | 482 |
| V3a | 719 | 309 |
| V3B/KO | 747 | 411 |
| V7 | 623 | 255 |
| hMT+/V5 | 606 | 212 |

Accuracies were represented in units of discriminability (*d'*) using the formula:

$$d' = 2 \cdot \mathrm{erfinv}(2p - 1) \quad \text{(Eqn. 2.10)}$$

where *erfinv* is the inverse error function and *p* the proportion of correct predictions. To prevent undefined solutions, an observed value of $p=1$ was set to $p=0.99$ ($d'=3.29$). Furthermore, when calculating ratio values, any bootstrapped values of *p* near 0.5 (i.e. $d'=0$) were rounded down to 0.49 or up to 0.51 to prevent divide-by-zero type behavior.

In addition to standard MVPA, where we test and train the SVM with the same condition (e.g. disparity), we also looked at transfer between disparity and shading cues in Chapter 5. The SVM was trained with one condition and tested on the other, and we used a Recursive Feature Elimination method (RFE) to analyse these results (De Martino *et al.*, 2008). This method allowed us to detect the minimum number of voxels with the highest prediction performance, where we started with the maximum set of voxels and discarded five

voxels in each feature elimination step. Hence, sparse patterns of multiple voxels can be detected when they provide meaningful information for discriminatation between classes, i.e. convex vs. concave.

## 2.3 Measuring Eye vergence and eye movements

In Chapter 5, we used supplementary methods to control for eye movements and vergence. Observer fixation is crucial for interpretation of fMRI results because the fixation centre of the visual field is represented in a much larger area of the cortex then the rest of the peripheral visual field (Cortical magnification, Qiu *et al.*, 2006), hence deviations in observers' fixations might cause large disruptions in the neural activity map aligned on the cortex.

To make sure that participants were fixating at the centre (radius = 1°), we presented a square crosshair target (side = 0.5°) during the experimental scans. A vertical Vernier target was briefly flashed (250 ms) on half of the trials, and participants were asked to judge its offset: "Left" *vs.* "Right". The Vernier task encouraged fixation and kept the observers alert while giving us a subjective measure of eye vergence (Popple, Smallman, & Findlay, 1998).

In a separate session, we measured eye movements with a CRS Limbus Eye Tracker (Cambridge Research Systems, CRS Ltd., UK). While the participant was lying in the scanner, the eye tracker was placed between the eye and the spectral filters. A subset of participants (n = 3) was asked to repeat the experiment during separate eye tracking sessions in the scanner. Eye position data was recorded with a spatial resolution of < 0.25° as a single voltage, and this was analysed by comparing the eye position for each experimental condition.

## 2.4 Statistical testing and reporting the data

Statistical analyses such as paired sampled t-tests and repeated measures ANOVAs were performed using SPSS (SPSS Inc.), and Greenhouse-Geisser correction was used when appropriate. Custom scripts in Matlab were used for bootstrapping for d-prime statistics

(minimum 1000 bootstrap samples) when calculating various indices defined in the empirical chapters. Figures were edited in Adobe Illustrator prior to publishing.

## 2.5 Observers

Experimental procedures received favourable ethical opinion by the University of Birmingham ethics committee. All observers were recruited from University of Birmingham students and they gave written informed consent before participating in the experiments. Participants had normal or corrected to normal vision. They were screened for stereo acuity using the Netherlands Organization for Applied Scientific Research (TNO) stereo test (Chapter 3), or in an RDS based stereo screening test developed in the Binocular Vision Laboratory, University of Birmingham (Lugtigheid, 2012). The monetary compensation for participants' time was £6 per hour for behavioural studies, and £15 per hour for fMRI sessions.

# CHAPTER 3:

# Associative learning of second-order cues to shape from shading.

Shading patterns on a corrugated, textured and illuminated surface comprise two signals: first-order modulations of luminance (LM) and correlated second-order modulations of local luminance amplitude (AM). Human vision is sensitive to both of these signals, and their alignment is beneficial as a cue to shape perception (Baker, 1999; Dakin & Mareschal, 2000; Ellemberg, Allen, & Hess, 2006; Fleet & Langley, 1994; Schofield & Georgeson, 1999, 2003). Observers see LM and AM gratings aligned in-phase (LM+AM) as shaded corrugations, and anti-phase (LM-AM) as flat reflectance changes, when two mixtures are presented together in a plaid. First, we trained naïve observers with strong, trial-by-trial feedback. LM/AM mixes were presented in separate spatial locations and feedback consisted of a disparity defined corrugated surface superimposed on the in-phase pairing; anti-phase stimuli were paired with a flat surface. Performance improved to a maximum over the first hour of training, after which there was no further improvement. Even though this could suggest that rapid perceptual learning had occurred, such a rapid performance increase could also arise from associative learning or the learning of labels applied to already discriminable stimuli. When the feedback was flipped to reinforce the anti-phase pair as corrugated, the observers flipped their responses without any deterioration in performance. In the first part of training, AM thresholds to discriminate the phase relationship of LM/AM mixes reduced in the first hour and saturated for the rest of the sessions. This result and the sudden reversal in the second phase of training suggest that the performance benefit resulted from the association of the stimulus with labels rather than changes at a perceptual level.

## 3.1 Introduction

The visual system is able to construct the 3D shape of a surface from the luminance variations in a 2D image. However straightforward this construction might seem in a natural environment, when isolated, the physical cause of variations in luminance can be ambiguous. For example, at first sight, a corrugated matte surface with uniform reflectance under a point light source can produce a similar luminance pattern to a flat surface that has non-uniform reflectance (e.g. painted stripes). Even though the two luminance patterns are the same in this example, the 3D surface geometry of the actual surfaces differs, i.e. corrugated vs. flat. The visual system's ability to disambiguate the cause of luminance variations, and accurately construct a 3D surface shape (shape-from-shading) can be enhanced when second-order or texture cues are available.

Shape judgements from luminance patterns are shown to improve when a reflectance texture is added to the surface (Schofield *et al.*, 2006; 2010; Todd & Mingolla, 1983) . When a surface has highly variable reflectance at a small scale (a reflectance texture) and is illuminated, local variations in luminance amplitude (the difference between the luminance of light and dark elements) arise. These amplitude variations are second-order cues, similar and sometimes identical to the contrast modulations used by many to study second-order vision (Baker, 1999; Dakin & Mareschal, 2000; Ellemberg *et al.*, 2006; Schofield *et al.*, 2006; Schofield *et al.*, 2010). The spatial relationship between modulations of local mean luminance (LM) and local luminance amplitude (AM) seems to be useful for disambiguating the physical cause of luminance variations (Schofield *et al.*, 2006; 2010). When LM and AM are added in-phase (LM+AM), such that the luminance peaks will coincide with the highest amplitude, perceived depth is enhanced. Conversely, when LM and AM are negatively correlated (anti-phase, LM-AM), the impression of depth is reduced, although some depth is still perceived. Moreover, if an anti-phase mix is shown together with an in-phase mix on the opposing

diagonals of a plaid, the anti-phase pairing is perceived as a flat surface with non-uniform reflectance (Figure 3.1c), that is, strips of material change laid across the undulations formed by the in-phase pair. These findings suggest that when an AM signal is negatively correlated with an LM signal, the resulting anti-phase luminance pattern is ambiguous in the sense that it might be interpreted as a corrugated uniform reflectance surface or as a flat non-uniform reflectance surface.

When visual signals are ambiguous, the visual system requires additional information in order to estimate surface shape and other properties. When shading is presented alone, the visual system employs prior knowledge, i.e. assumes that the illuminant direction is known (light-from-above, Kleffner & Ramachandran, 1992) in order to interpret the ambiguous shading pattern. Further, co-occurring visual cues can aid the construction of a coherent percept. For example, the shape implied by variations in luminance can be disambiguated by contour information (Knill, 1992). Some cues dominate others, such that, for example, shading information can be completely overridden by stereoscopic disparity (Bülthoff & Mallot, 1988).

Performance changes related to exposure to visual stimuli can be understood in several different ways. Observers can adjust their performance statistically: cues occurring together may interact, or, similarly, they can modify their perceptual responses in response to external feedback. The improvement in performance can be at a perceptual level (perceptual learning, Fahle & Poggio, 2002, see also Chapter 4), or it can be at a higher cognitive level such as associative learning. The disambiguation of visual signals has been studied in the context of associative learning, where repetitive exposure to the ambiguous stimulus is paired with explicit feedback to favour a coherent percept. In the case of bistable stimuli, such as the Necker cube, observers can very quickly acquire a stabilised percept with contextual feedback contingent on location (Haijiang, Saunders, Stone, & Backus, 2006; van

Dam & Ernst, 2010). The light-from-above prior is shown to be adaptable after only a few hours of laboratory training with haptic or disparity feedback (Adams *et al.*, 2004; 2010). More recently, Harding, Harris, and Bloj (2012) have used realistically lit 3D Mach card stimuli to show that luminance gradient information alone is not sufficient to discriminate convex from concave stimuli. However, training observers using a short video at the start of the session was enough to disambiguate the gradient cue, producing reliable shape estimates. Here, we explore whether observers can learn to disambiguate the anti-phase combination of LM and AM cues to indicate a flat surface by using trial-by-trial training with disparity feedback.

I used a plaid configuration of in- and anti-phase LM/AM mixes to measure discrimination performance for the phase relationship between the two components of the mix. Observers' ability in this regard remained at chance level, even when plaids were presented for 1 second (**Experiment 1**). I then trained observers to discriminate in- and anti-phase gratings when presented separately (not in a plaid). The training used disparity feedback to promote discrimination: in-phase gratings were paired with disparity modulations depicting a corrugated surface presented in a feedback interval immediately following each test trial. Anti-phase gratings were paired with flat, zero disparity, feedback surfaces. I hypothesised that learning to see in-phase gratings as corrugated and anti-phase gratings as flat, when presented alone, would transfer to improved discrimination of the phase relationships when presented in a plaid configuration. Hence, after training, I expected observers to achieve above chance level performance with plaid stimuli even when there had been no exposure to plaids since the initial test phase of the experiment. To test for long-term consolidation of learning, the post-test was repeated after a 20-day interval without training (post-test 2, **Experiment 2**). Finally, to further investigate the nature of learning, reversed feedback was given to reinforce the opposite responses to those previously learnt by the observers in **Experiment 1**. Resistance to reversed training would suggest that

learning had taken place at a perceptual level, whereas an immediate reversal of responses would suggest associative learning and a cognitive strategy involving the labelling of already discriminable percepts.

## 3.2 Methods

### 3.2.1 Stimuli

I presented sinusoidal modulations of luminance (LM) and amplitude (AM) superimposed on noise elements. As described in Chapter 2, noise elements consisted of Gabor patterns. For training experiments, luminance and amplitude modulations were imposed on the noise texture to create a single LM/AM grating (Equation 2.3). These modulations were aligned either in-phase (LM+AM, Figure 3.1a) or anti-phase (LM-AM, Figure 3.1b). I used plaid stimuli consisting of two LM/AM mixes in the tests before and after training (Equation 2.6). The left and right components of the plaid were separated by 90° throughout the tests (Figure 3.1c), *i.e.* each plaid stimulus had an in-phase and anti-phase grating pseudo-randomly located in the right (45°) and left (135°) diagonals.

For the feedback in training, we used a flat disparity defined surface matching the anti-phase grating, and a corrugated disparity defined surface for the in-phase grating superimposed on the in-phase grating. Disparity was acquired using sub-pixel displacements of the Gabor micro-patterns (101×101 sub-pixels per display pixel) presented to each eye. In this magnified grid, Gabors were drawn with an offset from the centre (maximum offset = 50.5 sub-pixels). Then, sub-pixel grey levels were averaged to produce the grey level for each display pixel, hence translating sub-pixel shifts into subtle changes in the grey level of the display pixel. Overall, 101 templates were created offline, and these were used in addition to whole pixel shifts of the Gabors. This method allowed the presentation of both monocular shading cues and binocular cues in the same stimulus.

The spatial frequency of the modulations was kept at 0.5c/deg within the experiment. Gabor noise contrast was fixed at 0.1 in all trials. The contrast of the LM signal was set to 0.2 in all trials, whereas in the main experiments, AM modulation depths values differed in 5 fixed steps equally spaced between 0 and 0.4. To measure AM detection thresholds in the post training tests, we used 9 logarithmic steps of AM modulation depth: AM = 0.040 to 0.244.

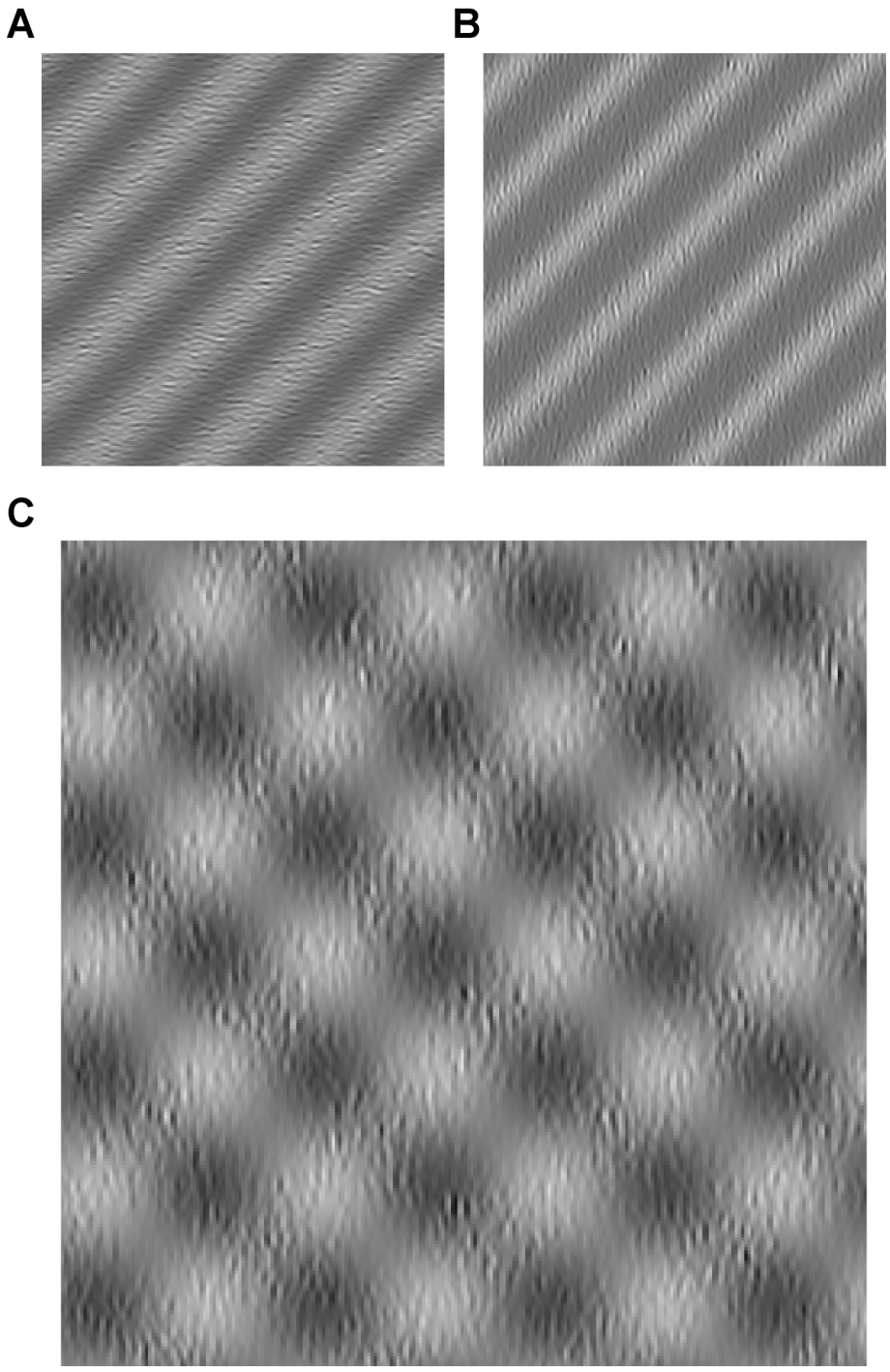



**Figure 3.1:** Stimulus examples. **(a)** A 45° orientated in-phase grating composed of LM and AM signals added to Gabor noise (amplitude modulation depth: 0.3) where pixels with the highest luminance values are superimposed with AM peaks (highest amplitude). **(b)** An anti-phase grating where LM troughs coincide with AM peaks. **(c)** Plaid stimuli consisting of an in-phase grating on the right diagonal and an anti-phase grating on the left diagonal.

In binocular displays, in-phase gratings were superimposed on a stereoscopically defined sinusoidal surface where the peaks of the 3D surface were aligned with the peaks of the shaded in-phase grating. The maximum depth amplitude in the stereo surfaces varied between d = 200 to 600 sub-pixels (100 sub-pixels = 0.039 degrees). These amplitude steps in disparity corresponded to the 5 steps of AM signal used in the shaded surface. The anti-phase grating present in the same trial was superimposed on a flat, zero disparity surface.

Luminance and amplitude modulations could be aligned either in-phase (LM+AM) or anti-phase (LM-AM) to form a single sine wave. Single sine wave gratings could be orientated at two different angles: 45 deg (left diagonal) and 135 deg (right diagonal) with respect to the vertical. During training trials, I used left diagonal sine waves, whereas pre- and post-training tests included both orientations. This use of different orientations allowed me to test stimulus-specific aspects of the training. In the test sessions before and after training, I used plaid stimuli consisting of in- and an anti-phase gratings presented orthogonally (Figure 1E, in-phase on the right diagonal, anti-phase on the left diagonal). Schofield and colleagues (2010) have shown that when in- and anti-phase gratings are presented in a plaid configuration, anti-phase is seen as flat; as opposed to separate presentation, where anti-phase grating can be seen as corrugated, although not as strong as in-phase grating.

Stimuli were generated in the frame store of a VSG2.5 graphics card (CRS Ltd, UK) with custom software written in C++ and version 8 of the VSG library. A 20-inch, 150Hz, Clinton Monoray monitor (CRS Ltd, UK) with a fast decay, yellow/green (DP104) phosphor was used for stimulus display with images intended for the two eyes presented in alternate video frames. This monitor was combined with FE-1 shutter goggles (CRS Ltd, UK), which were synchronised to the display by the VSG. The fast decay phosphor and low transmittance of the FE-1 goggles in their dark state greatly reduce cross talk between the eyes. The monitor's gamma nonlinearity was estimated using a ColourCal luminance meter (CRS Ltd,

UK) with a neutral density filter to dim the signal from the monitor, which otherwise overdrives the ColourCal device. Gamma was corrected using the VSG's lookup tables loaded with values determined by fitting a 4 parameter monitor model (Brainard *et al.*, 2002) to the luminance readings multiplied by the inverse of the attenuation factor of the neutral density filter. The stimuli subtended 13 by 13 degrees of visual angle (512x512 pixels). Binocular stimuli were not occluded in any way because of the 44 mm wide apertures on the goggles.

### 3.2.2 Participants

Six postgraduate researchers from the University of Birmingham participated in the experiments (mean age = 28 ± 4 years). Participants were all right handed and had normal or corrected to normal vision. They were screened for stereo acuity using the Netherlands Organization for Applied Scientific Research (TNO) stereo test. Observers were naïve to the purpose of the experiments; they gave written informed consent, and were paid £6 per hour of participation. The study was conducted under a protocol ethical approval by the University's ethics committee.

### 3.2.3 Procedure

In the test trials, stimulus presentation was limited to 1 second, but presentation time was unlimited in the training trials. In every trial, an in-phase and an anti-phase stimulus appeared either 1.5 deg above or below a fixation marker located in the centre of the screen. Throughout the study, top and bottom stimuli were orientated at the same angle. In all of the experiments, viewing distance was fixed to 0.6 m with a chin rest that also supported the FE-1 goggles. Experiments took place in a dark room where the experimental monitor was the only light source. The observers viewed the stimuli through the goggles even when the

presentation was monocular (pre- and post-training tests). This ensured the same mean luminance in all parts of the experiment.

The participants indicated whether the top or bottom stimulus seemed more corrugated in depth using a CB3 response box (CRS Ltd, UK); the observers were instructed to toggle a single key upwards if choosing the stimulus above fixation, and downwards for the stimulus below fixation.

The pre-training test consisted of 100 trials (2 orientations x 5 AM levels x 10 repetitions) and participants took approximately 40 minutes to complete this session. No feedback was provided at this stage.

During training, only LM/AM pairs on the right diagonal were used (45 deg). Participants first viewed the stimulus without any stereoscopic cues and then made their response. Feedback then appeared in the form of additional stereoscopic modulations superimposed onto the LM/AM pairs. Feedback reinforced the in-phase stimuli as the correct response in every trial (**Experiment 1**) by imposing stereoscopically defined corrugations onto this stimulus pairing; Schofield and colleagues (Schofield *et al.*, 2006; Schofield *et al.*, 2010) have previously shown that the LM+AM pairing is normally seen as more corrugated than the LM-AM pairing. After the feedback (duration = 1 s), participants pressed a key to start the next trial. Training continued for five days (800 trials per day), and was followed directly by a post-training test, which was identical to the pre-training session. Data from the first 400 trials of the first training session (IE, initial exposure) are reported separately in the results to demonstrate subsequent changes in performance with training.

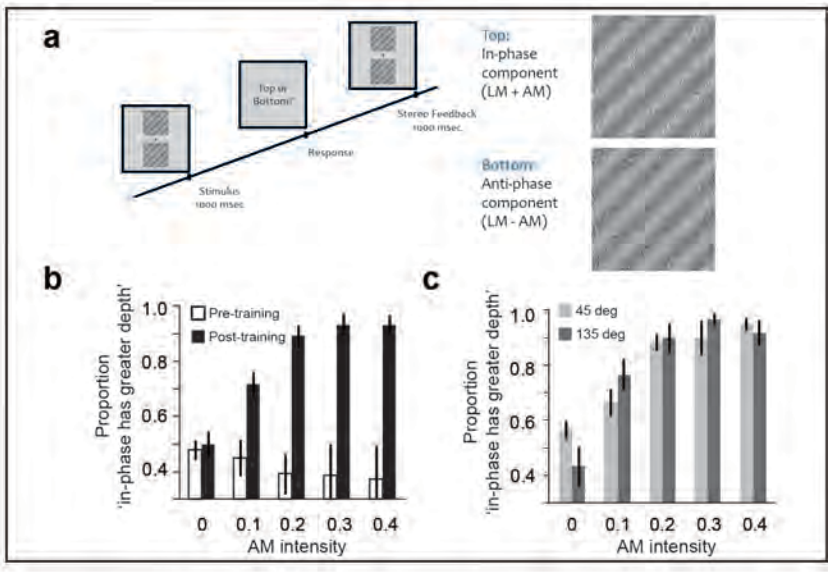
Twenty days after the post-training test, the observers were asked to repeat the test, followed by a reversed training session on the twenty-first day. In the reverse training trials, feedback was flipped: Corrugated stereo surfaces were superimposed on anti-phase stimuli,

while in-phase stimuli were paired with flat surfaces. This reversal tested the unlearning of the previously trained state, and helped me to explore the nature of learning that took place in the first part of the training. The participants continued reversed training until they reached the level of performance obtained in the initial training phase, but with 'correct' responses recorded when the anti-phase stimuli were chosen as more corrugated. Participants required between one and five reverse training sessions. Following the reverse training, a further post-test session was conducted. Finally, the observers undertook a further experiment to determine their AM detection thresholds following training. This session was based on the methods of Schofield & Georgeson (1999).

## 3.3 Results

### 3.3.1 Experiment 1: Disparity feedback training with single gratings

In Experiment 1, I examined accuracy in discriminating the phase relationships (in- and anti-phase) between LM/AM mixes. During training, observers were encouraged to see a 45 deg orientated anti-phase grating as flat with trial-by-trial feedback; in-phase gratings were seen as corrugated. On the feedback screen, a disparity defined corrugated surface was superimposed to the in-phase grating, while the anti-phase grating had a flat surface superimposed. Pre- and post-training tests consisted of plaid stimuli: an orthogonal combination of in- and anti-phase components orientated at 45 and 135 degrees. In all trials in **Experiment 1**, accuracy was assessed by counting "in-phase has greater depth" as correct.




igure 3.2: Plaid test results in Experiment 1 (a) Cartoon showing the training procedure used ı the experiment. (b) Data from pre- and post-training tests with plaids is shown. Before raining (white bars), mean performance is around chance but this increases after training olack bars). (c) Data from the post-training test shown separately for the two possible rientations of the LM+AM. Light grey bars indicate correct responses for trained orientation eft diagonal, 45 deg), and dark grey bars indicate responses for the untrained orientation 135 deg). All bars (b, c) show mean data from six participants, error bars indicate standard rror of the mean.

Figure 3.2b shows the proportion of correct values before (white bars) and after (black bars) training (mean of six observers, averaged over trained and untrained orientations). Before training, the mean proportion correct was around chance level, showing that participants were not able to see depth in LM/AM mixes or to discriminate them on the basis of a difference in perceived depth. After five days of training, all participants showed a benefit from training, with a mean performance of 93 per cent correct for the highest AM level (St. dev = 0.08). Repeated measures ANOVA showed a main effect of training (before and after training, $F_{1, 5} = 21.55$, $p < .01$) and AM level in the stimulus ($F_{4, 20} = 4.39$, $p < .05$). As might be expected, this performance increment was more strongly emphasised for the stronger AM signals (these being more visible) as shown by the interaction: $F_{4, 20} = 11.43$, $p < .001$.

The pre-training data (white bars, Figure 3.2b), seems to indicate that on average, naïve participants have a tendency to respond anti-phase has greater depth. However, at the group level (n=6), a repeated measures ANOVA show that these results are not significantly different from chance level performance ($F_{1,5}$ = 1.49, p = .28). Out of 6 participants, 3 participants' performance is at chance level, and the rest is showing a tendency towards anti-phase responses. Overall button responses show that 40 percent of the time participants pressed the button on the right (St. dev = 8.2 percent).

When I analyse the results separately for in-phase is on the *right* and on the *left* oblique, the group average does not show a trend towards an orientation. But two participants show different results, one having higher performance for left oblique and the other for the right oblique. This might be explained as the participant attending to a single component of the plaid (e.g. always the right oblique).

During training, participants were only exposed to 45 deg sine wave gratings, whereas the pre- and post-tests included in- and anti-phase gratings orientated orthogonally in a plaid configuration. In Figure 3.2c, the light grey bars indicate the proportion of correct values after training for the trained orientation (i.e. in-phase component of the plaid is orientated at 45 deg) and dark grey bars show data for untrained stimuli (i.e. in-phase component is orientated at 135 deg). I looked for an orientation-specific effect of training, however a repeated measures ANOVA showed no significant difference between trained and untrained orientations ($F_{1,5}$ = .034, p = .862), suggesting a full transfer to 90 deg rotation from trained to untrained orientation.

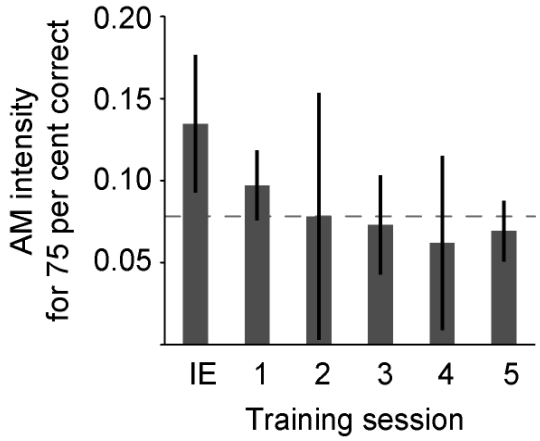



**Figure 3.3:** Performance during training in Experiment 1. Bars represent thresholds (75% correct) for AM signal intensity required to discriminate in- and anti-phase gratings in each training session and during the initial exposure phase of the first training session (first 400 trials). Gray dashed line indicates mean threshold (AM = 0.08) over the training sessions. All bars show mean data from 5 participants, error bars indicate standard error of the mean.

A cumulative Gaussian function was fitted to proportion correct data collected during training, and thresholds at 75% correct were acquired where available. Data from one participant, IF, where training thresholds were outside 2 standard deviations of the mean, were discarded from this analysis. Figure 3 shows AM thresholds for discriminating the phase of LM/AM mixes during training sessions (mean across five participants). Overall, the mean threshold in all training sessions (AM = 0.08 ± 0.01) was reached in the second session and saturated after this point. There was no significant training effect, as suggested by a repeated ANOVA on training sessions ($F_{4, 12} = 2.20$, $p = .131$). Learning took place very quickly and most of the improvement was seen within the first hour of training.

### 3.3.2 Experiment 2: Consolidation and then reversal of learning

In **Experiment 2**, first I explored the lasting effects of training and next, the nature of learning, by employing a reverse training paradigm. Twenty days after the first post-training test (**Experiment 1**), the six participants were invited to the lab for a follow up session of testing.

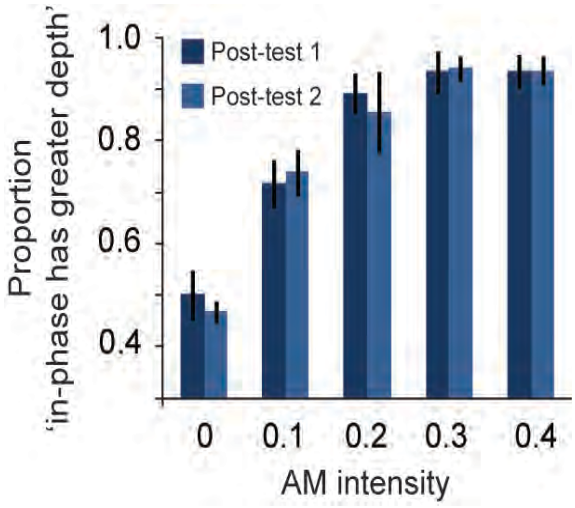



**Figure 3.4:** Performance in plaid test after a 20-day break (Post-test 2, **Experiment 2**) is compared to post-training test immediately after the training (Post-test 1, **Experiment 1**). All bars show mean data from 6 participants, error bars indicate standard error of the mean.

In Figure 3.4, light blue bars (Post-test 2) show the proportion of correct values after the 20-day break (mean of six observers, averaged over trained and untrained orientations). Data from **Experiment 1** is also shown for comparison (dark blue bars, Post-test 1). The benefit of training was still clearly observable after the 20-day interval: performance was no different from the immediate post-training test reported in **Experiment 1** (Repeated measures ANOVA, $F_{1,5} = .058$, $p = .819$).

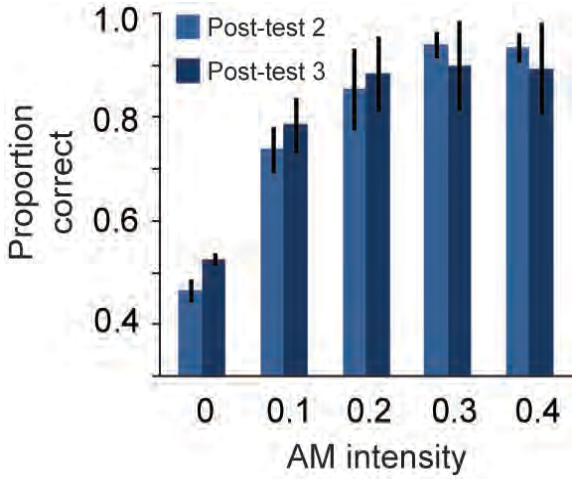



**Figure 3.5:** Post-test data before and after reversed training (**Experiment 2**). Proportion correct values before reversed training (Post-test 2) indicate matches for 'in-phase component has greater depth'; where after reversed training (Post-test 3), responses for 'anti-phase component has greater depth' were represented as correct. Mean data from 6 participants is shown; error bars indicate standard error of the mean.

Next, we ran a reverse training regime in which everything was the same as in **Experiment 1**, except for the feedback. In contrast to the previous training sessions, anti-phase components were reinforced as being corrugated by superimposing a corrugated stereo surface onto them during the feedback phase of each trial. Proportion correct data are shown in Figure 5 for the second post-test training session (light blue bars, Post-test 2) conducted before the reverse training was applied, in which "in-phase component has greater depth" was counted as correct, and for a final post-test session (dark blue bars, Post-test 3) conducted after the reversed training, and in which "anti-phase component has greater depth" was counted as correct. Reverse training with single gratings took effect immediately within the first session, and this effect was carried on to the post-test with plaids (Post-test 3).

## 3.4 Discussion

Constructing the 3D surface shape from the luminance variations in a 2D luminance pattern is a complex task. Luminance variations can be directly related to the orientation of the surface, with the brightest parts of the image representing surfaces orientated towards he light. On the other hand, changes in the reflectance properties of a flat surface might cause luminance variations, and these might look like the shading patterns produced by a corrugated surface with uniform reflectance. One way in which the visual system might disambiguate such luminance changes is by using the second-order (AM) information carried by reflectance textures. Schofield *et al.* (2006; 2010) have shown that the phase relationship between first-order luminance modulations and second-order AM signals is a useful cue to discriminate luminance changes due to shape from those due to material variations. In-phase combinations (LM+AM) cue for changes in surface shape, whereas anti-phase (LM-AM) combinations cue for material changes in a flat surface. They have also shown that naive observers are able to discriminate the phase relationship between first- and second-order cues, given long presentation times. Here I have shown that if the presentation time is restricted, observers cannot discriminate in-phase from anti-phase (pre-test, Experiment 1).

Observers learned to discriminate between in- and anti-phase gratings very quickly in this study. Although previously reported studies show evidence for fast visual perceptual learning (Fahle, Edelman, & Poggio, 1995; Karni & Sagi, 1993; Seitz *et al.*, 2005), I cannot argue that learning in Experiment 1 took place at a perceptual level. Firstly, perceptual learning in visual tasks is often identified by specificity to some low-level feature of the training stimulus, such as orientation, spatial frequency, or visual field position (Fahle *et al.*, 1995; Fiorentini & Berardi, 1981; Karni & Sagi, 1991). Here, I observed that training with 45 degree orientated gratings transferred to 135 degree orientated components of plaids (post-test 1, Experiment 1). Next, in perceptual learning examples, initial fast learning is followed by slow enhancement between days of training, which is frequently explained by

consolidation of learning with sleep (Karni, Tanne, Rubenstein, Askenasy, & Sagi, 1994; Stickgold, James, & Hobson, 2000). However, I did not observe improvement between training sessions. Rather, discrimination sensitivity saturated during the second day of training, showing a closer resemblance to associative learning behaviour based in an already discriminable but previously un-labelled stimulus set.

If observers learned to associate the 'correct' feedback with the in-phase grating, then I would expect them to be able to flip this behaviour immediately in a reverse feedback design, whereas for perceptual learning, the 'unlearning' phase would take more time. Here, observers were capable of adjusting their responses within a very short period in a reversed feedback design. Moreover, their discrimination sensitivity remained intact when the feedback was reversed to indicate that the in-phase component should be 'seen' as flat. This quick reversal in discrimination performance suggests that learning was not at a perceptual level but relied on a cognitive process similar to naming, i.e. associating in-phase (or anti-phase) with the correct response.

The observers retained the improvement from the first training phase, and they were able to interpret LM/AM mixes after a period of 20 days during which time, presumably, they had no direct exposure to the test stimuli. It has been reported that training effects can last for several last years when learning is at a perceptual level (Qu, Song, & Ding, 2010).

In summary, I have shown that observers learn to discriminate phase relationships with short presentation times rapidly when reinforced with trial-by-trial disparity defined feedback. Performance improves to a maximum and saturates within the first hour of training (**Experiment 1**). When feedback is reversed to reinforce the anti-phase component, observers' responses reverse rapidly without any deterioration in performance (**Experiment 2**). This training is more likely to be characterised as associative learning.

# CHAPTER 4:

# Perceptual learning of second-order cues for layer decomposition.[1]

---

Luminance variations are ambiguous: they can signal changes in surface reflectance or changes in illumination. Layer decomposition—the process of distinguishing between reflectance and illumination changes—is supported by a range of secondary cues including colour and texture. For an illuminated corrugated, textured surface, the shading pattern comprises modulations of luminance (first order, LM) and local luminance amplitude (second order, AM). The phase relationship between these two signals enables layer decomposition, predicts the perception of reflectance and illumination changes, and has been modelled based on early, fast, feed-forward visual processing (Schofield *et al*., 2010). However, while inexperienced viewers appreciate this scission at long presentation times, they cannot do so for short presentation durations (250 ms). This might suggest the action of slower, higher-level mechanisms. Here we consider how training attenuates this delay, and whether the resultant learning occurs at a perceptual level. We trained observers to discriminate the components of plaid stimuli that mixed in-phase and anti-phase LM/AM signals over a period of five days. After training, the strength of the AM signal needed to differentiate the plaid components fell dramatically, indicating learning. We tested for transfer of learning using stimuli with different spatial frequencies, in-plane orientations, and acutely angled plaids. We report that learning transfers only partially when the stimuli are changed, suggesting that benefits accrue from tuning specific mechanisms, rather than general interpretative processes. We suggest that the mechanisms which support layer decomposition using second-order cues are relatively early, and not inherently slow.

---

[1] This chapter has been published as: Dövencioğlu, D. N., Welchman, A. E.,. Schofield, A.J.; Perceptual learning of second order cues for layer decomposition, Vision Research, V. 77, 25/1/2013, P. 1–9, doi: 10.1016/j.visres.2012.11.005. Main text and figures were kept as in the manuscript. All authors contributed to the conceptualisation of the experiment and writing of the paper, DND collected data and ran the analyses.

## 4.1 Introduction

Interpreting the luminance variations in an image in terms of their underlying physical cause poses a significant challenge to the visual system. Specifically, luminance variations in an image can have two distinct causes: (i) they might arise from variations in 3D surface geometry so that different portions of a surface are differentially illuminated by the light source(s) and/or (ii) they might arise from variations in the surface albedo, such as different textures or paint on the surface. Somehow, the visual system must parse changes caused by the illumination with respect to the 3D surface (shape-from-shading) from changes in surface reflectance properties. This process is known as layer decomposition (Kingdom, 2008) or intrinsic image extraction (Barrow & Tanenbaum, 1978) and it can be achieved by considering the relationship between luminance variations and a range of other cues, including colour (Kingdom, 2003) and, as we review below, second-order cues that arise in objects with a textured surface.

A potentially informative cue to layer decomposition is provided by the spatial relationship between changes in local mean luminance (LM) and local variations in the range of luminance values that arise from an albedo texture: Local luminance amplitude (AM; Schofield *et al*., 2006). In particular, when the illumination varies across an albedo textured surface, changes in local mean luminance (LM) are positively correlated with changes in local luminance amplitude (AM). Adding an albedo texture to a shaded surface, such that LM and AM correlate positively (in-phase; LM+AM), enhances the impression of depth (Todd & Mingolla 1983; Schofield *et al*, 2006, 2010; compare Figures 4.1A and B). Moreover, if LM and AM are negatively correlated (anti-phase; LM-AM) the impression of depth is reduced (compare Figure 1D with A). If both relationships are present in a plaid configuration, the in-phase pairing appears as a shaded undulating surface whereas the anti-phase pairing appears as a flat material change (Schofield *et al*., 2006, 2010; Figure 4.1E). The enhanced shape-

from-shading in the in-phase case may be due to improved layer decomposition owing to the information provided by the relative phase of the AM cue.

The changes in local luminance amplitude described above are, mathematically, closely related to the contrast modulations typically used to study second-order vision. The human visual system is known to be sensitive to second-order signals and it is thought that they are detected separately from first-order cues (Baker, 1999; Dakin & Mareschal, 2000; Ellemberg *et al.*, 2006; Fleet & Langley, 1994; Schofield & Georgeson, 1999, 2003). First- and second-order information is correlated in natural images (Johnson & Baker, 2004) but the sign of this correlation varies (Schofield, 2000), suggesting that contrast/amplitude modulations are informative by virtue of their relationship with luminance variations.

Building on the physiological work of Zhou and Baker (1996), Schofield *et al*. (2010) developed the shading channel model in order to explain the role of AM in layer decomposition. In this model, LM and AM are initially detected separately and then recombined in an orientation / frequency specific additive sum that broadly mimics Zhou and Baker's (1996) envelope neurons. In-phase pairings sum to produce an enhanced output (greater perceived depth), whereas anti-phase pairings subtract, weakening the output / depth percept. However, AM components are given a relatively low weighting at the summation stage, such that their effect on single LM components is marginal. A competitive gain control mechanism working across orientations produces the dramatic scission found for plaid stimuli. The model has been used to describe a range of psychophysical results (Schofield *et al.*, 2010; Sun & Schofield, 2011), has been applied directly to natural images (Schofield *et al*., 2010), and has been used as the basis for a machine vision system for layer decomposition (Jiang, Schofield, & Wyatt, 2010).

The shading channel model relies on relatively low-level mechanisms, which might be considered comparable to the envelope neurons found in area 17/18 of the cat visual cortex (Zhou & Baker, 1996). Therefore we would expect layer decomposition based on LM and AM mixtures to be automatic and fast acting. However, while the cues are effective for naïve participants at relatively long presentation times (circa 1s), anecdotally they see no difference between LM+AM and LM-AM at short presentation times (250ms): even in the more robust plaid condition. We confirmed this failing in Experiment 1. Thus, layer decomposition is rather slow, implying the use of attentional mechanisms or at least multiple stages of processing beyond those implied by the shading channel model. However, it is also possible that early mechanisms exist for layer decomposition based on LM/AM mixtures but that they are either (i) relatively underused or (ii) not well engaged by plaid stimuli that are too different from everyday experience to allow fast layer decomposition. If this were the case, we would expect performance to improve with training. Furthermore, if low level mechanisms, such as those described in the shading channel model, are critical to the task, we would expect any benefits of learning to follow the stimulus-specific pattern observed in perceptual learning studies (Manfred, Fahle, & Morgan, 1996; Jeter, Dosher, Liu, & Lu, 2010; Jeter, Dosher, Petrov, & Lu, 2009).

Perceptual learning has been explored in various visual contexts. For instance, through repetitive training, humans improve in their ability to: detect luminance contrast (Fiorentini & Berardi, 1981; Sowden, Rose, & Davies, 2002), perform Vernier tasks (Fahle & Morgan, 1996; Spang, Grimsen, Herzog, & Fahle, 2010), discriminate between orientated stimuli (Jeter *et al.*, 2009), and discriminate between textures (Karni & Sagi, 1991). Such perceptual learning is often reported to be specific to ancillary stimulus features such as retinal location, spatial frequency or orientation. For example, Fiorentini and Berardi (1980) report rapid learning in a phase discrimination task where observers are asked to discriminate the two types of composite sinusoidal grating. This learning effect was specific

for the trained orientation and did not transfer across 90 deg stimulus rotations. We might expect to see similar cue-specific learning in the case of the relative phase discrimination required in our layer decomposition task.

We used a perceptual learning paradigm to examine the improvement in layer decomposition associated with LM/AM plaids at short presentation times. We then tested for transfer across stimulus dimensions as a marker for perceptual learning. We hypothesised that training would enable layer decomposition at short presentation times. Moreover, we tested the generalisation of learning that results from training, considering the stimulus dimensions of orientation and spatial frequency. In particular, we trained naïve observers to discriminate the phase relationship of LM and AM signals in briefly-presented plaid stimuli. We then conducted three tests to probe the learning, examining the transfer of depth discriminations based on LM+AM to different stimulus rotations (Experiment 1), different stimulus spatial frequencies (Experiment 2), and plaids that differed in the relative orientation of their compositions (Experiment 3). Poor transfer across these stimulus manipulations would indicate perceptual learning, whereas full transfer would suggest the learning of cognitive strategies such as labelling based on the LM/AM relationship regardless of the percept formed by the stimuli.

In Chapter 3, I have found evidence that naïve observers learnt to associate in-phase gratings with positive feedback, which indicates greater depth. Even though the results showed that naïve observers could be trained to discriminate the phase relationships, it is not sufficient to prove that the observers learnt to use first and second order signals to infer shape from shading. Instead, it might be explained with a symbolic association of a type of the stimulus with the feedback, in other words, they might have come up with a strategy to succeed in the task without any improvement in the sensitivity to discriminate the phase relationship of the signals. In Chapter 4, I am aiming to find evidence that naïve observers can

learn to benefit from LM/AM mixes' phase relationships to infer shape from shading. To do this, I provide intermittent information during training to induce a change in sensitivity."

## 4.2 Methods

### 4.2.1 Stimuli

To isolate shading and illumination cues from additional sources of shape information (e.g. boundaries, occlusions or recognisable object outlines), we imposed sinusoidal modulations of luminance (LM, Figure 4.1A) and amplitude (AM, Figure 4.1B) on binary noise textures (see Schofield, Hesse, *et al*., 2006 for full details of the stimulus preparation method). Luminance and amplitude modulations could be aligned either in-phase (LM+AM, Figure 4.1C) or out-of-phase (LM-AM, Figure 4.1D) and could be orientated at four different angles (22.5, 67.5, 112.5 and 157.5 deg) with respect to the vertical. Plaids were formed from combinations of an in-phase and an anti-phase grating presented at different orientations (Figure 4.1E, in-phase on the right diagonal and anti-phase on the left diagonal). The orientations of the plaid components were orthogonal except in Experiment 3.

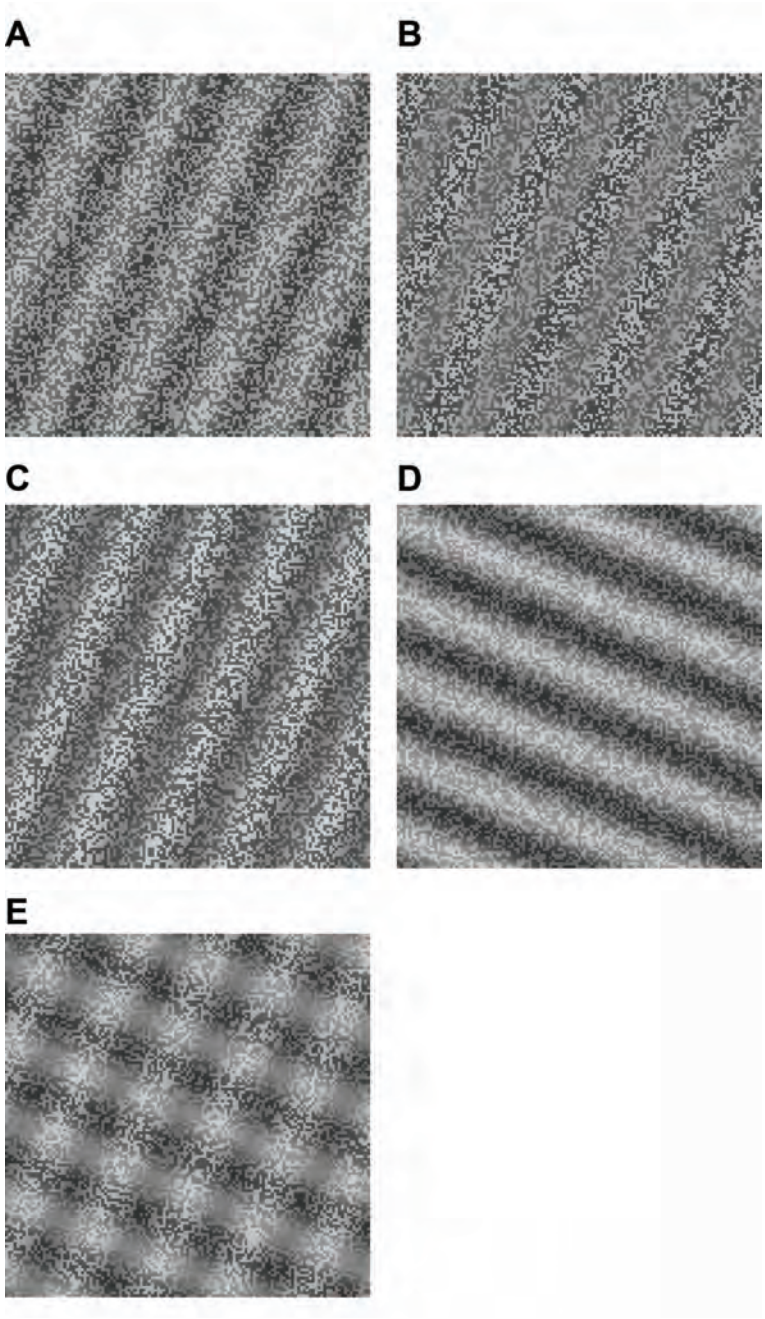



Figure 4.1: Stimulus examples (A) LM-only: A 45 deg oriented sine wave luminance grating added to binary noise (B) AM-only: Amplitude modulated binary noise pattern (modulation depth = 0.40). (C) An in-phase composite grating where peaks of LM (highest luminance) and AM (highest amplitude) are superimposed. (D) An anti-phase grating where LM troughs are superimposed with AM peaks. (E) A plaid consisting of an in-phase grating on the right diagonal (LM+AM) and an anti-phase grating on the left diagonal (LM-AM).

The contrast of all LM components was set to 0.2 in all experiments. The AM values in the main experiment were chosen from an interval that would bound individual AM detection thresholds based on previous work (Schofield & Georgeson, 1999) and a pilot study: AM = 0.040 to 0.244 in 5 logarithmic steps. The spatial frequency of the modulations was 0.5 c/deg, except in Experiment 2. The noise contrast was fixed at 0.1 and new noise samples were generated for each trial. The stimuli subtended 6.5 by 6.5 degrees of visual angle (256x256 pixels).

There were two training sets. Half of the participants were trained on Set 1 (22.5 deg and 112.5 deg plaids) and the other half on Set 2 (67.5 deg and 157.5 deg plaids); the allocation of participants to training sets was random. We describe the plaids with respect to the orientation of the LM+AM component; thus, the 22.5 deg plaid contained the LM+AM component on the left diagonal, orientated 22.5 deg, and an LM-AM component on the right diagonal, orientated 112.5 deg. Having two training sets allowed us to examine the transfer of learning to un-trained but otherwise similar stimuli. The component orientations were chosen to allow us to reasonably ask observers to judge whether the left or right tilted component was more corrugated.

Stimuli were generated in the frame store of a VSG graphics card (CRS Ltd, UK) with custom software written in C++ and were presented on a ViewSonic P225f monitor at a refresh rate of 160 Hz. The monitor's gamma nonlinearity was estimated using a ColourCal luminance meter (CRS Ltd., UK) and corrected using the VSG's lookup tables.

### 4.2.2 Participants

A total of 12 postgraduate students from the University of Birmingham took part in the study. Six participants (mean age = 29 ± 6 years) were tested to assess baseline measurements using the test stimuli of Experiments 1 and 3 without any training. Another set of six participants (mean age = 25 ± 3 years) undertook the plaid training followed by the three test experiments. One of the participants showed a reverse learning effect during the first training session. She gave the opposite responses to those that were reinforced by the feedback; this observer was excluded from further study and replaced by a new participant, given that we set out to study perceptual learning and this participant was unable to benefit from the feedback we provided. Participant GM was excluded from analysis in Experiment 3, as we could not estimate thresholds for him in two out of the three conditions. All of the participants were naïve to the purposes of the experiment, and had normal or corrected to normal vision. Participants gave written informed consent and were paid £6 per hour. They were debriefed after the last test session. The work was subject to ethical review prior to experimentation (University of Birmingham STEM ethics committee).

### 4.2.3 Procedure

The stimulus duration was 250 ms for each condition. Stimuli could appear in one of two locations either 1.5 deg above or below the fixation marker; the other location was filled with a binary noise pattern. This manipulation prevented the build up of afterimages, which can selectively reduce the visibility of the LM cue. The participants viewed the stimuli in a darkened room at a viewing distance of 0.6 m. Head position was stabilised with a chin rest. Participants indicated whether the right or the left oblique seemed more corrugated in depth by pressing one of two keys on a button box (CB3, CRS Ltd, UK). Given previous results (Kingdom, 2003, Schofield *et al.*, 2006; 2010), responses for "In-phase component has greater depth" were counted as hits. Symbolic, intermittent feedback was given: specifically, at the end of each block of 20 trials, observers were shown their per cent correct score for the last block via a written message on the display. The first 200 trials on Day 1 of training were

analysed separately to represent the performance levels for initial exposure to the stimuli. Training continued for five days (1000 trials per day). In cases where overall accuracy was below 75% correct after five days (participants GM, JB, and AAM), training continued until an overall accuracy of 75% correct was achieved (longest duration: 10 days). Trained participants undertook three experiments to test their newly learnt abilities in the days immediately following training.

### 4.2.4 Post-Training Experiments

After training, participants made forced-choice judgements ("Which orientation in the plaid is more corrugated in depth?") on three sets of test stimuli: 1) stimuli were orthogonal plaids differing from the training stimuli by a 45 deg rigid rotation (Figure 4.1E): participants who were trained on Set 1 stimuli were presented with Set 2 stimuli to establish performance on untrained stimuli, and vice versa; 2) spatial frequency (Figure 4.4A, s.f.=2 or 4 c/deg, angle between components 90 deg); or 3) shear (Figure 4.5A, angle between the LM+AM and LM-AM components varied while still allowing left / right judgements to be made). No feedback was given during the test phase. Experiment 1 took place one day after the final day of training; Experiment 2, 9-13 days post training; and Experiment 3, 15-19 days post training.

## 4.3 Results

### 4.3.1. Performance during training

As a first analysis, we considered the efficacy of the training paradigm on participants' behavioural performance. In particular, we considered trial-by-trial performance during training for the three observers who completed the training regime within five days (Figure 2). We calculated the proportion correct (later converted to per cent correct) as a running average, based on a window of the preceding 100 trials (1 = correct = "in-phase component has most depth2, 0 = incorrect) for each day of training. (Performance on each day is described from the 100th trial, therefore there are gaps in the traces between each day.) On

the first training day, performance improved up to a peak of around 80% correct and but fell dramatically in the last 200 trials, perhaps due to fatigue or reduced participant confidence due to a run of weak stimuli. Such dips occur elsewhere in the data and are not confined to the last trials of a session. Performance at the start of Day Two was above that at the outset of Day One but below the Day One peak. Performance on subsequent days showed progressively increasing initial performance with smaller lapses from the previous day. By Day Five, initial performance was consentient with the overall mean. The remaining observers showed similar training performance, but were slower to reach the asymptotic performance and showed greater initial drops in performance.

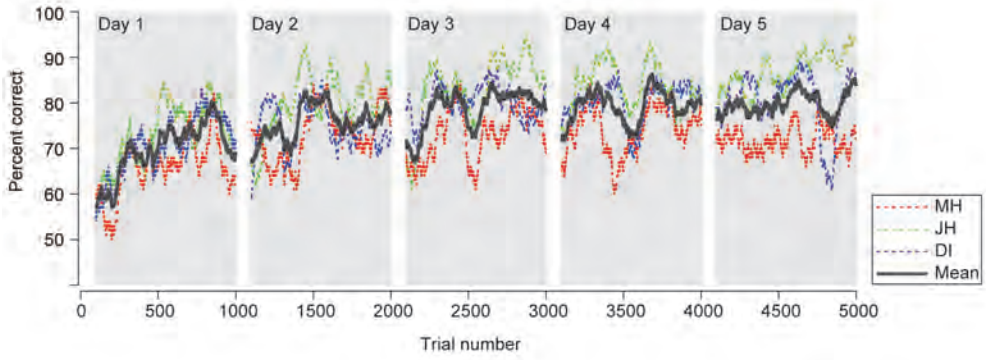
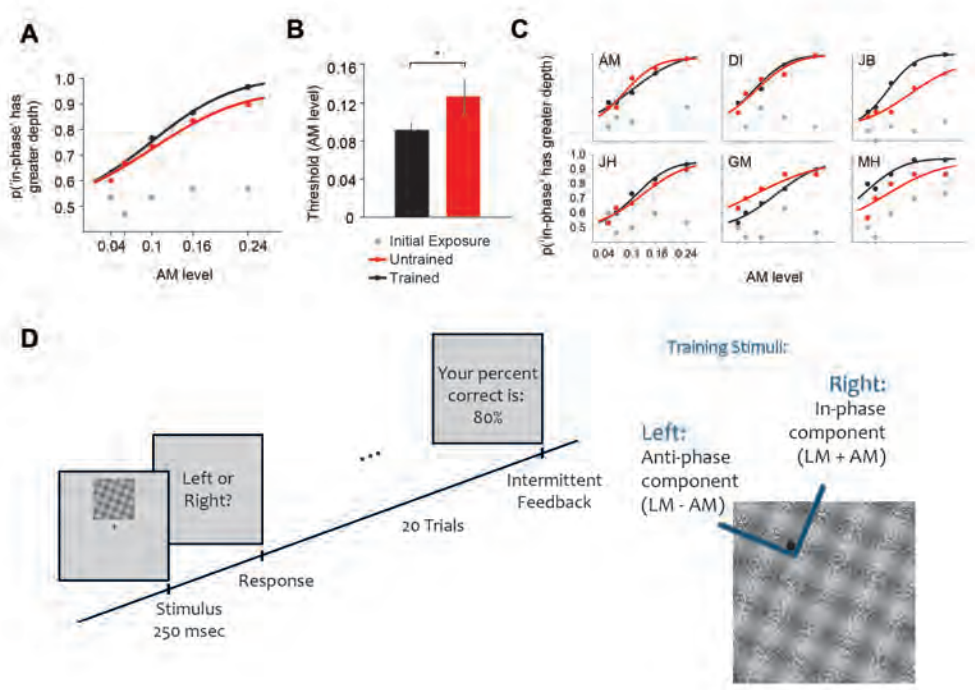



Figure 4.2: The time course of training. Lines show trial-by-trial percent correct scores calculated over the preceding 100 trials starting from the 100th trial in each session. Accuracy was assessed relative to 'in-phase has greater depth' this being deemed the correct response. Gray boxes show each day's training. Green, blue and red traces show results for participants MH, DI and JH respectively. The black line shows the mean performance of the three participants. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

### 4.3.2 Experiment 1: Specificity for orientation.

Following the training phase, we examined whether improvements in depth judgements were specific to the trained stimuli. Experiment 1 tested for transfer between different stimulus orientations. In particular, we tested for the transfer of performance between the trained and untrained stimulus sets, which differed in overall orientation by 45 degrees.




igure 4.3: Plaid training and Experiment 1 – specificity for trained orientation: (A) Mean sychometric functions (6 participants), grey data points indicate per cent correct for initial xposure to plaids, black symbols (line) data (fit) for trained stimuli, red symbols (line) data fit) for untrained stimuli. (B) Thresholds (75% correct) for AM level required to discriminate hase relationship for trained (black bar) and untrained (red bar) orientations. (C) Individual sychometric functions, each graph represents data from a single participant on three onditions as described in (A). (D) Cartoon showing the training procedure used in the xperiment.

Figure 4.3A shows the per cent correct values and a fitted cumulative Gaussian function (mean of six new observers; fits obtained using psignifit version 2.5.6; Wichmann and Hill, 2001) for the trained and untrained stimulus sets. Performance during initial exposure to stimuli (grey data points, see also Supplementary Figure 1) is around chance, suggesting that untrained observers cannot differentiate LM-AM from LM+AM at short presentation durations. This was also confirmed for six new, untrained observers (Supplementary Figure 2). However, after training (black and red lines), observers were able to determine that the LM+AM component had a greater corrugation than the LM-AM component, with average performance reaching up to 96±4% correct at the highest AM level for trained stimulus orientation (black dots). The difference between thresholds for trained and untrained stimuli shows that the learning effect is specific to the trained stimulus set

(Figure 2B). In particular, thresholds for the trained stimuli (black bar, AM = 0.09±0.01) were significantly lower than the untrained stimulus thresholds (red bar, AM = 0.13±0.02; $t(5)=2.11$, $p=.044$). However, post-training thresholds for untrained stimuli were better than pre-training thresholds (which were not measureable), suggesting a partial transfer of training. Indeed two participants (GM and AM) performed slightly better on the untrained stimuli, suggesting complete transfer for these individuals.

### 4.3.3 Experiment 2: Specificity for spatial frequency

In Experiment 1, the full benefits of training were specific to the trained orientation, with only partial transfer to 45 deg rigid rotations. Here we test transfer along another stimulus dimension: spatial frequency. Detection thresholds for LM and AM cues depend on spatial frequency in different ways (Schofield & Georgeson, 1999) but relative and absolute sensitivity for the two cues is approximately equal at 0.5 and 2 c/deg in the presence of binary noise. That is, LM (or AM) thresholds are similar at the two frequencies and the ratio of LM to AM sensitivity is also similar at the two frequencies. In this experiment, we tested whether training at 0.5 c/deg transfers to 2 c/deg plaids (Figure 4.4A, top image). We also tested for transfer to a higher spatial frequency (4 c/deg; Figure 4.4A, bottom image), where AM sensitivity is known to be relatively weak. Only the strongest AM level (0.244) was used in this experiment. The other stimulus dimensions were the same as the training stimuli and the task was the same as in the general methods.

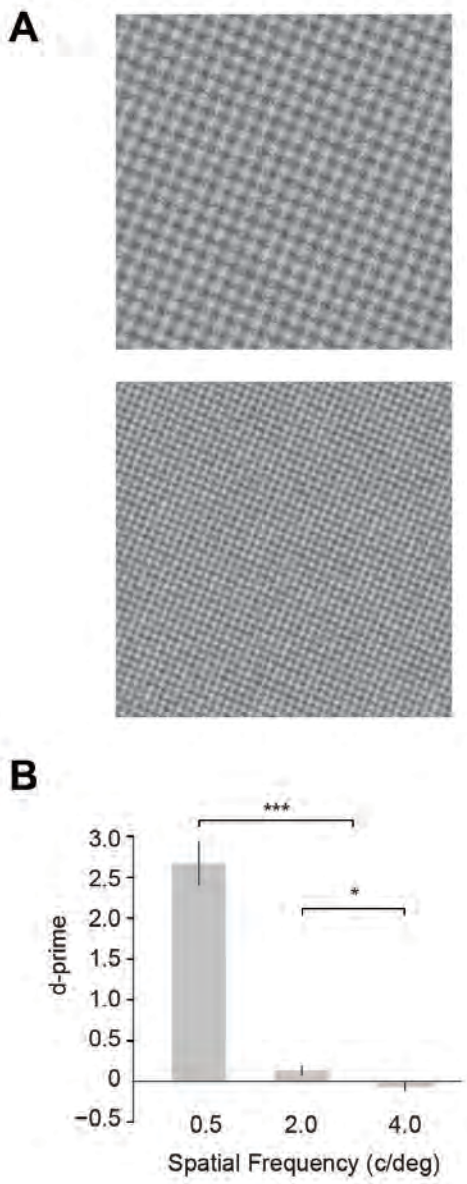



Figure 4.4**:** Experiment 2 – no transfer across spatial frequency: **(A)** Example stimuli for 2 c/deg (upper) and 4 c/deg (lower) plaid. Both plaids are oriented at 22.5 deg, i.e. in-phase component is on the right diagonal. **(B)**: Performance for high spatial frequency plaids (2 c/deg and 4c/deg) at the highest AM level compared to the equivalent perfromance for 0.5 c/deg for the trained stimuli in Experiment 1. Error bars indicate ± S.E.M. and asterisks indicate where the difference is significant (*** for $p < .001$ and * for $p < .05$).

Per cent correct values were converted to *d'* to indicate the discrimination sensitivity for LM/AM phase relationship (in- or anti-phase) at the highest AM level (0.244). Figure 4.4B

shows the mean performance across six observers. Data from Experiment 1 (0.5c/deg, trained orientation, AM=0.244) is shown for comparison. A repeated measures ANOVA showed that there was no transfer of training with 0.5 c/deg stimuli to 2 or 4 c/deg stimuli (main effect spatial frequency, Greenhouse - Geisser corrected, $F_{1.0,\ 5.0}$ = 100.7, $p$<.001). Bonferroni corrected comparisons also showed differences in mean *d'* values for 2 c/deg vs. 4c/deg ($p$<.05).

### 4.3.4 Experiment 3: Partial transfer to non-orthogonal plaids.

Experiment 1 showed that training for a single orientation of plaids did not fully transfer to 45 deg rigid rotations of orthogonal plaids. Here we investigate whether the training effect is specific to the angle between the components of the plaids in training sets. The plaids used in the training were all orthogonal; here we rotated the two components in a plaid separately so that their combination was no longer orthogonal. We introduce *shear angle* to define non-orthogonal combinations of the in- and anti-phase components. That is, if a plaid has a shear angle of +10 deg, the angle between the two components is 100 deg; whereas orthogonal plaids have 90 deg between their components and hence a shear angle of 0 deg. In Experiment 3, participants viewed non-orthogonal plaids with 6 shear angles (-50, -30, -10, 10, 30, 50 deg; Figure 4.5A) at all five levels of AM. All other stimulus parameters were as described in the general methods.

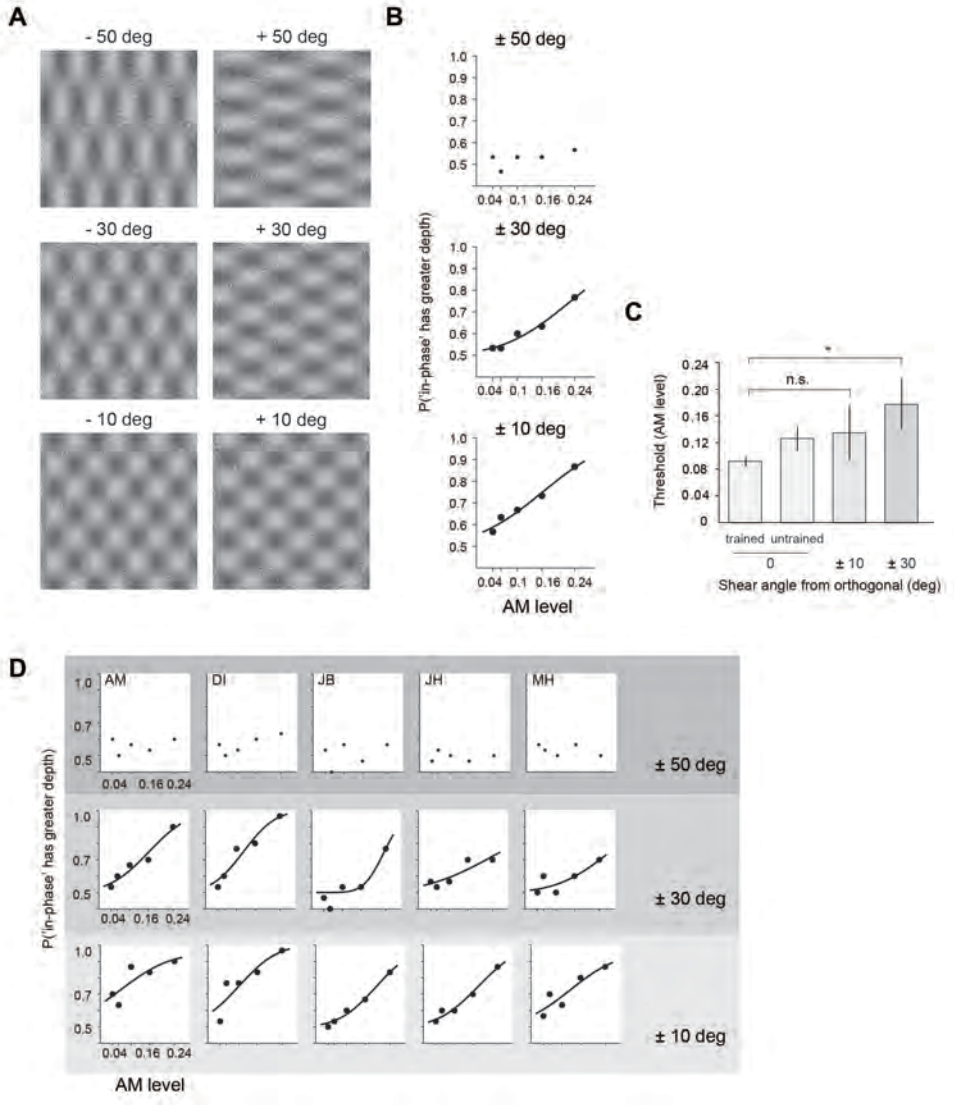



Figure 4.5: Experiment 3 – Partial transfer to non-orthogonal plaids: (A) Stimulus examples used in Experiment 3. Different shear angles are shown (rows) where left and right columns separate negative and positive shears, respectively. (B) Mean psychometric functions for six participants. Each plot shows per cent correct data points and psychometric functions (where available) as shear angle decreases from top to bottom. (C) Mean thresholds for AM level at 75% correct ('in-phase has a greater depth') rate from six participants are presented for ±30 deg shear (darker grey bar) and ±10 deg shear (lighter grey bar); it was not possible to extract threshold for ±50 deg shears. Data from Experiment 1 (white bars, 0 deg shears) are added for comparison. Error bars indicate ± S.E.M. and asterisk indicates a significant difference ($p < .05$). (D) Individual plots show psychometric functions for each participant (columns) on three groups of shear angles (rows). Shear angles decrease from top row to bottom row, hence becoming more similar to the training stimulus (shear = 0 deg).

It was not possible to fit psychometric functions for shear angles +50 or -50 deg, because participants performed around chance even for the highest AM signal in these

conditions (Figure 4.5B, top graph). A repeated measures ANOVA with 2 factors (shear sign and magnitude) showed a significant main effect of magnitude ($F_{2,8}$=31.93, $p$<.0001) but no effect of sign ($F_{1,4}$<1, $p$=.42) and no interaction ($F_{1.0,4.0}$<1, $p$=.729), therefore we grouped shear angles according to their magnitudes: ±10, ±30, ±50. Figure 4.5B shows average per cent correct (see Supplementary Figure 3 for *d'*) values at each AM level and fitted cumulative Gaussian psychometric functions (where possible) for three different groups of shear angle magnitudes of non-orthogonal plaids.

The thresholds decreased for smaller absolute shear angles but they were still higher than the thresholds obtained for trained stimuli in Experiment 1 (mean ± st. dev. for ±10 deg: AM = 0.14±0.04 and for ±30 deg: AM = 0.18±0.04). Figure 4.5C shows mean thresholds alongside data from Experiment 1: it can be seen that thresholds increase as the shear increases. These results show that the smallest shear angles did not differ from the trained plaids when we compare thresholds (trained plaid vs. ±10 deg shear: $t$(4)=1.81, $p$=.078), so the benefit of training transferred to ±10 deg shears. Experiment 3 was conducted last in the test sequence, so this result also shows that the lack of transfer in Experiment 2 cannot be due to a return to the untrained state over time. The thresholds for ±30 deg are significantly higher than the plaid thresholds for the trained stimulus set in Experiment 1 ($t$(4)=4.36, $p$=.006). However, statistical analysis showed that the thresholds for ±10 and ±30 deg were not significantly different ($t$(4)=<1, $p$=.373). Supplementary Figure 4.5 shows *d'* values for untrained observers in the shear task, showing again that the task is impossible without training.

## 4.4 Discussion

The luminance variations in a scene are potentially ambiguous. They could be caused by changes in the light source position, changes in illumination due to surface orientation or shadows, or they can represent intrinsic properties of the viewed surface such as albedo

reflectance. For example, a surface might have different colours or it may consist of different materials, so that its reflectance changes. Schofield *et al*. (2006; 2010) have shown that the phase relationship between first-order luminance modulations (LM) and second-order amplitude modulations (AM) can be used to discriminate luminance dependent changes from reflectance dependent changes: in-phase combinations give rise to the percept of a corrugated surface via shape-from-shading, while anti-phase combinations appear as reflectance changes. Schofield *et al*. (2010) proposed the shading channel model as a mechanism by which AM can influence the perceived role of luminance variations in an image. This model relies on early visual mechanisms and suggests that layer decomposition using these cues should be automatic and quick. However, whereas naive participants can use the relationship between LM and AM at relatively long presentations times (Schofield *et al*., 2006; 2010) they fail to do so at shorter presentation times.

Here we have shown that layer decomposition based on the phase relationships of LM and AM cues in plaid stimuli can be achieved at short presentation times (250ms) following training with intermittent feedback. This decomposition was specific to the trained stimulus and did not fully transfer to plaids at other orientations (Experiment 1). It transferred for small shear angles of non-orthogonal plaids (Experiment 3). However, training did not transfer at all to higher spatial frequency plaids (Experiment 2) or to larger shear angles for non-orthogonal plaids, even though the AM cue was as visible in such test stimuli as it was in the trained stimuli.

In the initial exposure phase of plaid training, we showed that observers are not able to differentiate LM+AM from LM-AM at brief presentation times. After five to ten days of training with intermittent feedback, performance improves and observers start judging LM+AM as more corrugated than LM-AM. This suggests that observers learn to make use of the AM cue and its alignment with LM (in- or anti-phase) as cues to shape from shading; they

learn to segment shading dependent illumination changes from material dependent changes. In other words, they learn to see the difference caused by the alignment of the AM cue; judging the anti-phase aligned LM/AM combination as a flat surface.

In Experiment 1, we also used novel stimuli to test whether the benefit of training transfers across rigid rotations. We found that the performance on novel plaids was better than the performance at initial exposure; however, thresholds remained significantly higher than those for trained plaids.

The results of Experiment 2 indicate that observers could not make use of the AM cue or phase relationship to differentiate shading from reflectance changes in high frequency stimuli. The ability to use the AM signal failed to transfer to 2 c/deg plaids. AM sensitivity varies with spatial frequency, as does that for LM, but, based on the sensitivity functions found by Schofield and Georgeson (1999), we would not expect any marked change in the visibility of either cue between 0.5 and 2 c/deg when binary noise is present. So the change in performance cannot be due to a lack of visibility for the AM cue at the higher frequencies, and must rather reflect an inability to combine the cues or make use of the relative phase information.

The spatial configuration of LM+AM and LM-AM components in a plaid seems to be important for layer decomposition. Specifically, the LM-AM component in a plaid with an orthogonal LM+AM component is seen as a very flat reflectance change, whereas it is seen as moderately corrugated when presented alone. We now show that the orthogonal configuration is itself important. Even when trained in the layer decomposition task, participants cannot discriminate the two components when the plaid is sheared by 50 deg, thus reducing the minimum angle between the two components to 40 deg. There is little transfer of training at a 30 deg shear (min angle 60 deg). This shows that the training is specific to the alignment of the two components in the plaid.

Overall, our findings provide evidence for stimulus-specific perceptual learning of the layer decomposition task based on the phase relationship of LM/AM mixtures at short presentation times. This supports the shading channel model proposed by Schofield *et al.* (2010). The ability to perform the tasks described in this paper at short presentation times strongly suggests that the task is supported by early, automatic, mechanisms. The failure of transfer across stimuli properties confirms that learning took place at a perceptual rather than cognitive level, again implicating low level mechanisms.

We should, however, consider why fast layer decomposition is available only after training. According to the shading channel model, cross-orientation gain control is fundamental to the perceptual scission between LM+AM and LM-AM cues in the plaid condition. One possibility is that this mechanism, which most likely relies on feedback loops, is normally quite sluggish, but that its action can be speeded up with training via a strengthening of the inhibitory links. Gain control mechanisms are known to be relatively broadband, so this reasoning may explain the partial transfer that we found in some conditions. It should also be noted that the gain control mechanism seems to be less useful in natural stimuli than in our plaid stimuli (Schofield *et al*., 2010) and that the machine vision system proposed by Jiang *et al*. (2010) dispenses with it altogether. Thus, the human visual system might not normally deploy the cross-orientation gain control mechanism implied by the shading channel model, but might engage it when repeatedly presented with the plaid decomposition task.

In summary, we have shown that layer decomposition based on the phase relationship of LM and AM cues can be achieved at short presentation times only after training and that this training is characterised by perceptual rather than cognitive learning. These findings support an account of layer decomposition based on early, automatic processes, although training may be required to tune these processes to deal with specific experimental stimuli.

# CHAPTER 5:

# Dorsal visual cortex integrates qualitatively different depth cues in a perceptually-relevant manner.[2]

---

The visual system's flexibility in estimating depth is remarkable: we readily perceive three-dimensional (3D) structure under diverse conditions from the seemingly random dots of a 'magic eye' stereogram to the aesthetically beautiful, but obviously flat, canvasses of the Old Masters. However, 3D perception is often enhanced when different cues specify the same depth. This perceptual process is understood as Bayesian inference that improves sensory estimates. Despite considerable behavioural support for this theory, insights into the cortical circuits involved are limited. Moreover, extant work has tested quantitatively similar cues, reducing some of the challenges associated with integrating computationally and qualitatively different signals. Here we address this challenge by measuring functional MRI responses to depth structures defined by shading, binocular disparity and their combination. We quantified information about depth configurations (convex 'bumps' *vs*. concave 'dimples') in different visual cortical areas using pattern-classification analysis. We found that fMRI responses in dorsal visual area V3B/KO were more discriminable when disparity and shading concurrently signalled depth, in line with the predictions of cue integration. Importantly, by relating fMRI and psychophysical tests of integration, we observed a close association between depth judgements and activity in this area. Finally, using a cross-cue transfer test, we found that fMRI responses evoked by one cue afford classification of responses evoked by the other. This reveals a generalised depth representation in dorsal visual cortex that integrates qualitatively different information in line with 3D perception.

---

[2] This chapter was published in Journal of Cognitive Neuroscience on 06/05/2012. Dövencioğlu, D., Ban, H., Schofield, A.J., and Welchman, A.E. (2013). "Perceptual Integration for Qualitatively Different 3-D Cues in the Human Brain." Journal of Cognitive Neuroscience: 1-15. Main text and figures were kept as in the submitted manuscript. All authors contributed to the conceptualisation of the experiment and writing of the paper, DND collected data and ran the analyses.

## 5.1 Introduction

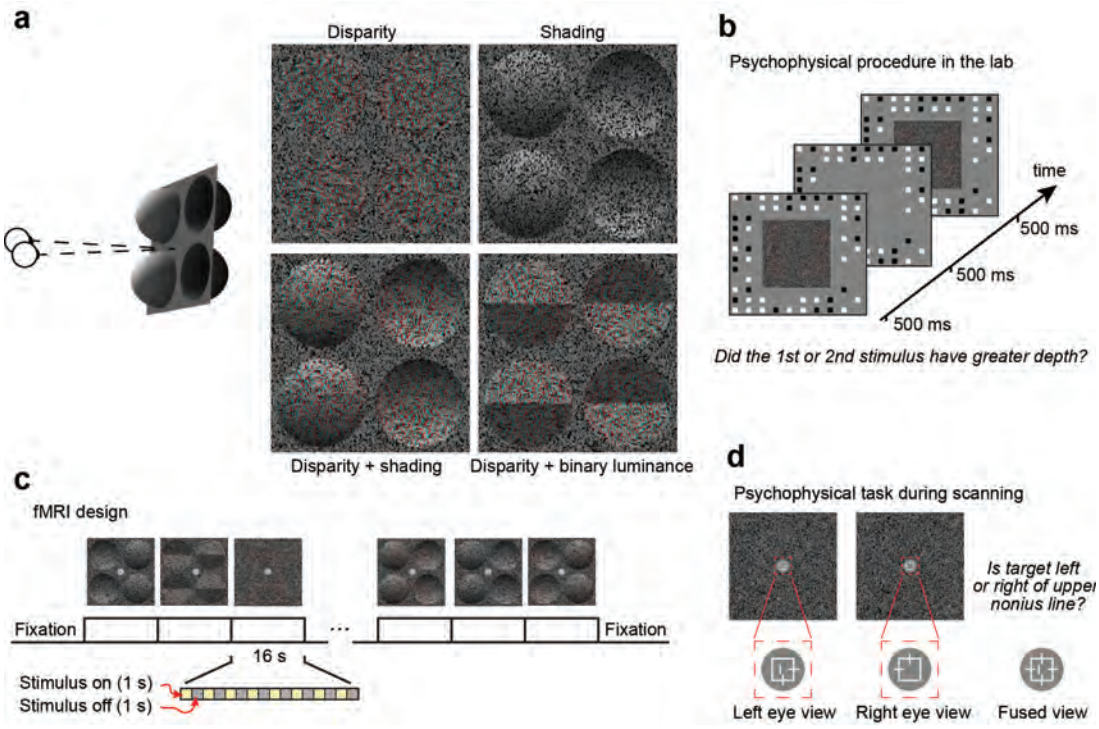
Many everyday tasks rely on depth estimates provided by the visual system. To facilitate these outputs, the brain exploits a range of inputs: from cues related to distance in a mathematically simple way (e.g., binocular disparity, motion parallax) to those requiring complex assumptions and prior knowledge (e.g. shading, occlusion) (Burge, Fowlkes, & Banks, 2010; Kersten, Mamassian, & Yuille, 2004; Mamassian & Goutcher, 2001). These diverse signals each evoke an impression of depth in their own right; however, the brain aggregates cues (Buelthoff & Mallot, 1988; Dosher, Sperling, & Wurst, 1986; Landy, Maloney, Johnston, & Young, 1995) to improve perceptual judgments (Knill & Saunders, 2003).

Here we probe the neural basis of integration, testing binocular disparity and shading depth cues that are computationally quite different. At first-glance these cues may appear so divergent that their combination would be prohibitively difficult. However, perceptual judgments show evidence for the combination of disparity and shading (Buelthoff & Mallot, 1988; Doorschot, Kappers, & Koenderink, 2001; Lovell, Bloj, & Harris, 2012; Schiller, Slocum, Jao, & Weiner, 2011; Vuong, Domini, & Caudek, 2006), and the solution to this challenge is conceptually understood as a two stage process (Landy et al., 1995) in which cues are first analyzed quasi-independently followed by the integration of cue information that has been 'promoted' into common units (such as distance). Moreover, observers can make reliable comparisons between the perceived depth from shading and stereoscopic, as well as haptic, comparison stimuli (Kingdom, 2003; Schofield, Rock, Sun, Jiang, & Georgeson, 2010), suggesting some form of comparable information.

To gain insight into the neural circuits involved in processing three-dimensional information from disparity and shading, previous brain imaging studies have tested for overlapping fMRI responses to depth structures defined by the two cues, yielding locations in

which information from disparity and shading converge (Georgieva, Todd, Peeters, & Orban, 2008; Nelissen *et al.*, 2009; Sereno, Trinath, Augath, & Logothetis, 2002). While this is a useful first step, this previous work has not established integration: for instance, representations of the two cues might be collocated within the same cortical area, but represented independently. By contrast, recent work testing the integration of disparity and motion depth cues, indicates that integration occurs in higher dorsal visual cortex (area V3B/Kinetic Occipital (KO)) (Ban, Preston, Meeson, & Welchman, 2012). This suggests a candidate cortical locus in which other types of 3D information may be integrated, however, it is not clear whether integration would generalize to (i) more complex depth structures and/or (ii) different cue pairings.

First, Ban and colleagues (2012) used simple fronto-parallel planes that can sub-optimally stimulate neurons selective to disparity-defined structures in higher portions of the ventral (Janssen, Vogels, & Orban, 2000) and dorsal streams (Srivastava, Orban, De Maziere, & Janssen, 2009) compared with more complex curved stimuli. It is therefore possible that other cortical areas (especially those in the ventral stream) would emerge as important for cue integration if more 'shape-like' stimuli were presented. Second, it is possible that information from disparity and motion are a special case of cue conjunctions, and thus integration effects may not generalize to other depth signal combinations. In particular, depth from disparity and from motion have computational similarities (Richards, 1985), joint neuronal encoding (Anzai, Ohzawa, & Freeman, 2001; Bradley, Qian, & Andersen, 1995; DeAngelis & Uka, 2003) and can, in principle, support metric (absolute) judgments of depth. In contrast, the 3D pictorial information provided by shading relies on a quite different generative process that is subject to different constraints and prior assumptions (Fleming, Dror, & Adelson, 2003; Horn, 1975; Koenderink & van Doorn, 2002; Mamassian & Goutcher, 2001; Sun & Perona, 1998; Thompson, Fleming, Creem-Regehr, & Stefanucci, 2011).




**igure 5.1:** Stimulus illustration and psychophysical results. **(a)** Left side: Cartoon of the isparity and/or shading defined depth structure. One of the two configurations is resented: bumps to the left, dimples to the right. Right side: stimulus examples rendered as ed-cyan anaglyphs. **(b)** Behavioral tests of integration. Bar graphs represent the between-ubjects mean slope of the psychometric function. * indicates $p<.05$. **(c)** Psychophysical esults as an integration index. Distribution plots show bootstrapped values: the center of he 'bowtie' represents the median, the colored area depicts 68% confidence values, and the pper and lower error bars 95% confidence intervals.

To test for cortical responses related to the integration of disparity and shading, we assessed how fMRI responses change when stimuli are defined by different cues (**Fig 1a**). We used multi-voxel pattern analysis (MVPA) to assess the information contained in fMRI responses evoked by stimuli depicting different depth configurations (convex vs. concave hemispheres to the left vs. right of the fixation point). We were particularly interested in how information about the stimulus contained in the fMRI signals changed depending on the cues used to depict depth in the viewed display. Intuitively, we would expect that discriminating fMRI responses should be easier when differences in the depicted depth configuration were defined by two cues rather than just one (i.e., differences defined by disparity and shading together should be easier to discriminate than differences defined by only disparity). The

theoretical basis for this intuition can be demonstrated based on statistically optimal discrimination (Ban, Preston, Meeson, & Welchman, 2012), with the extent of the improvement in the two-cue case providing insight into whether the underlying computations depend on the integration of two cues or rather having co-located but independent depth signals.

To appreciate the theoretical predictions for a cortical area that responds to integrated cues vs. co-located but independent signals, first consider a hypothetical area that is only sensitive to a single cue (e.g., shading). If shading information differed between two presented stimuli, we would expect neuronal responses to change, providing a source of information that could be decoded by the MVPA technique. By contrast, manipulating a non-encoded stimulus features such as disparity would have no effect on neuronal responses, meaning that our ability to decode the stimulus from the fMRI response profile of the area would be unaffected. Such a computationally isolated processing module is biologically rather unlikely, so next we consider a more plausible scenario where an area contains different subpopulations of neurons, some of which are sensitive to disparity and others to information from shading. In this case, we would expect to be able to decode stimulus differences based on changes in either cue. Moreover, if the stimuli contained differences defined by both cues, we would expect decoding performance to improve, where this improvement is predicted by the quadratic sum of the discriminabilities for changes in each cue. This expectation can by understood graphically by conceiving of discriminability based on shading and disparity cues as two sides of a right-angled triangle, where better discriminability equates to longer side lengths; the discriminability of both stimuli together equals the triangle's hypotenuse whose length is determined based on a quadratic sum (i.e., the Pythagorean equation) and is always at least as good as the discriminability of one of the cues (i.e., the hypotenuse can never be less than the length of the longest side).

The alternative possibility is a cortical region that responds on the basis of integrating

two different depth cues. Under this scenario, we would also expect better discrimination performance when two cues define differences between the stimuli. Importantly however, unlike the independence scenario, when stimulus differences are defined by only one cue, a fusion mechanism is adversely affected. For instance, if contrasting stimulus configurations differ in the depth indicated by shading but disparity indicates no difference, the fusion mechanism combines the signals from each cue with the result that it is less sensitive to the combined estimate than the shading component alone. By consequence, if we calculate a quadratic summation prediction based on measuring MVPA performance for depth differences defined by single cues (i.e., disparity; shading) we will find that empirical performance in the combined cue case (i.e., disparity + shading) actually exceeds the prediction (Ban, Preston, Meeson, & Welchman, 2012). Here we exploit this expectation to identify cortical responses to integrated depth signals, seeking to identify discrimination performance that is 'greater than the sum of its parts' due to the detrimental effects of presenting stimuli in which depth differences are defined in terms of a single cue.

To this end, we generated random dot patterns (**Fig. 5.1a**) that evoked an impression of four hemispheres, two concave ('dimples') and two convex ('bumps'). We formulated two different types of display that differed in their configuration: (1) bumps left – dimples right (depicted in **Fig. 5.1a**) *vs.* (2) dimples left – bumps right (turn the stereograms in **Fig. 5.1a** upside down). We depicted depth variations from: (*i*) binocular disparity, (*ii*) shading gradients, and (*iii*) the combination of disparity and shading. In addition, we employed a control stimulus (*iv*) in which the overall luminance of the top and bottom portions of each hemisphere differed (Ramachandran, 1988) (disparity + binary luminance). Perceived depth for these (deliberately) crude approximations of the shading gradients relied on disparity. We tested for integration using both psychophysical- and fMRI- discrimination performance for the component cues (*i*, *ii*) with that for stimuli containing two cues (*iii*, *iv*). We reasoned that a response based on integrating depth cues would be specific to concurrent cue stimulus (*iii*)

and not be observed for the control stimulus (*iv*).

In previous Chapters (3 and 4) I have provided evidence that human observers can use first and second order luminance signals as a cue to shape from shading, and they learn to discriminate luminance dependent changes in the shading pattern from material dependent changes. Here, I focus on the interaction of the shading cue with disparity cue. In this chapter, I focus on the underlying neural mechanisms of shading and disparity processing when observers are estimating 3D shape.

## 5.2 Materials and Methods

### 5.2.1 Observers

Twenty observers from the University of Birmingham participated in the fMRI experiments. Of these, five were excluded due to excessive head movement during scanning, meaning that the correspondence between voxels required by the MVPA technique was lost. Excessive head movement was defined as ≥ 4mm over an eight minute run, and we excluded participants if they had fewer than 5 runs below this cut-off as there was insufficient data for the multivoxel pattern analysis. Generally, participants were able to keep still: the average absolute maximum head deviation relative to the start of the first run for included participants was 1.2 mm vs. 4.5 mm for excluded participants. Moreover only one included participant had an average head motion of > 2 mm per run, and the mode of the head movement distribution across subjects was <1 mm. Six female and nine male participants were included; twelve were right-handed. Mean age was 26 ± 1.2 (S.E.M.) years. Authors AEW and HB participated, all other subjects were naïve to the purpose of the study. Four of the participants had taken part in the study of (Ban, Preston, Meeson, & Welchman, 2012). Participants had normal or corrected to normal vision and were prescreened for stereo deficits. Experiments were approved by the University of Birmingham Science and Engineering ethics committee; observers gave written informed consent.

### 5.2.2 Stimuli

Stimuli were random dot stereograms (RDS) that depicted concave or convex hemispheres (radius = 1.7°; depth amplitude 1.85 cm ≈ 15.7 arcmin) defined by disparity and/or shading. The individual dots subtended 0.06° and patterns had a density of 94 dots/deg$^2$. We depicted shading using the Blinn-Phong shading algorithm implemented in Matlab under a diffuse light source that was positioned directly overhead. For the disparity condition, the dots in the display followed the luminance distribution of the shaded patterns; however, their positions were randomised across the shape. For the shading condition, disparity specified a flat surface. To create the binary luminance stimuli, the luminance of the top and bottom portions of the hemispheres was held constant at the mean luminance of these portions of the shapes for the shaded stimuli. Four hemispheres were presented: two convex, and two concave, located either side of a fixation marker. Two types of stimulus configuration were used: (i) convex on the left, concave on the right, and (ii) vice versa. The random dot pattern subtended 8×8° and was surrounded by a larger, peripheral grid (18×14°) of black and white squares which served to provide a stable background reference. The other portions of the display were set to mid-level grey.

### 5.2.3 Psychophysics

Stimuli were presented to participants in a lab setting using a stereo set-up in which the two eyes viewed separate CRTs (ViewSonic FB2100x) through front-silvered mirrors at a distance of 50 cm. Linearisation of the graphics card outputs was achieved using photometric measurements. Images were displayed at 100Hz with a screen resolution of 1600 x 1200 pixels.

Under a two interval forced choice design, participants decided which stimulus had the greater depth profile (Fig 1b; presentation time = 500ms, interstimulus interval = 500ms). On

every trial, one interval contained a standard disparity-defined stimulus (± 1.85 cm / 15.7 arcmin), the other interval contained a stimulus from one of three conditions (disparity alone; disparity + shading; disparity + binary luminance) and had a depth amplitude that was varied using the method of constant stimuli. The shading cue varied as the depth amplitude of the shape was manipulated such that the luminance gradient was compatible with a bump/dimple whose amplitude matched that specified by disparity. Similarly, for the binary luminance case, the stimulus luminance values changed at different depth amplitudes to match the luminance variations that occurred for the gradient shaded stimuli. The order of the intervals was randomized, and conditions were randomly interleaved. On a given trial, a random jitter was applied to the depth profile of both intervals (uniform distribution within ±1 arcmin to reduce the potential for adaptation to a single disparity value across trials). Participants judged "did the first or second stimulus have greater depth" by pressing an appropriate button. On some runs participants were instructed to consider their judgment relative to the convex portions of the display, in others the concave portions. The spatial configuration of convex and concave items was randomized. A single run contained a minimum of 630 trials (105 trials × 3 conditions × 2 curvature instructions). We made limited measures of the shading alone condition as we found in pilot testing that participants' judgments based on shading 'alone' were very poor (maximum discriminability in the shading condition was $d' = 0.3 \pm 0.25$) meaning that we could not fit a reliable psychometric function so threshold estimates were unstable and uninformative, and participants became frustrated by a seemingly impossible task. Moreover, in the shading alone condition, stimulus changes could be interpreted as a change of light source direction, rather than depth, given the bas-relief ambiguity (Belhumeur, Kriegman, & Yuille, 1999). This ambiguity should be removed by the constraint imposed by the disparity signals available in the disparity + shading condition, although this does not necessarily happen (see Discussion section on Individual differences in integration).

### 5.2.4 Imaging

Data were recorded at the Birmingham University Imaging Centre using a 3 Tesla Philips MRI scanner with an 8-channel multi-phase array head coil. BOLD signals were measured with an echo-planar (EPI) sequence (TE: 35 ms, TR: 2s, 1.5×1.5×2 mm, 28 slices near coronal, covering visual, posterior parietal and posterior temporal cortex) for both experimental and localizer scans. A high-resolution anatomical scan (1 $mm^3$) was also acquired for each participant to reconstruct cortical surface and coregister the functional data. Following coregistration in the native anatomical space, functional and anatomical data were converted into standardized Talairach coordinates.

During the experimental session, four stimulus conditions (disparity; shading; disparity + shading; disparity + binary luminance) were presented in two spatial configurations (convex on left vs. on right) = 8 trial types. Each trial type was presented in a block (16s) and repeated three times during a run (Fig. 1c). Stimulus presentation was 500 ms on, 500 ms off, and different random dot stereograms were used for each presentation. These different stimuli had randomly different depth amplitudes (jitter of 1 arcmin) to attenuate adaptation to a particular depth profile across a stimulus block. Each run started and ended with a fixation period (16s), total duration = 416s. Scan sessions lasted 90 minutes and allowed us to collect 7 to 10 runs depending on the initial setup time and each individual participant's requirements for breaks between runs.

Participants were instructed to fixate at the centre of the screen, where a square crosshair target (side = 0.5°; horizontal and vertical nonius lines displayed) was presented at all times (Fig. 1d). This was surrounded by a mid-grey disc area (radius = 1°). A dichoptic Vernier task was used to encourage fixation and provide a subjective measure of eye vergence (Popple, Smallman, & Findlay, 1998). In particular, a small vertical Vernier target was flashed (250 ms) at the vertical center of the fixation marker to one eye. Participants judged whether this

Vernier target was to the left or right of the upper nonius line, which was presented to the other eye. We used the method of constant stimuli to vary Vernier target position, and fit the proportion of 'target on the right responses' to estimate whether there was any bias in the observers responses that would indicate systematic deviation from the desired vergence state. The probability of a Vernier target appearing on a given trial was 50%, and the timing of appearance was variable with respect to trial onset (during the first vs. second half of the stimulus presentation), requiring constant vigilance on behalf of the participants. In a separate session, a subset of participants (n = 3) repeated the experiment during an eye tracking session in the scanner. Eye movement data were collected with CRS limbus Eye tracker (CRS Ltd, Rochester, UK).

The vernier task was deliberately chosen to ensure that participants were engaged in a task orthogonal to the main stimulus presentations and manipulation. The temporal uncertainty in the timing of presentation, and its brief nature, ensured that participants had to constantly attend to the fixation marker. Thereby, we ensured that differences in fMRI responses between conditions could not be ascribed to attentional state, task difficulty or the degree of conflict inherent in the different stimuli.

Note also that the differences between stimuli presented during scanning were highly suprathreshold (i.e. convex vs. concave) to ensure that the depth configurations could be reliably decoded from the fMRI responses. This differed from the psychophysical judgments where we measured sensitivity to small differences in the depth profile of the shapes. We would expect benefits from integrating cues in both cases, however it is important to note these differences imposed by the different types of measurement paradigms (fMRI vs. psychophysics) we have used.

Stereoscopic stimulus presentation was achieved using a pair of video projectors (JVC D-ILA SX21), each containing separate spectral comb filters (INFITEC, GmBH) whose projected images were optically combined using a beam-splitter cube before being passed through a

wave guide into the scanner room. The INFITEC interference filters produce negligible overlap between the wavelength emission spectra for each projector, meaning that there is little crosstalk between the signals presented on the two projectors for an observer wearing a pair of corresponding filters. Images were projected onto a translucent screen inside the bore of the magnet. Subjects viewed the display via a front-surfaced mirror attached to the headcoil (viewing distance = 65 cm). The two projectors were matched and linearized for grey scale outputs using photometric measurements. The INFITEC filters restrict the visibility of the eyes for a remote monitoring system, making standard remote eye tracking equipment unsuitable for eye movement recording in our set up. We therefore employed a monocular limbus eye tracker located between the participants' eyes and the spectral comb filters. This eye tracking system has a stated accuracy of < 0.25 degrees of visual angle.

Functional and anatomical pre-processing of MRI data was conducted with BrainVoyager QX (BrainInnovation B.V.) and in-house MATLAB routines. We transformed anatomical scans into Talairach space, created inflated and flattened surface models for both hemispheres for each participant. For each functional run, data were corrected with slice time correction, 3D motion correction, high pass filtering, and linear trend removal. After motion correction, each participant's functional data were aligned to their anatomical scan and transformed into Talairach space. No spatial smoothing was performed. Retinotopic areas were identified in individual localizer scanning sessions for each participant.

### 5.2.5 Mapping Regions of Interest

We identified regions of interest within the visual cortex for each individual participant in a separate fMRI session prior to the main experiment. To identify retinotopically organized visual areas, we used rotating wedge stimuli and expanding/contracting rings to identify visual field position and eccentricity maps (DeYoe *et al.*, 1996; M. I. Sereno *et al.*, 1995). Thereby we identified areas V1, V2 and the dorsal and ventral portions of V3 (which we denote V3d and V3v). Area V4 was localized adjacent to V3v with a quadrant field

representation (R. B. H. Tootell & Hadjikhani, 2001) while V3A was adjacent to V3d with a hemi-field representation. Area V7 was identified as anterior and dorsal to V3A with a lower visual field quadrant representation (R. B. Tootell *et al.*, 1998; Tyler, Likova, Chen, Kontsevich, & Wade, 2005). The borders of area V3B were identified as based on a hemi-field retinotopic representation inferior to, and sharing a foveal representation with, V3A (Tyler et al., 2005). This retinotopically-defined area overlapped with the contiguous voxel set that responded significantly more ($p = 10^{-4}$) to intact vs. scrambled motion-defined contours which has previously been described as the kinetic occipital area (KO) (Dupont *et al.*, 1997; Zeki, Perry, & Bartels, 2003). We therefore denote this area as V3B/KO as we have previously found no good means of differentiating the response properties of this region (Ban et al., 2012) (see also (Larsson, Heeger, & Landy, 2010)). We provide Talairach coordinates for the centroids of this area in Table 1. We identified the human motion complex (hMT+/V5) as the set of voxels in the lateral temporal cortex that responded significantly more ($p = 10^{-4}$) to coherent motion than static dots (Zeki et al., 1991). Finally, the lateral occipital complex (LOC) was defined as the set of voxels in the lateral occipito-temporal cortex that responded significantly more ($p = 10^{-4}$) to intract vs. scrambled images of objects (Kourtzi & Kanwisher, 2001). The posterior subregion LO extended into the posterior inferiotemporal sulcus and was defined based on the overlap of functional activations and anatomy (Grill-Spector, Kushnir, Hendler, & Malach, 2000).

**Table 5.1.** Talairach coordinates of the centroids of the region we denote V3B/KO. We present data for all participants, and participants separated into the good and poor integration groups.

| | | Left hemisphere | | | Right hemisphere | | |
|---|---|---|---|---|---|---|---|
| | | *x* | *y* | *z* | *x* | *y* | *z* |
| V3B/KO (all participants) | Mean | –27.5 | –84.7 | 7.0 | 31.6 | –80.7 | 6.5 |
| | SD | 4.4 | 4.3 | 3.9 | 4.1 | 4.4 | 4.8 |
| Good integrators | Mean | –26.9 | –86.0 | 6.7 | 31.5 | –81.7 | 7.2 |
| | SD | 4.2 | 4.2 | 4.2 | 4.0 | 4.5 | 4.8 |
| Poor integrators | Mean | –28.1 | –83.4 | 7.2 | 31.8 | –79.9 | 5.8 |
| | SD | 4.6 | 4.4 | 3.7 | 4.1 | 4.4 | 4.8 |

### 5.2.6 Multi-voxel pattern analysis (MVPA)

To select voxels for the MVPA, we used a participant-by-participant fixed effects GLM across runs on grey matter voxels using the contrast 'all stimulus conditions vs. the fixation baseline'. In each ROI, we rank ordered the resultant voxels by their t-statistic (where $t>0$), and selected to the top 300 voxels as data for the classification algorithm (Preston, Li, Kourtzi, & Welchman, 2008). To minimize baseline differences between runs we *z*-scored the response timecourse of each voxel and each experimental run. To account for the hemodynamic response lag, the fMRI time series were shifted by 2 TRs (4 s). Thereafter we averaged the fMRI response of each voxel across the 16 sec stimulus presentation block, obtaining a single test pattern for the multivariate analysis per block. To remove potential univariate differences (that can be introduced after *z*-score normalization due to averaging across timepoints in a block, and grouping the data into train *vs*. test data sets), we normalized by subtracting the mean of all voxels for a given volume (Serences & Boynton, 2007), with the result that each volume had the same mean value across voxels, and differed only in the pattern of activity. We performed multivoxel pattern analysis using a linear support vector machine (SVM$^{light}$ toolbox) classification algorithm. We trained the algorithm

to distinguish between fMRI responses evoked by different stimulus configurations (e.g., convex to the left vs. to the right of fixation) for a given stimulus type (e.g., disparity). Participants typically took part in 8 runs of the experiment, each of which had 3 repetitions of a given spatial configuration and stimulus type, creating a total of 24 patterns. We used a leave-one-run out cross validation procedure whereby we trained the classifier using 7 of the 8 runs (i.e., 21 patterns) and then evaluated the prediction performance of the classifier using the remaining, non-trained data (i.e., 3 patterns). We repeated this, leaving a single run out in turn, and calculated the mean prediction accuracy across cross-validation folds. Accuracies were represented in units of discriminability ($d'$) using the formula:

$$d' = 2 \cdot \text{erfinv}(2p - 1) \qquad \text{(Eqn. 5.1)}$$

where erfinv is the inverse error function and $p$ the proportion of correct predictions.

For tests of transfer between disparity and shading cues, we used a Recursive Feature Elimination method (RFE) (De Martino et al., 2008) to detect sparse discriminative patterns and define the number of voxels for the SVM classification analysis. In each feature elimination step, five voxels were discarded until there remained a core set of voxels with the highest discriminative power. In order to avoid circular analysis, the RFE method was applied independently to the training patterns of each cross-validation fold, resulting in eight sets of voxels (i.e. one set for each test pattern of the leave-one-run out procedure). This was done separately for each experimental condition, with final voxels for the SVM analysis chosen based on the intersection of voxels from corresponding cross-validation folds. A standard SVM was then used to compute within- and between- cue prediction accuracies. This feature selection method was required for transfer, in line with evidence that it improves generalization (De Martino et al., 2008).

We conducted Repeated Measures GLM in SPSS (IBM, Inc.) applying Greenhouse-Geisser correction when appropriate. Regression analyses of the psychophysical and fMRI integration indices were also conducted in SPSS. For this analysis, we considered the use of repeated measures MANCOVA (and found results consistent with the reported regression results); however, the integration indices (defined below) we use are partially correlated between conditions because their calculation depends on the same denominator, violating the GLM model's assumption of independence. We therefore limited our analysis to the relationship between psychophysical and fMRI indices for the same condition, for which the psychophysical and fMRI indices are independent of one another.

Statistical analyses were performed in SPSS (SPSS Inc.), and Greenhouse-Geisser correction was used when appropriate.

### 5.2.7 Quadratic summation and integration indices

We formulate predictions for the combined cue condition (i.e., disparity + shading) based on the quadratic summation of performance in the component cue conditions (i.e., disparity; shading). As outlined in the Introduction, this prediction is based on the performance of an idea observer model that discriminates pairs of inputs (visual stimuli or fMRI response patterns) based on the optimal discrimination boundary. Psychophysical tests indicate that this theoretical model is the appropriate one to use in understanding performance when human observers combine cues (Hillis et al., 2002; Knill and Saunders, 2003).

To compare measured empirical performance in disparity + shading condition with the prediction derived from the component cue conditions, we calculate a ratio to formulate an index (Ban *et al.*, 2012; Nandy & Tjan, 2008) whose general form is:

$$\text{Integration} = \text{Combined} + \text{Quadratic}\,2 + \text{Component}\,2 - 1$$

If the responses of the detection mechanism to the disparity and shading conditions ($C_D$, $C_S$) are independent of each other, performance when both cues are available ($C_{D+S}$) should match the quadratic summation prediction, yielding a ratio of 1 and thus an index of zero. A value of less than zero suggests suboptimal detection performance, and a value above zero suggests that the component sources of information are not independent (Ban *et al.*, 2012; Nandy & Tjan, 2008). However, a value above zero does not preclude the response of independent mechanisms – for instance noise introduced during fMRI measurement (scanner noise, observer movement) can lead to a positive index based on co-located but independent responses (Ban et al, 2012 consider this issue in their Supplementary Figure 3; they use fMRI simulations to show that in low non-neural noise situations, collocated but independent signals can sometimes surpass the quadratic summation prediciton). Thus, the integration index alone cannot be taken as definite evidence of cue integration, and therefore needs to be considered in conjunction with the other tests. To assess statistical significance of the integration indices, we used bootstrapped resampling, as our use of a ratio makes distributions non-Normal, and thus a non-parametric procedure more appropriate.

## 5.3 Results

### 5.3.1 Psychophysics

To assess cue integration psychophysically, we measured observers' sensitivity to slight differences in the depth profile of the stimuli. Participants viewed two shapes sequentially, and decided which had the greater depth (that is, which bumps were taller, or which dimples were deeper). By comparing a given standard stimulus against a range of test stimuli, we obtained psychometric functions. We used the slope of these functions to quantify observers' sensitivity to stimulus differences (where a steeper slope indicates higher sensitivity). To

determine whether there was a perceptual benefit associated with adding shading information to the stimuli, we compared performance in the disparity condition with that in the disparity and shading condition. Surprisingly, we found no evidence for enhanced performance in the disparity and shading condition at the group level ($F(1,14)<1$, $p=.38$). In light of previous empirical work on cue integration this was unexpected (e.g. (Buelthoff and Mallot, 1988; Doorschot et al., 2001; Vuong et al., 2006; Schiller et al., 2011; Lovell et al., 2012)), and prompted us to consider the significant variability between observers ($F(1,14)=62.23$, $p<.001$) in their relative performance in the two conditions. In particular, we found that some participants clearly benefited from the presence of two cues, however others showed no benefit and some actually performed worse relative to the disparity only condition. Poorer performance might relate to individual differences in the assumed direction of the illuminant (Schofield, Rock, & Georgeson, 2011); ambiguity or bistability in the interpretation of shading patterns (Liu & Todd, 2004; Wagemans, van Doorn, & Koenderink, 2010); and/or differences in cue weights (Knill & Saunders, 2003; Lovell *et al.*, 2012; Schiller *et al.*, 2011) (we return to this issue in the *Discussion*). To quantify variations between participants in the relative performance in two conditions, we calculated a psychophysical integration index ($y$):

$$y = S_{D+S} - S_{D} - 1 \qquad \text{(Eqn. 5.2)}$$

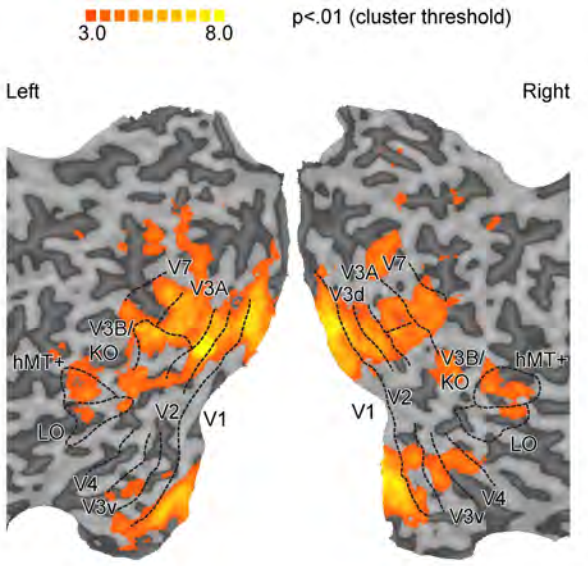
where $S_{D+S}$ is sensitivity in the combined condition and $S_D$ is sensitivity in the disparity condition. This index is based on the quadratic summation test (Ban *et al.*, 2012; Nandy & Tjan, 2008); see Methods for a description of the logic) where a value above zero suggests that participants integrate the depth information provided by the disparity and shading cues when making perceptual judgments. In this instance we assumed that $S_D \approx \sqrt{(S_D^2 + S_S^2)}$ because our attempts to measure sensitivity to differences in depth amplitude defined by shading alone in pilot testing resulted in such poor performance that we could not fit a reliable psychometric function. Specifically, discriminability of the maximum depictable depth difference was $d' = 0.3 \pm 0.25$ for shading alone, in contrast to $d' = 3.9 \pm 0.3$ for disparity, i.e. $S_D^2 >> S_S^2$.

We rank-ordered participants based on □, and thereby formed two groups (Fig. 1b, c): *good integrators* (the 7 participants for whom □ > 0) and *poor integrators* (the 8 participants where □ < 0). By definition, these post-hoc groups differed in the relative sensitivity to disparity and disparity + shading conditions. Our purpose in forming these groups, however, was to test the link between differences in perception and fMRI responses.

### 5.3.2 fMRI measures of integration

Before taking part in the main experiment, each participant underwent a separate fMRI session to identify regions of interest (ROIs) within the visual cortex (**Fig. 2**). We identified retinotopically organized cortical areas based on polar and eccentricity mapping techniques (DeYoe *et al.*, 1996; M. I. Sereno *et al.*, 1995; Tootell & Hadjikhani, 2001; Tyler, Likova, Chen, Kontsevich, & Wade, 2005). In addition we identified area LO involved in object processing (Kourtzi & Kanwisher, 2001), the human motion complex (hMT+/V5) (Zeki et al., 1991) and the Kinetic Occipital (KO) region which is localized by contrasting motion-defined contours with transparent motion (Dupont et al., 1997; Zeki et al., 2003). Responses to the KO localizer

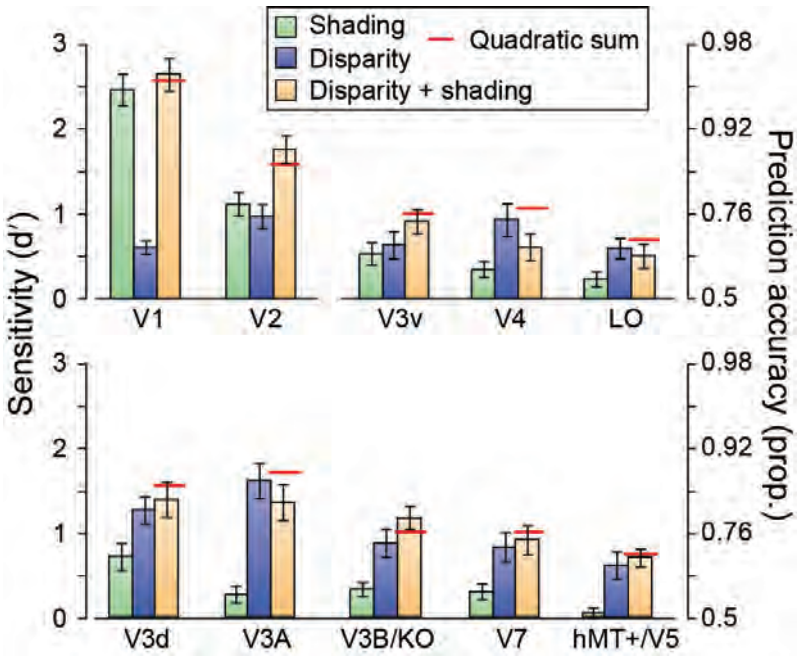
overlapped with the retinotopically-localized area V3B and were not consistently separable across participants and/or hemispheres (see also (Ban et al., 2012)) so we denote this region as V3B/KO. A representative flatmap of the regions of interest is shown in Fig. 2, and Table 1 provides mean coordinates for V3B/KO.




**igure 5.2:** Representative flat maps from one participant showing the left and right regions f interest. The sulci are depicted in darker grey than the gyri. Shown on the maps are etinotopic areas, V3B/KO, the human motion complex (hMT+/V5), and lateral occipital (LO) rea. The activation on the maps shows the results of a searchlight classifier analysis that noved iteratively throughout the measured cortical volume, discriminating between timulus configurations. The colour code represents the *t*-value of the classification ccuracies obtained. This procedure confirmed that we had not missed any important areas utside those localised independently.

We then measured fMRI responses in each of the independently localized ROIs, and were, *a priori*, particularly interested in responses from the V3B/KO region (Ban, Preston, Meeson, & Welchman, 2012; Tyler, Likova, Kontsevich, & Wade, 2006). We presented stimuli from four experimental conditions (**Fig. 1a**) under two configurations: (a) bumps to the left of fixation, dimples to the right or (b) bumps to the right, dimples to the left, thereby allowing us to contrast fMRI responses to convex *vs*. concave stimuli.

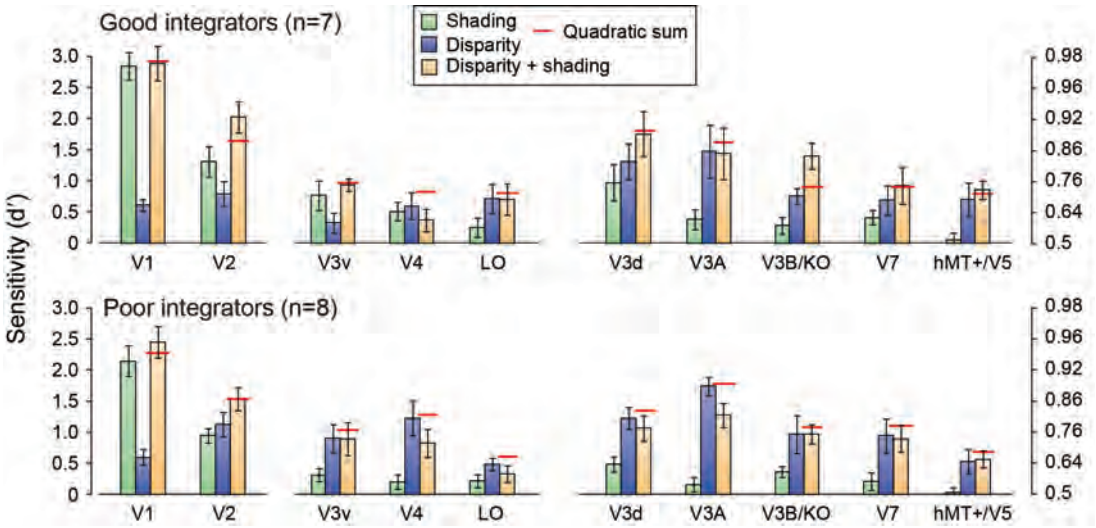
To analyze our data, we trained a machine learning classifier (support vector machine: SVM) to associate patterns of fMRI voxel activity and the stimulus configuration (convex *vs*. concave) that gave rise to that activity. We used the performance of the classifier in decoding the stimulus from independent fMRI data (i.e., a leave-one-run-out cross-validation procedure) as a measure of the information about the presented stimulus within a particular region of cortex.




**igure 5.3:** Performance in predicting the convex vs. concave configuration of the stimuli ased on the fMRI data measured in different regions of interest (n=15). The bar graphs how the results from the 'single cue' experimental conditions, the 'disparity + shading' ondition, the quadratic summation prediction (horizontal red line). Error bars indicate SEM.

We were able to reliably decode the stimulus configuration in the four different conditions in almost every region of interest (**Figure 5.3**), and there was a clear interaction between conditions and regions of interest ($F_{8.0,\ 104.2}$=8.92, p<.001). This widespread sensitivity to differences between convex *vs*. concave stimuli is not surprising, in that a range of image features might modify the fMRI response (e.g., distribution of image intensities,

contrast edges, mean disparity, etc.). The machine learning classifier may thus decode low-level image features, rather than 'depth' *per se*. We were therefore not interested in overall prediction accuracies between areas (which are influenced by our ability to measure fMRI activity in different anatomical locations). Rather, we were interested in the relative performance between conditions, and whether this related to between-observer differences in perceptual integration. We therefore considered our fMRI data subdivided based on the behavioural results (significant interaction between condition and group (*good vs. poor integrators*): $F_{2.0, 26.6}$=4.52, p=.02).



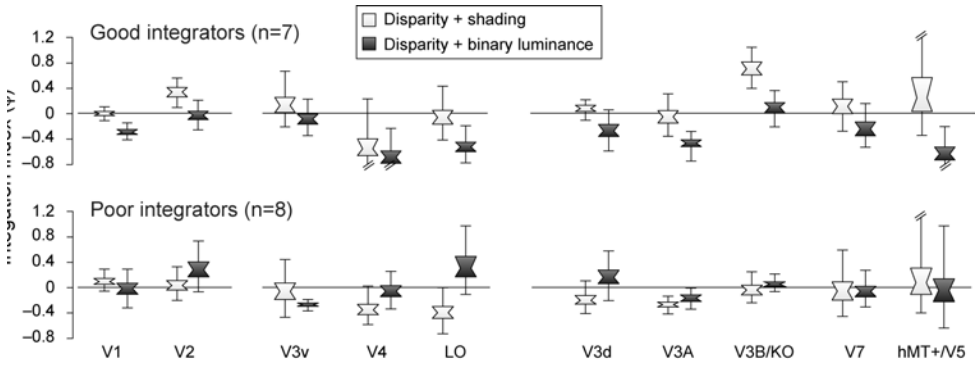

**igure 5.4:** Prediction performance for fMRI data separated into the two groups based on the sychophysical results ('good' vs. 'poor' integrators). The bar graphs show the results from he 'single cue' experimental conditions, the 'disparity + shading' condition, the quadratic ummation prediction (horizontal red line). Error bars indicate SEM.

First, we wished to determine whether fMRI decoding performance improved when the two depth cues were viewed concurrently. Prediction accuracies for the concurrent stimulus (disparity + shading) were statistically higher than the component cues in areas V2 ($F_{3, 39}$ =7.47, p<.001) and V3B/KO ($F_{1.6, 21.7}$=14.88, p<.001). To assess integration, we compared the extent of improvement in the concurrent stimulus relative to a minimum bound prediction (**Figures 5.3, 5.4,** red lines) based on the quadratic summation of decoding

accuracies for 'single cue' presentations (Ban *et al.*, 2012). This corresponds to the level of performance expected if disparity signals and shading signals are collocated in a cortical area, but represented independently. If performance exceeds this bound, it suggests that cue representations are not independent, as performance in the 'single' cue case was attenuated by the conflicts that result from 'isolating' the cue (e.g., responses to shading in the 'single cue' shading case are attenuated by conflicting disparity information that the surface was flat). We found that performance was higher (outside the SEM) than the quadratic summation prediction in areas V2 and V3B/KO. However, this result was only statistically reliable in V3B/KO (**Figure 5.4**). Specifically, there was a significant interaction between behavioural group and experimental condition ($F_{2,\ 26}$=5.52, p=.01), with decoding performance in the concurrent (disparity + shading) condition exceeding the quadratic summation prediction for good integrators ($F_{1,\ 6}$=9.27, p=.011), but not for poor integrators ($F_{1,\ 7}$<1, p=.35); **Figure 5.4**). ). In V2 the there was no significant difference between the quadratic summation prediction and the measured data in the combined cue conditions (F(2,26)<1, p=.62) nor an interaction between condition and behavioral group (F(2,26)=2.63, p=.091). We quantified the extent of integration using a bootstrapped index (□) that contrasted decoding performance in the concurrent condition (*d'D+S*) with the quadratic summation of performance with 'single' cues (*d'D* and *d'S*):

$$□ = d'D+Sd'D2+d'S2-1 \qquad \text{(Eqn. 5.3)}$$

Using this index, a value of zero corresponds to the performance expected if information from disparity and shading are collocated, but independent. We found that the integration index for the concurrent condition was only reliably above zero for the good integrators in areas V2 and V3B/KO (**Figure 5.5**; **Table 5.1**).




**igure 5.5:** fMRI based prediction performance as an integration index for the two groups of articipants in all the regions of interest. A value of zero indicates the minimum bound for usion as predicted by quadratic summation. Data represent the 'Disparity + shading' and Disparity + binary shading' conditions. Data are presented as notched distribution plots. The entre of the 'bowtie' represents the median, the coloured area depicts 68% confidence alues, and the upper and lower error bars 95% confidence intervals.

**Table 5.2:** Probabilities associated with obtaining a value of zero for the fMRI integration index in the (i) disparity + shading condition and (ii) luminance control condition. Values are obtained from a bootstrapped resampling of the individual participants' data using 10,000 samples. Bold formatting indicates Bonferroni corrected significance.

| Cortical area | Disparity + Shading | | Luminance control | |
|---|---|---|---|---|
| | Good integrators | Poor integrators | Good integrators | Poor integrators |
| V1 | 0.538 | 0.157 | 0.999 | 0.543 |
| V2 | **0.004** | 0.419 | 0.607 | 0.102 |
| V3v | 0.294 | 0.579 | 0.726 | 1.000 |
| V4 | 0.916 | 0.942 | 0.987 | 0.628 |
| LO | 0.656 | 0.944 | 0.984 | 0.143 |
| V3d | 0.253 | 0.890 | 0.909 | 0.234 |
| V3A | 0.609 | 1.000 | 0.999 | 0.961 |
| V3B/KO | **<0.001** | 0.629 | 0.327 | 0.271 |
| V7 | 0.298 | 0.595 | 0.844 | 0.620 |
| hMT+/V5 | 0.315 | 0.421 | 0.978 | 0.575 |

To provide additional evidence for neuronal responses related to depth estimation, we used the binary luminance stimuli as a control. We constructed these stimuli such that they contained a very obvious low-level feature that approximated luminance differences in the shaded stimuli but did not, *per se*, evoke an impression of depth. As the fMRI response in a given area may reflect low-level stimulus differences (rather than depth from shading), we wanted to rule out the possibility that improved decoding performance in the concurrent disparity + shading condition could be explained on the basis that two separate stimulus dimensions (disparity and luminance) drive the fMRI response. The quadratic summation test should theoretically rule this out; nevertheless, we contrasted decoding performance in the concurrent condition vs. the binary control (disparity + binary luminance) condition. We

reasoned that if enhanced decoding is related to the representation of depth, the superquadratic summation effects would be limited to the concurrent condition. We found this to be true for the good integrator subjects in area V3B/KO: sensitivity in the concurrent condition was above that in the binary control condition ($F_{1, 6}$=14.69, p=.004). By contrast, sensitivity for the binary condition in the poor integrator subjects matched that of the concurrent group ($F_{1, 7}$<1, p=.31) and was in line with quadratic summation. Results from other regions of interest (**Figure 5.5**) did not suggest the clear (or significant) differences that were apparent in V3B/KO. As a further line of evidence, we used regression analyses to test the relationship between psychophysical and fMRI measures of integration. While we would not anticipate a one-to-one mapping between them (the fMRI measure is likely to be more variable) our group-based analysis suggested a correspondence. We found a significant relationship between the fMRI and psychophysical integration indices in V3B/KO (**Figure 5.6**) for the concurrent ($R$=0.57, $p$=.026) but not the binary luminance ($R$=0.10, $p$=.731) condition. This result was specific to area V3B/KO (**Table 5.2**), and, in line with the preceding analyses, suggests a relationship between activity in area V3B/KO and the perceptual integration of disparity and shading cues to depth.

**Table 5.3:** Results for the regression analyses relating the psychophysical and fMRI integration indices in each region of interest. The table shows the Pearson correlation coefficient (R) and the signficance of the fit as a *p* value for the 'Disparity + shading' and 'Disparity + binary luminance' conditions.

| | **Disparity + shading** | | **Disparity + binary luminance** | |
|---|---|---|---|---|
| **Cortical area** | **R** | **p-value** | **R** | **p-value** |
| V1 | -0.418 | 0.121 | -0.265 | 0.340 |
| V2 | 0.105 | 0.709 | -0.394 | 0.146 |
| V3v | -0.078 | 0.782 | 0.421 | 0.118 |
| V4 | 0.089 | 0.754 | -0.154 | 0.584 |
| LO | 0.245 | 0.379 | -0.281 | 0.311 |
| V3d | 0.194 | 0.487 | -0.157 | 0.577 |
| V3A | 0.232 | 0.405 | -0.157 | 0.577 |
| V3B/KO | **0.571** | **0.026** | 0.097 | 0.731 |
| V7 | 0.019 | 0.946 | -0.055 | 0.847 |
| hMT+/V5 | 0.411 | 0.128 | -0.367 | 0.178 |

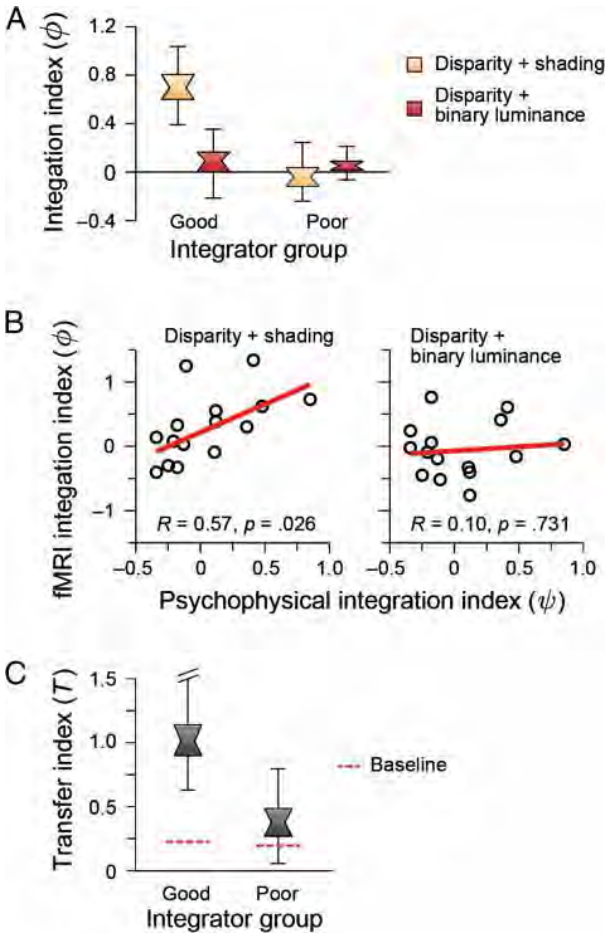



**Figure 5.6:** (A) fMRI based prediction performance as an integration index for the two groups of participants in area V3B/KO. A value of zero indicates the minimum bound for fusion as predicted by quadratic summation. The index is calculated for the "disparity + shading" and "disparity + binary shading" conditions. Data are presented as notched distribution plots. The center of the"bowtie" respresents the median, the coloured area depicts 68% confidence values, and the upper and lower error bars represent 95% confidence intervals. (b) Correlation between behavioral and fMRI integration indices in area V3B/KO. Psychophysics and fMRI integration indices are plotted for each participant for disparity + shading and disparity + binary luminance conditions. The Pearson correlation coefficient (R) and *p*-value are shown. (c) The transfer index values for V3B/KO for the good and poor integrator groups. Using this index, a value of 1 indicates equivalent prediction accuracies when training and testing on the same cue vs. training and testing on different cues. Distribution plots show the median, 68% and 95% confidence intervals. Dotted horizontal lines depict a bootstrapped chance baseline based on the upper 95$^{th}$ centile for transfer analysis obtained with randomly permuted data.

As a final assessment of whether fMRI responses related to depth structure from different cues, we tested whether training the classifier on depth configurations from one cue (e.g. shading) afforded predictions for depth configurations specified by the other (e.g.

disparity). To compare the prediction accuracies on this cross-cue transfer with baseline performance (i.e., training and testing on the same cue), we used a bootstrapped transfer index:

$$\square = 2\square'\square\ \square'\square + \square'\square \quad \text{(Eqn. 5.4)}$$

where $\square'\square$ is between-cue transfer performance. A value of one using this index indicates that prediction accuracy between cues equals that for testing within cues. To provide a baseline for the amount of transfer that might occur by chance, we calculated the transfer index on randomly shuffled data (1000 tests per ROI). We used the 95th centile of the resulting distribution of transfer indices as the cut-off for statistical significance. We found reliable evidence for transfer between cues in area V3B/KO (**Figure 5.7; Table 3**) for the good, but not poor, integrator groups. Moreover, this effect was specific to V3B/KO and was not observed in other areas. Together with the previous analyses, this result suggests a degree of equivalence between representations of depth from different cues in V3B/KO that is related to an individual's perceptual interpretation of cues.

**Table 5.4:** Probabilities associated with the transfer between disparity and shading producing a Transfer index above the random (shuffled) baseline. These p-values are calculated using bootstrapped resampling with 10,000 samples. Bold formatting indicates Bonferroni-corrected significance.

| Cortical area | Good integrators | Poor integrators |
|---|---|---|
| V1 | 0.247 | 0.748 |
| V2 | 0.788 | 0.709 |
| V3v | 0.121 | 0.908 |
| V4 | 0.478 | 0.062 |
| LO | 0.254 | 0.033 |
| V3d | 0.098 | 0.227 |
| V3A | 0.295 | 0.275 |
| V3B/KO | **<0.001** | 0.212 |
| V7 | 0.145 | 0.538 |
| hMT+/V5 | 0.124 | 0.302 |

To ensure we had not missed any important loci of activity outside the areas we sampled using our region of interest localizer approach, we conducted a searchlight classification analysis (Kriegeskorte, Goebel, & Bandettini, 2006) in which we moved a small aperture (9 mm) through the sampled volume performing MVPA on the difference between stimulus configurations for the concurrent cue condition (Fig. 2). This analysis indicated that discriminative signals about stimulus differences were well captured by our region of interest definitions.

Our main analyses considered MVPA of the fMRI responses partitioned into two groups based on psychophysical performance. To ensure that differences in MVPA prediction performance between groups related to the pattern of voxel responses for 3D processing,

rather than the overall responsiveness of different regions of interest, we contrasted the average fMRI activations (% signal change) in each ROI for the two groups of participants. Reassuringly, we found no evidence for statistically reliable differences between groups in any of the measured ROIs. Further, we ensured that we had sampled from the same cortical location in both groups by calculating the mean Talairach location of V3B/KO subdivided by groups (Table 5.1). This confirmed that we had localized the same cortical region in both groups of participants.

To guard against artefacts complicating the interpretation of our results, we took specific precautions during scanning to control attentional allocation and eye movements. First, the participants performed a demanding Vernier judgement task at fixation. This ensured equivalent attentional allocation across conditions, and, as the task was unrelated to the depth stimuli, psychophysical judgements and fMRI responses were not confounded and could not thereby explain between-subject differences. Second, the attentional task served to provide a subjective measure of eye vergence (Popple *et al.*, 1998). In particular, participants judged the relative location of a small target flashed (250 ms) to one eye, relative to the upper vertical nonius line (presented to the other eye). We fit the proportion of "target is to the right" responses as a function of the target's horizontal displacement. Bias (i.e. deviation from the desired vergence position) in this judgement was around zero. Using a repeated measures ANOVA, we found that there was no effect of condition ($F_{3, 42}$=2.59, $p$ =0.07) or curvature sign ($F_{1, 14}$=1.43, $p$=0.25), and no interaction ($F_{3, 42}$=1.95, $p$=0.14). Moreover, there were no differences in the slope of the psychometric functions: no effect of condition ($F_{3, 42} < 1$, p =0.82) or curvature ($F_{1, 14} < 1$, p =0.80), and no interaction ($F_{3, 42} < 1$, p =0.85). Third, our stimuli were constructed to reduce the potential for vergence differences: disparities to the left and right of the fixation point were equal and opposite, a constant low spatial frequency pattern surrounded the stimuli, and participants used horizontal and vertical nonius lines to monitor their eye vergence. Finally, we recorded horizontal eye movements for three

participants inside the scanner bore. Analysis of the eye position signals suggested that the participants were able to maintain steady fixation: in particular, deviations in mean eye position were < 1 degree from fixation. A repeated measures ANOVA showed no significant difference between conditions in mean horizontal eye position ($F_{3, 6} < 1$, $p = 0.99$), number of saccades ($F_{3, 6} < 1$, $p = 0.85$), or saccade amplitude ($F_{3, 6} = 1.57$, $p = 0.29$).

## 5.4 Discussion

Here we provide three lines of evidence that activity in the dorsal visual area V3B/KO reflects the integration of disparity and shading depth cues in a perceptually-relevant manner. First, we used a quadratic summation test to show that performance in concurrent cue settings improves beyond that expected if depth from disparity and shading are collocated but represented independently. Second, we showed that this result was specific to stimuli that are compatible with a three-dimensional interpretation of shading patterns. Third, we found evidence for cross-cue transfer. Importantly, the strength of these results in V3B/KO varied between individuals in a manner that was compatible with their perceptual use of integrated depth signals.

These findings complement evidence for the integration of disparity and relative motion in area V3B/KO (Ban et al., 2012), and importantly suggest both a strong link with perceptual judgments and a more generalized representation of depth structure. Such generalization is far from trivial: binocular disparity is a function of an object's 3D structure, its distance from the viewer and the separation between the viewer's eyes; by contrast, shading cues (i.e., intensity distributions in the image) depend on the type of illumination, the orientation of the light source with respect to the 3D object, and the reflective properties of the object's surface (i.e., the degree of Lambertian and Specular reflectance). As such disparity and shading provide complementary shape information: they have quite different

generative processes, and their interpretation depends on different constraints and assumptions (Blake, Zisserman, & Knowles, 1985; Doorschot, Kappers, & Koenderink, 2001). Taken together, these results indicate that the 3D representations in the V3B/KO region are not specific to specific cue pairs (i.e., disparity-motion) and generalize to more complex forms of 3D structural information (i.e., local curvature). This points to an important role for higher portions of the dorsal visual cortex in computing information about the 3D structure of the surrounding environment.

### 5.4.1 Individual differences in disparity and shading integration

One striking, and unexpected feature of our findings was that we observed significant between-subject variability in the extent to which shading enhanced performance, with some subjects benefitting, and others actually performing worse. What might be responsible for this variation in performance? While shading cues support reliable judgments of ordinal structure (Ramachandran, 1988), shape is often underestimated (Mingolla & Todd, 1986) and subject to systematic biases related to the estimated light source position (Curran & Johnston, 1996; Mamassian & Goutcher, 2001; Pentland, 1982; Sun & Perona, 1998) and composition (Schofield, Rock, & Georgeson, 2011). Moreover assumptions about the position of the light source in the scene are often esoteric: most observers assume overhead lighting, but the strength of this assumption varies considerably (Liu & Todd, 2004; Thomas, Nardini, & Mareschal, 2010; Wagemans, van Doorn, & Koenderink, 2010), and some observers assume lighting from below (e.g., 3 of 15 participants in Schofield et al, 2011). Our disparity + shading stimuli were designed such that the cues indicated the same depth structure to an observer who assumed lighting from above. Therefore, it is quite possible that observers experienced conflict between the shape information specified by disparity, and that determined by their interpretation of the shading pattern. Such participants would be 'poor integrators' only inasmuch as they failed to share the assumptions typically made by observers (i.e., lighting

direction, lighting composition, and Lambertian surface reflectance) when interpreting shading patterns. In addition, participants may have experienced alternation in their interpretation of the shading cue across trials (i.e., a weak light-from-above assumption which has been observed quite frequently, Schofield et al, 2011; (Thomas et al., 2010; Wagemans et al., 2010)); aggregating such bimodal responses to characterize the psychometric function would result in more variable responses in the concurrent condition than in the 'disparity' alone condition which was not subject to perceptual bistability. Such variations could also result in fMRI responses that vary between trials; in particular, fMRI responses in V3B/KO change in line with different perceptual interpretations of the same (ambiguous) 3D structure indicated by shading cues (Preston, Kourtzi, & Welchman, 2009). This variation in fMRI responses could thereby account for reduced decoding performance for these participants.

An alternative possibility is that some of our observers did not integrate information from disparity and shading because they are inherently poor integrators. While cue integration both within and between sensory modalities has been widely reported in adults, it has a developmental trajectory and young children do not integrate signals (Gori, Del Viva, Sandini, & Burr, 2008; Nardini, Bedford, & Mareschal, 2010; Nardini, Jones, Bedford, & Braddick, 2008). This suggests that cue integration may be learnt via exposure to correlated cues (Atkins, Fiser, & Jacobs, 2001) where the effectiveness of learning can differ between observers (Ernst, 2007). Further, while cue integration may be mandatory for many cues where such correlations are prevalent (Hillis, Ernst, Banks, & Landy, 2002), inter-individual variability in the prior assumptions used interpret shading patterns may cause some participants to lack experience of integrating shading and disparity cues (at least in terms of how these are studied in laboratory settings).

These different possibilities are difficult to distinguish from previous work that has looked at the integration of disparity and shading signals. This work indicated that perceptual judgments are enhanced by the combination of disparity and shading cues (Buelthoff & Mallot, 1988; Doorschot *et al.*, 2001; Lovell *et al.*, 2012; Schiller *et al.*, 2011; Vuong *et al.*, 2006). However, between-participant variation in such enhancement is difficult to assess given that low numbers of participants were used (mean per study = 3.6, max = 5) a sizeable proportion of whom were not naïve to the purposes of the study. Here we find evidence for integration in both authors H.B. and A.W., but considerable variability among the naïve participants. In common with Wagemans et al (2010), this suggests that interobserver variability may be significant in the interpretation of shading patterns in particular, and integration more generally, providing a stimulus for future work to explain the basis for such differences between individuals.

### 5.4.2 Responses in other regions of interest

When presenting the results for all the participants, we noted that performance in the disparity + shading condition was statistically higher than for the component cues in area V2 as well as in V3B/KO (**Fig. 5.3**). Our subsequent analyses did not provide evidence that V2 is a likely substrate for the integration of disparity and shading depth cues. However, it is possible that the increased decoding performance—around the level expected by quadratic summation—is due to parallel representations of disparity and shading information. It is unlikely that either signal is fully elaborated, but V2's more spatially extensive receptive fields may provide important information about luminance and contrast variations across the scene that provide signals important when interpreting shape from shading (Schofield et al., 2010).

Previous work (Georgieva et al., 2008) suggested that the processing of 3D structure from shading is primarily restricted in its representation to a ventral locus near the area we

localize as LO (although (Gerardin, Kourtzi, & Mamassian, 2010) suggested V3B/KO is also involved and (Taira, Nose, Inoue, & Tsutsui, 2001) reported widespread responses). Our fMRI data supported only weak decoding of depth configurations defined by shading in LO, and more generally across higher portions of both the dorsal and ventral visual streams (**Figs. 5.3, 5.4**). Indeed, the highest prediction performance of the MVPA classifier for shading (relative to overall decoding accuracies in each ROI) was observed in V1 and V2 which is likely to reflect low-level image differences between stimulus configurations rather than an estimate of shape from shading *per se*. Nevertheless, our findings from V3B/KO make it clear that information provided by shading contributes to fMRI responses in higher portions of the dorsal stream. Why then is performance in the 'shading' condition so low? Our experimental stimuli purposefully provoked conflicts between the disparity and shading information in the 'single cue' conditions. Therefore, the conflicting information from disparity that the viewed surface was flat is likely to have attenuated fMRI responses to the 'shading alone' stimulus. Indeed, given that sensitivity to disparity differences was so much greater than for shading, it might appear surprising that we could decode shading information at all. Previously, we used mathematical simulations to suggest that area V3B/KO contains a mixed population of responses, with some units responding to individual cues and others fusing cues into a single representation (Ban et al., 2012). Thus, residual fMRI decoding performance for the shading condition may reflect responses to non-integrated processing of the shading aspects of the stimuli. This mixed population could help support a robust perceptual interpretation of stimuli that contain significant cue conflicts: for example, the reader should still be able to gain an impression of the 3D structure of the shaded stimuli in **Fig. 5.1**, despite conflicts with disparity).

In summary, previous fMRI studies suggest a number of locations in which three-dimensional shape information might be processed (Nelissen *et al.*, 2009; Sereno, Trinath, Augath, & Logothetis, 2002). Here we provide evidence that area V3B/KO plays an important

role in integrating disparity and shading cues, compatible with the notion that it represents 3D structure from different signals (Tyler et al., 2006) that are subject to different prior constraints (Preston, Kourtzi, & Welchman, 2009). Our results suggest that V3B/KO is involved in 3D estimation from qualitatively different depth cues, and its activity may underlie perceptual judgments of depth.

# CHAPTER 6:

# Learning to integrate shading and disparity cues to estimate 3D shape

---

To estimate a coherent 3D shape, the visual system exploits multiple sources of information. The cues to shape estimation differ in their nature. Shape estimates from binocular disparity and motion parallax are quantitative and metric, and can be understood with easily tractable mathematical models. Pictorial cues such as shading, on the other hand, provide only qualitative estimates of shape and are not mathematically tractable: to make sense of such cues, observers must make additional assumptions (e.g. that the light is coming from above) or apply constraints from other, more metric, cues. Despite their computationally different nature, shading and disparity cues are integrated to produce coherent shape estimation, and it has been shown that using these cues together can improve perceptual judgements. However, the extent to which an observer benefits from the integration of disparity and shading varies idiosyncratically. Here, I explore individual differences in cue integration. First, I measure sensitivity for slight changes in depth profiles of convex ('bump') and concave ('dimple') surfaces defined by shading, disparity, and combination of shading and disparity. I find that for some observers, depth structures are more discriminable when shading and disparity are both present. In contrast, other observers show either no benefit or poorer performance in the combined shading and disparity condition than they do for disparity or shading alone. I used a learning paradigm to probe experience dependent changes in shading and disparity integration, and show that non-integrating observers can learn to integrate the two cues after training with feedback.

## 6.1 Introduction

Estimating shape is a crucial ability of the visual system that aids interaction with the environment. There are multiple sources of information on shape, and the visual system exploits these various cues to come up with the most likely interpretation (Bülthoff & Mallot, 1988; Dosher *et al.*, 1986; Landy *et al.*, 1995), using multiple cues to improve shape estimates (Knill & Saunders, 2003). Regardless of the computational differences between the shape estimates provided by disparity and shading, it has been shown that the human visual system integrates these cues when they signal the same shape (Bülthoff & Mallot, 1988; Lovell *et al.*, 2012; Schiller *et al.*, 2011; Vuong *et al.*, 2006). In the previous chapter, we demonstrated that observers can benefit from having shading and disparity information together by considering perceptual judgements of shape and their correlated neural activity patterns (Chapter 5). Additionally, we encountered a systematic variation in the extent to which individuals integrate these cues. Some integrated the two cues super-optimally, going beyond the predicted boundary for independent processing; others did not integrate the cues at all. Here, we investigate these individual differences further while probing experience-related changes in people's ability to integrate disparity and shading cues.

When multiple cues to shape are available, the interaction between the cues may result in a fusion, where redundant information is discarded to efficiently come up with a coherent percept. During this process, more reliable cues moderate the influence of less reliable cues by adjusting their weights in the integration process (Adams *et al.*, 2010; Atkins, Fiser, & Jacobs, 2001; Ernst, Banks, & Bulthoff, 2000; Landy *et al.*, 1995). When the cues occur in a statistically meaningful fashion, the visual system calibrates the cue weights in the integration process according to the reliability of each cue, how frequently they occur, and also the consistency between the available cues. Binocular disparity is a robust cue to depth (and hence shape) and can be understood via mathematical models which produce a metric

shape geometry. In contrast, qualitative (pictorial) shape cues such as shading do not provide straightforward estimates of shape: additional calculations and assumptions must be applied obtain a coherent estimate of shape from shading. The shading cue is particularly intriguing because of the prior assumptions necessary to infer shape from it. The light-from-above prior is one such assumption (Kleffner & Ramachandran, 1992; Schofield *et al.*, 2011; Sun & Schofield, 2012) and is consistent with the natural statistics of the environment (Dror, Willsky, & Adelson, 2004). Individuals' different assumptions of the lighting prior may result in a discrepancy between shape inferences or, similarly, an observer's inability to adopt a consistent lighting source might result in ambiguous perception of the shading signal. In such cases, the shading cue would be very unreliable and hence this cue would receive a low weight in any cue integration process. Thus for these individuals shading would not contribute to estimates of shape when a more reliable cue such as disparity was available.

Although reliable, perhaps even mandatory, cue integration is established in adults, recent studies show that infants lack the ability to combine visual signals to form an improved estimate (Burr & Gori, 2012; Nardini *et al.*, 2010). This finding suggests that cue integration might be a learnt ability for the human visual system.

In Chapter 5, we observed large individual differences for the integration of disparity and shading when estimating 3D surface shape. Our results showed that while some observers clearly benefit from fusing these cues, about half of the population perform at the level of a single cue, or even worse, when disparity and shading concurrently signal the same shape. Here, I focus on poor integrators and ask if more optimal integration of disparity and shading can be learnt via appropriate training. To investigate this, I first measured discrimination thresholds in shape profiles for each cue alone and for their combination. The cues depicted concave and convex surfaces (Figure 5.1a). Next, I quantified the benefit from cue integration using a method based on the quadratic summation test (see Chapter 5).

Finally, I trained observers who performed sub-optimally with combined disparity and shading conditions, and also with a control condition combining disparity with binary luminance. I hypothesise that repeated exposure improves sensitivity for both disparity and shading. Moreover, I aim to demonstrate that after training, sub-optimal integrators achieve cue integration, with performance exceeding the minimum limit implied by independent processing of the two cues.

In Chapter 5, we have seen significant variability between subjects in terms of their benefit from having both disparity and shading cues signaling the same shape. While some observers benefitted from having both cues present, others performed better when disparity cue alone signaled the shape. In this chapter, I look into the behavioural paradigm closer with a group of naïve participants; and investigate whether observers who initially perform worse in the combined cue condition can be trained to benefit from having the two cues present at the same time.

## 6.2 Materials and Methods

### 6.2.1 Observers

Twelve experimentally naive observers from the University of Birmingham participated in the experiments. The participants had normal or corrected to normal vision. The experiments were approved by the local ethics committee; all observers gave written informed consent.

### 6.2.2 Stimuli and Methods

The stimuli were similar to those described in Chapter 5. The random dot stereograms had disparity and/or shading cues to depict two concave and two convex hemispheres of radius 1.7° and depth amplitude of 6 arcmin. The spatial configuration of the hemispheres was either convex on the left of fixation and concave on the right, or vice versa, this configuration being randomised from trial to trial to avoid adaptation effects. Elements of the RDS subtended 0.06° and there were approximately 94 dots per degree. I wanted to

increase the benefit from the shading cue by using dense patterns with smaller dots. The shading model followed the Blinn-Phong shading algorithm, with a directional light source that was positioned at infinity above the observer's head, making a 45° angle with the surface normal. We used three component cues: (i) disparity only: where the intensity of dots in the display followed the luminance distribution of the shaded patterns, but their positions were randomised across the shape; (ii) shading only: where disparity indicated a flat surface; (iii) binary luminance only, where luminance levels were binarised and disparity indicated a flat surface. To create the binary luminance stimuli, the luminance of the top and bottom portions of the hemispheres was held constant at the mean luminance of the corresponding portions in the shaded stimuli. We also created two composite cues from: (iv) disparity + shading; and (v) disparity + binary luminance, where disparity and the shading or binary luminance cue specified the same shape. The random dot pattern subtended 8×8° and was surrounded by a larger, peripheral grid (18×14°) of black and white squares which served to provide a stable background reference. All other parts of the display were set to mid-grey.

Stimuli were presented to the participants in a lab setting using a stereo set-up in which the two eyes viewed separate CRTs (ViewSonic FB2100x) through front-silvered mirrors at a distance of 50 cm. Linearisation of the graphics card outputs was achieved using photometric measurements. Images were displayed at 100Hz with a screen resolution of 1600 x 1200 pixels.

Participants sequentially viewed two images and decided which stimulus had the greater depth profile (presentation time = 500 msec, interstimulus interval = 500 msec). In two randomly ordered intervals, one interval contained a standard stimulus (± 6 arcmin) defined by one of the five cues and the other interval contained a stimulus defined by the same cue varying in depth magnitude. In the test blocks, the QUEST method was used to estimate 82% correct discrimination thresholds: the QUEST algorithm started at a maximum

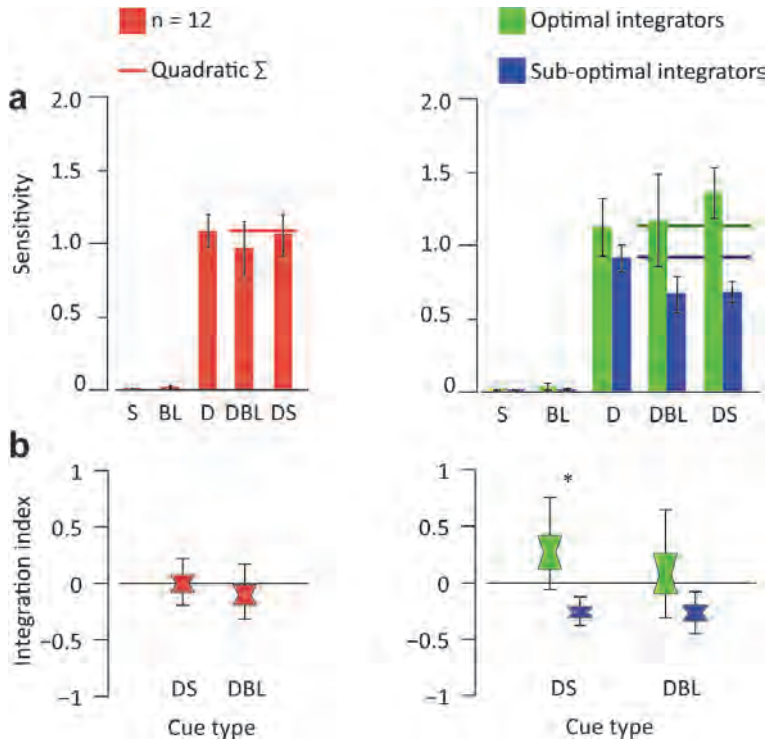
intensity difference of 4.5 arcmin in the first trial. Threshold values were converted to sensitivity (sensitivity = 1/threshold) before interpreting results.

To screen for idiosyncrasies in integrating the two cues, an initial test block containing 150 trials (15 trials × 5 conditions × 2 curvature configurations) was applied and an integration index was calculated for each observer. First, participants completed a screening test where they viewed 3 single cues (shading, disparity, binary luminance) and 2 composite cues (disparity+shading, disparity+binary luminance). This was done to screen for idiosyncrasies in integrating the two cues, and it contained 150 trials (15 trials × 5 conditions × 2 curvature configurations). After this session, an integration index calculated for each observer and the group of subjects were divided into two: those who benefit from having both disparity and shading signalling the same shape (good integrators), and those who perform worse in disparity+shading condition (poor integrators). This was followed by the training sessions. Because there was a gap of 1-5 days between screening test and first day of training, poor integrators repeated a follow up test, which was used to establish pre-training integration levels for Experiment 2. During training, participants completed 2-3 blocks (120 trials per block) for 3 days. Participants viewed composite cues (disparity + shading and disparity + binary luminance) in training blocks, which included 120 trials (30 trials × 2 conditions × 2 curvature configurations). After every trial, we presented a high pitch tone 'beep' for correct, and a low pitch tone 'boop' for incorrect responses. During training blocks, the threshold and standard deviation estimates from the previous block were fed into the subsequent block as the initial values. Conditions in a block were randomly interleaved. One day after the last training session, participants completed post-training test (2-3 blocks) which was the same as the screening test.

## 6.3 Results

### 6.3.1 Experiment 1: Quantifying individual differences in cue integration

We used QUEST to measure discrimination thresholds in a task where the observers judged the depth profiles of the stimuli. Within a block, both the component cues (shading, disparity, binary luminance) and the composite cues ('disparity and shading', 'binary luminance and shading') were shown. A block included five separate but interleaved QUEST procedures (30 trials each) to estimate a threshold for discriminating the depth thresholds for each cue / cue combination. Out of twelve participants, six showed higher sensitivity for the composite cue 'disparity + shading', suggesting better than optimal integration of disparity and shading (optimal integrators); the remaining six observers performed less well with the composite cue as compared to the disparity only condition (sub-optimal integrators). Measures of the shading- and binary-luminance-only cases were limited in the sense that both the thresholds, and their variance were very high, making it difficult to compare sensitivity for shading to that for disparity (i.e. $\square\square 2 \gg \square\square 2$).




igure 6.1: Observers' performance in discriminating slight differences in five cues: shading S), binary luminance (BL), disparity (D), disparity + binary luminance (DBL), and disparity + hading (DS). (a) The bar graphs show sensitivity based on the threshold estimate for each ue. Error bars indicate ±SEM. Data averaged across all participants is shown on the left (red ars), and on the right, group means are separately shown for optimal (n = 6, green bars) and ub-optimal integrators (n = 6, blue bars). Quadratic summation prediction is marked with a orizontal line in each plot. (b) Distribution plots represent discrimination performance as an ıtegration index calculated for two composite cues, disparity + shading (DS) and disparity + inary luminance (DS). Bow-tie plots mark mean, 68% and 90% confidence intervals

We quantified the differences in interpreting disparity + shading cues with an integration index ($\psi$) originating from a quadratic summation test (see also Chapter 5, Equation 5.2).

$$\psi = \psi + \psi - 1$$

A value of $\psi$ around zero would indicate that disparity and shading are processed independently, while positive $\psi$ values suggest the fusion of these two cues. After rank

ordering participants based on $\psi$ value (Figure 6.1c), I formed two groups: For six participants, the disparity + shading indices fell above zero (optimal integrators) while for the remaining six participants the integration index was below zero (sub-optimal integrators). A repeated measures ANOVA between 5 cues revealed a significant main effect ($F_{1.4,\ 13,7} = 51.57$, $p<10^{-6}$) and an interaction with the two groups as a between-subjects factor ($F_{1.4,\ 13.7} = 4.11$, $p<0.05$). Group means across the participants showed higher sensitivity measures for optimal integrators (Figure 6.1a, green bars) when compared to sub-optimal integrators (blue bars, post-hoc pairwise comparisons, $p<0.05$).

The difference was also confirmed using a quadratic summation test: Optimal integrators showed a super-quadratic summation performance for 'disparity and shading' condition, but this was not the case for sub-optimal integrators (green and blue lines indicate quadratic summation prediction for optimal and sub-optimal integrators respectively). The integration index for disparity + shading was significantly higher for optimal integrators (Figure 6.1b, right column, 10,000 bootstrapped samples, $p<0.05$). Overall, these results suggest that participants in the optimal integrators group benefitted from having both disparity and shading together, while sub-optimal integrators performed better when disparity was presented alone. Even though sub-optimal integrators seem to be less sensitive to the disparity cue presented alone, the difference between group means for disparity sensitivity remains statistically non-significant ($t_{10} = 0.73$, $p = 0.42$). The sub-optimal performance in this group might be relevant to the interpretation of the shading cue: idiosyncrasies in shading priors (Liu & Todd, 2004; Wagemans *et al.*, 2010) could conflict with the fixed overhead lighting in our stimuli, or the internal lighting assumption might be bistable, leading to an ambiguous percept. If the latter case were true, then practice with external feedback would disambiguate the percept, either by altering the light source prior and interpretation of shading (Adams *et al.*, 2004) or merely associating a type of feedback

with the type of cues presented (Haijiang *et al.*, 2006). To investigate this, in Experiment 2, we trained the participants in the sub-optimal group with external feedback.

### 6.3.2 Experiment 2: Training with composite cues

I reasoned that if sub-optimal integration results were due to the individual differences in shading inferences that conflict with our fixed shading interpretation, then reinforcing a single interpretation with training would alter the percept, consequently resulting in an increased benefit from co-occurring disparity and shading cues. Alternatively, training might help to disambiguate bistable interpretations of the shading cue throughout a testing block, hence causing improved sensitivity for disparity + shading after training. Five days after Experiment 1, the six poor integrators were asked to complete a further test block (pre-training test), repeating the same procedure described in Experiment 1. After this, we ran training with trial-by-trial symbolic feedback for three days. During the training sessions, participants viewed two cues randomly interleaved in a block: disparity + shading and disparity + binary luminance.

One participant (OO) was excluded from further study because, unlike the sub-optimal integration she performed in Experiment 1, one week later, and with no additional exposure to the stimuli, she showed super-quadratic summation in the follow-up test (> 2 standard deviations above group mean = -.20). Although we observed increased sensitivities for all other observers (n = 5) in the follow-up test, their integration performance remained sub-optimal.

We found a significant effect of training day ($F_{2, 8} = 5.98$, $p<0.05$), but there was no significant difference between disparity + shading and disparity + binary luminance ($F_{1, 4} <1$, $p = 0.73$). After three days of training, participants undertook a post-training test (2-3 blocks) where they judged all of the five cues in every block.

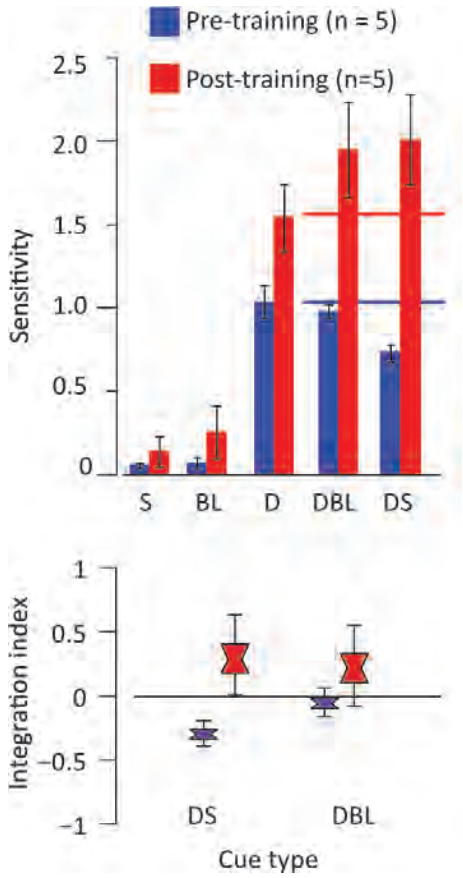



Figure 6.2: Sub-optimal observers' performance, before and after training, in discrimination sensitivity for five cues. (a) The bar graphs show sensitivity based on the threshold estimate for each cue in the tests before (blue bars) and after (red bars) training. Quadratic summation prediction is marked with a horizontal line in the corresponding colour of testing. Data averaged across 5 participants, error bars indicate ±S.E.M. (b) Distribution plots represent discrimination performance as an integration index calculated for two composite cues disparity + shading (DS) and disparity + binary luminance (DS) before and after training (blue and red respectively). Bow tie plots mark mean, 68% and 90% confidence intervals.

Figure 6.2 shows the training effect in terms of sensitivity for each cue before (blue bars) and after (red bars) training. Compared to pre-training, post-training sensitivities improved largely for all the cues ($F_{4, 16} = 73.92$, $p<10^{-9}$). The results showed a significant difference between pre- and post-training tests ($F_{1, 4} = 7.09$, $p<0.05$) and a significant interaction between the cue and test conditions ($F_{4, 16} = 6.08$, $p<0.005$). After training, the disparity and shading integration index for all sub-optimal integrators ($\psi$) increased above 0 (Figure 6.2b, red distributions, $p<0.05$). Integration indices for disparity + binary luminance

also increased for three participants, but the difference remained non-significant at the group level (Figure 6.2b, blue distributions, $p = 0.84$). Overall, after the training, the observers seem to benefit from both composite cues but this effect is more pronounced for disparity + shading cues when compared to disparity + binary luminance cues.

## 6.4 Discussion

In this study, I explored the integration of disparity and shading and experience related changes in their integration. In line with our previous results (Chapter 5), I observed systematic differences between individuals' ability to discriminate slight changes in depth profiles defined by disparity and shading. Here, I used an adaptive staircase method to show that for some, but not all, observers, when disparity and shading are presented together, discrimination sensitivities are reliably higher than when disparity is presented alone. Next, I tested this increase in sensitivity using a quadratic summation test that sets a minimum prediction boundary for the independent processing of component cues.

First, in **Experiment 1**, I showed that half of the participants (optimal integrators) showed sensitivity to disparity + shading above the minimum quadratic prediction, suggesting fusion between these cues. For the remaining observers, performance remained close to that for disparity alone, or in some cases worse than for disparity alone.

When presented alone, disparity and shading provide different estimates of shape: disparity is generally a more reliable cue from which an observer can infer a metric estimate of depth structure. Shading, on the other hand, is more ambiguous and the observer has to make assumptions about the light source. Less reliable cues such as perspective may indicate bistable percepts (van Ee, Adams, & Mamassian, 2003); similarly, the same shading pattern can be interpreted as convex if a light-from-above prior is adopted, or concave if the lighting is assumed to come from below. Even when observers adopt a single light source assumption

to make an inference about depth throughout a testing session, estimates relying solely on shading information are shown to be of lower magnitude than estimates from disparity (Mingolla & Todd, 1986).

It is worth noting that, as with the results from the previous chapter, there was a clear division in individual performances when judging depth from shading, disparity, and their combination. Half of the participants performed better than an optimal integration mechanism would predict when both cues were present, whereas the other half showed better discrimination sensitivity when disparity was presented independently. Considering that the participant populations were completely different in the two studies, the integration index ($\psi$) based on the quadratic summation test seems to be a useful measure with which to exploit the individual variation in integrating disparity and shading cues.

Next, in **Experiment 2**, I trained the sub-optimal integrators with symbolic external feedback. I observed an increase in discrimination sensitivities for composite cues after training, and this can be explained in the light of the following concepts. Here, instead of quantitatively comparing depth estimates from separately presented cues, I examined how disparity and shading cues complement each other by using only composite cues during training. In other words, I aimed to quantify the benefit gained from an additional shading cue when the perceived surface already conveys disparity information.

Idiosyncrasies related to light source assumption in interpreting shape from shading have often been reported (Adams *et al.*, 2004; Liu & Todd, 2004; Lovell *et al.*, 2012; Wagemans *et al.*, 2010). Furthermore, Adams and colleagues have shown that the light source prior can be modified during interaction with haptic cues (Adams *et al.*, 2004). In this regard, our improved integration after training might be explained by a modification of the light source assumptions towards a light-from-above prior. In other words, individual observers normalise their lighting priors to match the only shading interpretation reinforced by our

stimuli. This explanation would also explain the increased sensitivities to shading after training. Stabilisation of the interpretation of shape from shading could produce such an outcome. However, I also observed improvement for the disparity and binary luminance condition. This is an unexpected result, because the percept in this composite cue mainly relies on the disparity signal, as binary luminance is not supposed to signal 3D shape *per se*. Then again, during training, observers are equally exposed to this cue as to the disparity and shading cues. Even though binary luminance alone does not signal 3D shape, it still provides a symbolic cue (e.g. upper portion brighter for convex) that might have complemented the shape estimate from the disparity cue.

Additionally, training may compensate for the underestimation of depth estimates from the shading cue compared to disparity estimates, and this would yield an increase in gain from shading when the cues are available together. If we reconsider the findings in the light of the quadratic summation test, initially some observers perform below the limit for optimal integration, suggesting an independent processing mechanism. After training, their performance improves beyond the minimum predicted boundary for fusion of these cues. The fusion mechanism indeed predicts an increase in gain from the cues, as well as an improvement in their discrimination ability.

Given the points discussed above, our findings suggest that between-observer differences are inevitable for integration of shading and disparity to estimate 3D shape. In addition, sub-optimal integration performance can be improved via training, indicating that observers can learn to fuse shading and disparity cues. This finding has important implications for our understanding of cue integration processes, which are often considered to be mandatory but based on the instantaneous reliability of cues. My results suggest that fusion is not mandatory, but that it can be learnt based on associations between cues over time.

# CHAPTER 7:

# General Discussion

In this thesis, I have presented psychophysical and fMRI studies aiming to provide further insight into the role of shading and binocular disparity cues in shape estimation. Specifically, the goal of this thesis was to answer the following questions: (i) how does the visual system resolve ambiguities in the luminance signal to separate shading cues from material changes, and can tasks such as layer decomposition be learnt at a perceptual level; (ii) when both shading and disparity are available, how do these cues interact in the estimate 3D surface shape; (iii) what are the neural correlates for the integration of shading and disparity? In the following sections, first, I summarise the main findings of each experimental chapter. Then I list the contributions of these findings to the literature, and I close the thesis with a conclusion.

## 7.1 Summary of Findings

### 7.1.1 Chapter 3: Adaptive learning to use first- and second-order signals when inferring shape from shading

Luminance variations in a scene are inherently ambiguous, as they can be caused by changes in the light source position, different surface geometries, or different surface materials. The visual system can use the relationship between first-order luminance variations (LM) and second-order amplitude modulations (AM) as a cue to discriminate the potential causes of a luminance change (layer decomposition, Kingdom, 2008). However, naïve observers do not make use of the relationship between these cues at short presentation times. In the first experimental chapter I looked at whether the relationship between first- and second-order cues can be used for layer decomposition following training with binocular disparity feedback. It has been previously shown that in-phase combinations of LM and AM are seen as corrugated surfaces whereas anti-phase combinations are perceived as flat material changes (Schofield *et al.*, 2006; Schofield *et al.*, 2010). The human visual system is sensitive to small variations in local luminance amplitude (AM, second order) as well as modulations in luminance (LM, first order). Here I asked whether naïve observers could use the relation between LM/AM combinations to disambiguate the nature of luminance variations in a shading pattern, and infer a coherent percept of surface geometry.

The stimuli I present in this thesis enforces the observer to make judgments about illumination and reflectance properties of the surface with only one input that is the intensity of each pixel. The generative model assumes uniform Lambertian surfaces with a texture on. When there is directional lighting onto this surface, the luminance at a point L(x,y) is the multiplication of shading S(x,y) and the reflectance R(x,y). The training reinforces in-phase

grating to be seen as a corrugated surface. In other words, when local mean luminance changes (LM) positively correlate with the changes in local luminance amplitude (AM), then the shading pattern is more likely to be caused by a change in surface shape. This is mainly because when there is uniform texture on the surface, the difference between dark and light pixels will vary with the change in mean luminance (i.e. LM+AM). On the other hand, when this correlation is negative (anti-phase, LM-AM), feedback suggests this to be interpreted as texture changes on a flat surface, so that the variations in the luminance pattern would be seen as painted stripes on a flat surface.

Initially, a plaid consisting of an in-phase and an anti-phase grating was presented for 1 s in a single interval while the observers judged which grating seemed more corrugated (Chapter 3, Experiment 1). In this part of the study, the observers' performance was at chance level, showing no ability to discriminate phase relationships of LM/AM mixes. This initial test was followed by training with feedback, where in- and anti-phased gratings were shown separately above and below fixation instead of a plaid configuration. A disparity defined corrugated surface was superimposed on the in-phase grating, and a flat disparity surface was superimposed on the anti-phase grating in the feedback. This training was followed by a post-training test using plaid stimuli. I reasoned that if observers are trained to use phase information when LM/AM mixes are presented separately, then they would transfer this ability to a post-training test for judging LM/AM mixes in a plaid configuration because it was previously shown that the discrimination performance improved when LM/AM mixes are presented in a plaid. Training took effect very quickly and saturated after the second training session for most of the observers. In the post-training test, I observed that performance was similar for LM/AM mixes in the trained orientation (in-phase component orientated at 45°), and a novel orientation (in-phase component orientated at 135°). The rapid learning effect

and full transfer of learning to 90° rotation to untrained stimulus is atypical of learning at a perceptual level. Even though there are examples of fast perceptual learning (e.g. Seitz *et al.*, 2005), the reported effects are stimulus- and task-specific. In a follow-up test 20 days afterwards (Chapter 3, Experiment 2), I observed that performance remained at a same level as the immediate post-training test. This time, to further investigate the nature of learning, observers were trained with a reversed feedback to reinforce the in-phase grating being seen as flat. I hypothesised that if learning were at a perceptual level, reverse reinforcement would not affect observers' phase discrimination sensitivity, or at least it would take time to flip the perception of in-phase gratings from corrugated to flat. The results revealed a sudden reversal (within 1 hour) of depth judgements, and there was no deterioration of discrimination sensitivity after this training with reversed feedback.

In summary, I showed that naïve observers could learn to benefit from initially ambiguous phase relationships of first- and second-order signals to discriminate shading patterns caused by luminance changes from those caused by material changes. This learning generalises to other orientations, and discrimination performance is susceptible to reversed training. The overall results imply that the observed learning here can be characterised as an association between the stimulus and feedback, such as labelling; rather than as perceptual learning.

### 7.1.2 Chapter 4: Perceptual learning of layer decomposition when inferring shape from shading

In Chapter 4, we aimed to gain further insight onto how observers can use shading information from an albedo textured surface to infer shape. Following the results from the previous chapter, we used first- and second-order signals to analyse shading patterns, this time aiming to demonstrate changes at a perceptual level. When two LM/AM mixes are presented in an orthogonal plaid configuration for long enough, the visual system can benefit

from phase relationship information to disambiguate the cause of shading pattern, i.e. to decompose two layers of the plaid: The in-phase mix being seen as a uniform reflectance corrugated surface and anti-phase mix being seen as a flat surface with varying reflectance.

When in- and anti-phase combinations are presented briefly (250 ms), naive observers cannot use the phase relationship to decompose the layers of a plaid. The studies reported in Chapter 4 do not present a systematic analysis of the effect of presentation time per se, but they rely on previously reported studies using the same stimuli. In a depth mapping experiment by Schofield et al. (2006), naïve observers (5 out of 6) were given unlimited time to view LM/AM mixes and indicate which of the two probe dots appeared closer to them in depth. In the following two experiments, the authors used a two-interval forced choice design to present LM/AM mixes, and single signals (LM-only, AM-only) for 1400 ms (interstimulus interval = 1400 ms) where they asked naïve observers to indicate whether the first or second stimulus appeared to have greater depth. Overall from these three experiments, authors conclude that naïve observers tend to see in-phase LM/AM combination as having the greater depth followed by LM-only signal, and anti-phase LM/AM combination. They also report that when anti-phase grating is presented alone, naïve observers tend to see it corrugated but when it is presented together with the in-phase grating in a plaid configuration, anti-phase component is seen as almost flat. These findings were duplicated in a more recent study (Schofield et al., 2010), where naïve participants (4 out of 5) adjusted the depth amplitude of the haptic stimulus (sinusoidal undulations) to match the visual stimulus (LM/AM mixes) with no limitation on viewing time. These findings show that naïve observers can discriminate phase relationships of LM and AM gratings for long stimulus durations. In Chapter 4, I use 250 ms as the presentation time for the plaid stimulus, and as can be seen from the initial exposure data reported in Experiment 1 (Section 4.3.2), I show that naïve observers cannot discriminate in-phase grating from anti-phase grating before they go through training. However, after training observers can discriminate

phase relationships to match in-phase grating with greater depth in short presentation durations.

We trained naive observers with intermittent feedback for five to ten days (Chapter 4, Experiment 1). Consequently, the observers were able to benefit from the phase relationships of LM/AM combinations to disambiguate the cause of luminance variations in shading patterns. Next, when the training effect was tested with novel stimuli differing in overall orientation (Chapter 4, Experiment 1) and spatial frequency (Chapter 4, Experiment 2), we found performance to be significantly lower (orientation) and equivalent to initial exposure (spatial frequency). These results showed that the learning was specific to the stimulus dimensions to which observers were exposed during training. Finally, we manipulated the combination angle of the LM/AM components superimposed as a plaid (Chapter 4, Experiment 3). The results showed a learning effect for small shear angles, but we did not observe any generalisation to higher shear angles. This partial transfer indeed emphasises that the orthogonal layout of the plaid is crucial to the layer decomposition process.

When considered on its own, results reported in Chapter 4, Experiment 3, might not fully convince that learning is specific to the spatial frequency of the training stimulus. I show post-test data on two levels of spatial frequency (2 and 4 c/deg) as evidence for a lack of transfer after training with low spatial frequency (0.5 c/deg) plaids. Having said that, previous evidence (Experiment 4, Georgeson & Schofield, 2002) show that in- and anti-phase gratings are discriminable at a high spatial frequency (1 c/deg). Furthermore, Sun and Schofield (2011) look at how the ratio of modulation (LM) and carrier (AM) frequencies correlates with the layer decomposition process, i.e. discriminating the phase relationship of LM/AM mixes. In a depth comparison task, they test a range of 1 – 16 c/deg carrier frequencies, and report very similar perceived depth amplitudes for 1 and 2 c/deg stimuli (Experiment 1, Figure 3 in the paper) suggesting that observers can use the phase

relationship as a cue to layer decomposition even for high frequency (2 c/deg) AM signal. In the light of these two examples, I concluded that observers failed to transfer the effects of training with 0.5 c/deg to higher spatial frequencies.

Nevertheless, to fully understand the specificity of learning to the spatial frequency, one can directly investigate the effect of modulation and carrier frequency on learning. One way to do this might be to exploit the effects of training with high frequency stimuli and ask whether the training transfers to low spatial frequencies.

In these results, the performance after training can be explained by a computational model where first- and second-order signals are processed through separate channels at first, and then combined to produce a shading signal (Schofield *et al.*, 2010). The model accounts for the processing of the shading signal as automatic and quick, such that the earlier visual mechanisms would be involved. The specificity of learning effect to the trained stimulus is also in line with a perceptual change at the early stages of visual processing.

Plaid stimuli are defined by correlated first and second order luminance signals, and these signals are reported to be correlated in natural images (Johnson & Baker, 2004). Jiang, Schofield, and Wyatt (2010), present an algorithm based on the relationship between intensity, luminance amplitude, texture and colour. The image is split into components in the frequency domain. Each of the components is given a weight according to their corresponding relationship to the above cues. Weighted combinations of these components are then used to construct shading and reflectance images from a single image. Their tests using photographs of surfaces under different lighting conditions (Image database: www.bold.bham.ac.uk) suggest that luminance amplitude is a useful cue to extract intrinsic images. Undoubtedly, the stimuli used in this thesis are artificial where I try to isolate luminance variations in the shading pattern by eliminating any other cue to shape. For this reason, in- and anti-phase gratings might look indiscriminable to the naïve observer at first

instance. However, as previously reported studies suggest that under limited viewing, -even naïve observers see in-phase grating to have greater depth than the anti-phase grating.

Overall, this chapter demonstrates that AM signal's spatial phase relation to LM signal plays a crucial role in resolving potential ambiguities caused by a shading pattern. Although at long presentation times, naïve observers can benefit from these signals to interpret shape from shading, training is necessary to accomplish this at shorter presentations. A computational shading model that assumes automatic processes can explain the finding that trained observers can learn to decompose layers to judge whether luminance variations are caused by geometry or material changes. Specificity to the trained stimulus suggests that training can result in changes at a perceptual level.

In its broadest sense, practice dependent changes in a perceptual task are referred to as perceptual learning, whether it is explicitly changing the decision criterion, or implicitly changing the (discrimination) sensitivity. Although perceptual learning was initially reported as quite stimulus specific (Fiorentini & Berardi, 1980; Karni & Sagi, 1991), more recently there has been growing evidence suggesting that learning transfers, for instance if the task is easy or low precision (Ahissar & Hochstein, 1997; Jeter, Dosher, Petrov, Lu, 2009). Generally, if learning is described by the change in sensitivity, an unbiased criterion is assumed and changes in decision criterion are ignored. However, more recently, Aberg and Herzog (2012) suggested that changes in sensitivity and decision learning are very different, they interact with each other and these can be distinguished by looking at the type of feedback provided, maintenance of the changes between sessions, and changes with consolidation. According to their study, the learning we observe in Chapter 3 is more likely to be changes in decision criterion: There is a tendency towards 'anti-phase has greater depth' during pre-test. This is rapidly reversed to 'in-phase has greater depth' as indicated by the feedback, but following sessions with feedback fail to show improvement in discrimination sensitivities. Moreover, in

the reverse training phase (when feedback is reversed), observers change their criteria almost immediately to match the 'correct response'.

I have reported two studies demonstrating perceptual learning. First, in Chapter 3, observers undertook training with trial-by-trial feedback and they improved their performance in the in- and anti-phase discrimination task as seen by the increase in proportion correct in pre- and post-training tests. The feedback enforced in-phase component to be matched with 'has-greater-depth' responses at every trial. During training, sensitivity change was observed only in the first two sessions, where it saturated for the following three days. Change in the sensitivity is usually the benchmark for perceptual learning; however, it has been shown that criterion changes can signal learning without any improvement in sensitivity (Aberg, Herzog, 2012). In the light of this, the results observed in Chapter 3 can be explained by a change in criterion: observers were associating the feedback with each trial, i.e. voluntarily updating the decision criterion by the explicit information provided by the trial-by-trial feedback. This can be done very quickly (e.g. criterion can be reversed within a trial after a negative feedback), as opposed to changes in sensitivity, which are reported to be dependent on consolidation (Karni, et al., 1994 -Science). Therefore, it is difficult to argue that the observer learnt to see in-phase aligned LM/AM signal as corrugated, but more likely that it was cognitive learning, where the observer generates a strategy to match the positive feedback with in-phase stimuli.

Next, in Chapter 4, feedback was provided intermittently during training. After every 20 trials, the percent of correct responses for the most recent block was presented on the screen. In addition to sensitivity improvement before and after training, change in sensitivity was also observed during training performance this time. The intermittent feedback did not provide the complete information for the observers to directly match one type of stımulus (e.g. the in-phase) with a specific feedback (e.g. the positive feedback). Since they did not

have the complete information to change their criterion, one might argue that the observed changes were involuntary and they were seen as a change in sensitivity . In this case, the nature of learning is better explained with early processes in the visual system rather than a cognitive strategy.

### 7.1.3 Chapter 5: Neural correlates of estimating shape from shading and disparity cues

In Chapter 5, we explored 3D shape estimation when both shading and disparity cues were present, aiming to assess their integration using psychophysical and fMRI techniques. Despite the different nature of depth estimates that shading and disparity can signal (pictorial vs. metric), observers benefit from their integration, as suggested by behavioural results (Bülthoff & Mallot, 1988; Lovell *et al.*, 2012; Schiller *et al.*, 2011; Vuong *et al.*, 2006). Human brain imaging studies also indicate the cortical locations involved in both processing disparity and shading (Georgieva *et al.*, 2008; Nelissen *et al.*, 2009; M. E. Sereno *et al.*, 2002), but these studies do not explore the locus for cue integration. A recent finding (Ban *et al.*, 2012) predicts that the dorsal visual region V3B/KO is a crucial locus for cue integration when processing 3D shape from disparity and motion parallax. Here, we ask whether this brain region is also involved in the integration of disparity and shading.

Using psychophysical methods, observers judged slight differences in depth profiles between sequentially presented stimuli which depicted convex and concave surfaces using random dot stereograms, Blinn-Phong shading gradients and their combination. A binary luminance and disparity defined stimulus condition was included to control for having multiple signals in the stimulus. Here we found that half of the observers benefitted from having disparity and shading signalling the same shape, showing less variance in their performance, but the other half did not benefit or even performed worse when compared to disparity defined stimuli. The amount of benefit obtained from using two cues together was

quantified using a psychophysics integration index that is based on the quadratic summation test to set a minimum bound for cue integration.

Next, in an fMRI experiment, we measured shading and disparity related cortical activation in independently localised regions of the occipital cortex. One of these regions (V3B/KO) was previously reported to be involved in processing the integration of cues to 3D depth, so a specific aim of this chapter was to probe this region with disparity and shading cues. The results were analysed with an MVPA method where a support vector machine was trained to classify the spatial configuration of the stimulus by decoding multivoxel patterns of fMRI activity. Two classes were defined: convex on the left and concave on the right, and vice versa; i.e. the classification performance at chance level was 50 per cent. We reasoned that the classification would improve when the two cues signalled the same shape together. The results revealed that in V3B/KO, for disparity and shading combined stimulus classification, performance was above the limit predicted by quadratic summation test. This suggests that the neural responses were not merely related to co-processing of disparity and shading, but rather that fMRI patterns essentially reflect the fusion of these cues.

The data were then re-analysed after separating the observers into two groups (good and poor integrators) based on their psychophysical performance. The cue integration effect in V3B/KO was carried by those observers who showed a benefit from having disparity and shading together in the psychophysics results (good integrators). In the poor integrators group, classification performance in the combined cue condition was no better than for the disparity alone condition.

Using an fMRI based integration index (similar to the quadratic summation test), we quantified the extent of improvement in classification performance for the combined cue condition for each observer. The quantifications for fMRI patterns in V3B/KO exhibited a

correlation with the individual differences in the psychophysics integration index. This relation was only observable in V3B/KO, out of the ten occipital regions examined.

All these comparisons were also tested for the binary luminance plus disparity stimuli, but no evidence for integration was observed in this case. This adds weight to the notion that our predictions for cue integration are not solely based on having multiple signals in the stimuli; on the contrary, the findings are limited to the disparity and shading combination, where both of the cues provide an interpretation of 3D shape.

Furthermore, we probed the fMRI patterns related to disparity and shading with a transfer analysis where the machine learning algorithm was trained with one cue, and tested with the other cue. Classifier performance for cross-cue transfer stayed at chance level in all of the regions tested except V3B/KO. In V3B/KO, for the good integrators, performance in the cross-cue MPVA comparison was on par with the with-in cue MVPA (where both training and testing was conducted on the same cue). We concluded that, in addition to our previous findings, this result implies that shading and disparity are represented similarly in this region. This representation might be an entity of the observers' interpretation of 3D shape isolated from shading and disparity.

At a first glance, the psychophysical results described above suggest that poor integrators lack the ability to benefit from fusion between disparity and shading, and that they base their shape estimates on only one cue. However, individual differences in our results might be caused by several reasons, including different prior assumptions made by each observer while inferring shape from shading. Poor integrators might be capable of combining the two cues under other circumstances, but the constrained lighting condition in this study might conflict with their prior assumptions. Alternatively, these observers' performance might be compromised by an unstable, alternating interpretation of shading during an experimental session.

### 7.1.4 Chapter 6: Learning to estimate shape from shading and disparity in an optimal fashion

In Chapter 6, I reported behavioural results relating to observers' ability to discriminate slight differences in the depth profiles of surfaces defined by disparity, shading, and their combination. The aim was first to gain further insight onto idiosyncrasies between individuals' shape-from-shading inferences, and the extent of the benefit they gain from the integration of shading and disparity. Moreover, I attempted to improve sub-optimal performance for cue integration inducing learning using feedback.

First, I used a QUEST procedure adaptive staircase method to measure discrimination thresholds for three component cues (shading, disparity, binary luminance) and two composite cues (disparity and shading, disparity and binary luminance). To measure this, observers were asked to judge differences in the depth profiles of two sequentially presented with-in cue stimuli. A psychophysics index (see also Chapter 5) was used to rank order the participants with respect to the extent of their benefit from viewing composite cues. The results from this test were used as an initial screening, where half of the participants showed performance above the minimum bound for optimal cue integration, and the other half of the participant group showed sub-optimal performance for cue integration. Even though there were small alterations in the methods, and the observers were naïve to the study, the idiosyncrasies showed a remarkably similar grouping to that found in Chapter 5.

Next, I asked whether the sub-optimal performance was due to ambiguous interpretation of the shading cue, and whether observers could learn to maintain a coherent percept through training. To investigate this, the sub-optimal integrators undertook three training sessions where they judged the depth profiles defined by composite cues and received symbolic feedback for every trial. Previous studies imply that cue integration can be improved with repeated exposure to multi-modal cues (Adams *et al.*, 2010; Atkins *et al.*, 2001; Burge, Girshick, & Banks, 2010; Ernst *et al.*, 2000) or in multiple visual cues to shape

(Knill, 2007). Here, I investigated the effect of repeated exposure to combined stimuli in the training. After training, discrimination thresholds improved for both trained composite cues, and the sensitivity exceeded the minimum bound for optimal cue integration. Disparity and binary luminance was again used as a control cue and, as expected, the improvement in performance was not as prominent as for the disparity and shading combination. However, a small super-quadratic summation effect was still observed for the control condition. One reason for this might be that participants were exposed to as much training with the control combination as they were the disparity and shading combination; hence their sensitivity for the control cue is also increased. After all, the binarised luminance pattern in this control condition can still act as a cue to shape: an approximation to binary shading can occur in extreme lighting conditions. In Chapter 6, I trained participants who did not initially benefit from having disparity and shading signals combined. Training took three days, during which observers judged small depth differences within disparity+shading (DS) and disparity+binary luminance (DBL) defined stimuli, and they received symbolic trial-by-trial feedback. Following this, a post-test with single (binary luminance, shading, disparity) and composite (disparity+shading and disparity+binary) cues revealed a significant improvement in sensitivity for both composite cues. I expected improvements in sensitivity for shading and disparity condition as a sign of learning to integrate these cues to estimate shape. At first glance, binary luminance cue does not seem to be as informative as the shading cue, but as can be seen from the integration indices (Figure 6.2b, p.110) before training, observers perform better for DBL (almost around the level of sensitivity for disparity) when compared to DS (worse than sensitivity for disparity). One reason might be that, however rough the binary luminance cue is, it still signals the shape; e.g. in the case of DBL, upper portion is bright when disparity signals convex, and vice versa. Considering that during training observers were exposed to co-occurring disparity and binary luminance repetitively, and explicit feedback was available in every trial, it is not difficult to suggest that they have learnt

to integrate binary luminance with disparity solely by their statistical co-occurrence (Ernst, 2007). At first glance, one might expect that disparity and binary luminance cues are combined as a linear summation showing a simple co-occurred visual event integration, where disparity and more naturalistic shading cue are fused. But the results show that after training both composite cues are improved to the same level. This might suggest a more flexible joint neural coding mechanism, where fusion can be established after training.

We correlated fMRI metrics with perceptual metrics for the integration of disparity and shading in Chapter 5. However this relationship was not observed in the disparity and binary luminance condition. It would be interesting to investigate the classification patterns from trained integrators to see if arbitrarily associated cues also improve decoding performance from fMRI patterns as they improve perceptual discrimination sensitivities.

In summary, this chapter shows that while some people initially fail to integrate shading and disparity cues, they can be trained to integrate these two cues. This apparent change in integration performance via training has important implications for our understanding of cue integration.

## 7.2 Contributions

Throughout this thesis, I have explored how shading and disparity information is processed to estimate 3D shape. This section summarises the contribution of the findings presented in this thesis to our current understanding of the human visual system.

First, in Chapter 3, I addressed how observers could use first- and second-order luminance signals in a shading pattern to overcome the ambiguities related to the cause of luminance variations (Schofield *et al.*, 2006; Schofield *et al.*, 2010). In Chapters 3 and 4, we provided evidence supporting that the human visual system is capable of using the phase relation between these signals to accomplish layer decomposition and discriminate

luminance variations caused by surface shape (related to in-phase signal) from those that are caused by surface material (related to in-phase signal). Furthermore, I studied the effect of presentation duration on this layer decomposition task. To the untrained eye, the difference between in- and anti-phase signals is invisible or takes a long time to discriminate. However, my results reveal that layer decomposition can be achieved at very short presentations after training. A computational model that adopts a quick and automatic mechanism could explain this shape-from-shading process (Schofield *et al.*, 2006; Schofield *et al.*, 2010). Perceptual learning was specific to trained stimulus dimensions, again mostly explained as changes in the early steps of perceptual processes (Fahle & Poggio, 2002; Jeter *et al.*, 2010; Karni & Sagi, 1991). Together, these results suggest that shape from shading occurs in the early stages of visual processing, and the relation of first and second-order signals can serve as a useful tool to understand this process.

Second, I reported behavioural and fMRI results to establish further understanding into the interaction of shading and disparity cues. When presented together, shading and disparity have been shown to improve perceptual judgements of 3D shape (Bülthoff & Mallot, 1988; Lovell *et al.*, 2012; Schiller *et al.*, 2011), but there is a lack of evidence on the neural correlates of this improvement. Recently, the dorsal visual area V3B/KO has been identified as an important cortical locus for integration of disparity and motion parallax cues to 3D shape (Ban *et al.*, 2012). In Chapter 5, we demonstrate evidence for V3B/KO also being involved in fusion of shading and disparity. The results presented in this chapter repeatedly imply that V3B/KO is not only important for 3D shape processing from computationally similar cues, but is also involved in the estimation of shape from qualitatively different visual cues such as disparity and shading. Moreover, we present a correlation between idiosyncrasies in behavioural estimates of shape and patterns of cortical activation in V3B/KO during combined disparity and shading processing. Finally, in Chapter 6, I quantified individual differences for cue integration between disparity and shading. The results here

show that observers who show sub-optimal cue integration improve their sensitivity for composite cues after training with feedback. That is, they learn to integrate the cues at a super-quadratic level, suggesting that fusion of disparity and shading can be learned.

## 7.3 Conclusions

In summary, the experimental findings presented in this thesis have two main implications. First, the human visual system can learn to decompose stimuli into shading and reflectance cues, and thus derive shape from shading at very short display durations, and the computational mechanisms behind this process can be explained as a layer decomposition process that takes place in the early steps of visual processing. Second, integration of disparity and shading cues improves estimates of shape when compared to estimates from either cue alone. Despite the computational discrepancy between binocular disparity and shading, the fusion of these cues is observable in patterns of cortical activation and correlating behavioural judgements. However, not all people integrate shading and disparity automatically. About half of the population fail to integrate these two cues, as has been verified using both psychophysical and neural metrics. Nonetheless, poor integrators can be trained to integrate the cues in a super-optimal fashion with relatively little training.

## References


Adams, W. J. (2007). A common light-prior for visual search, shape, and reflectance judgments. *J Vision, 7*(11), 1-7.

Adams, W. J. (2008). Frames of reference for the light-from-above prior in visual search and shape judgements. *Cognition, 107*(1), 137-150.

Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the 'light-from-above' prior. *Nat Neurosci, 7*(10), 1057-1058.

Adams, W. J., Kerrigan, I. S., & Graf, E. W. (2010). Efficient visual recalibration from either visual or haptic feedback: the importance of being wrong. *J. Neurosci., 30*(44), 14745-14749.

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol, 14*(3), 257-262.

Anzai, A., Ohzawa, I., & Freeman, R. D. (2001). Joint-encoding of motion and depth by visual cortical neurons: neural basis of the Pulfrich effect. [10.1038/87462]. *Nat Neurosci, 4*(5), 513-518.

Atkins, J. E., Fiser, J., & Jacobs, R. A. (2001). Experience-dependent visual cue integration based on consistencies between visual and haptic percepts. *Vision Res, 41*(4), 449-461.

Backus, B. T., Banks, M. S., van Ee, R., & Crowell, J. A. (1999). Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Res, 39*(6), 1143-1170.

Baker, C. L. (1999). Central neural mechanisms for detecting second-order motion. [doi: 10.1016/S0959-4388(99)80069-5]. *Current Opinion in Neurobiology, 9*(4), 461-466.

Ban, H., Preston, T. J., Meeson, A., & Welchman, A. E. (2012). The integration of motion and disparity cues to depth in dorsal visual cortex. *Nat Neurosci, 15*(4), 636-643.

Barron J.T., Malik J. (2011) High-Frequency Shape and Albedo from Shading using Natural Image Statistics. *Computer Vision and Pattern Recognition (CVPR).*

Barron, J. T., Malik, J. (2012). Shape, Albedo, and Illumination from a Single Image of an Unknown Object, *Computer Vision and Pattern Recognition (CVPR)*.

Barrow, H. G., & Tanenbaum, J. M. (1978). *Recovering intrinsic scene characteristics from images*. AI Center, SRI International, CA.

Belhumeur, P. N., Kriegman, D. J., & Yuille, A. L. (1999). The bas-relief ambiguity. *International Journal of Computer Vision, 35*(1), 33-44.

Bradley, D. C., Qian, N., & Andersen, R. A. (1995). Integration of motion and stereopsis in middle temporal cortical area of macaques. *Nature, 373*(6515), 609-611.

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

Brainard, D. H., Pelli, D. G., & Robson, T. (2002). Display characterization. In J. Hornak (Ed.), *Encyclopedia of Imaging Science and Technology* (pp. 172-188): New York: Wiley.

Brewster, D. (1826). On the optical illusion of the conversion of cameos into intaglios, and of intaglios into cameos with an account of other analogous phenomena. *Edinburgh Journal of Science, 4*, 99-108.

Buelthoff, H. H., & Mallot, H. A. (1988). Integration of depth modules - stereo and shading. *J Opt Soc Am: A, 5*(10), 1749-1758.

Burge, J., Fowlkes, C. C., & Banks, M. S. (2010). Natural-scene statistics predict how the figure-ground cue of convexity affects human depth perception. *J Neurosci, 30*(21), 7269-7280.

Burge, J., Girshick, A. R., & Banks, M. S. (2010). Visual-haptic adaptation is determined by relative reliability. *J Neurosci, 30*(22), 7714-7721.

Burr, D., & Gori, M. (2012). *Multisensory Integration develops late in humans: the neural bases of multisensory processes*. Boca Raton FL: LLC.

Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage, 19*(2 Pt 1), 261-270.

Cumming, B. G., & DeAngelis, G. C. (2001). The physiology of stereopsis. *Annu Rev Neurosci, 24*, 203-238.

Curran, W., & Johnston, A. (1996). The effect of illuminant position on perceived curvature. *Vision Res, 36*(10), 1399-1410.

Dakin, S. C., & Mareschal, I. (2000). Sensitivity to contrast modulation depends on carrier spatial frequency and orientation. [doi: DOI: 10.1016/S0042-6989(99)00179-0]. *Vision Research, 40*(3), 311-329.

De Martino, F., Valente, G., Staeren, N., Ashburner, J., Goebel, R., & Formisano, E. (2008). Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *NeuroImage, 43*(1), 44-58.

DeAngelis, G. C., & Uka, T. (2003). Coding of horizontal disparity and velocity by MT neurons in the alert macaque. *J Neurophysiol, 89*(2), 1094-1111.

DeYoe, E. A., Carman, G. J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., . . . Neitz, J. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. *Proc Natl Acad Sci U S A, 93*(6), 2382-2386.

Dosher, B. A., Sperling, G., & Wurst, S. A. (1986). Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vision Res, 26*(6), 973-990.

Dror, R. O., Willsky, A. S., & Adelson, E. H. (2004). Statistical characterization of real-world illumination. *J Vis, 4*(9), 821-837.

Ellemberg, D., Allen, H. A., & Hess, R. F. (2006). Second-order spatial frequency and orientation channels in human vision. [doi: DOI: 10.1016/j.visres.2006.01.028]. *Vision Research, 46*(17), 2798-2803.

Ernst, M. O. (2006). A Bayesian view on multimodal cue integration. In Knoblich, G., Thornton, I.M., Grosjean, M., & Shiffrar, M. (Ed.s), Human body perception from inside out. New York, NY: Oxford University Press. Chapter 6, 105-131.

Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision, 7*(5).

Ernst, M. O., & Banks, M. S. (2002). Using visual and haptic information for discriminating objects. *Perception, 31*, 147-147.

Ernst, M. O., Banks, M. S., & Bulthoff, H. H. (2000). Touch can change visual slant perception. [10.1038/71140]. *Nat Neurosci, 3*(1), 69-73.

Fahle, M., Edelman, S., & Poggio, T. (1995). Fast perceptual learning in hyperacuity. [doi: 10.1016/0042-6989(95)00044-Z]. *Vision Research, 35*(21), 3003-3013.

Fahle, M., & Morgan, M. (1996). No transfer of perceptual learning between similar stimuli in the same retinal position. *Current biology : CB, 6*(3), 292-297.

Fahle, M., & Poggio, T. (2002). *Perceptual learning.* Cambridge,MA.: MIT Press.

Fiorentini, A., & Berardi, N. (1980). Perceptual learning specific for orientation and spatial frequency. [10.1038/287043a0]. *Nature, 287*(5777), 43-44.

Fiorentini, A., & Berardi, N. (1981). Learning in grating waveform discrimination: Specificity for orientation and spatial frequency. [doi: 10.1016/0042-6989(81)90017-1]. *Vision Research, 21*(7), 1149-1158.

Fleet, D. J., & Langley, K. (1994). Computational analysis of non-Fourier motion. *Vision Res, 34*(22), 3057-3079.

Friston, K. J., Rotshtein, P., Geng, J. J., Sterzer, P., & Henson, R. N. (2006). A critique of functional localisers. *NeuroImage, 30*(4), 1077-1087.

Georgeson, M. A., & Schofield, A. J. (2002). Shading and texture: separate information channels with a common adaptation mechanism? *Spatial Vision, 16*(1), 59-76.

Georgieva, S. S., Todd, J. T., Peeters, R., & Orban, G. A. (2008). The extraction of 3D shape from texture and shading in the human brain. *Cerebral Cortex, 18*(10), 2416-2438.

Gerardin, P., Kourtzi, Z., & Mamassian, P. (2010). Prior knowledge of illumination for 3D perception in the human brain. *Proceedings of the National Academy of Sciences, 107*(37), 16309-16314.

Gilchrist,A.L. (1977). Perceived lightness depends on perceived spatial arrangement. *Science*, **195**, 185-187.

Gilchrist,A.L. (1988). Lightness contrast and failures of constancy: A common explanation. *Perception and Psychophysics*, **43**, 415-424.

Haijiang, Q., Saunders, J. A., Stone, R. W., & Backus, B. T. (2006). Demonstration of cue recruitment: change in visual appearance by means of Pavlovian conditioning. *Proc Natl Acad Sci U S A, 103*(2), 483-488.

Harding, G., Harris, J. M., & Bloj, M. (2012). Learning to use illumination gradients as an unambiguous cue to three dimensional shape. [doi:10.1371/journal.pone.0035950]. *PLoS One, 7*(4), e35950.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and Overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425-2430.

Hillis, J. M., Ernst, M. O., Banks, M. S., & Landy, M. S. (2002). Combining sensory information: Mandatory fusion within, but not between, senses. *Science, 298*(5598), 1627-1630.

Horn, B. K. P. (1975). Obtaining shape from shading information. In P. H. Winston (Ed.), *The Psychology of Computer Vision* (pp. 115-155). New York: Mcgraw-Hill.

Howard, I. P. (2003). Neurons that respond to more than one depth cue. [doi: 10.1016/S0166-2236(03)00238-8]. *Trends in Neurosciences, 26*(10), 515-517.

Howard, I. P., & Rogers, B. J. (2002). *Seeing in depth*. Toronto: I. Porteous.

Hubel, D. H., & Wiesel, T. N. (1970). Stereoscopic vision in macaque monkey: cells sensitive to binocular depth in area 18 of the macaque monkey cortex. [10.1038/225041a0]. *Nature, 225*(5227), 41-42.

Huettel, S. A. (2009). fMRI: BOLD Contrast. In L. S. Squire (Ed.), *Encyclopedia of Neuroscience* (pp. 273-281). Oxford: Academic Press.

Huk, A. C., Dougherty, R. F., & Heeger, D. J. (2002). Retinotopy and functional subdivision of human areas MT and MST. *J Neurosci, 22*(16), 7195-7205.

Humphrey, G. K., Goodale, M. A., Bowen, C. V., Gati, J. S., Vilis, T., Rutt, B. K., & Menon, R. S. (1997). Differences in perceived shape from shading correlate with activity in early visual areas. *Curr Biol, 7*(2), 144-147.

Jeter, P. E., Dosher, B. A., Liu, S. H., & Lu, Z. L. (2010). Specificity of perceptual learning increases with increased training. [Article]. *Vision Research, 50*(19), 1928-1940.

Jeter, P. E., Dosher, B. A., Petrov, A., & Lu, Z.-L. (2009). Task precision at transfer determines specificity of perceptual learning. *J Vision, 9*(3).

Jiang, X., Schofield, A., & Wyatt, J. (2010). Correlation-Based intrinsic image extraction from a single image. In K. Daniilidis, P. Maragos & N. Paragios (Eds.), *Computer vision – ECCV 2010* (Vol. 6314, pp. 58-71): Springer Berlin Heidelberg.

Johnson, A. P., & Baker, J. C. L. (2004). First- and second-order information in natural images: a filter-based approach to image statistics. *J. Opt. Soc. Am. A, 21*(6), 913-925.

Julesz, B. (1971). *Foundations of cyclopean perception*. Cambridge, MA.: MIT Press.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. [10.1038/nn1444]. *Nat Neurosci, 8*(5), 679-685.

Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. *P Natl Acad Sci USA, 88*(11), 4966-4970.

Karni, A., & Sagi, D. (1993). The time course of learning a visual skill. [10.1038/365250a0]. *Nature, 365*(6443), 250-252.

Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J., & Sagi, D. (1994). Dependence on REM sleep of overnight improvement of a perceptual skill. *Science, 265*(5172), 679-682.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annu Rev Psychol, 55*, 271-304.

Kingdom, F. A. A. (2003). Color brings relief to human vision. [10.1038/nn1060]. *Nat Neurosci, 6*(6), 641-644.

Kingdom, F. A. A. (2008). Perceiving light versus material. [doi: DOI: 10.1016/j.visres.2008.03.020]. *Vision Research, 48*(20), 2090-2105.

Kleffner, D. A., & Ramachandran, V. S. (1992). On the perception of shape from shading. *Percept Psychophys, 52*(1), 18-36.

Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoolbox-3? *Perception, 36*.

Knill, D. C. (1992). Perception of surface contours and surface shape: from computation to psychophysics. *J. Opt. Soc. Am. A, 9*(9), 1449-1464.

Knill, D. C. (2007). Learning Bayesian priors for depth perception. *J Vision, 7*(8).

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? [doi: 10.1016/S0042-6989(03)00458-9]. *Vision Research, 43*(24), 2539-2558.

Koenderink, J. J., Pont, S. C., van Doorn, A. J., Kappers, A. M., & Todd, J. T. (2007). The visual light field. *Perception, 36*(11), 1595-1610.

Koenderink, J. J., van Doorn, A. J., Kappers, A. M., & Todd, J. T. (2001). Ambiguity and the 'mental eye' in pictorial relief. *Perception, 30*(4), 431-448.

Koenderink, J. J., Van Doorn, A. J., & Pont, S. C. (2007). Perception of illuminance flow in the case of anisotropic rough surfaces. *Percept Psychophys, 69*(6), 895-903.

Kourtzi, Z., Erb, M., Grodd, W., & Bülthoff, H. H. (2003). Representation of the perceived 3-D object shape in the human lateral occipital complex. *Cerebral Cortex, 13*(9), 911-920.

Kourtzi, Z., & Kanwisher, N. (2000). Cortical regions involved in perceiving object shape. *J Neurosci, 20*(9), 3310-3318.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. [doi: 10.1016/0042-6989(94)00176-M]. *Vision Research, 35*(3), 389-412.

Liu, B., & Todd, J. T. (2004). Perceptual biases in the interpretation of 3D shape from shading. [doi: 10.1016/j.visres.2004.03.024]. *Vision Research, 44*(18), 2135-2145.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature, 412*(6843), 150-157.

Logothetis, N. K., & Pfeuffer, J. (2004). On the nature of the BOLD fMRI contrast mechanism. *Magn Reson Imaging, 22*(10), 1517-1531.

Lovell, P. G., Bloj, M., & Harris, J. M. (2012). Optimal integration of shading and binocular disparity for depth perception. *J Vis, 12*(1).

Lugtigheid, A. (2012). *Psychophysics and modelling of depth perception.* (PhD PhD Thesis), University of Birmingham, Birmingham, UK.

Lyon, R. F. (1993). *Phong Shading reformulation for hardware renderer simplification* (A. T. Group, Trans.): Apple Computer, Inc.

Mamassian, P., & Goutcher, R. (2001). Prior knowledge on the illumination position. [doi: 10.1016/S0010-0277(01)00116-0]. *Cognition, 81*(1), B1-B9.

Mamassian, P., Jentzsch, I., Bacon, B. A., & Schweinberger, S. R. (2003). Neural correlates of shape from shading. *Neuroreport, 14*(7), 971-975.

Mingolla, E., & Todd, J. T. (1986). Perception of solid shape from shading. *Biol Cybern, 53*(3), 137-151.

Moore, C., & Engel, S. A. (2001). Neural response to perception of volume in the lateral occipital complex. *Neuron, 29*(1), 277-286.

Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science, 257*(5075), 1357-1363.

Nandy, A. S., & Tjan, B. S. (2008). Efficient integration across spatial frequencies for letter identification in foveal and peripheral vision. *J Vision, 8*(13).

Nardini, M., Bedford, R., & Mareschal, D. (2010). Fusion of visual cues is not mandatory in children. *Proceedings of the National Academy of Sciences*.

Nelissen, K., Joly, O., Durand, J.-B., Todd, J. T., Vanduffel, W., & Orban, G. A. (2009). The extraction of depth structure from shading and texture in the macaque brain. [doi:10.1371/journal.pone.0008306]. *PLoS One, 4*(12), e8306.

Ogle, K. N. (1938). Induced size effect I A new phenomenon in binocular space perception associated with the relative sizes of the images of the two eyes. [Article]. *Archives of Ophthalmology, 20*(4), 604-623.

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

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis, 10*(4), 437-442.

Pentland, A. P. (1982). Finding the illuminant direction. *J. Opt. Soc. Am., 72*(4), 448-455.

Pentland, A. P. (1984). Local Shading Analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, PAMI-6*(2), 170-187.

Popple, A. V., Smallman, H. S., & Findlay, J. M. (1998). The area of spatial integration for initial horizontal disparity vergence. *Vision Res, 38*(2), 319-326.

Preston, T. J., Kourtzi, Z., & Welchman, A. E. (2009). Adaptive estimation of three-dimensional structure in the human brain. *Journal of Neuroscience, 29*(6), 1688-1698.

Preston, T. J., Li, S., Kourtzi, Z., & Welchman, A. E. (2008). Multivoxel pattern selectivity for perceptually relevant binocular disparities in the human brain. *Journal of Neuroscience, 28*(44), 11315-11327.

Qiu, A., Rosenau, B. J., Greenberg, A. S., Hurdal, M. K., Barta, P., Yantis, S., & Miller, M. I. (2006). Estimating linear cortical magnification in human primary visual cortex via dynamic programming. [doi: 10.1016/j.neuroimage.2005.11.049]. *NeuroImage, 31*(1), 125-138.

Qu, Z., Song, Y., & Ding, Y. (2010). ERP evidence for distinct mechanisms of fast and slow visual perceptual learning. [doi: 10.1016/j.neuropsychologia.2010.01.008]. *Neuropsychologia, 48*(6), 1869-1874.

Ramachandran, V. S. (1988). Perception of shape from shading. [10.1038/331163a0]. *Nature, 331*(6152), 163-166.

Read, J. C., & Cumming, B. G. (2004). Understanding the cortical specialization for horizontal disparity. *Neural Comput, 16*(10), 1983-2020.

Richards, W. (1985). Structure from stereo and motion. *J Opt Soc Am A, 2*(2), 343-349.

Rittenhouse, D. (1786). Explanation of an optical deception. *Transactions of the American Phylosophical Society, 2*, 37-42.

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

Schiller, P. H., Slocum, W. M., Jao, B., & Weiner, V. S. (2011). The integration of disparity, shading and motion parallax cues for depth perception in humans and monkeys. [Article]. *Brain Res, 1377*, 67-77.

Schofield, A. J. (2000). What does second-order vision see in an image? *Perception, 29*(9), 1071-1086.

Schofield, A. J., & Georgeson, M. A. (1999). Sensitivity to modulations of luminance and contrast in visual white noise: separate mechanisms with similar behaviour. *Vision Research, 39*(16), 2697-2716.

Schofield, A. J., & Georgeson, M. A. (2003). Sensitivity to contrast modulation: the spatial frequency dependence of second-order vision. *Vision Research, 43*(3), 243-259.

Schofield, A. J., Hesse, G., Rock, P. B., & Georgeson, M. A. (2006). Local luminance amplitude modulates the interpretation of shape-from-shading in textured surfaces. *Vision Research, 46*(20), 3462-3482.

Schofield, A. J., Rock, P. B., & Georgeson, M. A. (2011). Sun and sky: Does human vision assume a mixture of point and diffuse illumination when interpreting shape-from-shading? *Vision Res, 51*(21-22), 2317-2330.

Schofield, A. J., Rock, P. B., Sun, P., Jiang, X., & Georgeson, M. A. (2010). What is second-order vision for? Discriminating illumination versus material changes. *J Vision, 10*(9).

Seitz, A. R., Yamagishi, N., Werner, B., Goda, N., Kawato, M., & Watanabe, T. (2005). Task-specific disruption of perceptual learning. *P Natl Acad Sci USA, 102*(41), 14895-14900.

Serences, J. T., & Boynton, G. M. (2007). The representation of behavioral choice for motion in human visual cortex. *J Neurosci, 27*(47), 12893-12899.

Sereno, M. E., Trinath, T., Augath, M., & Logothetis, N. K. (2002). Three-dimensional shape representation in monkey cortex. [doi: 10.1016/S0896-6273(02)00598-6]. *Neuron, 33*(4), 635-652.

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., . . . Tootell, R. B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science, 268*(5212), 889-893.

Sowden, P. T., Rose, D., & Davies, I. R. L. (2002). Perceptual learning of luminance contrast detection: specific for spatial frequency and retinal location but not orientation. *Vision Research, 42*(10), 1249-1258.

Spang, K., Grimsen, C., Herzog, M. H., & Fahle, M. (2010). Orientation specificity of learning vernier discriminations. [doi: DOI: 10.1016/j.visres.2009.12.008]. *Vision Research, 50*(4), 479-485.

Stevenson, S. B., & Schor, C. M. (1997). Human stereo matching is not restricted to epipolar lines. *Vision Res, 37*(19), 2717-2723.

Stickgold, R., James, L., & Hobson, J. A. (2000). Visual discrimination learning requires sleep after training. *Nat Neurosci, 3*(12), 1237-1238.

Sun, J., & Perona, P. (1998). Where is the sun? [10.1038/630]. *Nat Neurosci, 1*(3), 183-184.

Sun, P., & Schofield, A. J. (2011). The efficacy of local luminance amplitude in disambiguating the origin of luminance signals depends on carrier frequency: Further evidence for the active role of second-order vision in layer decomposition. [doi: 10.1016/j.visres.2011.01.008]. *Vision Research, 51*(5), 496-507.

Sun, P., & Schofield, A. J. (2012). Two operational modes in the perception of shape from shading revealed by the effects of edge information in slant settings. *J Vision, 12*(1).

Taira, M., Nose, I., Inoue, K., & Tsutsui, K. (2001). Cortical areas related to attention to 3D surface structures based on shading: an fMRI study. *NeuroImage, 14*(5), 959-966.

Todd, J. T., & Mingolla, E. (1983). Perception of surface curvature and direction of illumination from patterns of shading. *Journal of Experimental Psychology: Human Perception and Performance, 9*(4), 583-595.

Tsutsui, K., Jiang, M., Yara, K., Sakata, H., & Taira, M. (2001). Integration of perspective and disparity cues in surface-orientation-selective neurons of area CIP. *J Neurophysiol, 86*(6), 2856-2867.

Tsutsui, K., Sakata, H., Naganuma, T., & Taira, M. (2002). Neural correlates for perception of 3D surface orientation from texture gradient. *Science, 298*(5592), 409-412.

Tyler, C. W., Likova, L. T., Kontsevich, L. L., & Wade, A. R. (2006). The specificity of cortical region KO to depth structure. *NeuroImage, 30*(1), 228-238.

van Dam, L. C. J., & Ernst, M. O. (2010). Preexposure disrupts learning of location-contingent perceptual biases for ambiguous stimuli. *J Vision, 10*(8), -.

van Ee, R., Adams, W. J., & Mamassian, P. (2003). Bayesian modeling of cue interaction: bistability in stereoscopic slant perception. *J Opt Soc Am A, 20*(7), 1398-1406.

Vuong, Q. C., Domini, F., & Caudek, C. (2006). Disparity and shading cues cooperate for surface interpolation. *Perception, 35*(2), 145-155.

Wagemans, J., van Doorn, A. J., & Koenderink, J. J. (2010). The shading cue in context. *1*(3), 159-177.

Wallis, S. A., Baker, D. H., Meese, T. S., & Georgeson, M. A. (2013). The slope of the psychometric function and non-stationarity of thresholds in spatiotemporal contrast vision. [doi: 10.1016/j.visres.2012.09.019]. *Vision Research, 76*(0), 1-10.

Watson, A. B., & Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychometric method. *Percept Psychophys, 33*(2), 113-120.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Percept Psychophys, 63*(8), 1293-1313.

Yamamoto, H., Ban, H., Fukunaga, M., Tanaka, C., Umeda, M., & Ejima, Y. (2009). Large- and small-scale functional organization of visual field representation in the human visual cortex. In T. A. Portocello & R. B. Velloti (Eds.), *Visual cortex: new research* (pp. 195-226). New York: Nova Science Publisher.

Zhou, Y. X., & Baker, C. L., Jr. (1996). Spatial properties of envelope-responsive cells in area 17 and 18 neurons of the cat. *J Neurophysiol, 75*(3), 1038-1050.