Eccles Institute for Neuroscience, John Curtin School of Medical Research, Australian National University

---

A Computational Neuroscience Approach to Higher-Order Texture Perception

John W. G. Seamons

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Submission Date:

Declaration

This is to certify that:

   I.   **The thesis comprises only my original work towards the PhD except where indicated.**

   II.   **Due acknowledgement has been made in the text to all other material used.**

During the course of this thesis, I participated in a study which was also a mTurk based study of learning texture discrimination. Thus far, that material on the learning aspect has only been presented as an invited platform presentation. It is in preparation for publication and I will be a co-author on that submitted paper. Some of the data from that study, which did not involve learning, and the collection of which was overseen by me was used in Chapter 3.

Dominique Coy, Marconi Barbosa, John Seamons, Ted Maddess. *Learning isotrigon textures*. Presented at Weill Cornell Medical College, Dept of Neurology and Neuroscience. September 2014.

_____

John W. G. Seamons

Acknowledgements

I would like to extend my deepest thanks and gratitude to Professor Ted Maddess. The opportunity to learn from him has been invaluable and I will remember my time at ANU with great fondness.

I would also like to extend thanks to my colleagues in the Maddess group, Dr Corinne Carle and Dr Marconi Soares Barbosa, terrifically talented people from whom I have learned a great deal.

I give warm thanks to my mother and father; to my sister Hana, my new nephews Art and Edwyn, and to my partner Renae. Your love and support have kept me going.

This doctoral thesis is dedicated to my father, G. R. Seamons. From an early age, he instilled in me a passion for learning, something that I will always hold dear.

Thesis Abstract

Natural images contain large amounts of structural information characterised by higher-order spatial correlations. Neurons have limited capacities, so the visual system must filter out non-salient information, but retain that which is behaviourally relevant. Previous research has concentrated on two-point correlations; there has been less research into higher-order correlations, although the visual system is sensitive to them. Isotrigon textures can be used for this purpose. Their salient structure is exclusively due to fourth- and higher-order spatial correlations and they have the same structural features that create salience in natural images.

In Chapter 2, we evaluated human texture discrimination using 10 novel isotrigon textures (VnL2) and 17 standard V3L2 isotrigon textures. Factor analysis revealed that as few as 3 mechanisms may govern the detection of fourth- and higher-order image structure. The Maddess group has previously published evidence that the number of independent mechanisms is less than 10 and perhaps as small as 3-4. The computation of higher-order correlations by the brain is neuro-physiologically plausible via nonlinear combinations of recursive and/or rectifying processes.

In Chapter 3, we utilised the crowdsourcing platform "mTurk" to implement a large texture discrimination study. Under laboratory conditions, we showed that the testing modality was robust across a range of browsers, resolutions, contrasts and screen sizes. Texture discrimination data was gathered from 121 naïve subjects and compared to 2 independent laboratory data sets. Factor analysis indicated the presence of 3-4 factors, consistent with previous studies. Based on Pearson's correlation and coefficients of

repeatability, mTurk is capable of producing data of comparable quality to laboratory studies. This is significant as mTurk has not previously been systematically evaluated for visual psychometric research.

In Chapter 4, we employed a set of statistically controlled ternary textures. The textures were constrained (spatial correlations from 1st to 4th order) and their salience could be independently controlled by the addition of noise. To the ideal observer, all textures defined by a given amount of noise are equally detectable. However, humans are *not* ideal observers; their visual perceptual resources are restricted.

Because of the number of textures available, we used mTurk to gather performance functions from 928 subjects for a subset of the texture space. Perceptual salience varied for each image statistic, with rank order: gamma > beta_hv > beta_diag > alpha > theta. This supports the order previously published for the related *binary* stochastic textures. The two least salient directions were consistently white:black and grey-bias (for gammas and betas), and black:grey and grey:white (for thetas and alphas). Such differences reflect the sensitivities and limitations of neural processing and are a manifestation of efficient coding.

We hypothesised that the grey token conferred non-salience. Indeed, for gammas and betas, the grey-bias was consistently the second least salient. However, this did not hold for thetas or alphas. Counter-intuitively, the *order* of texture presentation *did not* significantly affect discrimination performance. An analysis of 31 repeat Workers found evidence of learning for beta textures, whereas performance for other textures was already maximal. This thesis concludes by considering future research.

Publications Arising from this Thesis

o Seamons, J. W. G., Bubna-Litic, A., Barbosa, M. S., & Maddess, T. 2015. A Lower Bound on the Number of Mechanisms for Discriminating Fourth and Higher Order Spatial Correlations. Vision Research, 108: 41-48.

o Seamons, J. W. G., Bubna-Litic, A., Barbosa, M. S., & Maddess, T. 2013. Isotrigon Discrimination by a Small Number of Underlying Neural Mechanisms. In Clinical and Experimental Ophthalmology, 41: 122-122.

o Seamons, J. W. G., Bubna-Litic, A., Barbosa, M. S., & Maddess, T. 2015. The Lower Bound of Neurological Mechanisms for Discriminating Binary Isotrigon Tetures. BMC Neuroscience (Vol 16, Suppl. 1).

o Seamons, J. W. G., Barbosa, M. S., Coy, D., & Maddess T. 2015. A Validation of Crowdsourced Methods for Visual Psychometric Tasks Using Isotrigon Textures. BMC Neuroscience (Vol 16, Suppl. 1).

o Seamons, J. W. G., Barbosa, M. S., Coy, D., & Maddess T. 2015. Developing and Validating an Isotrigon Texture Discrimination Task Using Crowdsourcing. (In Preparation).

Contents

Chapter 1: General Introduction

1.1 Prologue

I will begin this General Introduction by discussing image and texture analysis methods and forms of Factor analysis; these concepts are critical to understanding the material which follows. I will then review cortical processing and receptive fields. I will move on to give an overview of texture perception, before considering how isotrigon textures can be used to probe the sensitivities and limitations of the human visual system. I will discuss which brain regions are involved in texture perception and then review the use of Amazon Mechanical Turk to conduct crowdsourced visual psychometric experiments. Lastly, I will outline the structure of this thesis.

1.2 Texture Analysis and Factor Analysis Techniques

1.2.1 Texture Analysis Algorithms

At the heart of this thesis is the evaluation of texture perception and textures using texture analysis algorithms. Textures can be defined as complex visual patterns composed of regular, and possibly self-similar or repeating, elements on a surface. Digital images are composed of pixels; therefore digital textures can be defined as entities consisting of mutually related pixels and groups of pixels. These pixel groupings may be called texture "primitives" or texture elements ("Texels") (Srinivasan and Shobha, 2008). Here we broadly define texture analysis *functions* to be a class of

13

mathematical methods that characterize spatial variations within an image. They allow us to view textures as a primarily mathematical constructs and are therefore central to our understanding of texture perception.

According to Materka (Materka and Strzelecki, 1998), four types of texture analysis approaches are commonly defined. *Structural* approaches (Haralick, 1979) represent textures based on the hierarchical arrangement of primitives. *Model based* texture analysis (Strzelecki and Materka, 1997) attempts to define a texture by use of a stochastic model. *Transform methods* of texture analysis, such as Gabor (Bovik et al., 1990) and wavelet transforms (Lu et al., 1997), use image spaces where coordinates are related to the characteristics of the texture (such as frequency or size). *Statistical* approaches (Julesz, 1975) define textures based on the distributions and relationships between grey levels within an image. In this thesis, we will be primarily concerned with statistical approaches. They include conjoint space and spatial frequency methods, convolution filters, and spatial autocorrelation (Srinivasan and Shobha, 2008).

1.2.2 Spatial Frequency

What is a spatial frequency? Spatial frequency is a measure of how often sinusoidal components repeat per unit of distance. To illustrate this concept, consider one dimensional spatial frequency as defined by a simple sine wave. We can create a spatial sine wave by defining the grey value (G) as a function of the x-direction, as shown in Equation 1:

$$G(x) = 127 + 127sin(x) \qquad \text{(Eq. 1)}$$

When this function is repeated vertically, it defines a series of vertical stripes (Figure 1). We can modify this function, and thereby produce a rectangular waveform, by adding odd harmonic waves from a Fourier series. A Fourier series decomposes any periodic function into the sum of a set of simple sines and cosines. This is illustrated in Equation 2:

$$G(x) = \sin(x) + \frac{1}{3}\sin(3x) + \frac{1}{5}\sin(5x) + \frac{1}{7}\sin(7x) \dots \quad \text{(Eq. 2)}$$

As an increasing number of elements of the Fourier series are added, the approximation of the rectangular wave becomes better (Figure 1). This illustrates that any wave pattern can be created by the superposition of sine waves, a fundamental principle of the Fourier Transform (discussed below in Section 1.3). The concept of spatial frequency can be readily extended to two-dimensional images. In this case, both the location and the direction of the trace affect the spatial frequency content recorded (Heilbronner and Barrett, 2013). Sinusoidal gratings are frequently used during the study of visual perception to probe the capabilities of the visual system. In this case, spatial frequency is commonly expressed as *cycles per degree* of the visual angle subtended (Snowden et al., 2012).
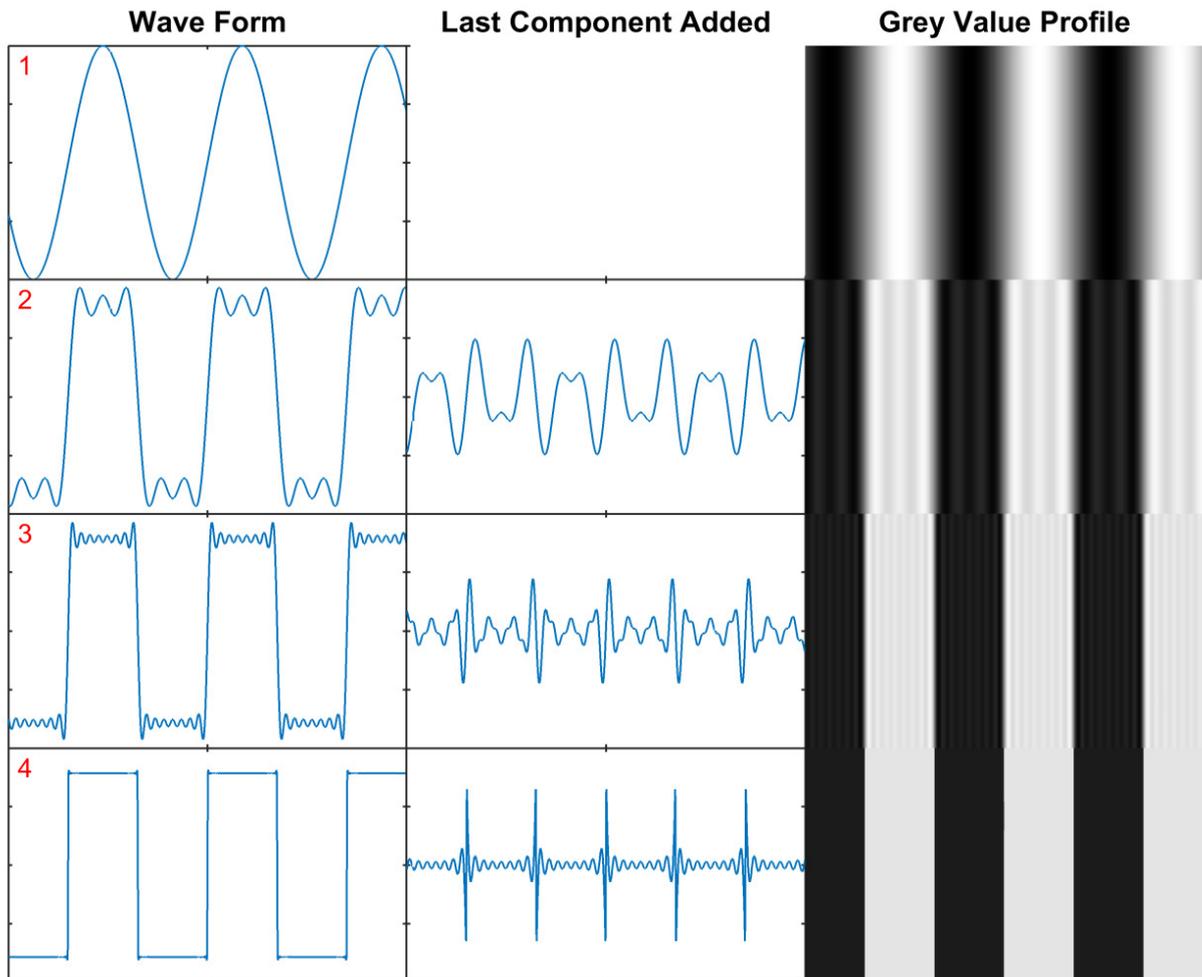
Figure 1. An illustration of one dimensional spatial frequency using the superposition of different sine waves. Row 1 is derived from the fundamental frequency sin(x). Row 2 is derived from sin(x) plus the first 2 odd harmonics from the Fourier series (given in Eq. 2). i.e.: up to and including 1/5sin(5x). Row 3 shows sin(x) plus harmonics up to and including 1/15sin(15x). Row 4 shows sin(x) plus 450 odd harmonics, which approximates a *square wave*.

This example also illustrates the wave property *phase* and the phenomenon of *phase cancellation*. If we consider the time domain, the phase of a wave is the fraction of a complete wave cycle which has occurred since some

specified time point (normally taken to be the origin, time t = 0). The relationship between amplitude, frequency and phase can be shown as follows in an equation for simple harmonic motion (Eq. 3):

$$x(t) = A.\sin(2\pi ft + \Theta(t)) \qquad \text{(Eq. 3)}$$

…where *A* is the signal amplitude, *f* is frequency, *t* is the time elapsed and *θ* is the phase (Reddy et al., 1994).

When two out of phase waves are combined, phase cancellation may result in a reduction in the overall amplitude of the combined signal; the opposite effect is phase reinforcement, which occurs when waves are in phase. In the context of the Fourier series discussed above (and shown in Eq. 2), successive phase cancellation and phase reinforcement to local regions of the base waveform (sin(X)) leads to the production of an approximated square wave (Heilbronner and Barrett, 2013).

1.2.3 Fourier Transforms

The Fourier transform is an image processing tool which can be used to decompose an image into its constituent sine and cosine waveforms. In the case of a 2D image, whereas the input image is in the spatial domain, the output of the transformation represents the image in the *Fourier* or frequency domain. Each point in the Fourier domain image corresponds to a particular frequency in the spatial domain image. A one-dimensional Fourier transform is given by Equation 4:

$$F(u) = \int f(x)\, e^{-2\pi i u x} dx \qquad \text{(Eq. 4)}$$

…where e^{iux} is cos(*ux*) + i * sin(*ux*) and *x* is length and *u* is spatial frequency (Heilbronner and Barrett, 2013).

The Fourier transform has real and imaginary components, the cos and sin parts of Eq. 2. These can be related to the magnitude and phase of the wave by Pythagoras' rule. In image processing, the phase information is often discarded as the magnitude contains the geometric structural information of the spatial domain image (note that the phase information *is required* for inverse Fourier transforms). Fourier transforms are computationally intensive, especially for large images, so the *Fast Fourier Transform* (FFT) is commonly used instead.

An image in the Fourier domain is decomposed into its sinusoidal components; therefore, it is easier to examine and analyse the geometric structure in the spatial domain. Two-dimensional Fourier transforms of some more complex and realistic images are shown in Figure 2. For example, the Fourier transform of the image of the fern shows a maximum at approximately 70 degrees and a frequency of 15 per image. This corresponds to the leaflets highlighted in the greyscale image above. Also note the vertical and horizontal lines through the centre, which correspond to the image boundaries (Heilbronner and Barrett, 2013).
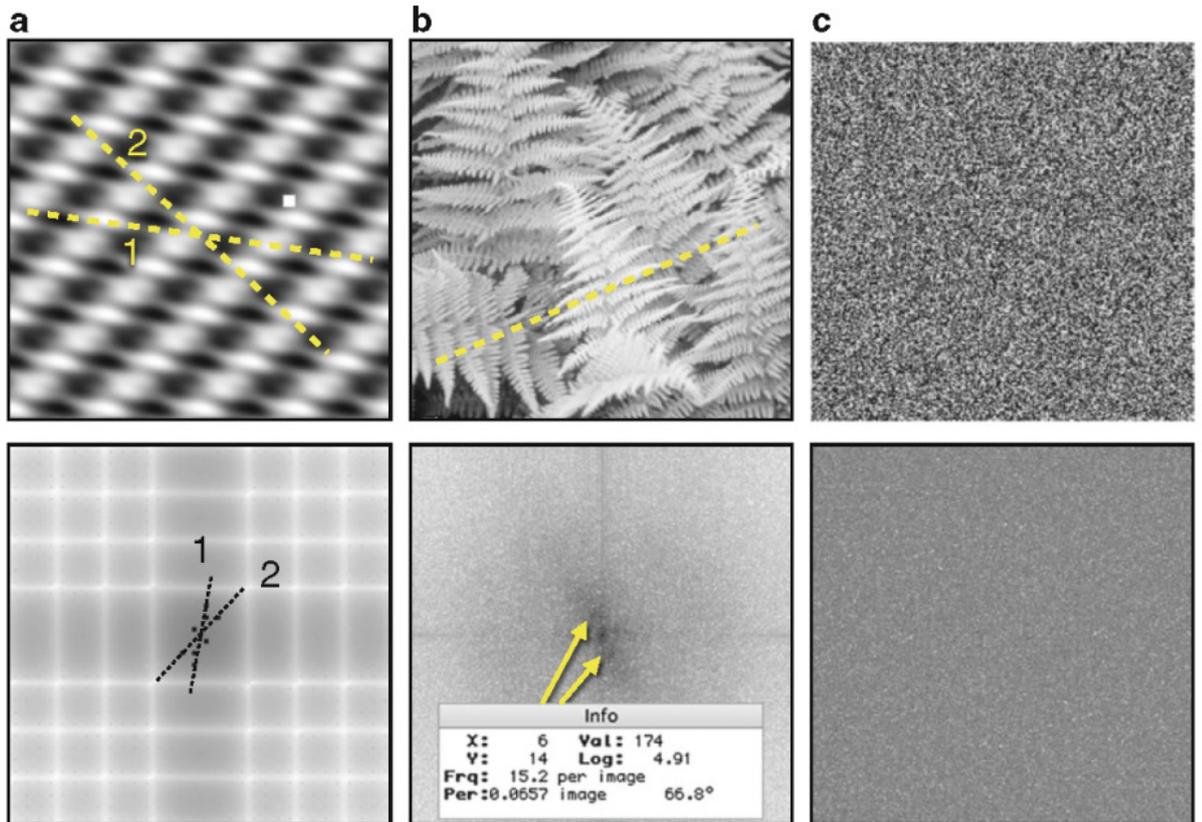
Figure 2. Two-dimensional Fourier transforms of complex and realistic images. The top row shows greyscale images and the bottom row shows the corresponding Fourier transforms. (a) a superposition of four sine waves with a cut-out square. (b) a natural image of a fern. (c) ternary noise. Selected orientations and frequencies are highlighted. The darker tones correspond to higher contrasts of the constituent spatial frequencies. Note that the each of the images is a quarter of the size of the original for reasons of presentation (Heilbronner and Barrett, 2013).

Fourier transforms can also be used to produce filter masks. High-pass filtering blocks low frequencies, which results in the sharpening of an image. Low pass filtering is the inverse and results in image blurring. Manipulations

19

to the Fourier domain are reflected in the spatial domain following inverse an Fourier transform, and vice versa (Heilbronner and Barrett, 2013).

1.2.4 Autocorrelation Function

A texture cannot be defined by single point measurements of brightness intensity nor colour. Rather, texture descriptors must be based on *relationships* between neighbouring features. One measure of these relationships is called spatial autocorrelation. Griffith defines spatial autocorrelation as "…the correlation among values of a single variable strictly attributable to their relatively close locational positions on a two-dimensional surface" (Griffith, 2003). The degree to which such relationships exit can be mathematically quantified using the autocorrelation function (ACF). The ACF measures the spatial frequency content of a texture and, in so doing, defines its regularity and coarseness (Srinivasan and Shobha, 2008). To put it another way, the ACF describes the correlation between image primitives as a function of displacement. Equivalently, in a greyscale image, the ACF describes the correlation between the grey values of every pixel with its neighbours as a function of relative location. Heilbronner presents a thought experiment which provides an intuition of the ACF, adapted as follows (Heilbronner and Barrett, 2013).

The image in Figure 3 shows a white circle on a black background. This image is duplicated on two transparent films, so that the circles are fully transparent and the background is fully opaque. These transparencies are then held up to a light source. Maximum light transmission is achieved when the circles are precisely aligned. This starting position corresponds to zero

displacement (x' = 0, y' = 0). Now, keeping one film fixed in place, the second film can be displaced in any direction. As the films overlap, the amount of light transmitted decreases. The total light transmitted for a given displacement can be regarded as a measure of the value of the ACF at that position (x', y'). The brightness analogue of the ACF is the total light transmitted for all possible displacements (Figure 3) (Heilbronner and Barrett, 2013).
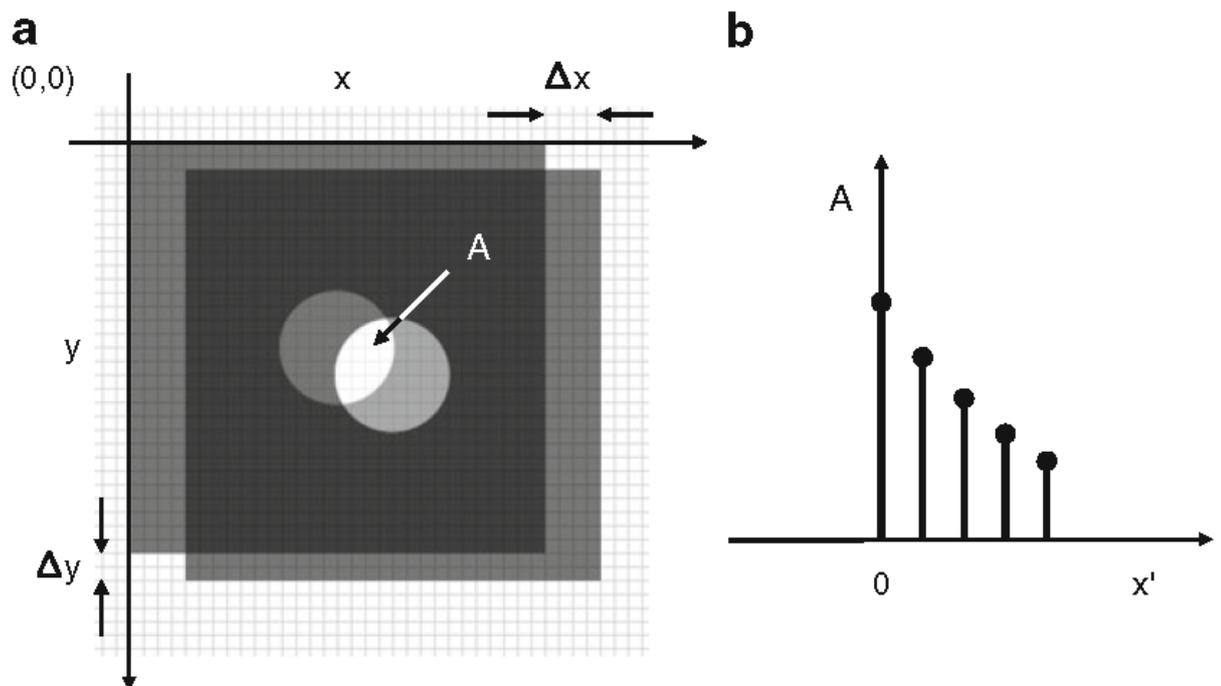
Figure 3. A visualization of the ACF. (a) Two copies of the circle on transparent films are superimposed. One copy is then displaced in the x and y-planes by Δx and Δy. A is the area of overlap of the circles. (b) the correlation between the two image copies is measured as A(x', y'), the area of overlap as a function of the x and y displacement; here, only A(x') is shown (Heilbronner and Barrett, 2013).

Based on the autocorrelation function, single-point correlations, or first-order statistics, inform the visual system about the mean brightness of an image. Two-point correlations, or second-order statistics, inform the visual system about the spatial frequency content of an image. Higher-order correlations capture salient features such as object contours. Expressed mathematically, an example of the autocorrelation function for second-order statistics (two point correlational structure) is shown as follows (Equation 5) (Taylor, 2008). The discrete zero mean second-order spatial correlation function for an image ($I(x,y)$) of size N pixels is:

$$C_{2,f(h,v)} = 1/N \ \Sigma \ \Sigma \ I(x, y) \ I(x+h, y+v) \ \Delta x \Delta y \qquad \text{(Eq. 5)}$$

This function results from the pointwise multiplication of each pixel in x and y, belonging to the set of all $I(x,y)$, by all possible fixed displacements ($\Delta x \Delta y$), in the horizontal (h) and vertical (v) planes (Taylor, 2008). As we would expect, if texture primitives are large, the ACF decreases slowly with increasing displacement, and vice versa. If the primitives are positioned periodically, the ACF will increase and decrease periodically with displacement (Srinivasan and Shobha, 2008).

The ACF as defined thus far registers correlations between *pairs* of pixels. Figure 3 and Eq. 5 define pairs of planes sliding past each other, with all pairs of overlapping pixels being multiplied and cumulative sums taken. However, we can also evaluate correlations between larger cohorts of pixels, such as triplets or quadruplets, to thereby generate third- and fourth-order correlation functions. In reference to the number of pixels per cohort being

considered, second-order statistics are sometimes referred to as *dipole statistics* and higher orders as involving "N-gons" (for example, where N = 3, "trigons"). Natural images are rich in these higher-order spatial correlations; the spatial correlation content of natural images at different orders is shown in Figure 4.
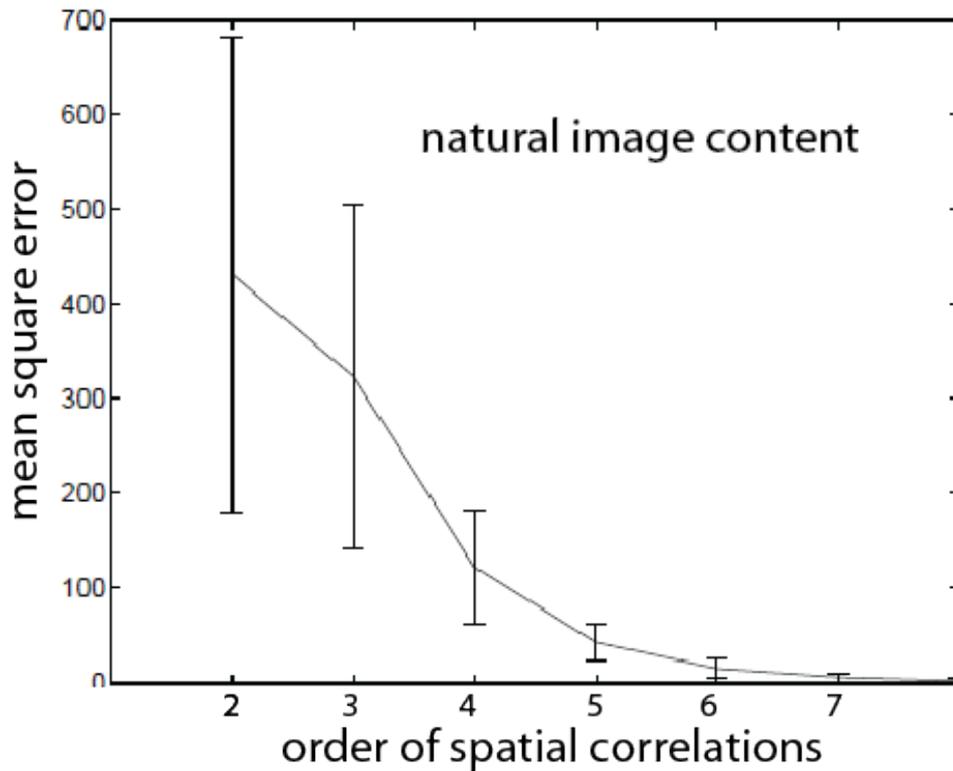
Figure 4. The content of natural images in terms of spatial correlations. The content is expressed in terms of mean square reconstruction error (MSE), where only correlations up to a particular order are considered. The data is obtained from 10 natural images. Note the high proportion of third- and higher-order spatial correlations (Franz and Schölkopf, 2005).

Correlations between triplets of pixels (trigons) can be computed by considering the displacement of three shifting planes: h1 and h2, v1, and v2. Therefore, the three correlation function is 4 dimensional (Eq. 6).

$$S3, f(h1, v1, h2, v2) = \frac{1}{N} \iint f(x,y) f(x + h1, y + v1) \, f(x + h2, y +$$

$$v2) dx dy \qquad\qquad\qquad\qquad \text{(Eq. 6)}$$

Another important concept for texture analysis is the *power spectrum*. The power spectrum of a signal is the power of that signal at each of its constituent frequencies; that is, the square of the Fourier transform. White noise, for example, has a flat power spectrum as all frequencies are equally represented at the same power. The power spectrum, Fourier transform, and autocorrelation function are closely related. The Fourier transform of an autocorrelation function is the power spectrum, or equivalently, the autocorrelation is the inverse Fourier transform of the power spectrum (Bracewell, 1965). Note that the square root of the power spectrum is the amplitude (magnitude) spectrum.

1.2.5 Gabor Transform

The Gabor transform is a special type of Fourier transform which can be used to determine the sinusoidal frequency and phase content of discrete signal sections as they change over patches of time or space (Sandler and Lindenbaum, 2006). The Gabor transform was originally derived from Gabor's theory of communication, which showed that signals can be represented in terms of functions that are localized in both time and frequency (Gabor, 1946). Image filters based on Gabor transforms are

particularly pertinent as they have been found to resemble those of the human visual system (Marčelja, 1980, Daugman, 1985).

Whereas the Fourier transform supplies all the average frequency content across the whole image, the Gabor transform does a piecewise analysis of the frequency content within patches of a signal. The key element of the Gabor transforms are so called *Gabor functions*, which are Gaussian envelopes of different sizes multiplied by a sinusoidal carrier frequency. Note that to encode phase, the carriers must be of two types; typically cosine and sine pairs are used.

For a 2D signal, the Gaussians are 2D and the 1D carriers need to have two orientations to span orientation space. The bank of functions is convolved, with the image centred on a number of sample locations. Importantly, as Gabor pointed out, the number of coefficients (i.e. Gabor functions) and sample locations should be equal to the number of samples in the original signal. For a digital image, the number of Gabor functions required is equal to the number of pixels. This makes the Gabor transform lossless. Thus, the gap between the Gabor function sampling positons are wider apart than the original time or space domain samples, growing larger as the number of Gabor functions per position increases.

The number of Gabor functions per position, and the number of different carriers, defines the number of frequency bands analysed at each position. Having more Gabor functions than the minimum number required is often described as *over-complete* or *redundant*. Several modern variants of the

Gabor transform that use different envelope functions fall under the heading of *wavelet transforms*.

Gabor space has been widely used in applications such as iris (Ma et al., 2002) and fingerprint recognition (Jain et al., 2001). An example of a two-dimensional Gabor filter g(x,y) is defined in Equation 7 (Yin et al., 2009):

$$g(x,y) = 1/(2\pi\sigma_x\sigma_y) * \exp[-1/2(x^{2/}\sigma^2 x + y^2/\sigma^2 y)] * \exp(j2\pi Wx)$$

(Eq. 7)

…where *W* is the frequency of the sinusoidal plane wave; the Gaussian envelope in the x and y planes is specified by *σx* and *σy* (and this acts as the window function, thus determining the bandwidth of the Gabor filter) (Yin et al., 2009). Figure 5 shows an example of a Gabor filter in two dimensions.

1.2.6 Factor Analysis, Principal and Independent Component Analysis

Principal Components Analysis (PCA) was developed in 1901 by Karl Pearson (Pearson, 1901) and later independently by Hotelling (Hotelling, 1933). PCA is a variable reduction technique that allows a complex data set to be reduced to one of lower dimensionality (Fabrigar et al., 1999, Suhr, 2005, Meglen, 1992, Jeong et al., 2009).

In the first stage of PCA, a covariance matrix is produced from the original data set. The *covariance* of two variables is their tendency to vary together. From this covariance matrix, *eigenvectors* and *eigenvalues* are determined.

Singular value decomposition (SVD) is commonly used for this purpose, particularly for non-square matrices (Jeong et al., 2009).
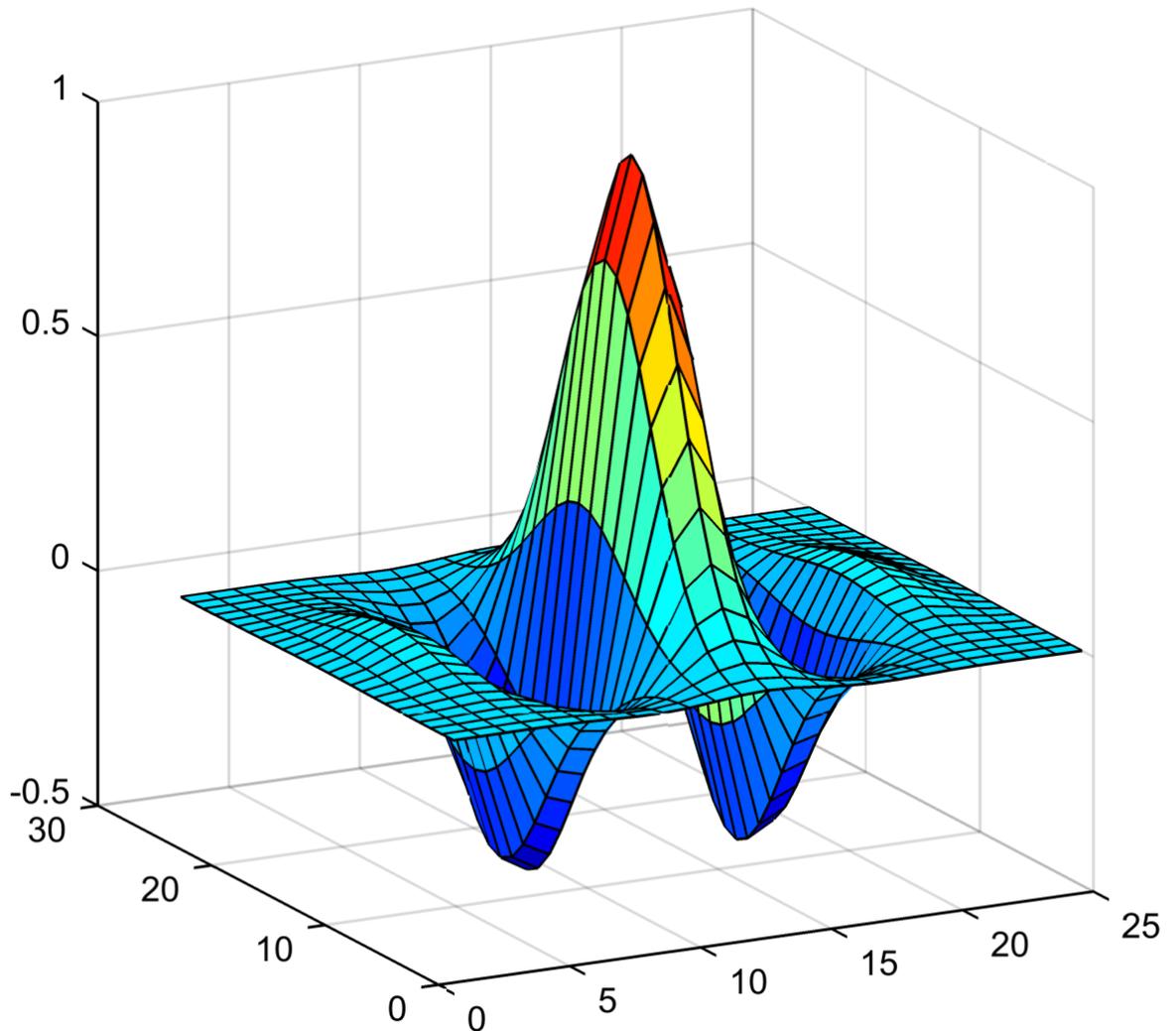
Figure 5. Example of a two-dimensional Gabor filter. The units are relative.

Eigenvectors are vectors which do not change *direction* during a transformation. The magnitude of an eigenvector may change during a transformation; this is known as the *eigenvalue*. In simple terms, for a square matrix *A*, a scalar λ, and a vector *v*, if Equation 8 is satisfied:

$$Av = \lambda v \qquad \text{(Eq. 8)}$$

…then λ is an eigenvalue of v, where v is the eigenvector (Press, 2007). The set of all values of λ and v for the matrix A defines an *eigenspace*. The number of eigenvectors, and the dimensionality of the eigenspace, is equal to the size of the original covariance/cross correlation matrix. The size of the *Wigner values* is proportional to the proportion of the covariance accounted for by each orthogonal component (eigenvector). In PCA, the original data set is projected onto a new coordinate system, defined by the eigenvectors and eigenvalues. A second step, *Factor analysis*, is to reconstruct the data with a subset of the eigenvector/value pairs that gives a good account of the data with the aim of reducing its dimensionality (Jeong et al., 2009).

When sorted in order of magnitude, the eigenvalues indicate the relative significance of the eigenvectors, i.e. the significant dimensions. A scree plot is commonly used to sort and display the eigenvalues. Scree plots are accurate and reliable, although prone to some variations between assessors (Cattell, 1966, Hayton et al., 2004, DeVellis, 2012, Zwick and Velicer, 1986). The eigenvector with the highest eigenvalue contributes the most to the observed variation in the data set. The adequacy of a PCA factor model can be evaluated by calculating its reconstruction error versus the original data set. i.e.: evaluating what proportion of the data set is accounted for by the k-factor model derived (Reyment and Joreskog, 1996). We will use this technique in Chapters 2 to 4 and it is discussed in detail in Chapter 2 (Section 2.4.2).

An alternative data reduction method is *independent component analysis* (ICA). ICA is a method for separating a signal into a linear mixtures of independent additive components (Hyvärinen and Oja, 2000). ICA consists of searching for a linear transformation that minimizes the statistical dependence between constituent components. Whereas PCA is normally used where the data being analysed has an assumed Gaussian distribution, ICA can recover basis functions from more complex signals.

There are many situations in which data is not Gaussian and the most famous example in the context of ICA is the "cocktail party effect", described as follows (Bronkhorst, 2000). Two microphones record two people speaking simultaneously, from different locations within a room. Each microphone will therefore record a particular linear combination of the two voices. Whereas PCA could reduce the voice data into a smaller set of orthogonal variables, it would *not separate* the voices into coherent streams. In contrast, ICA would be able to separate the voice of each speaker from the mixed recording (Langlois et al., 2010, Comon, 1994, Stone, 2004).

ICA can be formally defined as follows. Let us denote the random observed vector:

$$X = [X_1, X_2, ..., X_m]^T \qquad \text{(Eq. 9)}$$

…whose *m* elements are mixtures of *m* independent elements from a random vector:

$$S = [S_1, S_2, \ldots, S_m]^T \qquad\qquad \text{(Eq. 10)}$$

…given by:

$$X = AS \qquad\qquad \text{(Eq. 11)}$$

…where *A* represents a mixing matrix, the sample value of *Xj* is denoted by *xj* and *j* = 1, 2, ..., m. The goal of ICA is to find the un-mixing matrix *W* (i.e. the inverse matrix of *A*); this can be used to derive *Y*, the best possible approximation of *S* (Langlois et al., 2010, Comon, 1994) (Stone, 2004). i.e.:

$$Y = WX \cong S \qquad\qquad \text{(Eq. 12)}$$

ICA has a wide variety of applications, including data compression, deconvolution and separation. In the context of vision research, several authors have derived functions resembling simple cell receptive fields from ICA of natural images (Bell and Sejnowski, 1997, Hurri, 1997, Hurri et al., 1996, van Hateren and van der Schaaf, 1998, Wachtler et al., 2001). Some examples of this specific application will be discussed in Section 1.4.2 below. The contrast between ICA and PCA is illustrated in Figure 6 (Zetzsche et al., 1999).
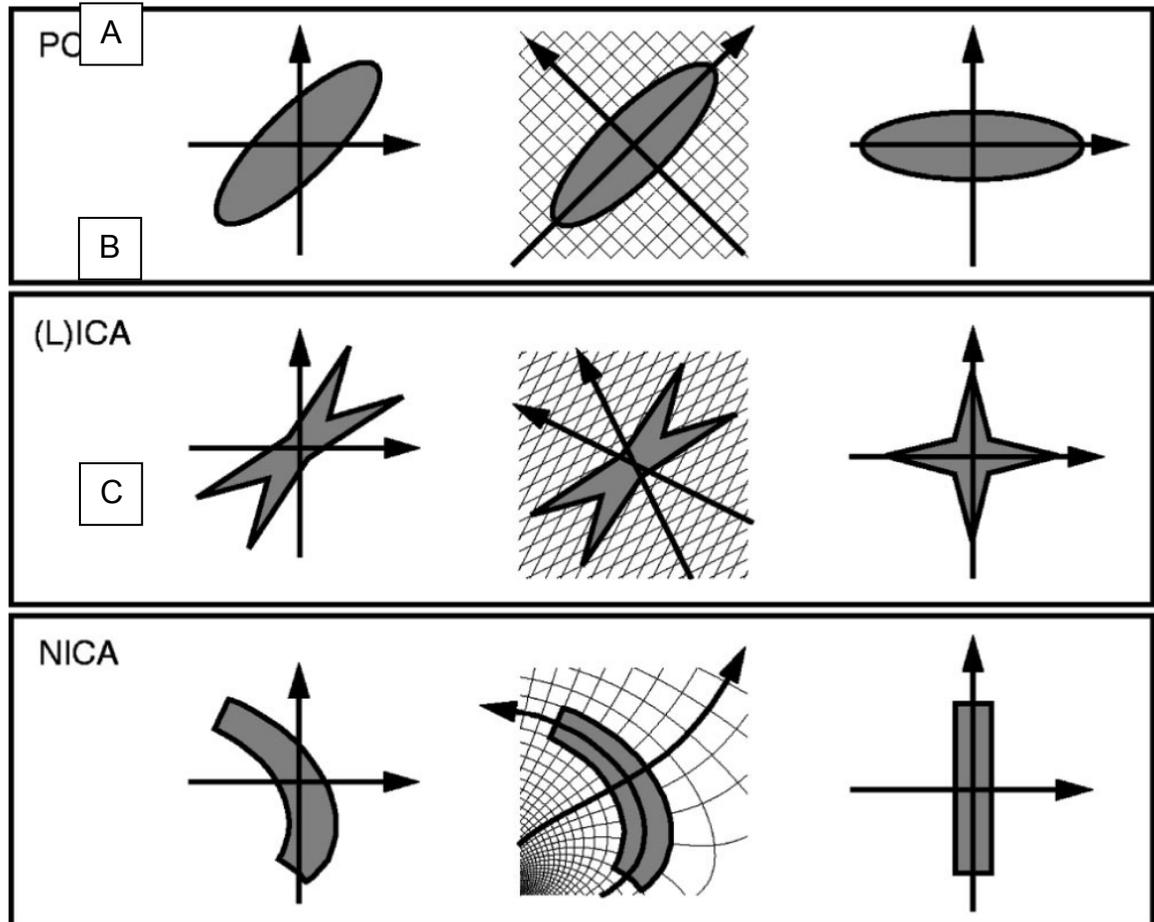
Figure 6. Illustration of the transformation of a probability density function (PDF) using three different methods. A: In PCA, the PDF is projected onto a new coordinate system with rotated Cartesian coordinates where the principal components are uncorrelated; for Gaussian statistics, the components are independent. B: In linear ICA, the PDF is transformed into a feature space with sheared Cartesian coordinates. This results in the separation of a linear mixture of independent sources. C: Separation of nonlinear mixtures via nonlinear ICA (Zetzsche et al., 1999).

While models like ICA or nonlinear ICA can handle more complex data types, they are more complicated and so require cross-validation. In a psychophysical setting, where the data are often hard won in terms of time,

we have chosen in this thesis to confine ourselves to more parsimonious PCA based Factor analysis.

1.3 Central Visual Pathways

1.3.1 Central Projections of Retinal Ganglion Cells

Information supplied by the retina interacts with multiple brain regions and results in visual perception. Ganglion cell axons exit the retina through the optic disk, where they aggregate and form the optic nerve. Optic nerve axons advance directly to the optic chiasm at the base of the diencephalon and then form the optic tract (Figure 7). Decussation at the optic chiasm means that visual information from each retina is processed by corresponding cortical sites in each hemisphere (Purves et al., 2001).

Ganglion cell axons in the optic tract target structures in the midbrain and diencephalon (Figure 7). In the diencephalon, the primary target is the dorsal lateral geniculate nucleus (LGN) of the thalamus. LGN neurons send their axons to the cerebral cortex via the internal capsule. These axons pass through the optic radiations and terminate in the primary visual cortex (V1), also referred to as the visual striate cortex or Brodmann's area 17. In humans, V1 resides along and within the calcarine fissure of the occipital lobe (Purves et al., 2001).

Optic nerve

Optic chiasm

Lateral geniculate nucleus

Optic radiation

Striate cortex

Optic tract

Hypothalamus: regulation of circadian rhythms

Pretectum: reflex control of pupil and lens

Superior colliculus: orienting the movements of head and eyes
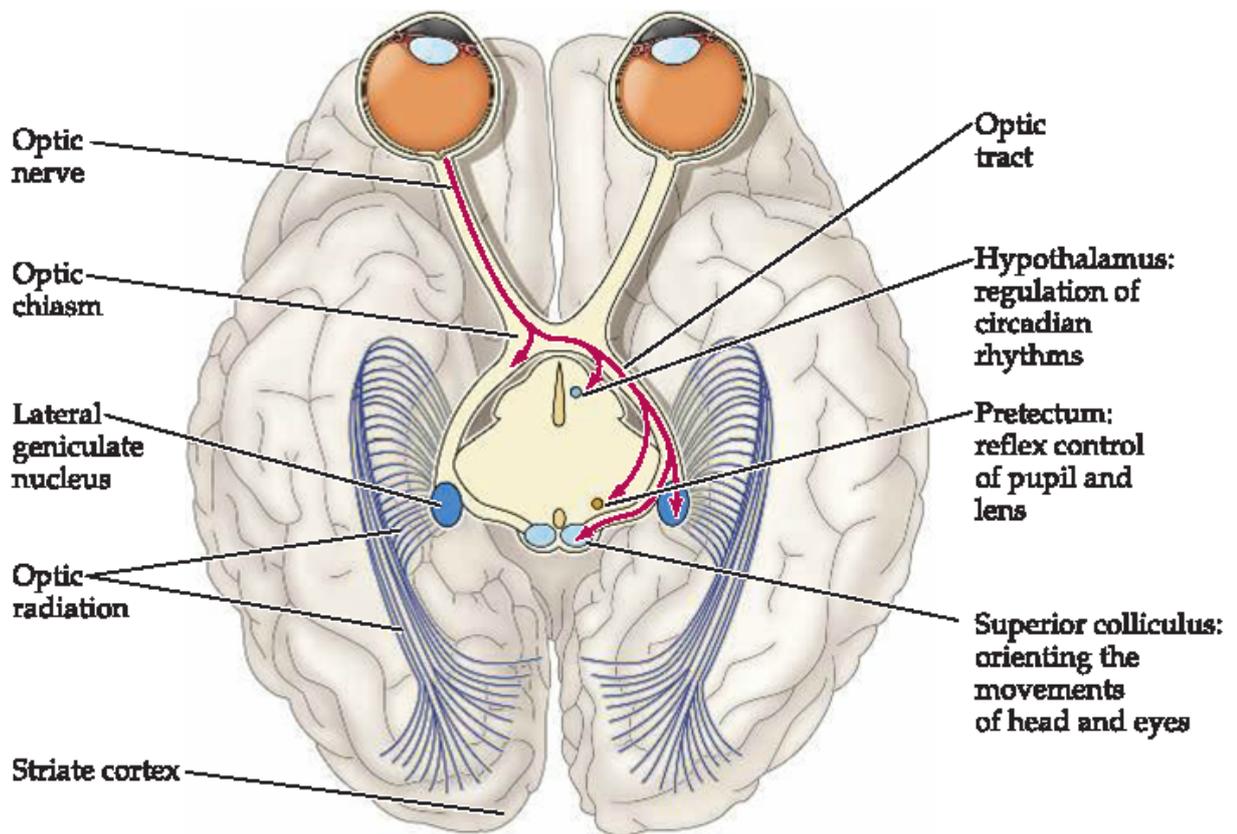
Figure 7. Central projections of retinal ganglion cells. Ganglion cell axons terminate in the LGN, superior colliculus, pretectum, and hypothalamus. Only crossing axons, arising from the right eye, are shown (Purves et al., 2001).

Cortical processing of visual information begins in area V1, which is located in both hemispheres. V1 has a retinotopic organization; this means that it contains a complete map of the visual field. V1 in the left hemisphere receives input from the left LGN; this information derives from the right visual field, which is acquired by the left portion of the two retinas. Conversely, V1 in the right hemisphere processes information from the left visual field. The two hemispheres are connected via the corpus callosum (Purves et al., 2001). The retinotopic mapping is continuous and surprisingly free from deleterious distortions (Adams and Horton, 2003). The region of V1

representing the fovea is larger than those regions that represent more peripheral areas (Tootell et al., 1988, Azzopardi and Cowey, 1996) and mapping in V1 can be considered approximately log-polar (Tootell et al., 1988, Schwartz, 1980, Adams and Horton, 2003).

V1 processes visual inputs from the LGN and relays the results forward to subcortical structures and higher visual areas. In primates, these include V2, V3, MT, MST, and FEF (Maunsell and Newsome, 1987, Van Essen et al., 1992). Reciprocal feedback connections also exist, from extrastriate areas to V1, and in turn from V1 to the LGN (Alitto and Usrey, 2003, Briggs and Usrey, 2008). V1 receives input from other brain regions including the nucleus basalis, which may modulate alertness (Harris and Thiele, 2011).

Structurally, like most neocortical areas, V1 is divided into six functionally distinct layers (Shepherd, 1998, Hubel, 1988). Layer 4 receives most visual input from the LGN and is further divided into 4 layers (4A, 4B, 4Cα, and 4Cβ). In primates, the primary LGN inputs arrive in 4Cα and segregate depending on the source: magnocellular inputs to 4Cα, parvocellular inputs to 4Cß, and koniocellular inputs to layers 1, 3, and 4A (Casagrande and Xu, 2004). The name "striate cortex" derives from a visible band, called the line of Gennari, which represents myelinated axons from the lateral geniculate body terminating in Layer 4.

1.3.2 Functional Organisation of Extrastriate Visual Areas

Anatomical and electrophysiological studies in macaques have elucidated the anatomy and functional significance of the extrastriate areas involved in

visual processing (see (Orban, 2008) for a review). Functional imaging studies of the human extrastriate cortex point to a similar structural arrangement (Sereno et al., 1995) (Figure 8). Each of these higher areas is primarily dependent on V1 for its activation and feed-forward input. The response properties of the neurons in extrastriate areas suggest that they are specialized for different types of visual information, although they are far from fully understood.

The extrastriate areas are believed to be organized into two distinct visual streams. The first evidence for this came from lesion studies in macaques. Lesions within the inferior temporal area were found to cause deficits in visual discrimination, but not visuospatial tasks; conversely, parietal lesions were found to adversely affect visuospatial performance, whilst visual discrimination was spared (Mishkin et al., 1983). On this basis, Mishkin et al. proposed the existence of two processing streams. The ventral stream is involved in object recognition; it includes V4 and leads from the striate cortex into the inferior part of the temporal lobe. The dorsal stream is involved in spatial aspects of vision, such as the analysis of motion; it leads from the striate cortex to the parietal lobe (Mishkin et al., 1983). Goodale and Milner subsequently proposed a modification which emphasized *perception versus action* for the ventral and dorsal processing streams respectively (Goodale and Milner, 1992).
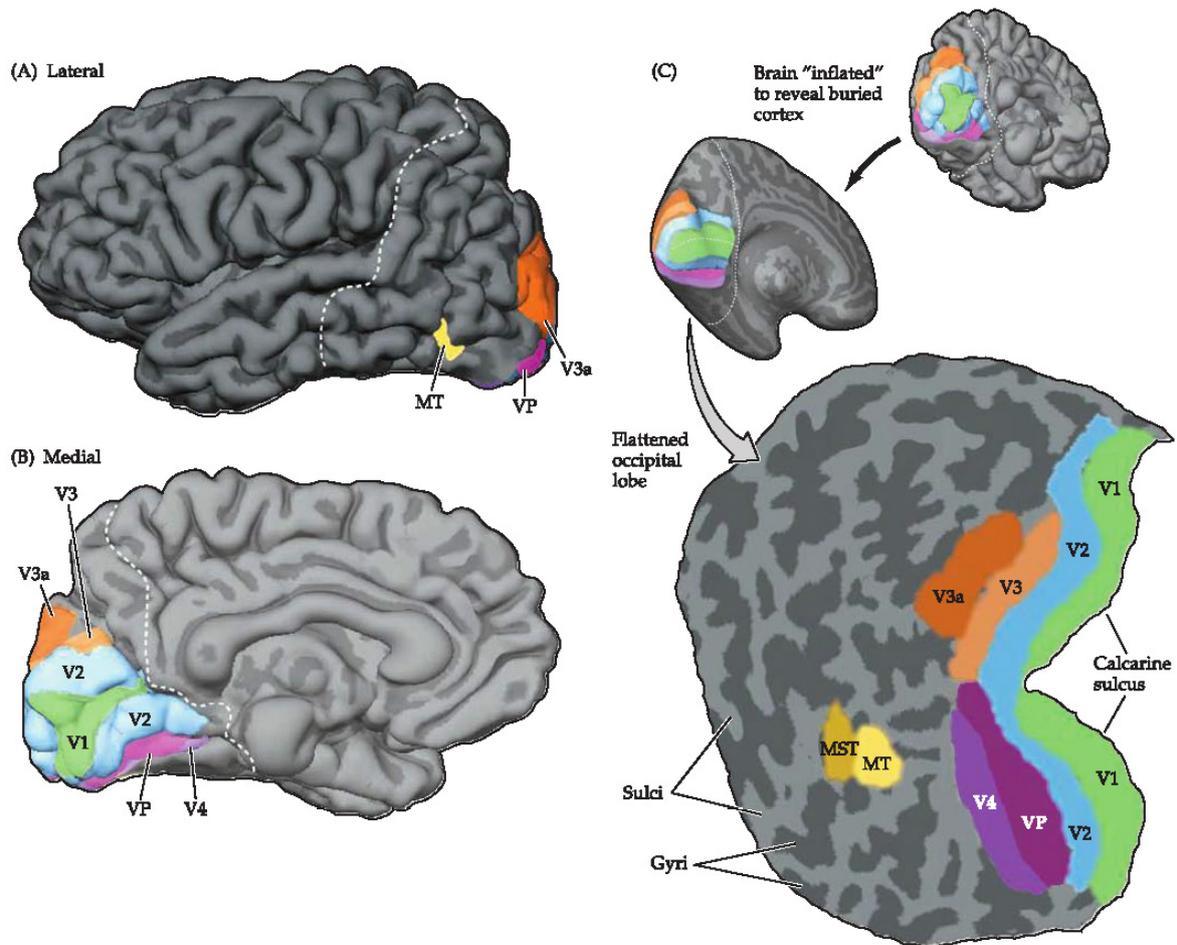
Figure 8. Striate and extrastriate areas in the human brain based on fMRI studies. (A/B) lateral and medial views (respectively) of the human brain derived from fMRI data. The location of V1 is indicated, along with the higher visual areas V2, V3, VP (ventral posterior area), V4, MT (middle temporal area), and MST (medial superior temporal area). (C) flattened view of retinotopically defined visual areas in the occipital lobe. Dark grey indicates cortical regions buried in sulci. The human visual areas closely resemble those of the macaque (Purves et al., 2001) (adapted from (Sereno et al., 1995)).

1.3.3 Spatiotemporal Tuning Properties of Visual Neurons

1.3.3.1 Surround Cells of the Retina

The response profiles of visual neurons have been explored using single cell recording (Boulton et al., 1990) and point or bar stimuli. If a microelectrode is placed near to the target cell, we can directly observe its response to a stimulus, based on changes in its firing rate. By varying the nature and the location of the stimulus, the response behaviour of the target cell can be mapped.

Among the earliest proponents of single cell recording were Hartline et al. They identified a region of frog retina where spots of light altered the firing rate of retinal ganglion cells. They called this region a *receptive field* (Hartline, 1938). In the 1950s, Kuffler et al. became the first to record visual responses from mammalian retinal ganglion cells (Kuffler, 1953). They found that Hartline's receptive field had a specific spatial structure termed a *centre-surround*. On-centre cells were found to have receptive fields composed of a central excitatory region, with a concentric inhibitory region. These two regions interact in an antagonistic way. Off-centre cells were also found, exhibiting the opposite architecture and response behaviour (Kuffler, 1953). The receptive fields of most lateral geniculate nucleus (LGN) cells were subsequently found to have a very similar centre-surround architecture to those of retinal ganglion cells (Hubel and Wiesel, 1961) (Figure 9).
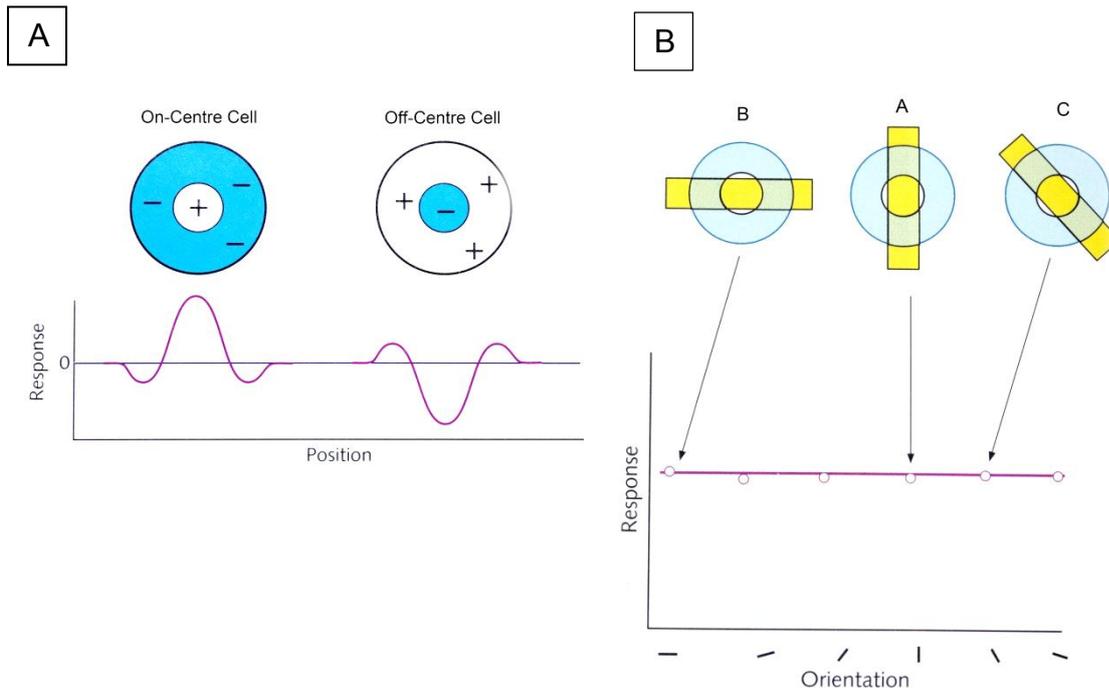
Figure 9. (A) Representations of On- and Off-centre receptive fields in the retina. The upper images show the maps of areas of excitation (white) and inhibition (blue) for an On-centre (left) and Off-centre cell (right). The graph below shows the response of each cell to a spot of light, presented at various positions across the centre of each cell. (B) The concentric receptive field of a LGN cell will respond in the same way to bar stimuli of all orientations (Snowden et al., 2012).

1.3.3.2 Cortical cells

The receptive fields of simple cells have been studied extensively (Hubel and Wiesel, 1968, De Valois et al., 1982). They are oriented and have band-pass characteristics in the spatial and temporal frequency domains. Simple cells have receptive fields with a characteristic spatial structure. Like the neurons of the retina and LGN, they can be divided into distinct excitatory and inhibitory regions. However, their On- and Off- subregions are elongated and

parallel, rather than circular and concentric (Hubel and Wiesel, 1962). There is spatial summation between these subregions; therefore, when the receptive field of a simple cell has been mapped, its response to visual stimuli can be predicted as the linear sum of regional responses (Movshon et al., 1978b).

Based on the architecture of simple cell receptive fields, it is unsurprising that small spots of light illicit only weak responses; instead, these cells respond more strongly to bar stimuli (Hubel and Wiesel, 1962, Hubel and Wiesel, 1968, De Valois et al., 1982). The magnitude of the response is dependent on the orientation of the bar. The cell is most highly stimulated by bars whose aspect ratio and orientation match the elongated subregions. The cell will respond less strongly as the orientation of the bar stimuli diverges from the optimal orientation. This phenomenon is known as *orientation tuning*.

V1 neurons are also typically selective for the spatial frequency of a stimulus. As with orientation, this selectivity arises naturally from the shape of the receptive field. Neurons in V1 tend to be much more selective for spatial frequency than LGN neurons (DeValois and both Professors, 1988) (Figures 10 and 11). Indeed it is clear that simple cell receptive fields resemble Gabor functions, as first pointed out by Marcelja, and as such these cells collectively appear to compute a Gabor-like transform of visual space, i.e. one localised in space and spatial frequency.
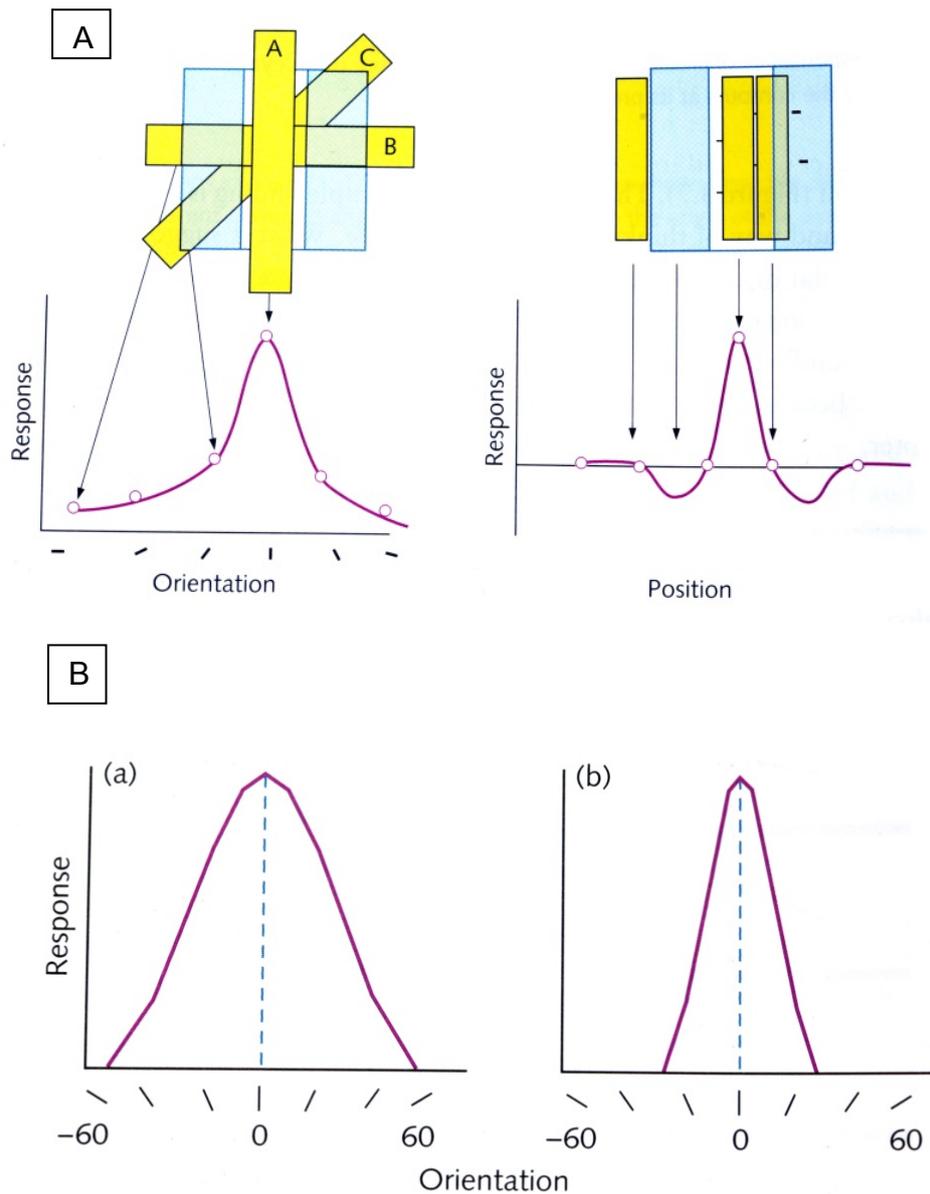
Figure 10. (A) The elongated receptive field of a cell in the primary visual cortex shows a strong preference for bar stimuli of a given orientation. In this example, the cell responds optimally to a vertical bar stimulus. (B) Orientation-selective cells with different spatial preferences, i.e. the width of their spatial frequency tuning relative to their centre frequency. Cell (a) responds optimally to a vertical bar stimulus. If we rotate the bar stimulus and plot the cell's response, we produce a bell curve, with no response recorded for a bar at 60

40

degrees orientation from vertical. Cell (b) also responds optimally to a vertical bar stimulus. Again, rotating the bar stimulus and plotting its response, we produce a bell curve. However, this time no response is recorded for a bar at 30 degrees orientation from vertical. Cell (b) is more narrowly tuned for orientation than cell (a).
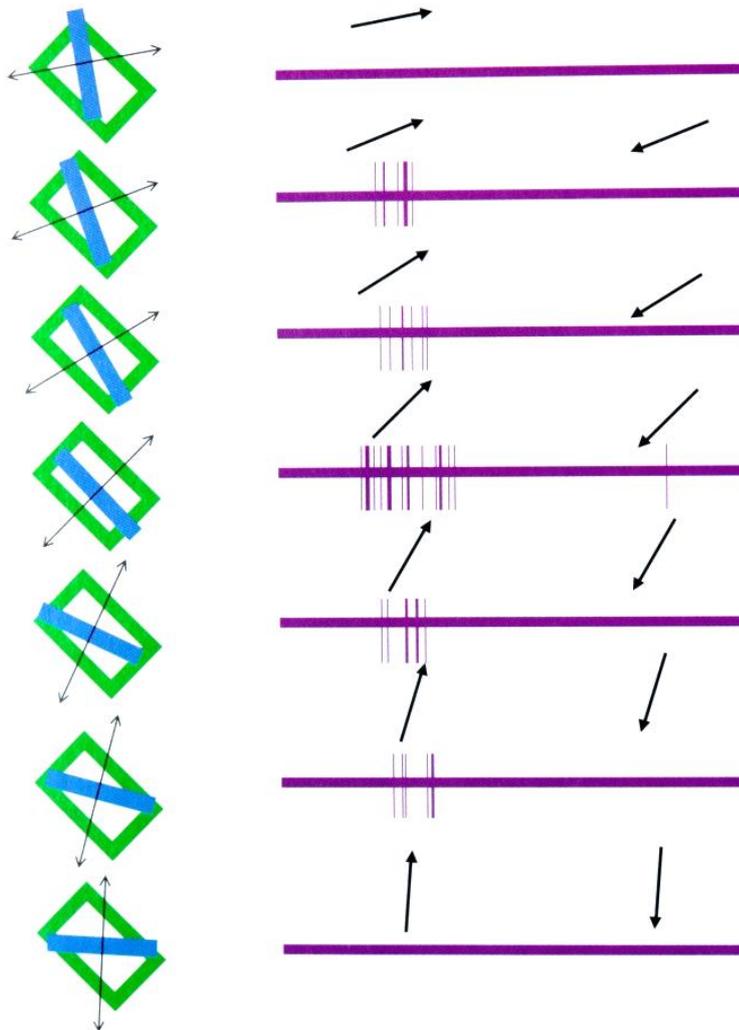
Figure 11. Response of a V1 neuron to bar stimuli of different orientations. (a) The receptive field of the cell is represented by the green rectangle and the bar stimulus by the blue bars. The bar was moved through the receptive field and back again, indicated by

arrows. (b) The firing rate of the cell was rarely elevated in response to the bar stimulus, unless it was close to the preferred orientation. Also note that this cell was selective for the direction of motion (Snowden et al., 2012).

However, it should be noted that some modern studies using different stimuli have called into question the canonical receptive field architecture of visual cortical neurons (Victor and Knight, 2003). There is some evidence that individual neurons are not restricted to a single orientation (Purpura et al., 1994b). Furthermore, some aspects of early visual processing appear to utilise two-dimensional information, including texture (Victor and Brodie, 1978) and T-junctions (Rubin, 2001). Victor et al. have proposed that receptive fields may instead be viewed as two-dimensional profiles, which are simultaneously space and bandwidth limited; such representations are based on Hermite functions (Victor and Knight, 2003).

Some cells of the primary visual cortex also exhibit *directional* selectivity. In the cat visual cortex, Hubel and Wiesel (Hubel and Wiesel, 1959) showed that many neurons responded more strongly to a bar stimulus moved in a particular direction. During the same period, Reichardt et al. (Reichardt and Rosenblith, 1961) began studying optomotor responses in the fly. Fly optomotor responses are used for course stabilization; when a fly is presented with a rotating wide-field stimulus that mimics its surrounding visual space, it reacts by turning in the direction of stimulus motion (Heisenberg and Wolf, 1984). Reichardt et al. proposed a motion-opponent directionally selective model (Reichardt and Rosenblith, 1961) which has

become the canonical approach to motion modelling. The basic Reichardt model consists of pairs of motion-detecting subunits. When each subunit receives an input, a signal is sent to the other subunit. Concurrently, a time delay is introduced to the signal within the subunit. Thus, within each subunit, two inputs are multiplied, one received directly from the receptor with a time delay and the other received from the adjacent subunit. The multiplied values from the two subunits are subtracted and the sign of the output determines the preferred direction (Reichardt and Rosenblith, 1961). In this way, motion is computed via cross correlation. Van Santen and Sperling subsequently modified this model for human motion perception (Van Santen and Sperling, 1985). The modified Reichardt model features spatial frequency-tuned receptive fields (such as Gabor functions). In this model, the subunits are formed from pairs of filters that differ in phase by 90 degrees (Van Santen and Sperling, 1985).

Another model, closely related to the Reichardt model, has been proposed by Adelson and Bergen (Adelson and Bergen, 1985). In this model, the linear filtering stage is followed by a nonlinear filter which generates a phase-independent signal (or *Motion Energy*). Emerson et al. repeated the bar tests of directionally selective cells in the cat's striate cortex (Emerson et al., 1992), as originally studied by Hubel et al. (Hubel and Wiesel, 1962); their responses were then compared to predicted responses based on Reichardt and Motion Energy. They identified clear distinctions between the two models and found that the physiological responses aligned most closely with the Motion Energy model (Emerson et al., 1992).

Although V1 neurons appear to have some selectivity for stimulus speed, more typically they are thought to be selective for temporal frequency (Movshon et al., 1978b). Temporal frequency is defined as the inverse of the period between temporal dark and light oscillations. Mante and Carandini point out that the selectivity of simple cells can be thought of in integrated terms by considering a three-dimensional *frequency space* (Mante and Carandini, 2005). In frequency space, the *Fx* and *Fy* axes define spatial frequency and the *Fz* axis, temporal frequency. In this space, the distinct preferences of a given receptive field can be represented by an ellipsoid; its *z*-displacement indicates preferred temporal frequency, *xy* displacement indicates preferred spatial frequency, and angle to the ground plane indicates its preferred direction of motion and orientation (Mante and Carandini, 2005). If all V1 neurons are sampled and analysed in this manner, they would presumably tile the frequency space (or at least the region of frequency space to which humans are perceptually sensitive).

1.3.3.3 Complex Cells

Whilst simple cells share relatively similar receptive field architectures, those of complex cells are more diverse. Complex cell receptive fields do not usually have distinct subregions (any regions responding to *on* or *off* in a rectified fashion). In response to stimuli, complex cells exhibit non-linearities of spatial summation. Therefore, the optimal stimulus for a complex cell cannot be predicted based on receptive field architecture alone (Hubel and Wiesel, 1962, Movshon et al., 1978a).

Moving sinusoidal gratings may be used to classify simple and complex cells: simple cells respond to such stimuli with oscillatory responses, but the responses of complex cells are less modulated (De Valois et al., 1982, Movshon et al., 1978a). Fourier analysis can be used to extract the mean response (F0) and amplitude of response at the fundamental frequency ($F$1) (Movshon et al., 1978a). Based on this criterion, simple cells can be defined to have $F$1/F0 ratios greater than 1, whilst complex cells have $F$1/F0 ratios less than 1 (Movshon et al., 1978a, Ibbotson et al., 2005). However, other investigators have pointed out that the bimodality of *F1/F0* ratios does not *necessarily* indicate the existence of two discrete populations of cells (Mechler and Ringach, 2002, Abbott and Chance, 2002). In one part of their detailed analysis, Mechler and Ringach (Mechler and Ringach, 2002) showed that a simple cell receptive field can also produce complex-like responses.

The discovery of cells in the visual processing pathway with increasingly complex properties led Hubel and Wiesel to propose the *hierarchical model* (Hubel and Wiesel, 1962). According to this model, simple cell receptive fields are constructed from the convergence of LGN inputs, aligned in visual space. In turn, the output of spatially offset simple cells provides the input to a complex cell.

This model received support from a study by Van Kleef et al. (Van Kleef et al., 2010). They used high and low contrast sinusoidal gratings, covering a range of spatial and temporal frequencies, to analyse complex cell receptive fields. The authors found that the responses of complex cells to low contrast stimuli resembled those of simple cells. The model that best reproduced their

45

findings was a variation of the hierarchical model wherein spatially offset simple cells (each with different contrast response functions) provided input to a complex cell. At low contrasts, only a single simple cells was sufficiently stimulated to respond; however, as contrast was increased, additional spatially offset simple cells with higher contrast thresholds responded, and the resultant combinatorial response was "complex" (Van Kleef et al., 2010).

Despite being appealingly simple, several other studies have challenged the hierarchical model. Dreher et al. found that some simple and complex cells exhibit end-stopping, undermining the concept of a three-step (simple-complex-hypercomplex) hierarchy (Dreher, 1972) (see Section 2.3.4). Some complex cells were also identified which received direct geniculate input (Martin and Whitteridge, 1984). An alternative model attributes complex cell activity to lateral recurrent connections between cortical neurons. i.e.: the response properties of complex cells might result from mutual excitation (Chance et al., 1999, Tao et al., 2004). A number of techniques have been employed to determine the nature of complex cell inputs (Movshon et al., 1978a, Martinez and Alonso, 2001, Anzai et al., 1999). However, there is limited evidence for the presence of excitatory connections from simple to complex cells (Toyama et al., 1981a, Toyama et al., 1981b).

The functional significance of complex cells is unclear. One theory proposes that complex cells provide feedback to earlier stages in the visual hierarchy. In so doing, they may mediate a gain control process, allowing visual neurons to adapt to variations in local contrast (Hubel, 1988). Another model proposes that complex cells represent a step up from a linear Gabor transform to a nonlinear, sparse representation based on the Wigner

distribution; this representation provides invariance to tilt, rotation and position in the visual field (Jacobson and Wechsler, 1984). The Wigner distribution function was first proposed by Eugene Wigner in 1932 (Wigner, 1932). It has been found to be useful in the field of signal analysis where is can be used to define a time-dependent frequency spectrum (Claasen and Mecklenbrauker, 1980).

Alternatively, it has been suggested that different brain areas and their diverse cell types represent a single linear/nonlinear hierarchical representation of visual space based on the related idea of the Wigner transform (Victor and Knight, 2003).

1.3.3.4 Hypercomplex Cells

Hubel and Wiesel also described neurons in the primary visual cortex which were not only orientation-tuned, but length-tuned (Hubel and Wiesel, 1965, Hubel and Wiesel, 1968). These are known as *hypercomplex*, *end-stopped*, or *end-inhibited cells*. A hypercomplex cell produces an excitatory response when a bar stimulus of a given length and orientation moves across its receptive field. Lengthening the bar stimulus improves the excitatory response (Hubel and Wiesel, 1965, Hubel and Wiesel, 1968). However, if the length of the bar exceeds a predefined limit, there is a marked decrease in the response (Figure 12). Curiously, this occurs in a part of visual space beyond where a response can be evoked from the cell.
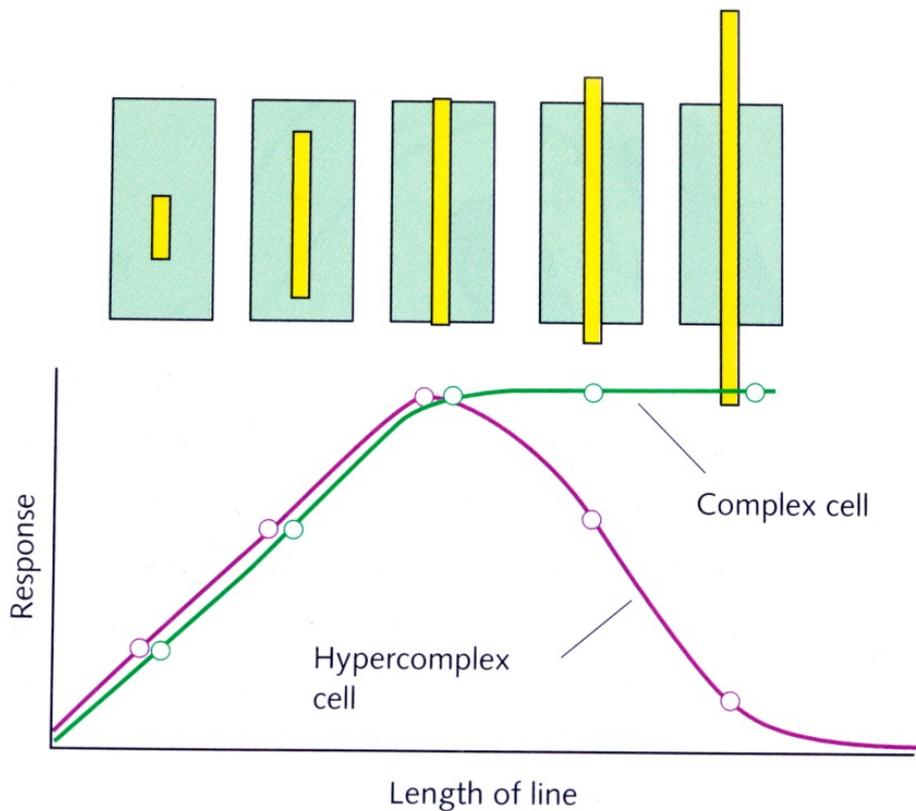
Figure 12. Hypercomplex (or end-stopped) cells prefer bar stimuli of a certain length, in addition to having an orientation preference. Studies have found that end-stopping is also present in many simple and complex cells (Dreher, 1972), which suggests that these cells are a subset of simple and complex types (Murphy and Sillito, 1987, Snowden et al., 2012).

A visual stimulus which falls *outside* of the receptive field of a visual cortical neuron does not trigger spiking; however, it may still *modulate* the response of the neuron (Adesnik et al., 2012, Gilbert et al., 1996). For example, it may attenuate the neuron's response to subsequent visual stimuli which fall within the receptive field. This phenomenon is called *surround suppression* (Hubel and Wiesel, 1965, Levitt and Lund, 1997a). Surround suppression has been

suggested to contribute to perceptual effects such as curvature detection (Dobbins et al., 1987) and orientation discrimination (Mareschal and Shapley, 2004). It is not only observed in the cortex, but also in the retina (Solomon et al., 2006) and thalamus (Alitto and Usrey, 2008), although the latter are not orientation tuned.

It is not known whether all of the surround suppression observed in the cortex has been relayed from earlier stages of visual processing. Some experimental findings and theoretical models suggest that cortical circuits contribute to surround suppression (Bolz and Gilbert, 1986, Ozeki et al., 2009, Haider et al., 2010). A cortical circuit has been described that directly contributes to surround suppression in the superficial layers of V1 (Adesnik et al., 2012).

End-stopping has subsequently been observed in both simple and complex cells (Dreher, 1972). Thus, hypercomplex cells are now considered to be a subset of simple and complex cells (Murphy and Sillito, 1987). The functional significance of end-stopping is unclear. Significantly for this thesis, hypercomplex cells may be important in removing redundant information (see Section 1.4 below) in natural images at higher-order correlations (Zetzsche and Rohrbein, 2001).

1.3.4 Explanatory Models of Visual Cortical Neurons

Although we know much about *how* early visual information processed, the reasons *why* the visual system operates in this manner are more opaque. The characterisation of simple cell receptive fields and their spatial frequency

tuning led to the hypothesis that the visual cortex acts as a spatial frequency analyser (Maffei and Fiorentini, 1973). This would mean that the visual cortex performs a two-dimensional Fourier transformation of the visual scene. However, the broad spatial-frequency bandwidths, and localized distribution of cortical cells, make them unsuitable for extracting Fourier coefficients (Field, 1987) and this hypothesis is no longer favoured. Therefore, how does the visual system implement frequency selective mechanisms?

Several studies have proposed that the behaviour of cortical simple cells can be modelled in terms of Gabor functions (see Section1.4) (Watson, 1983, Palmer et al., 1985, Marčelja, 1980, Daugman, 1985, Jones and Palmer, 1987). For example, Jones and Palmer showed that the real component of a Gabor function matches the receptive field functions of feline simple cells (Jones and Palmer, 1987). According to Marcelja et al. (Marčelja, 1980) Gabor functions provide the best local description of the most valuable image elements (in informational terms). A typical image contains lines and edges, with interspersed regions of uniform texture. Intuitively, the most important image information pertains to the locations and orientations of key lines and edges. Whereas a frequency analysis of a localized edge stimulus would elicit responses from many cells, a *Gabor* representation would trigger responses from only a few cells. The Gabor response is thus suited to the extraction of *significant* image features (Marčelja, 1980).

There is some experimental evidence to support a Gabor representation. For example, De Valois et al. measured spatial frequency responses from macaque simple cells (De Valois et al., 1978). As described by Marcelja et al. (Marčelja, 1980) there is a strong resemblance between the cell

responses recorded by De Valois et al. and Gabor elementary signals (Figure 13).
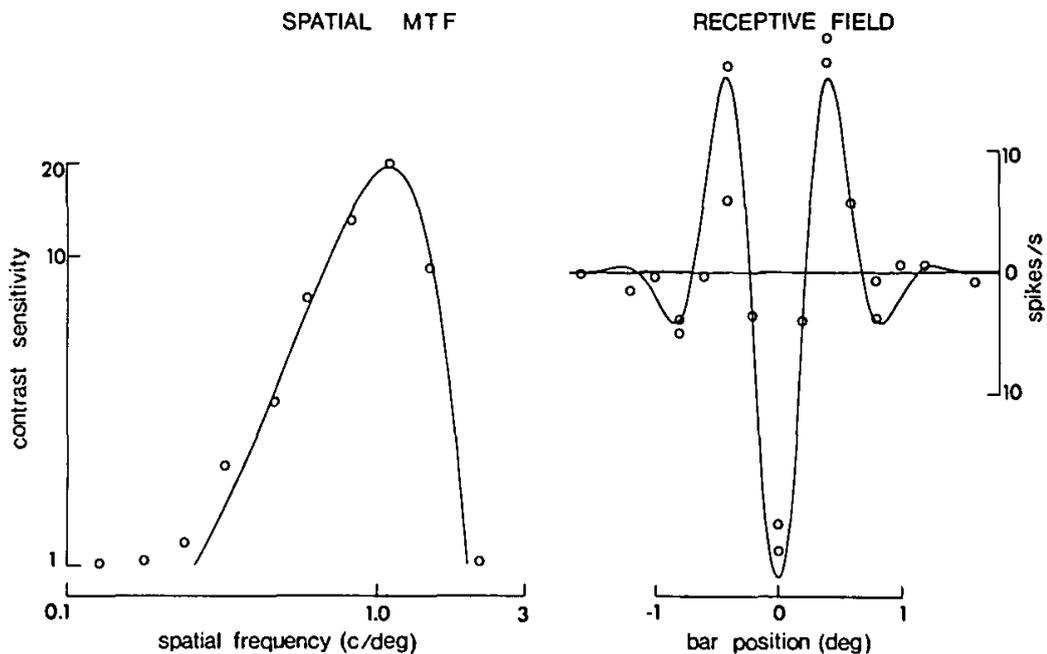
Figure 13. Comparison of the response of a macaque simple cell (De Valois et al., 1978) and a Gabor elementary signal. The figure on the left shows a least-squares-fit of the contrast sensitivity data as a function of spatial frequency. The figure on the right shows a Fourier transform of the elementary signal that fits the contrast sensitivity data, along with the measurements to a response to a narrow bar stimulus (Marčelja, 1980).

A Gabor representation captures a second significant aspect of cortical cell receptive fields - that they are part of a scale invariant self-similar hierarchy. I.e. at each cortical location, several receptive fields occur that are scaled versions of each other. By consequence, these receptive fields can process visual information across a variety of spatial scales. Evidence of this scale

51

invariant arrangement has been found in macaque cortex (De Valois et al., 1982) as well as psychometric studies (Polat and Sagi, 1993, Teichert et al., 2007). In a series of contrast detection tasks, Polat and Sagi found evidence of scale-independent inhibitory and excitatory subregions in the human visual cortex (Polat and Sagi, 1993).

Teichert et al. attempted to characterise how the receptive field properties of V1 neurons varied with preferred spatial wavelength (Teichert et al., 2007). The sizes of grating summation fields and inhibitory surrounds increased with preferred spatial wavelength. This supports the hypothesis that processing in V1 supports scale-invariant visual performance. The scale-invariant self-similar hierarchy of receptive fields may help with perspective judgements and looming (the sudden increase in size of an image, often resulting from rapid approach).

As discussed in Section 1.3.1, the mapping of visual cortical cells in V1 is approximately log-polar in radius (centred on the fovea), with the cortical region representing the fovea being larger (Tootell et al., 1988, Schwartz, 1980, Adams and Horton, 2003). Thus, if a person fixates on an object that looms towards them, such that its visual field area increases by r, the cortical representation of the object increases by $\log(r)$. Therefore, the relative size and shape of its cortical representation is maintained (Abdullah et al., 2012). In this context, Abdullah et al. investigated the effects of image contrast, stimulus density, and viewing distance upon a multifocal VEP method called MSV (Abdullah et al., 2012), which uses annular stimuli centred on the fovea. They found that viewing distance did not greatly affect response amplitude, suggesting that scaled annular stimuli can be used to study central and

peripheral disease (Abdullah et al., 2012). Taken together, these findings have implications for visual psychometric studies where stimuli size (and pixel size) may vary.

1.4 Natural Images and Information Theory

1.4.1 Information Theory

Information theory was originally developed in 1948 by Claude Shannon (Shannon and Weaver, 1948) (reviewed in (Shannon, 2001)). Information theory was originally conceived to address problems in the signal transmission (for example, finding the limit of signal compression for a given level of transmission reliability). More broadly, the theory has implications for the quantification of other types of information; therefore, it now plays an important role in modelling neurobiological systems and as such is directly pertinent to this thesis (Shannon and Weaver, 1948, Shannon, 2001).

There are two primary theorems of information theory, both of which relate to its canonical problem: the transmission of information over a noisy channel (Shannon and Weaver, 1948, Shannon, 2001). Firstly, according to *source coding theorem*, the number of bits needed to represent the result of an uncertain event is given by its entropy. Entropy quantifies the uncertainty involved in predicting the value of a random variable. Secondly, according to *noisy-channel coding theorem*, reliable communication is possible over a noisy channel, provided that the rate of communication is below a threshold; this threshold is called the *channel capacity* (Shannon and Weaver, 1948, Shannon, 2001).

Motivated by information theory, Attneave proposed that the goal of the visual system is to generate an efficient representation of the incoming signal (Attneave, 1954). In the neurological context, Barlow hypothesized that efficient representation must involve the removal of statistically redundant information during early vision (Barlow, 1961, Barlow, 2001). Variants of this *efficient coding hypothesis* have been formulated by several authors (Laughlin, 1981, Van Hateren, 1992c, Field, 1994, Atick and Redlich, 1992).

The efficient coding hypothesis states that neurons have limited energy budgets and information capacities; therefore, they should not waste energy nor bandwidth by transmitting *redundant* information (Barlow, 1961, Barlow, 2001). To put it another way, in order to maximise available computing resources, neurons should encode as much *functional* information as is structurally and energetically sustainable. Reducing informational redundancy potentially increases the amount of information that can be transmitted through the channels of the visual system.

Natural images contain a considerable amount of redundant information, so redundancy reductions are readily achievable. Image regions are spatially correlated, as discussed in Section 1.2 (Griffith, 2003, Srinivasan and Shobha, 2008, Heilbronner and Barrett, 2013). Natural images contain large amounts of structural information in the form of higher-order spatial correlations (Figure 4) (Franz and Schölkopf, 2005). Images may also be considered to be correlated in time, depending on their rate of change. Kersten demonstrated informational redundancy experimentally by asking human subjects to restore missing grey levels from 128 pixel square images (Kersten, 1987). Based on the resulting performance functions, Kersten

estimated that the images contained informational redundancies of 46-72% (Kersten, 1987). Consistent with efficient coding, the human visual system must employ mechanisms that remove redundant information, whilst retaining that image statistics that are behaviourally relevant.

Experimental approaches to efficient coding can be divided into two types, which will be discussed in Section 1.4.2. The most direct approach is to record the responses of visual neurons when exposed to naturalistic stimuli. This approach is analogous to that discussed in the preceding sections. i.e.: it is a bottom-up approach, where basic neurophysiology is examined and models of visual processing are built up from those observations.

An alternative, complementary, approach is to consider vision from a top-down perspective; that is, analyse the informational content of natural images and consider the challenges it presents for visual processing. The evolution of sensory systems, including vision, has been driven by 3 factors: the tasks that the organism must perform, its environmental constraints, and the computational limitations of neurons (both structural and metabolic) (Simoncelli and Olshausen, 2001). Visual processes have been shaped by evolutionary pressures and can therefore be assumed to be exquisitely adapted to their function. It therefore follows that the computational analysis of image statistics can reveal the underlying properties of neurons themselves.

In this context, note that the efficient coding hypothesis, as originally stated, does not guarantee the accuracy of neural representation, nor the effects of contaminating noise (Simoncelli and Olshausen, 2001). Furthermore, neural

efficiency may be assumed to depend both on the statistics of the input and on the transformation which maps it to neural responses. Therefore, because coding efficiency may vary between different inputs, analyses based on information theory should be based on *large populations* of images.

1.4.2 Neurological Studies related to Information Theory

Direct experimental support for efficient coding has emerged from studies of large monopolar cells (LMCs), first-order interneurons in the fly compound eye (Laughlin, 1981) that are analogues to the bipolar cells of the vertebrate retina. The dynamic range and signal to noise ratios of neurons are necessarily limited so, in order to maximise their information capacity, their sensitivity must be set at the correct level. If their sensitivity is too high, inputs will saturate and information will be lost through clipping. Conversely, if their sensitivity is too low, a proportion of the response range will not be utilized. Laughlin found that LMCs maximise their informational capacity by incorporating an optimised intensity-response function (Laughlin, 1981).

The encoding of a high quality image, such as one illuminated by a bright light, benefits from redundancy reduction. However, redundancy reduction may be problematic at low light levels, where is can enhance spatial and temporal fluctuations due to noise (Van Hateren, 1992c). Therefore, a mechanism is required which can adapt to different signal to noise ratios (SNRs) and reduce image redundancy only when the SNR is high. Van Hateren proposed a model which maximizes information flow through a noisy channel of limited dynamic range (Van Hateren, 1992a, Van Hateren,

1992b). At low SNRs, the channel filter has low-pass characteristics. At high SNRs, the filter has band-pass characteristics. The author found that neural images obtained by recording from second-order fly neurons were consistent with theoretical predictions based on this information maximisation model (Van Hateren, 1992a, Van Hateren, 1992b).

Srinivasan et al. provided the first evidence that the visual system takes advantage of the correlational structure of natural images to implement redundancy reduction (Srinivasan et al., 1982). Receptive field architecture exploits the degree of correlation in images by taking a weighted mean of the signals in neighbouring receptors. Thereby, a statistical prediction of the redundant component of the signal at the centre is produced. The predicted value is subtracted from the actual centre signal. The predictive part of the model incorporated the autocorrelation function (ACF). In this manner, the entire dynamic range of a neuron can be devoted to encoding non-redundant intensities; this also minimizes the effects of intrinsic noise. Srinivasan termed this process *predictive encoding* and it has the effect of reducing spatial redundancy. Note that the redundant information removed is second-order (Srinivasan et al., 1982).

Field et al. undertook a series of studies which related the properties of visual cortical cells to the statistics of natural images (Field, 1987, Field, 1994). The authors evaluated various coding schemes; for example, an optimization over the parameters of a Gabor function derived from a range of response properties that were characteristic of cortical simple cells (Field, 1987). Based on this analysis, the authors concluded that visual neurons are optimized for the coding of natural images. Furthermore, the spatial-

frequency and orientation tuning of simple cells is optimized for a *sparse* coding representation. In sparse coding, most visual neurons have small amplitude responses. To put it another way, the goal of sparse coding is to have a minimum number of cells highly active at any one time; this may in fact mean that the total number of cells is greater than the absolute minimum required to represent the image. As described in Section 1.2.5 such a representation is over-complete. By contrast, a true Gabor representation uses the minimum number of cells per image, and so all cells are often active. Sparse coding simplifies visual processing as it promotes the detection of redundant image statistics (Field, 1987).

Direct evidence of sparse coding was subsequently obtained by Vinje and Gallant (Vinje and Gallant, 2000). They recorded V1 responses from conscious macaques whilst they viewed natural scenes. The authors found that, during natural vision, interactions between classical and non-classical receptive fields resulted in the formation of sparse representations of the visual scene (Vinje and Gallant, 2000).

An interesting demonstration of sparse coding was published by Olshausen et al. They used learning algorithms to derive sparse linear codes from natural images (Olshausen and Field, 1996, Olshausen and Field, 1997). Starting from random initial conditions, the set of functions that emerged after hundreds of thousands of training sets strongly resembled the canonical simple receptive field architecture. I.e. they were oriented, spatially localized, and band-pass for different spatial frequencies (Olshausen and Field, 1996, Olshausen and Field, 1997).

Several other authors have taken a similar approach using independent component analysis (ICA) of natural images and have observed the emergence of functions resembling simple cell receptive fields (Bell and Sejnowski, 1997, Hurri, 1997, Hurri et al., 1996, van Hateren and van der Schaaf, 1998, Wachtler et al., 2001). For example, Wachtler et al. used ICA to study the efficient representation of colour in natural scenes. When applied to image patches, they observed that some basis functions were achromatic and others indicated colour opponency (although the latter were not strictly orthogonal). The authors suggest that non-orthogonal opponent encoding of photoreceptor signals produces higher coding efficiency. Compared to PCA analysis, ICA resulted in a more sparse representation and a higher coding efficiency (Wachtler et al., 2001). The PCA derived functions did not resemble V1 receptive fields.

As pointed out by van Hateren, the ICA model is abstracted and ignores many aspects of simple cells, such as nonlinearities and contrast adaptation. Nevertheless, the model has interesting information theoretic implications. If one of the main functions of simple cells is the decomposition of the visual scene into independent components, it is not surprising that the distribution of properties (spatial frequency, orientation, etc.) is determined by the statistics of the visual environment (van Hateren and van der Schaaf, 1998).

In evolutionary terms, sparse coding makes logical sense. Degrees of coding sparseness can be thought of as a continuum (Foldiak, 1995, Olshausen and Field, 2004). At one extreme are dense codes, where many visual neurons respond to most stimuli. For example, a strict Gabor representation would have only as many cells as retinal inputs (a sort of minimal sampling rate

representation). Due to the large numbers of constantly active neurons, this representation would be metabolically inefficient, if compact. At the other extreme are over-complete sparse codes (Olshausen and Field, 1996), which contain more cells than inputs. In this case, neuronal responses can be more selective for particular image features. As a result, relative few cells are active for any one image (in the limit one cell for every possible image). Thus, an *unsustainably large* number of neurons would be required to encode visual information, but would provide a low metabolic cost to represent any particular image. That being said, the costs of maintaining an overly large brain, together with capacity of the cranium, are limiting. The previously mentioned Wigner distribution is an example of such an over-complete sparse code, in which the main cell types are like complex cells. The ICA analysis of natural scenes discussed above tends to generate over-complete sparse codes (Bell and Sejnowski, 1997).

Thus, the extremes are inefficient, albeit in different ways. Conversely, if neurons encode only the most informative visual information, the resultant sparse codes can transmit information with minimal redundancy (Foldiak, 1995, Olshausen and Field, 2004). Evolutionary pressure can reasonably be assumed to have selected coding strategies that are optimally efficient in both informational and metabolic terms (if we assume the absence of *conflicting* evolutionary pressures). This is conceptually consistent with the observation by Balasubramanian, that sensory systems invest their resources in relation to the expected information gain (Balasubramanian and Sterling, 2009).

The majority of research into the visual responses of single striate cortical neurons has focused on two-point spatial correlations, as encapsulated by spatial frequency and orientation tuning. As mentioned above, the Fourier transform of the power spectrum is the auto-correlation function, so the view of early cortical processing being like a Gabor representation is equivalent to considering only second-order correlations. There has been less research into the processing third- and higher-order correlations, despite their ubiquity (Figure 4). Purpura et al. noted that while two-point correlations provide the visual system with information about spatial frequency content, encoding features such as contours *requires* mechanisms that are sensitive to third- and higher-order correlations (Purpura et al., 1994a). Higher-order image statistics may have been traditionally neglected as they are more challenging to analyse (Geisler, 2008), or because the second-order characteristics of visual neurons are better understood. However, higher-order image statistics are central to our perception of natural image features such as textures and contours (Oppenheim and Lim, 1981). In this context, the study by Tkacik et al. is particularly relevant (Tkacik et al., 2010).

Tkacik et al. analysed the fourth-order spatial correlations and intensity histograms of natural images. They then classified these image statistics according to how informative they were about the local structure of natural images. Fourth-order correlations that were most informative were also found to be the most visually salient (Victor and Conte, 1991, Julesz et al., 1978). Sensitivity to the latter high-order correlations is known be encoded by the visual cortex (Victor and Conte, 1991, Purpura et al., 1994b). The authors suggest that the visual system filters out predictable (and therefore

61

perceptually non-salient) high-order structural information from natural images, whilst retaining that which is not predictable (and therefore most salient) (Tkacik et al., 2010). This study suggests that efficient coding applies to high-order spatial correlations. Zetzsche et al. have proposed that end-stopping may be important in the filtering out of redundant higher-order structural information (Zetzsche and Rohrbein, 2001).

In the context of the above discussion, it is clear that investigating higher-order correlations - and how they are processed - is important to understanding the human visual system. Such investigations may also tell us a great deal about cortical functioning (Victor, 1995, Purpura et al., 1994b, Victor and Conte, 1996). One way to approach the problem is to use artificially generated textures (as opposed to natural images) to probe the sensitivities and limitations of visual perception.

1.5 Textures and Texture Perception

This thesis makes extensive use of isotrigon textures; this section will introduce the concept of texture, before considering the nature of isotrigon textures and how they are perceived.

The natural world is rich with texture and the surface of any visible object is textured at some scale. We have an intuitive sense of what texture means, but it is harder to define in precise terms. Texture can be considered the property of *stuff* in an image, as opposed to distinct features such as edges (Adelson and Bergen, 1991). Alternatively, texture is an "organised area phenomenon"' which can be decomposed into *primitives* with specific spatial

distributions (Haralick, 1979). However, textures are often highly variable and demonstrate contradictory properties, which challenge a strict definition (Landy and Graham, 2004, Bergen and Adelson, 1991, Adelson, 2001).

We intuit that texture is a visual indicator of surface properties, but it can also be a visual cue to planar orientation and depth (Landy and Graham, 2004, Bergen and Adelson, 1991). In a given scene, boundaries may be delineated by colour-opponent mechanisms or differences in contrast (Landy and Graham, 2004). However, humans are also adept at performing image segregation based on *textural* information (Landy and Graham, 2004). Adelson posits that texture allows observers to determine whether two regions are composed of different substances (Adelson, 2001). In this manner, differently textured regions may lead to the perception of a segregating border. Continuous changes in texture may also result in the perception of three-dimensional shape (Gibson, 1950, Blake et al., 1993, Todd et al., 2004). This naturally leads to the question: what aspects of image structure lead to image segregation based on texture?

As Victor points out, "…basic visual judgments are fundamentally statistical in nature…" and image statistics are central to our understanding of texture perception (Victor and Conte, 2012). In a series of pioneering studies, Julesz et al. hypothesised that humans could not discriminate textures whose second-order correlation functions were the same (so-called *isodiople* textures) (Julesz, 1962). The second-order hypothesis was initially difficult to test due to the lack of textures with controlled image statistics (Victor, 1988, Victor and Conte, 2012). Although initially supported (Rosenblatt and Slepian, 1962), subsequent studies identified discriminable isodiople textures

63

and the hypothesis was proved false (Caelli and Julesz, 1978, Caelli et al., 1978). A subsequent study by Julesz et al. identified a class of textures which, although discriminable, had similar first- to third-order correlation functions and these were termed *isotrigon* (Julesz et al., 1978).

The word isotrigon derives from the fact that the average first- to third-order correlation functions of small collections these textures is zero, as is the case for uniform noise (Maddess et al., 2007). The salient structure in these textures is therefore exclusively due to fourth- and higher-order spatial correlations (Maddess et al., 2004, Maddess et al., 2007, Julesz et al., 1978, Victor, 1994, Gilbert, 1980). Thus, in order to discriminate a particular isotrigon texture from noise patterns, it is necessary to identify its average complex, higher-order structure; isotrigon textures cannot be discriminated based on luminance or other lower-order properties. Whereas some textures are clearly structured and readily recognisable, others cannot be easily differentiated from noise (Maddess and Nagai, 2001). A wide variety of isotrigon ensembles have been identified, including *binary* (e.g. black and white) and *ternary* or *multi-level* (black, white, and grey) (Julesz et al., 1978, Maddess et al., 2004, Maddess et al., 2007).

Isotrigon textures are an important visual psychometric tool for studying texture perception. Whereas the complexity of natural stimuli makes it difficult to determine which image statistics are relevant, the image statistics of isotrigon textures can be carefully, independently controlled (Victor and Conte, 2012, Maddess et al., 2004, Victor and Conte, 1991). Furthermore, it has been shown that what confers visual salience in isotrigon textures also confers visual salience in natural images (Tkacik et al., 2010). That humans

and primates are sensitive to the higher-order correlations in isotrigon textures has been demonstrated in a number of psychophysical (Maddess and Nagai, 2001, Maddess et al., 2007, Victor and Conte, 2005), VEP (Victor and Conte, 1991), fMRI (Beason-Held et al., 1998a, Beason-Held et al., 2000), and PET (Beason-Held et al., 1998b) studies, in addition to single cell recordings (Purpura et al., 1994b).

1.5.1 Binary Isotrigon Textures

Most of the isotrigon textures used to date have been created using a recursion procedure (or *cellular automata* method). In this method, a recursively applied combinatorial rule operated on *N* pixels selected by a 4x4 pixel gliding template (or *glider*), determines the value of an output pixel. This is the *deterministic* method of isotrigon texture generation (Victor and Conte, 1991). The recursive process was originally described by Gilbert (Gilbert, 1980). The initial condition for texture generation is random binary noise (50% black and white pixels). The process is *deterministic* because the set of initial conditions defines the textures obtained: all of the possible textures generated are completely determined by the first few rows and columns of the boundary pixels within the noise pattern (Maddess and Nagai, 2001, Maddess et al., 2007, Taylor et al., 2008).

Although particular patterns are created, the *statistics* of deterministic textures do not depend on the initial pixel values or in which direction the recursion commences. Binary isotrigon textures, and other texture types, may also be produced via a *stochastic* method, which allows for more

complex refinement of image statistics (Victor and Conte, 2012). Examples of the first ensemble of binary isotrigon textures, the VnL2 textures, are shown in Figure 14. According to this nomenclature, *V* refers to the number of input variables in the glider, and *L* refers to the number of levels (2 for binary textures and 3 for ternary textures). In this thesis, *binary* isotrigon textures are used extensively in Chapters 2 and 3.
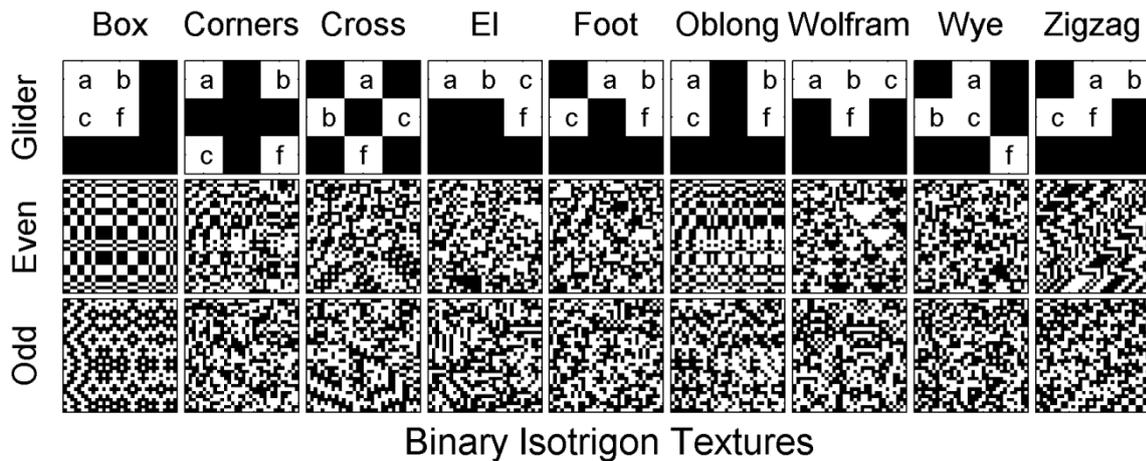
Figure 14. The upper row shows the gliders (gliding templates) that were used to create the textures. The gliders shown here have been greatly magnified; in reality, their pixels are the same size as those of the textures below. Each glider can generate two related texture types termed "Even" and "Odd" (Gilbert, 1980). Pixels marked *abc* are the input pixels, and *f* is the position of the output pixel. Each texture originates from a matrix of binary noise where each colouring (black or white) is equally likely. The glider is then moved in steps of one pixel at a time, across the matrix recursively, modifying the entries using the selected rule until the texture is complete (Maddess and Nagai, 2001).

1.5.2 Ternary Textures

A large number of ternary textures have been discovered (Maddess et al., 2004, Maddess et al., 2007, Taylor et al., 2008). Ternary textures can be produced via both a deterministic method (Maddess et al., 2004, Maddess et al., 2007) and the stochastic method, which has previously been described for binary textures (Victor and Conte, 2012). Whereas the binary textures previously discussed are composed of black and white pixels, the ternary textures have three levels: black, white and the mean luminance grey (contrasts -1, 0 and 1). The full range of ternary textures is larger than that of the binary isotrigon textures.

In the deterministic process for producing ternary texture, outlined by Maddess et al. (Maddess et al., 2007), each texture ensemble is generated via 24 arithmetic rules and a cellular automata method, combining a 3x3 pixel glider, which defines three input pixels (grey) and one output pixel (white). These produce 3V3L ternary textures (with three input variables and three levels). In combination with 20 commonly used glider patterns, this process generates 216 texture classes (Maddess et al., 2007).

The recursive process is similar to that discussed above for binary textures. The gliders and rules together specify how the values of each input pixel are combined to dictate the value of the output pixel. Each texture is initialized with ternary noise. The glider is then moved across the matrix recursively, interacting with matrix elements according to the rules. Therefore, newly instantiated pixel values may in turn feed back into the process (be selected by the input pixels) as the glider moves across the nascent texture (Maddess

67

et al., 2004, Maddess et al., 2007, Taylor et al., 2008). As with the binary isotrigon textures, the process is *deterministic* because the set of initial conditions (random boundary pixels) defines the textures obtained (Maddess et al., 2004, Maddess et al., 2007) (Figure 15).
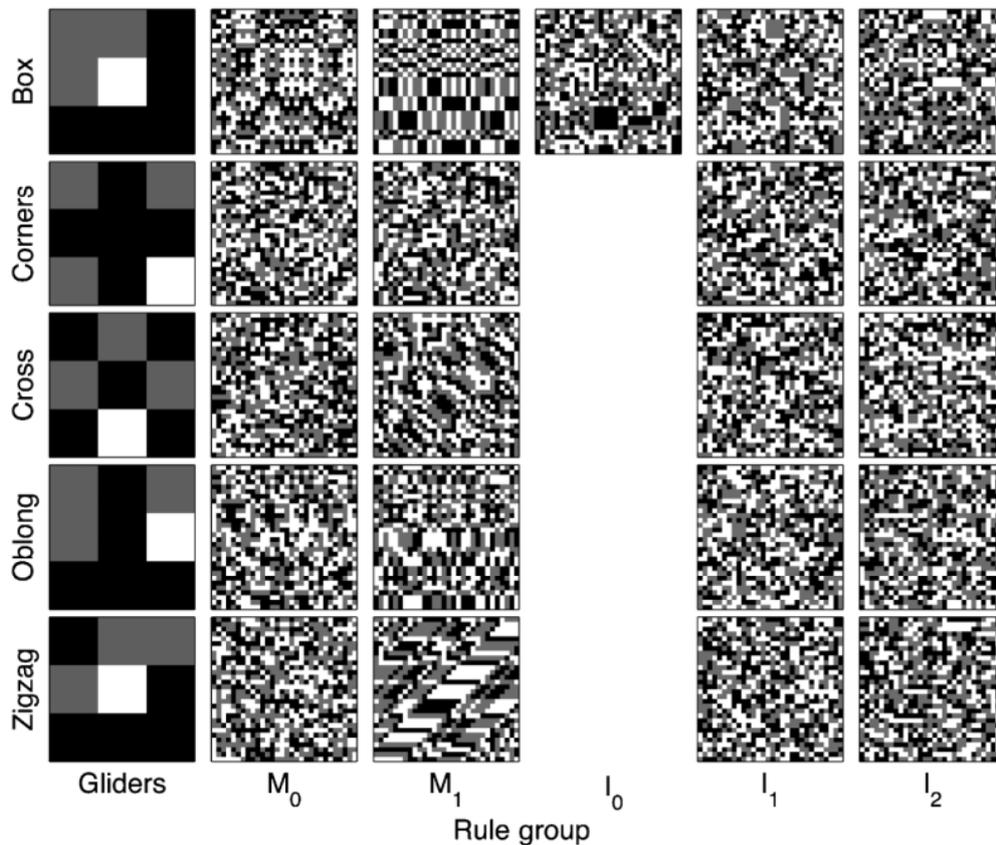
Figure 15: Single ternary texture examples from 21 ensembles (Taylor et al., 2008). Textures within each ensemble are generated by a set of arithmetic rules. The gliders are shown in the leftmost column. The grey squares indicate input pixels and the white square, the output. The ternary texture is formed as the glider moves recursively across the surface of a ternary noise pattern. When the glider interacts with a randomly assigned pixel in the ternary noise pattern, the value of that pixel may be changed. The result of the interaction

depends on a) the pixel value, b) the structure of the glider and c) one of a set of arithmetic rules. These rules are labelled on the x axis (M0, M1, I0, I1, and I2) and they insure the higher-order properties of the ternary textures (Maddess et al., 2007).

Various types of statistically constrained textures, including isotrigon textures, can also be produced using a *stochastic* method, which has been developed by Victor et al. and previously described for binary textures (Victor and Conte, 2012). This process is readily adaptable for producing ternary textures. The statistics of these stochastic textures can be controlled in more sophisticated ways, but they share the same statistics as their deterministic counterparts. The relationship between the stochastic and deterministic ternary texture cohorts was investigated by analysing *histograms of primitives* (HoPs). HoPs count the number of unique texture primitives (typically a 3x3 pixel sample) within a texture example and thereby uniquely identify a texture type (Maddess et al., 2007). Discounting rotations and reflections, the stochastic and deterministic textures were found to be the same (Maddess, 2015). The use of HOPs will be discussed further in Appendix 6.6.

Palettes of stochastic ternary textures are created in a space, where the central location contains random ternary noise; the three corners of the space are texture archetypes and the remaining space is filled with quantitatively derived mixtures (Victor and Conte, 2012). The simplest stochastic manipulation is illustrated in Figure 16: controlling the statistics of single pixels without reference to any other pixels; these ternary textures have been called "gamma" (Victor and Conte, 2012). This planar

arrangement of ternary textures is referred to as a *trico plane* and the 3 principle axes the *trico axes*.

The trico plane is a contrivance which enables what is a three-dimensional image space to be displayed in two-dimensions. The three orthogonal colour axes (black, grey and white) are projected onto a triangular plane via use of the three complex cube roots of -1 as the axis directions; this forms the axes of an equilateral triangle, where the length of each axis vector is 1 (Victor and Conte, 2012). Lines in the space that traverse from the centre point to the borders are referred to as *rays*. In order to explore the trico space, we have selected 6 rays across each trico axis: the 3 principle axes and the perpendicular bisectors of the borders. Along each ray that borders the trico axis, the frequency of one of the three pixel levels increases to the mid-point, where it becomes the dominant colouring. At the centre of the trico axis, all three colourings of pixels are equally likely; the central texture is therefore uniform ternary noise (Figure 16).

The ternary textures are classified according to the number of pixels that are constrained within a 2x2 pixel grid. If a single pixel is constrained, they are called "gamma" (as shown in Figure 16). If 2 pixels are constrained, they are termed "beta_hv" or "beta_diag" (depending on the orientation of the constrained pixels). If triplets of pixels are constrained they are termed "theta" and quadruplets of pixels constrained are "alpha" (Victor et al., 2013, Victor and Conte, 2012).

A further subdivision can be applied which depends on the *orientation* in which the pixels are constrained; this is known as the "linear combination".

Based on simple combinatorics, the ternary texture families have the following number of linear combinations: gamma (1), beta_hv (2), beta_diag (2), theta (4), and alpha (8); these correspond to constraints on first, second, third, and fourth order special correlations respectively. Texture types can also be combined (for example, gamma-theta). Therefore, a large number of these ternary textures are available for study (Victor et al., 2013, Victor and Conte, 2012).
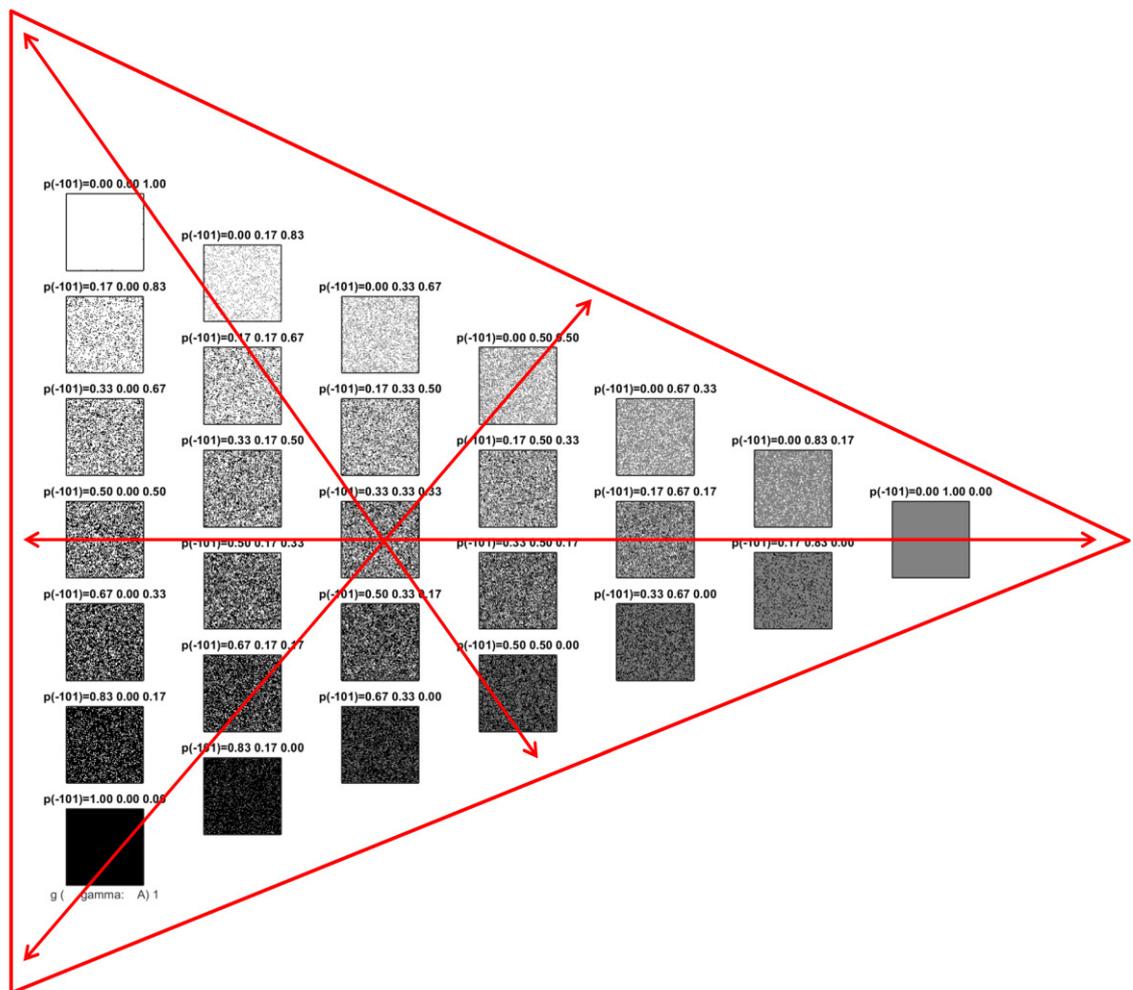
Figure 16. A trico plane showing the gamma ternary textures. The 3 orthogonal colour axes (black, grey and white) are projected onto a triangular plane. Along each outside edge, the frequency of one of the three pixel levels increases to the mid-point. The central space

contains ternary noise. The six rays are shown in red; note that they extend from the central position, which represents ternary noise. The captions above each ternary texture indicate the pixel probabilities for black, grey, and white respectively.

Along each ray, the number of subdivisions (and hence the number of textures), are referred to as *steps*. In broad terms, the distance of the step from the centre of the trico axis (ternary noise) affects how difficult the ternary texture is to discriminate. To put it another way, the step is a measure of decorrelation between 1 and 0, 0 being ternary noise (the central axis position) and 1 indicating no decorrelation (at the ends of each ray). Thus, in order to identify a given ternary texture, a specific nomenclature has been adopted. For example, "alpha_ABCD1111_B_0.0433" indicates that the ternary texture family is "alpha"; the arrangement of the pixels constrained are "ABCD1111", where "1111" indicates the linear combination; the pixel level (colour) bias is "B" (black); and the step is 0.0433, which indicates a high level of decorrelation, i.e. the pattern occurs 0.0433 of the distance from the centre along the principle axis (ray) pointing to the black-biased corner.

An alternative way to visualize the ternary textures, which emphasizes the effect of the step (decorrelation), is shown in Figure 17. For each texture family (in this case gamma) we can display a grid of textures for the 6 different predefined rays; these rays are the same as those shown in Figure 16 above. The textures are displayed with increasing levels of decorrelation (steps). So, in the example shown in Figure 17, 10 levels of decorrelation are displayed, from far left (no decorrelation) to far right (most decorrelation).

This corresponds to gamma textures that are obtained by starting from an outer position on the trico axis and then working in towards the middle, taking 10 steps, along each ray.
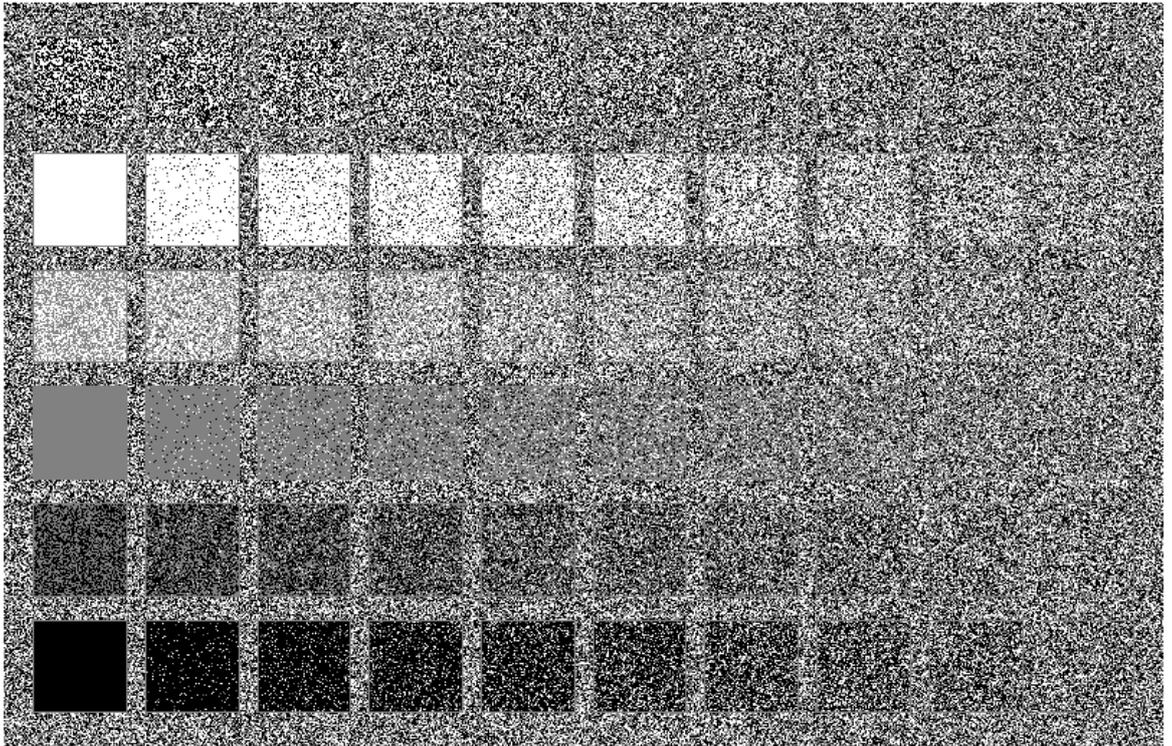
Figure 17. Examples of ternary gamma textures, displayed to emphasise the effect of decorrelation on discrimination performance. Five rays are represented by the y axis, with ternary noise at the top (the rays are shown in Figure 16 above on a trico axis). Each step is represented by displacement along the x axis and in this case 10 steps are shown. Step 1 is equivalent to zero decorrelation (far left) and step 10 is equivalent to complete decorrelation (far right). The textures are presented on a background of ternary noise. Compare the effect of increasing decorrelation on visual salience.

This method of display indicates the primary strength of the stochastic method: by adjusting the level of decorrelation, the discrimination difficulty of each ternary texture can be independently controlled. When used in psychometric studies, this allows for the calculation of performance functions, whilst avoiding performance *saturation*. That is, by adjusting the level of decorrelation one can find a desired noise level that yields a criterion probability correct (e.g. a limen of 70%). This, and the ability to mix different types of ternary texture, is the primary motivation for using the stochastic method of ternary texture generation. After gathering performance functions for the reasonable dimensions within the trico axes, they can be used to evaluate perceptual sensitivity within three-dimensional ternary texture space (Victor et al., 2005, Victor et al., 2013, Victor and Conte, 2012). In this thesis, *ternary* textures are used in Chapter 4, which also includes a detailed discussion of their production and analysis (Section 4.3.3).

1.5.3 Is Grey Perceived Differently?

The ternary isotrions differ from their binary counterparts by the presence of a third level, grey. Therefore, a natural questions is whether the grey token of ternary textures has special properties with regard to texture perception. i.e: according to the current literature on lightness/brightness perception, is *grey* perceived differently to black and white?

According to the standard terminology, *lightness* refers to the apparent reflectance of a surface within a visual scene, whereas *brightness* refers to the apparent luminance of a patch within the image itself (Evans and Bartley,

1948, Kingdom, 2011). Lightness constancy in natural scenes is generally good. Although we rarely experience lightness errors, as pointed out by Kingdom (Kingdom, 2011) "…the study of lightness perception…has been dominated by an exhaustive examination of its errors". Such errors, elicited by synthetic stimuli, provide clues as to how the visual system functions (Kingdom, 2011).

It is now well established that the perceived brightness of a region of visual space is not solely related to that regions' luminance; it also depends on the luminance of adjacent regions. This phenomenon is known as brightness induction and includes both brightness contrast and assimilation effects. Brightness contrast effects occur when the brightness of a test region shifts away from the brightness of adjacent regions. The canonical example of this is the simultaneous brightness contrast illusion (SBC), first reported by Chevreul (Chevreul, 1855). In this illusion, a grey patch on a white background looks darker than a grey patch *of equal luminance* on a black background (Figure 19) (Heinemann, 1955, Williams et al., 1998).
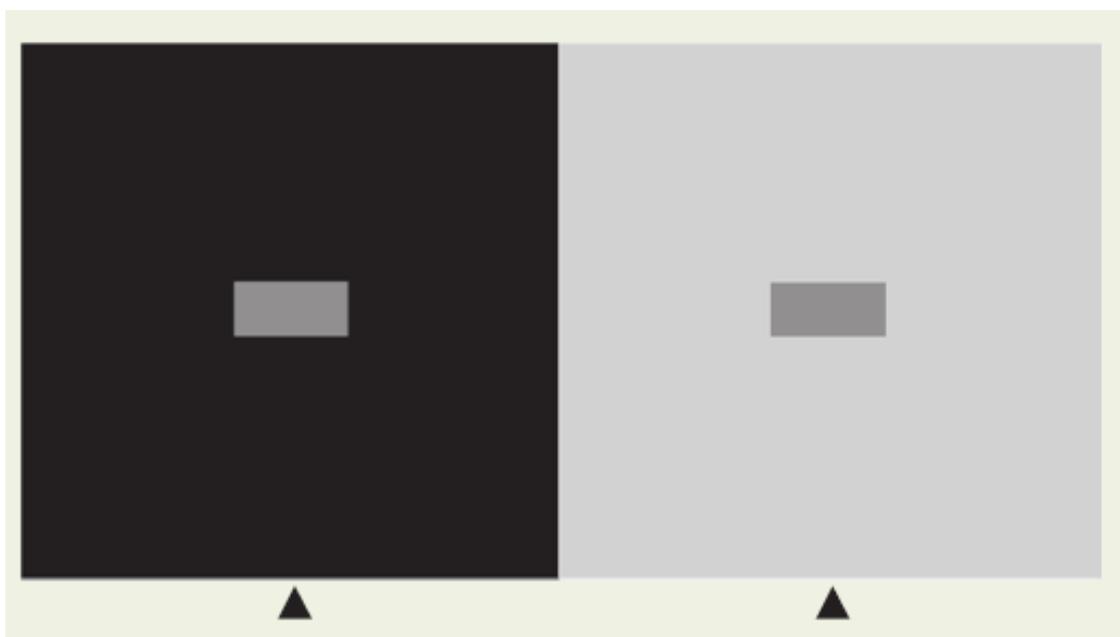
Figure 19. An example of the simultaneous brightness contrast illusion (SBC). Although they appear to differ in brightness, in reality the luminance of the two grey patches is equal (Purves et al., 2001).

Another influential class of lightness error is brightness assimilation. Brightness assimilation is the opposite of brightness contrast, in that lightness appears to shift *towards* that of the surround. Thus, a grey patch on a black background appears *darker* than a grey patch of equal luminance on a white background. One of the most intensely researched brightness assimilation illusions is White's Effect (White, 1979). In White's Effect, the two sets of grey bars have the same luminance, but differ markedly in their perceived brightness, which shifts towards that of the flanking bars (Figure 20) (Anderson, 2003), seemingly opposite to the result observed in Figure 19.
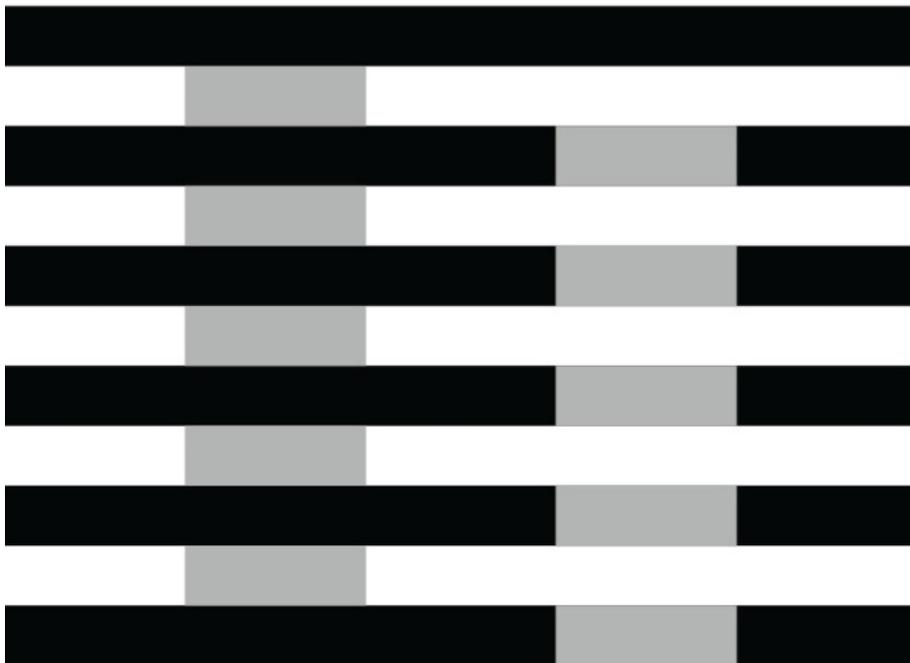
Figure 20. An example of White's Effect, a brightness assimilation illusion. The perceived brightness of the two grey bars appears to shift

towards that of the flanking bars; the ground truth is that both bars are equiluminant (Anderson, 2003).

First described by Cornsweet, a third lightness illusion is the Craik-O'Brien-Cornsweet (CCOB) (Cornsweet, 1970). In this case, illusory brightness values are afforded to regions based on the perception of edges (Figure 21) (Purves et al., 1999). Note how the region to the right of the CCOB edge looks slightly lighter than the region to its left. The ground truth is that the brightness of both areas is the same. Thus, the CCOB is a brightness induction phenomenon in which the central edge of an opposing pair of luminance gradients makes adjoining equiluminant regions appear dissimilar (Cornsweet, 1970).

The CCOB effect has traditionally been interpreted as evidence for the *filling-in* of lightness information (Kingdom, 2011). Filling-in describes a process that begins with edge detection and is followed by a propagation of neural activity which fills in the intervening regions. According to this model, brightness values are determined by what happens at the edges of image elements (Cohen and Grossberg, 1984).

Examples of brightness induction and brightness assimilation demonstrate the profound influence that *context* plays on brightness perception. Research into their neurophysiological basis has uncovered details regarding how the visual system functions.
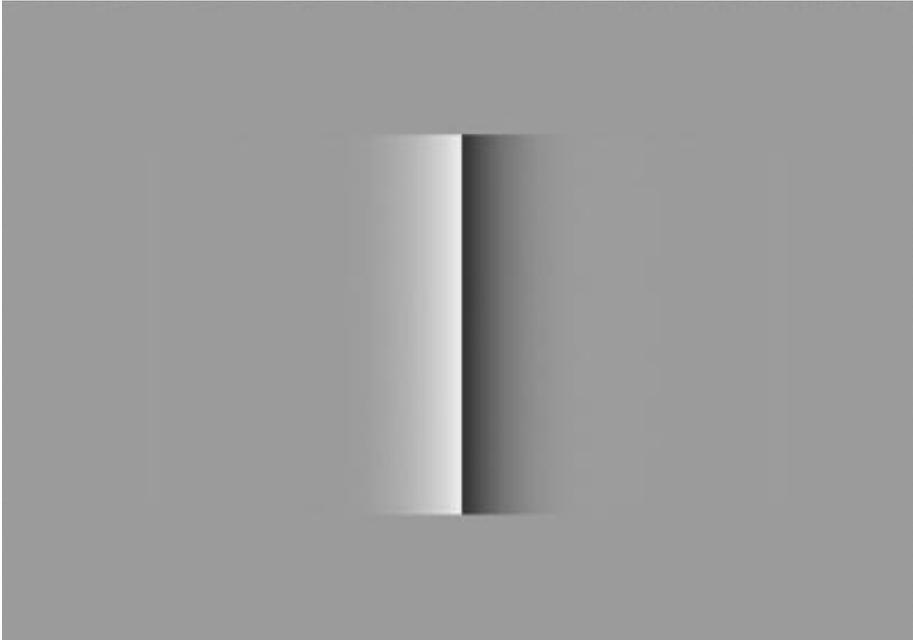
Figure 21. An example of the Craik-O'Brien-Cornsweet effect (CCOB) (Cornsweet, 1970). In this brightness induction phenomenon, the regions to the left and to the right of the CCOB edge have equal luminance, but they appear to be different (Purves et al., 1999).

The prevailing wisdom was that lightness was encoded in the retina. For example, Cornsweet championed the idea that SBC resulted from reciprocal interactions between retinal neurons (Cornsweet, 1970). By this account, SBC is caused by lateral inhibition by neighbouring receptive fields at the contrast boundaries; this has the secondary effect of enhancing edge detection (Purves et al., 2001, Sekular and Blake, 1994, Coren, 2003). Therefore, any target surrounded by an area of higher luminance should be perceived as darker than the ground truth, and vice-versa (Cornsweet, 1970).

However, according to more modern accounts, lightness and brightness perception is a multi-stage process involving both retina and cortex. Whereas

the retina normalizes luminance variations and encodes rudimentary contrast information, cortical neurons in the ventral stream explicitly signal lightness and brightness. In Section 1.5.4 below, we discuss various studies involving gain control mechanisms (Graham and Sutter, 2000, Heeger, 1992b, Maddess et al., 1988, Shapley and Enroth-Cugell, 1984).

The illusions discussed above can be explained by models based on low-level spatial filtering (Dakin and Bex, 2003, Blakeslee and McCourt, 2008, Blakeslee and McCourt, 2004, Blakeslee and McCourt, 2012). One such model has been described in a series of papers by Blakeslee and McCourt. Their model, called the Oriented Difference of Gaussians (ODOG), can explain most brightness illusions (Blakeslee and McCourt, 2008, Blakeslee and McCourt, 2004, Blakeslee and McCourt, 2012). The core idea of the ODOG model is as follows: although the model cells possess filters tuned to very low spatial frequencies, their range is necessarily finite so they need to incorporate a gain control mechanism. This means that, in the case of artificial images whose low spatial frequency content has been greatly reduced, the model cells increase their gain and thereby restore low spatial frequency content to physiologically expected, prior values; this has the side-effect of inducing lightness effects.

Note that in these high pass filtered images, the residual low frequency information is located near high contrast borders. Like cortical cells, the units in ODOG are orientation specific (Figure 22) so the gain changes can also be orientation specific. This can explain illusions like White's effect (White, 1979, Blakeslee and McCourt, 2004). Consequently, reconstituted images lack some of the low spatial frequency information which is present in the

ground truth image. The loss of these frequencies results in brightness induction (Blakeslee and McCourt, 2008, Blakeslee and McCourt, 2004, Blakeslee and McCourt, 2012).

Blakeslee and McCourt posit that the most responsive filters to White's Effect stimuli are appropriately tuned high spatial frequency, vertically-oriented filters. Contrast normalization attenuates the response of these vertical filters, relative to those tuned to horizontal orientations. In response to White's Effect stimuli, horizontally-oriented filters pool the luminance of the flanking bars with those of the test patches. However, because the responses of the vertical filters have been disproportionately attenuated, the relative contribution of the flanking bars is enhanced. This results in the brightness assimilation illusion observed in White's Effect (Figures 20 and 22) (Blakeslee and McCourt, 2008, Blakeslee and McCourt, 2004). A closely related model has been described by Dakin and Bex's (Dakin and Bex, 2003), wherein the contrast normalization stage equates filter responses across spatial-frequency, not orientation.
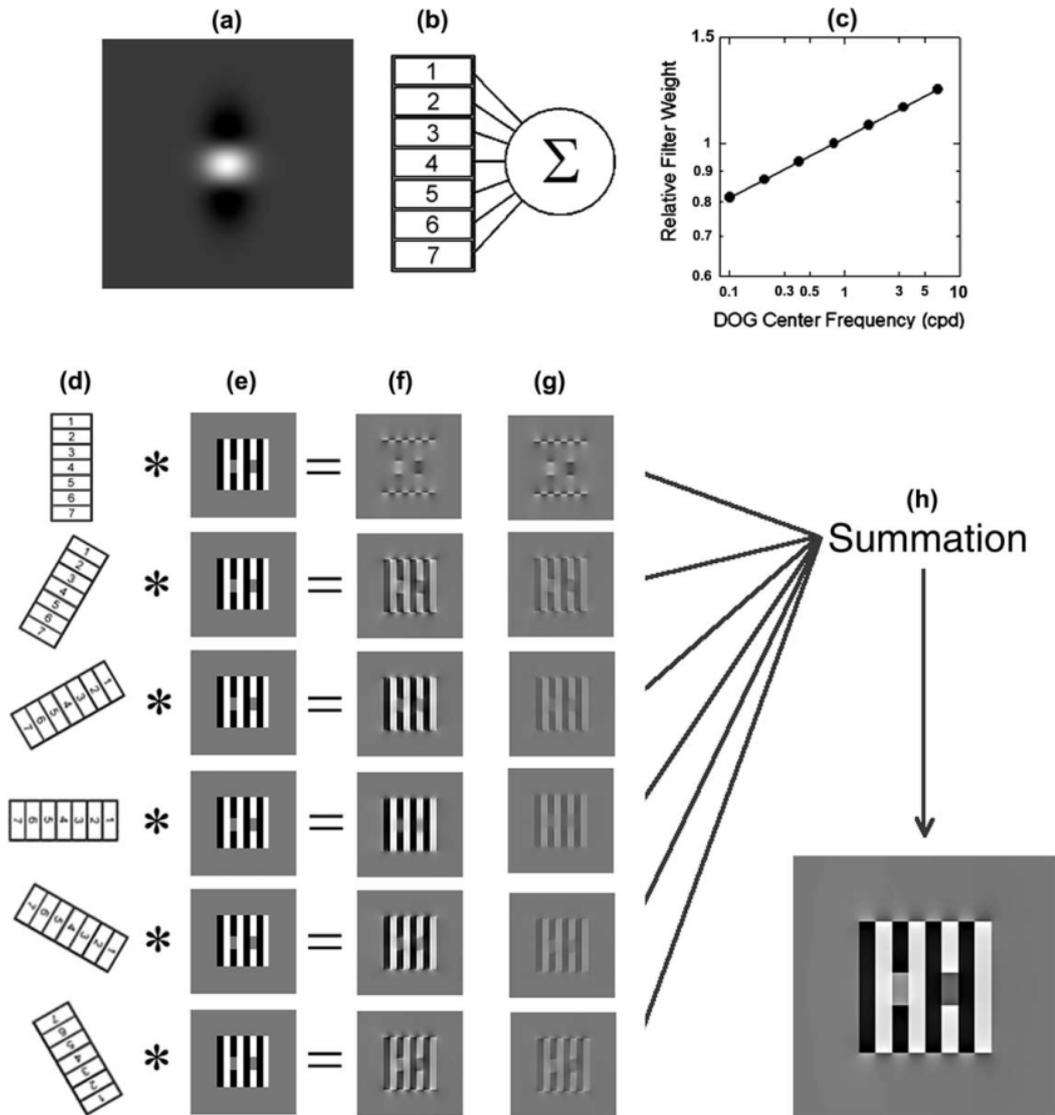
Figure 22. ODOG model as applied to White's Effect. (a) filter profile; (b) filters of different spatial scales and orientations are summed at a given orientation; (c) filter gains as a function of centre spatial frequency; (d) orientations of combined filters; (e) White's Effect stimuli; (f) the result of convolving each stimulus with d; (g) contrast normalization equates root mean squares of filter outputs; (h) outputs are summed across filter orientation (Blakeslee and McCourt, 2004).

Taken together, these studies indicate that grey regions are perceived differently in that *they frequently describe broad regions of no structure*, while

high contrasts near borders of images define structure. Thus, we may expect ternary textures on trico axes with a grey bias to produce different performance functions than those for the more traditional black and white contrasts. To put it another way, in the human colour space, grey represents the origin and not one of the cardinal colour and luminance sensory axes. Therefore, grey may be something of an invalid token for colouring a texture and not comparable to tokens selected from the principle sensory axes. These are questions we are interested in answering and they will be pursued in more depth in Chapter 4 in the context of ternary texture discrimination.

1.5.4 The Neurophysiology of Texture Perception

Whereas simple cells can identify edges of distinct luminance, the mechanisms which underpin the detection of *texture-defined boundaries* are less clear. How might texture segregation performance be modelled?

A basic model has emerged, which consists of three stages (Landy and Graham, 2004, Wolfson and Landy, 1995). The first stage involves a set of linear spatial filters, akin to simple cells. The second stage involves the nonlinear transformation of the output from the first stage (by gain control, rectification and/or squaring). In the third stage, a band pass and orientation-tuned linear filter is applied which enhances texture-defined edges (Bergen and Adelson, 1991, Bergen and Landy, 1991, Landy and Graham, 2004, Wolfson and Landy, 1995, Victor and Conte, 1991). This sequence of spatial filter, nonlinear filter, and second spatial filter has been variously called the

back pocket model (Chubb and Landy, 1991) or *second-order processing* (Chubb et al., 2001).

It should be noted that, in natural images, textural cues often co-occur with other cues such as depth, colour, and motion. Therefore, the current model is necessarily a simplification. In natural images, edge detection is likely to be inferred from a weighted average of multiple cues (Rivest and Cabanagh, 1996). Furthermore, the presence of multiple cues may improve edge detection accuracy (Landy and Kojima, 2001).

There is some evidence that this model is evolutionarily conserved across species. For example, bees can learn to discriminate between patterns based on orientation (Srinivasan et al., 1994). The orientation selective neurons of bees were investigated by Yang and Maddess (Yang and Maddess, 1997). Using narrow (5 degree) bar stimuli, they identified neurons with two different orientation tunings in the mid-brain of the honey bee. These neurons were found to be functionally similar to those required by the texture-extraction model (Yang and Maddess, 1997). Orientation-sensitive neurones have also been identified in the dragonfly brain using intracellular recordings (O'Carroll, 1993). Interestingly, the physiological properties of dragonfly small-target motion detectors are similar to those of mammalian hypercomplex cells (O'Carroll, 1993).

In a subsequent study, Maddess et al. investigated complex texture discrimination by bees (Maddess et al., 1999). To do this, they tested bee discrimination performance using isodipole textures; these textures can only be discriminated from binary noise based on their third- and higher-order

correlation functions. Bees were found to be capable of learning to discriminate isodipole texture pairings. Furthermore, electrophysiological experiments showed the presence of large field neurons that were the sum of many small, orientation tuned and rectified sub-units, thus fulfilling a requirement for the texture segregation model discussed above (Maddess et al., 1999).

The texture segregation model is predicated on the basis of two types of channels. Linear filters can be applied by spatial frequency and orientation selective channels (also called simple or first order channels). Also required are nonlinear channels (complex, second-order, or non-Fourier channels). Evidence for the existence complex channels is now well accepted and comes from a number of studies (Chubb and Landy, 1991, Graham et al., 1992, Graham and Sutter, 1996, Graham and Sutter, 1998, Graham and Sutter, 2000).

In one of the earlier studies, Chubb et al. examined texture segregation performance using filtered noise textures. They found that human performance functions could be modelled using a set of linear channels, tuned for different orientations, followed by an edge detector, which incorporated a nonlinearity (Chubb and Landy, 1991). A subsequent study by Graham et al. was based on texture segregation performance functions (Graham et al., 1992). The authors found that performance functions could not be explained by a purely linear model, based on spatial frequency and orientation channels. The requirement for a nonlinearity led the authors to the hypothesise the presence of complex channels (Graham et al., 1992).

84

A series of studies by Graham et al. have further characterised the properties of complex channels (Graham et al., 1992, Graham and Sutter, 1996, Graham and Sutter, 1998, Graham and Sutter, 2000). Complex channels contain two stages of linear filtering, with an intermediate nonlinearity. The intermediate nonlinearity discriminates complex from simple channels and is understood to be a full- or half-wave rectification, applied to the output of the first filter. At least two types of nonlinearity have been utilized in modelling texture segregation; Graham et al. refer to these as being *intensive* and *spatial* in character. The complex channel nonlinearity is believed to be spatial in nature (Graham and Sutter, 1996).

Graham et al. have also investigated the nature of the intensive nonlinearity (Graham and Sutter, 1996). They found it to be already strong at low contrasts, indicating rectification. It also occurred *after* sensitivity to spatial frequency and background luminance had been determined (Graham and Sutter, 1996, Graham and Sutter, 1998). In their 2000 study, they showed that the compressive effect could be explained by a gain control process, resulting from inhibition between channels (Graham and Sutter, 2000). Gain control allows channels to retain their selectivity over a wide range of contrasts (Graham and Sutter, 2000) and, as discussed in Section 1.4, has important ramifications for efficient coding (Van Hateren, 1992a, Van Hateren, 1992b).

Gain control is such an important process that some authors have proposed it to be the primary function of V1. Cortical cells have a limited dynamic range so, without a gain control mechanism, their responses would saturate

at high contrasts. Additional evidence for cortical gain control mechanisms has come from a variety of sources.

A study by Maddess et al. found evidence of gain control in area-17 of the cat visual cortex (Maddess et al., 1988). Heeger et al. proposed that normalization was fundamental to all neurological functions, not only in V1 (Heeger, 1992b). The authors presented a model in which cortical cell normalization was achieved by mutual inhibition between striate cells, thereby normalizing their responses with respect to stimulus contrast. In a companion paper, this model was compared to physiological measurements of cat striate cell responses and they were found to be congruent (Heeger, 1992a). In particular, the model emphasised the importance of the full- or half-wave rectification (Heeger, 1992a).

Studies on texture-defined motion also indicate the presence of a *texture grabber* (Chubb and Sperling, 1991, Werkhoven et al., 1993, Werkhoven et al., 1994, Turano and Pantle, 1989). Werkhoven et al. explored the perception of apparent motion between dissimilar gratings (Werkhoven et al., 1993). Their findings may be explained in terms of a single-channel motion computation, consisting of one pre-processing stage followed by standard motion analysis. The pre-processing stage is composed of a linear spatial filter, followed by a rectification. The authors called this pre-processing transformation a *texture grabber* (Werkhoven et al., 1993). A follow-up study by the same authors (Werkhoven et al., 1994) also supported the single channel motion computation model, as did an independent study by Chubb et al. which used stimuli called *texture quilts* (Chubb and Sperling, 1991).

To date, the functional roles of areas downstream of V1 have been difficult to characterize with regard to texture perception. After its participation in the early stages of texture segregation, V1 may leave more complex processes to higher cortical areas (Levitt and Lund, 1997b, Freeman et al., 2001, Lennie, 1998). Some authors have suggested that regions of similar structure may not even be grouped until area V4 (Lennie, 1998). Indeed, one fMRI study of texture segregation observed only marginal responses to texture borders in V1 and V2, but more significant responses in V3, V4 and TEO (Kastner et al., 2000) (Figure 18). Lennie has proposed that functions such as figure-ground segregation occur in areas V2 to V4 (Lennie, 1998). Single-unit studies have also implicated V4 neurons in another higher process, the perception of shape from shading (Hanazawa and Komatsu, 2001, Pasupathy, 2006, Arcizet et al., 2009).

The role of V2 has historically been particularly enigmatic. V2 accepts feedforward input from V1 and its response profile is often indistinguishable from that of V1 (Hegdé and Van Essen, 2000, Mahon and De Valois, 2001, El-Shamayleh and Movshon, 2011). However, a recent study by Freeman et al. indicates that V2 may have a role in the discrimination of natural versus synthetic textures (Freeman et al., 2013). The authors observed increased responses in V2, but not V1, on presentation of naturalistic stimuli; these observations were based on single-unit macaque recordings and human fMRI data, the latter of which was predictive of human discrimination performance (Freeman et al., 2013). Their stimuli were *naturalistic* in the sense of mimicking the higher-order correlation content of natural images.

Which brain regions are involved in isotrigon texture discrimination? In an attempt to answer this question, Beason-Held et al. used fMRI to measure responses to random textures and Box-Even binary isotrigon textures (Figure 14) (Beason-Held et al., 1998a, Victor, 1985). The viewing of random textures increased activity in the striate cortex, with slight involvement of the cuneus and middle occipital, lingual and fusiform gyri. The viewing of Box-Even binary isotrigon textures resulted in activation of the same areas, but to a greater extent. In addition, activation was observed in the middle temporal region. Based on these findings, the authors suggest the presence of receptive field mechanisms in the ventral visual pathway that are sensitive to higher-order spatial correlations (Figure 18) (Beason-Held et al., 1998a). Indeed, a subsequent fMRI study by the same authors identified a linear relationship between activation in the striate cortex and the *density* of higher-order spatial correlations in the textures presented (Beason-Held et al., 2000).

In another study by Beason-Held et al. (Beason-Held et al., 1998b) PET was used to evaluate regional cerebral blood flow (rCBF) in subjects viewing random or Box-Even binary isotrigon textures (as used in (Beason-Held et al., 1998a) and (Beason-Held et al., 2000)). The viewing of the isotrigon textures resulted in increased rCBF along the occipito-temporal pathway, versus the viewing of random textures. Significant activation was identified in "…striate, extrastriate, lingual, and fusiform cortices as well as the hippocampus and brain stem…" (Beason-Held et al., 1998b). Notably, increases in rCBF migrated from the occipito-temporal to the medial temporal areas (hippocampus) and frontal lobes following increased exposure to the

isotrigon texture stimuli. This suggests the recruitment of higher-order brain areas during learning (Beason-Held et al., 1998b).
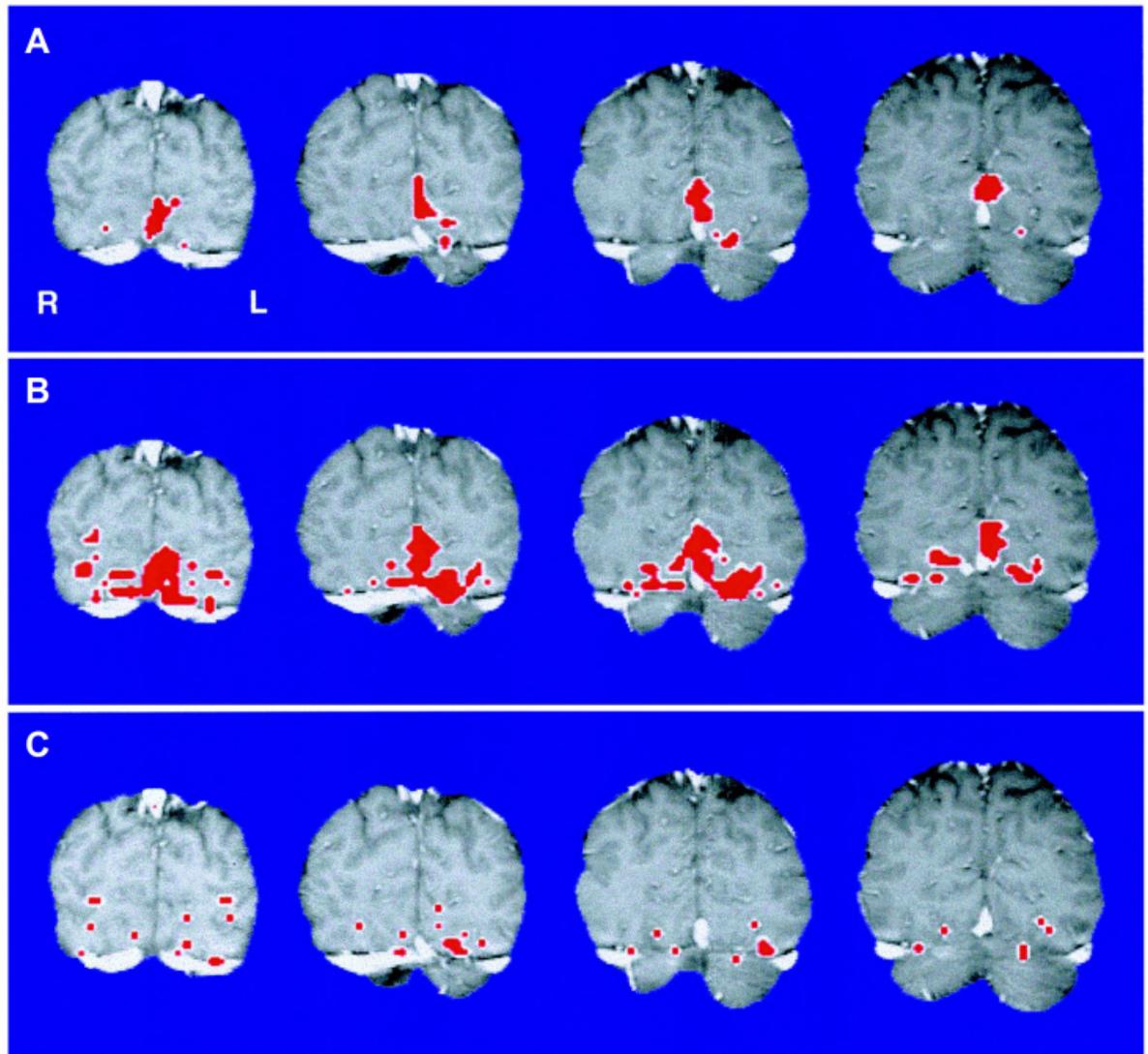
Figure 18. Differences in brain activation when viewing random and binary isotrigon textures. Significant response increases are shown in red. A: random texture stimulation (striate cortex, with slight involvement of the cuneus and middle occipital, lingual and fusiform gyri). B: Box-Even binary isotrigon texture stimulation (significantly greater activation of the above areas, plus middle temporal activation).

C: differences in activation between A and B (Beason-Held et al., 1998a).

Interestingly, these observations suggest that isotrigon texture discrimination may also have *clinical* significance. There is evidence that texture discrimination is negatively affected in multiple sclerosis (MS) patients, whilst other visual functions are spared (Regan and Simpson, 1995, Regan and Hong, 1994). A pattern of higher-order vision impairment has also been observed in Alzheimer's disease (AD) patients (Rizzo et al., 2000). The brain regions affected during AD vary considerably, but typically include higher-order association visual areas such as the IT cortex and the hippocampus (Hof et al., 1997, von Gunten et al., 2006). This idea is discussed further, in the context of future research directions, in Chapter 5 (Section 5.2.2).

1.6 Crowdsourcing Platforms

We have discussed the large number of binary and ternary textures available for study (Maddess et al., 2004, Maddess et al., 2007, Victor and Conte, 2012). An efficient way of exploring the properties of these textures is to use a crowdsourcing service to carry out remote visual psychometric experiments (Paolacci et al., 2010). Crowdsourcing websites coordinate the supply and demand of tasks that require human input. In Chapter 3 of this thesis, we will validate the use of crowdsourcing platforms for such experiments. In Chapter 4, crowdsourcing will be used to carry out a large study using ternary textures.

Launched in 2005, Amazon Mechanical Turk (mTurk) has developed into the largest crowdsourcing platform (see (Mason and Suri, 2012) and (Paolacci et al., 2010) for reviews on using mTurk for conducting experiments). It now has in excess of 200,000 Workers registered from over 100 countries (Ross et al., 2010, Pontin, 2007). MTurk provides the elements required to conduct research studies: a large, persistent pool of research subjects, an integrated payment system, a streamlined process for study design, recruitment, and data collection (Paolacci et al., 2010). Therefore, mTurk has developed into a labour market for tasks which vary from surveys and language translations, to psychometric experiments (Mason and Suri, 2012).

Anyone with internet access can register to use mTurk and as a Requester or a Worker. Requesters create Human Intelligence Tasks (HITs). HITs expire after a predefined time and/or when the available pool has been exhausted. Workers are presented with a list of HITs, which they can browse and complete. Completed HITs are reviewed by the Requester and accepted or rejected. Accepted HITs trigger payment to be transferred from the Requester to the Worker's account. Amazon charges Requesters 10% of the total pay issued (minimum $0.005 USD per HIT) (Mason and Suri, 2012). At the time of writing, the number of active HITs on mTurk was 403,336 (AWS, 2014).

MTurk has several advantages over traditional methods of subject recruitment and study implementation. It provides access a large, established workforce, which facilitates rapid recruiting with very little effort (Ross et al., 2010, Paolacci et al., 2010). The service is available every day of the year and subject availability appears to remain quite stable over time,

with minor seasonal fluctuations (Ipeirotis, 2010). Its established technical infrastructure allows studies to be developed relatively quickly and easily (Mason and Suri, 2012). It also includes an integrated payment system, which eliminates third-party services, such as PayPal, that have been linked to lower response rates (Goritz et al., 2008). It allows for pre-screening and can also incorporate quality control measures, such as catch trials (easy questions which can be used to gauge subject attentiveness) (Paolacci et al., 2010, Kittur et al., 2008). Oppenheimer et al. have also experimented with ways of checking that Workers are following instructions and remaining engaged (Oppenheimer et al., 2009).

There have been reports of Workers using programs (*bots*) to automatically complete HITs, although this appears to be rare (McCreadie et al., 2010). Nonetheless, it is important to validate Worker performance. To that end, mTurk has a built-in reputation system so Requesters can block Workers whose rejection rate exceeds a given threshold (Paolacci et al., 2010, McCreadie et al., 2010). Qualification tasks may also be used to enforce practice trials and careful reading of experimental procedures (Heer et al., 2010).

The majority of mTurk Workers are based in the USA, although there is considerable international diversity; the number of Workers in India in particular is increasing (Ross et al., 2010, Ipeirotis, 2010). The requirement for an Internet connection and English language proficiency restricts the majority of mTurk Workers to highly industrialized societies (Ross et al., 2010). MTurk Workers are highly educated, with 57% being educated to Bachelor's degree level or above; a slight majority are female (55%) and

aged 18-30 years (62%) (Ross et al., 2010). MTurk Workers are slightly more demographically diverse than standard Internet samples and significantly more diverse than US college samples (Buhrmester et al., 2011, Gosling et al., 2004). Therefore, mTurk allows researchers to access subjects that would be difficult to access by other means (Paolacci et al., 2010).

One of the advantages offered by crowdsourcing platforms such as mTurk is its low cost. One study found that Workers have a reservation wage (the minimum pay rate for which they would complete a HIT) of $1.38 USD per hour (Ipeirotis, 2010, Mason and Suri, 2012, Buhrmester et al., 2011). Compensation rates appear to have less effect on data quality than on the *rate* of data collection (Buhrmester et al., 2011, Paolacci et al., 2010, Mason and Watts, 2009). This is consistent with the idea that Workers are not primarily financially motivated, but also derive secondary benefits from mTurk, such as entertainment and a sense of being productive (Buhrmester et al., 2011, Ipeirotis, 2010, Paolacci et al., 2010, Ross et al., 2010). In addition to cost savings and reduced recruiting effort, crowdsourcing can scale to levels which would be prohibitive in a laboratory setting.

Given the low levels of compensation, it is interesting to consider what motivates mTurk Workers. The number of Workers that rely on mTurk for their primary income is quite low: 12% of US and 27% of Indian Workers (Ipeirotis, 2010). Nonetheless, only 12% of US and 10% of Indian Workers indicated that the money they derive from mTurk was "irrelevant" (Ross et al., 2010). Similarly, Paolacci et al. reported that although only 13.8% of the US Workers derived their primary income from mTurk, 61.4% reported that

earning additional money was an important motivating force (Paolacci et al., 2010).

Many Workers also use mTurk for entertainment (40.7%) and "killing time" (32.3%) (Ipeirotis, 2010, Paolacci et al., 2010). Buhrmester et al. also reported that Workers used mTurk because they found the tasks enjoyable (Buhrmester et al., 2011). The majority of US Workers (69%) reported that they thought of mTurk as a productive way to spend free time and make some extra money (Ipeirotis, 2010, Paolacci et al., 2010). This suggests that mTurk Workers may produce good quality data, despite receiving relatively low wages, as financial gain is not their sole motivation (Buhrmester et al., 2011, Ipeirotis, 2010, Paolacci et al., 2010, Ross et al., 2010). In this context, Crump has suggested that tasks which have an entertainment component may be more successful at recruiting Workers (Crump et al., 2013).

Several authors have suggested that, on ethical grounds, the pay of mTurk Workers should be comparable to that of laboratory subjects (Mason and Suri, 2012, Crump et al., 2013). However, because of the entertainment value derived from mTurk (Paolacci et al., 2010, Buhrmester et al., 2011, Ipeirotis, 2010), and the low numbers of Workers that rely on mTurk for their primary income (Ross et al., 2010, Paolacci et al., 2010, Ipeirotis, 2010), that case is not entirely clear.

An obvious concern with a novel platform such as mTurk is whether the data produced is of high quality. Workers are unsupervised and therefore may be less attentive than supervised subjects in a laboratory setting (Oppenheimer et al., 2009). The anonymity afforded to Workers may increase deceptive

94

responding (Skitka and Sargis, 2006) and rates of non-completion (Crump et al., 2013). Some financially motivated Workers may be more concerned with completing HITs rapidly than by the quality of their work (Mason and Suri, 2012). Therefore, it is important to consider how compensation levels affect Worker behaviour.

HIT participation rates are affected by the level of compensation and how long the HIT takes to complete (Buhrmester et al., 2011). Buhrmester et al. offered $0.02 USD for a 30 minute task, but still accumulated 25 completed HITs within 5 hours of posting. When the compensation was increased to $0.50 USD, they obtained the same number of completed HITs in less than 2 hours of posting (Buhrmester et al., 2011).

In a second study, Buhrmester et al. posted a set of personality questionnaires at three levels of compensation ($0.02, $0.10, and $0.50 USD). The alpha reliabilities for the data collected were within one hundredth of a point across all three compensation levels (Buhrmester et al., 2011). Mason et al. observed a similar relationship (Mason and Watts, 2009). They altered the compensation levels of two mTurk tasks, whilst simultaneously measuring the HIT participation rate and the data quality. As the compensation level increased (from $0.01 to $0.10 USD), so did the number of HITs completed; however, there was no difference in the data quality (Mason and Watts, 2009). Taken together, these findings suggest that, even at the lowest levels, compensation has less effect on data quality than the *rate* of data collection (Buhrmester et al., 2011, Paolacci et al., 2010, Mason and Watts, 2009). This may be explained by Workers deriving secondary

benefits from participating in mTurk studies (Buhrmester et al., 2011, Ipeirotis, 2010, Paolacci et al., 2010, Ross et al., 2010).

With regard to data quality, several studies have carried out replications of laboratory studies using mTurk. Paolacci et al. replicated a series of classic judgment and decision-making experiments on mTurk at a cost of just $1.71 USD per hour per subject. Quantitatively, there were only very minor differences between the mTurk and laboratory data sets (Paolacci et al., 2010). Similarly, Horton et al. and Crump et al. replicated some classic behavioural psychology experiments using mTurk, including the Prisoners' Dilemma, and found it to be congruent with previously published laboratory data (Horton et al., 2011, Crump et al., 2013). In the latter study, it is interesting to note that the introduction of instructional checks had a greater effect on data quality than increasing the level of compensation (Crump et al., 2013).

At the time of writing, only three studies have used mTurk to administer *visual* psychometric studies (Heer et al., 2010, Cole et al., 2009, Freeman et al., 2013). Cole et al. recruited 550 mTurk Workers to evaluate three-dimensional line drawings and indicate surface normals. They gathered 275,000 observations and used them to rate rendering techniques. However, the data gathered was not compared to supervised laboratory data and collection statistics were not reported (Cole et al., 2009). Heer et al. (Heer et al., 2010) carried out a series of graphical perception experiments, including a replication of a laboratory alpha contrast experiment (Stone and Bartram, 2008). Freeman et al. used mTurk to evaluate the discrimination of natural versus synthetic textures (Freeman et al., 2013). They gathered over "300

hours of behavioural data from thousands of human observers" (exact numbers were not reported). Each Worker was paid $0.40 USD for ~5 minutes of work ($4.80 USD per hour). Neither demographic data nor platform data were reported, and the authors did not report using catch trials or qualification tasks (Freeman et al., 2013). In this context, there is a clear need for studies which investigate the suitability of mTurk for performing visual psychometric testing.

1.7 Thesis Structure

We will now briefly review the structure of this thesis. Note that each Chapter has been written to be *self-contained*. Therefore, some introductory material appears in several Chapters and/or in the General Introduction.

1.7.1 Chapter 2

Humans use a number of neurophysiological mechanisms to capture visually salient higher order structure. The Maddess group has previously provided evidence that the number of independent mechanisms is less than 10 (Taylor et al., 2008, Barbosa et al., 2013), and is perhaps as small as 3 to 4 (Maddess and Nagai, 2001, Maddess et al., 2007, Seamons et al., 2015). One statistically well principled method to infer the number and form of the underlying independent mechanisms is to use Factor analysis of human performance functions; performance functions can be derived from many individuals, or many repeats from single individuals (Rosli et al., 2009, Maddess and Kulikowski, 1999, Sekuler et al., 1984, Simpson and McFadden, 2005).

In Chapter 2, we evaluate the underlying mechanisms of texture discrimination. We describe binary isotrigon texture discrimination experiments using 10 novel isotrigon textures (VnL2, Figure 15) and 17 standard V3L2 isotrigon textures. After calculating performance functions for each subject (probability correct across the different textures) and their d-primes, two forms of Factor analysis are performed based on rotated principal components and maximum likelihood estimates (Reyment, 1996, Norris and Lecavalier, 2010).

We then analyse the number of neurological mechanisms which govern the detection of higher-order image structure in two ways: using a Scree plot and by analysing communalities (Cattell, 1966, Hayton et al., 2004, DeVellis, 2012, Zwick and Velicer, 1986). Communalities indicate the proportion of variance accounted for within the data for each texture for an N-factor model. The possible nature of the underlying mechanisms of texture perception are discussed (Maddess and Nagai, 2001, Maddess et al., 2007, Seamons et al., 2015).

1.7.2 Chapter 3

In Chapter 3, we continue our exploration of the mechanisms which underpin higher-order texture discrimination using a complementary approach to that used in Chapter 2. We describe the use of the crowdsourcing platform, Amazon Mechanical Turk (mTurk), to design, test, and implement a binary isotrigon texture discrimination task using a large number of naïve subjects. Due to perceived technical limitations, very few visual psychometric studies

have utilised mTurk (Heer et al., 2010, Cole et al., 2009, Freeman et al., 2013) so an important secondary aim of this chapter is to evaluate its suitability for such studies.

We begin by describing the design and development of the mTurk isotrigon texture discrimination HIT. The HIT utilizes a four-alternative forced choice task testing modality, developed by Victor et al. (Victor et al., 2013). This protocol is intended to provide a more complex and naturalistic stimuli than those traditionally used in visual perception experiments based on gratings (Victor et al., 2013).

We discuss initial laboratory testing of the HIT using 6 different platforms (270 HITs) and establish whether it is fit for purpose. We evaluate the effects of variations between platforms, using correlational analysis and coefficients of repeatability (Bland and Altman, 1986, Bland and Altman, 1999, Bland, 2000, Vaz et al., 2013).

After testing and validation, we use this HIT to gather crowdsourced data from over one hundred naïve subjects. The resulting *Live* data is then compared to 2 laboratory data sets using correlational analysis, coefficients of repeatability and Factor analysis. The similarities and differences between the Live and the laboratory data sets are discussed in detail and the potential of mTurk for carrying out visual psychometric studies is reviewed. Recommendations are also made for future mTurk studies, including the use of safeguards such as catch trials and qualification tasks. This chapter lays the foundations for a second mTurk study, as described in Chapter 4.

1.7.3 Chapter 4

A large number of multi-level textures have been discovered both using a deterministic method (Maddess et al., 2007) and a modified stochastic method, based on that described for binary textures (Victor and Conte, 2012). Chapter 4 begins by reviewing the origins and properties of these ternary textures. The relationship between the deterministic and stochastic cohorts is examined using histograms of primitives. Histograms of primitives uniquely identify a texture type based on its constituent unique texture subregions (Maddess et al., 2007).

The full range of ternary textures is larger than that of the binary textures, even taking into the account redundancy between deterministic and stochastics ensembles (Maddess et al., 2007, Victor et al., 2013, Victor and Conte, 2012). Therefore, crowdsourcing is a good way to evaluate human discrimination performance for a subset of the possible ternary textures. We discuss the development of a ternary texture discrimination HIT, based on the binary HIT previously described in Chapter 3, its testing and vallidation.

The trico planes allow what are defined on three sensory dimensions to be displayed coherently in two dimensions. The objective of this study is to determine which dimensions of the trico plane are independently processed. As noted by Victor, an advantage of the trico plane is that changes in all directions are equally detectable to the ideal observer (Victor and Conte, 2012). However, the informational resources of the human visual system are *constrained* (Victor and Conte, 2012, Franz and Schölkopf, 2005, Barlow, 2001, Barlow, 1963). In this chapter, we investigate whether discrimination

performance varies along certain rays of the trico planes. We then consider what such variations tell us about the discrimination mechanisms involved. In this chapter, particular emphasis will be placed on what makes the ternary textures unique from their binary counterparts - the grey component.

1.7.4 Summary and Future Directions

In the final chapter, we will consider future research directions that follow naturally from the findings outlined in this thesis. Four such research directions will be considered: a study of material perception using 3D stimuli derived from binary isotrigon textures; a study utilizing isotrigon discrimination as a clinical diagnostic aid in neurological disease; an investigation of 3D illusions in ternary textures; and the application of the ternary texture testing modality to convolutional neural networks.

1.8 References

ABBOTT, L. & CHANCE, F. S. 2002. Rethinking the taxonomy of visual neurons. *Nature neuroscience,* 5**,** 391-392.

ABDULLAH, S. N., ALDAHLAWI, N., ROSLI, Y., BOON, M. Y. & MADDESS, T. 2012. Effect of contrast, stimulus density, and viewing distance on multifocal steady-state visual evoked potentials (MSVs). *Investigative ophthalmology & visual science,* 53**,** 5527-5535.

ADAMS, D. L. & HORTON, J. C. 2003. A precise retinotopic map of primate striate cortex generated from the representation of angioscotomas. *The Journal of neuroscience,* 23**,** 3771-3789.

ADELSON, E. H. On seeing stuff: the perception of materials by humans and machines.   Photonics West 2001-Electronic Imaging, 2001. International Society for Optics and Photonics, 1-12.

ADELSON, E. H. & BERGEN, J. R. 1985. Spatiotemporal energy models for the perception of motion. *JOSA A,* 2**,** 284-299.

ADELSON, E. H. & BERGEN, J. R. 1991. The plenoptic function and the elements of early vision. *Computational models of visual processing,* 1.

ADESNIK, H., BRUNS, W., TANIGUCHI, H., HUANG, Z. J. & SCANZIANI, M. 2012. A neural circuit for spatial summation in visual cortex. *Nature,* 490**,** 226-31.

ALITTO, H. J. & USREY, W. M. 2003. Corticothalamic feedback and sensory processing. *Current opinion in neurobiology,* 13**,** 440-445.

ALITTO, H. J. & USREY, W. M. 2008. Origin and dynamics of extraclassical suppression in the lateral geniculate nucleus of the macaque monkey. *Neuron,* 57**,** 135-46.

ANDERSON, B. L. 2003. Perceptual organization and White's illusion. *PERCEPTION-LONDON-,* 32**,** 269-284.

ANZAI, A., OHZAWA, I. & FREEMAN, R. D. 1999. Neural mechanisms for processing binocular information II. Complex cells. *Journal of Neurophysiology,* 82**,** 909-924.

ARCIZET, F., JOUFFRAIS, C. & GIRARD, P. 2009. Coding of shape from shading in area V4 of the macaque monkey. *BMC Neurosci,* 10**,** 140.

ATICK, J. J. & REDLICH, A. N. 1992. What does the retina know about natural scenes? *Neural computation,* 4**,** 196-210.

ATTNEAVE, F. 1954. Some informational aspects of visual perception. *Psychological review,* 61**,** 183.

AWS, A. W. S. 2014. *Amazon Mechanical Turk (mTurk)* [Online]. Available: https://www.mturk.com/mturk/welcome [Accessed 4 November 2014.

AZZOPARDI, P. & COWEY, A. 1996. The overrepresentation of the fovea and adjacent retina in the striate cortex and dorsal lateral geniculate nucleus of the macaque monkey. *Neuroscience,* 72**,** 627-639.

BALASUBRAMANIAN, V. & STERLING, P. 2009. Receptive fields and functional architecture in the retina. *The Journal of physiology,* 587**,** 2753-2767.

BARBOSA, M. S., BUBNA-LITIC, A. & MADDESS, T. 2013. Locally countable properties and the perceptual salience of textures. *JOSA A,* 30**,** 1687-1697.

BARLOW, H. 2001. Redundancy reduction revisited. *Network,* 12**,** 241-53.

BARLOW, H. B. 1961. Possible principles underlying the transformations of sensory messages.

BARLOW, H. B. 1963. The Information Capacity of Nervous Transmission. *Kybernetik,* 2**,** 1.

BEASON-HELD, L. L., PURPURA, K. P., KRASUSKI, J. S., DESMOND, R. E., MANGOT, D. J., DALY, E. M., OPTICAN, L. M., RAPOPORT, S. I. & VANMETER, J. W. 2000. Striate cortex in humans demonstrates the relationship between activation and variations in visual form. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale,* 130**,** 221-6.

BEASON-HELD, L. L., PURPURA, K. P., KRASUSKI, J. S., MAISOG, J. M., DALY, E. M., MANGOT, D. J., DESMOND, R. E., OPTICAN, L. M., SCHAPIRO, M. B. & VANMETER, J. W. 1998a. Cortical regions involved in visual texture perception: a fMRI study. *Brain research. Cognitive brain research,* 7**,** 111-8.

BEASON-HELD, L. L., PURPURA, K. P., VAN METER, J. W., AZARI, N. P., MANGOT, D. J., OPTICAN, L. M., MENTIS, M. J., ALEXANDER, G. E., GRADY, C. L., HORWITZ, B., RAPOPORT, S. I. & SCHAPIRO, M. B. 1998b. PET reveals occipitotemporal pathway activation during elementary form perception in humans. *Visual neuroscience,* 15**,** 503-10.

BELL, A. J. & SEJNOWSKI, T. J. 1997. The "independent components" of natural scenes are edge filters. *Vision research,* 37**,** 3327-3338.

BERGEN, J. R. & ADELSON, E. 1991. Theories of visual texture perception. *Spatial vision,* 10**,** 114-134.

BERGEN, J. R. & LANDY, M. S. 1991. Computational modeling of visual texture segregation. *Computational models of visual processing,* 1, 253-271.

BLAKE, A., BULTHOFF, H. H. & SHEINBERG, D. 1993. Shape from texture: ideal observers and human psychophysics. *Vision Res,* 33, 1723-37.

BLAKESLEE, B. & MCCOURT, M. E. 2004. A unified theory of brightness contrast and assimilation incorporating oriented multiscale spatial filtering and contrast normalization. *Vision research,* 44, 2483-2503.

BLAKESLEE, B. & MCCOURT, M. E. 2008. Nearly instantaneous brightness induction. *Journal of Vision,* 8, 15.

BLAKESLEE, B. & MCCOURT, M. E. 2012. When is spatial filtering enough? Investigation of brightness and lightness perception in stimuli containing a visible illumination component. *Vision research,* 60, 40-50.

BLAND, J. M. 2000. *An Introduction into Medical Statistics,* Oxford, Oxford University Press.

BLAND, J. M. & ALTMAN, D. G. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet,* 1, 307-10.

BLAND, J. M. & ALTMAN, D. G. 1999. Measuring agreement in method comparison studies. *Stat Methods Med Res,* 8, 135-60.

BOLZ, J. & GILBERT, C. D. 1986. Generation of end-inhibition in the visual cortex via interlaminar connections. *Nature,* 320, 362-5.

BOULTON, A. A., BAKER, G. B. & VANDERWOLF, C. H. 1990. *Neurophysiological techniques: applications to neural systems*, Humana Press.

BOVIK, A. C., CLARK, M. & GEISLER, W. S. 1990. Multichannel texture analysis using localized spatial filters. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* 12**,** 55-73.

BRACEWELL, R. 1965. The Fourier Transform and IIS Applications. *New York*.

BRIGGS, F. & USREY, W. M. 2008. Emerging views of corticothalamic function. *Current opinion in neurobiology,* 18**,** 403-407.

BRONKHORST, A. W. 2000. The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica united with Acustica,* 86**,** 117-128.

BUHRMESTER, M., KWANG, T. & GOSLING, S. D. 2011. Amazon's Mechanical Turk : A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science,* 6**,** 3-5.

CAELLI, T. & JULESZ, B. 1978. On perceptual analyzers underlying visual texture discrimination: Part I. *Biological Cybernetics,* 28**,** 167-175.

CAELLI, T., JULESZ, B. & GILBERT, E. 1978. On perceptual analyzers underlying visual texture discrimination: Part II. *Biological Cybernetics,* 29**,** 201-214.

CASAGRANDE, V. A. & XU, X. 2004. 31 Parallel Visual Pathways: A Comparative Perspective.

CATTELL, R. B. 1966. The Scree Test For The Number Of Factors. *Multivariate Behavioral Research,* 1**,** 245-276.

CHANCE, F. S., NELSON, S. B. & ABBOTT, L. 1999. Complex cells as cortically amplified simple cells. *Nature neuroscience,* 2**,** 277-282.

CHEVREUL, M. E. 1855. *The principles of harmony and contrast of colours, and their applications to the arts*, Longman, Brown, Green, and Longmans.

CHUBB, C. & LANDY, M. S. 1991. Orthogonal distribution analysis: A new approach to the study of texture perception. *Computational models of visual processing,* 12**,** 394.

CHUBB, C., OLZAK, L. & DERRINGTON, A. 2001. Second-order processes in vision: introduction. *JOSA A,* 18**,** 2175-2178.

CHUBB, C. & SPERLING, G. 1991. Texture quilts: Basic tools for studying motion-from-texture. *Journal of Mathematical Psychology,* 35**,** 411-442.

CLAASEN, T. & MECKLENBRAUKER, W. 1980. The Wigner distribution—A tool for time-frequency signal analysis. Part I: Continuous-time signals. *Philips J. Res,* 35**,** 217-250.

COHEN, M. A. & GROSSBERG, S. 1984. Neural dynamics of brightness perception: Features, boundaries, diffusion, and resonance. *Perception & psychophysics,* 36**,** 428-456.

COLE, F., SANIK, K., DECARLO, D., FINKELSTEIN, A., FUNKHOUSER, T., RUSINKIEWICZ, S. & SINGH, M. How Well do Line Drawings Depict Shape? ACM Transactions on Graphics, 2009.

COMON, P. 1994. Independent component analysis, a new concept? *Signal processing,* 36**,** 287-314.

COREN, S. 2003. *Sensation and perception*, Wiley Online Library.

CORNSWEET, T. 1970. *Visual Perception,* New York, Academic Press.

CRUMP, M. J., MCDONNELL, J. V. & GURECKIS, T. M. 2013. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS One,* 8**,** e57410.

DAKIN, S. C. & BEX, P. J. 2003. Natural image statistics mediate brightness 'filling in'. *Proc Biol Sci,* 270**,** 2341-8.

DAUGMAN, J. G. 1985. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *JOSA A,* 2**,** 1160-1169.

DE VALOIS, R. L., ALBRECHT, D. G. & THORELL, L. G. 1978. Cortical cells: bar and edge detectors, or spatial frequency filters? *Frontiers in visual science.* Springer.

DE VALOIS, R. L., ALBRECHT, D. G. & THORELL, L. G. 1982. Spatial frequency selectivity of cells in macaque visual cortex. *Vision research,* 22**,** 545-59.

DEVALOIS, R. L. & BOTH PROFESSORS, B. K. K. D. 1988. *Spatial vision*, Oxford University Press.

DEVELLIS, R. F. 2012. Factor Analysis. *Scale Development Theory and Applications.* 3rd ed. University of North Carolina, Chapel Hill, USA: SAGE Publications, Inc.

DOBBINS, A., ZUCKER, S. W. & CYNADER, M. S. 1987. Endstopped neurons in the visual cortex as a substrate for calculating curvature. *Nature,* 329**,** 438-41.

DREHER, B. 1972. Hypercomplex cells in the cat's striate cortex. *Investigative ophthalmology,* 11**,** 355-6.

EL-SHAMAYLEH, Y. & MOVSHON, J. A. 2011. Neuronal responses to texture-defined form in macaque visual area V2. *The Journal of Neuroscience,* 31**,** 8543-8555.

EMERSON, R. C., BERGEN, J. R. & ADELSON, E. H. 1992. Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision research,* 32**,** 203-218.

EVANS, R. M. & BARTLEY, S. H. 1948. Introduction to color. *American Journal of Physics,* 16**,** 491-491.

FABRIGAR, L. R., WEGENER, D. T., MACCALLUM, R. C. & STRAHAN, E. J. 1999. Evaluating the use of exploratory Factor analysis in psychological research. *Psychological methods,* 4**,** 272.

FIELD, D. 1994. What is the goal of sensory coding? *Neural computation,* 6**,** 559-601.

FIELD, D. J. 1987. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America. A, Optics and image science,* 4**,** 2379-94.

FOLDIAK, P. Y., M. P. 1995. Sparse Coding in the Primate Cortex. *In:* ARBIB, M. A. (ed.) *The Handbook of Brain Theory and Neural Networks.* Cambridge, MA: MIT Press.

FRANZ, M. O. & SCHÖLKOPF, B. 2005. Implicit Volterra and Wiener series for higher-order image analysis. *In:* SAUL, L. K., WEISS, Y. & BOTTOU, L. (eds.) *Advances in neural information processing systems.* Cambridge MA: MIT Press.

FREEMAN, J., ZIEMBA, C. M., HEEGER, D. J., SIMONCELLI, E. P. & MOVSHON, J. A. 2013. A functional and perceptual signature of the second visual area in primates. *Nature neuroscience,* 16**,** 974-981.

FREEMAN, R. D., OHZAWA, I. & WALKER, G. 2001. Beyond the classical receptive field in the visual cortex. *Progress in brain research,* 134**,** 157-170.

GABOR, D. 1946. Theory of communication. Part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering,* 93**,** 429-441.

GEISLER, W. S. 2008. Visual perception and the statistical properties of natural scenes. *Annu. Rev. Psychol.,* 59**,** 167-192.

GIBSON, J. J. 1950. The perception of the visual world.

GILBERT, C. D., DAS, A., ITO, M., KAPADIA, M. & WESTHEIMER, G. 1996. Spatial integration and cortical dynamics. *Proceedings of the National Academy of Sciences of the United States of America,* 93**,** 615-22.

GILBERT, E. N. 1980. Random Colorings of a Lattice of Squares in the Plane. *SIAM Journal on Algebraic and Discrete Methods,* 1**,** 152-159.

GOODALE, M. A. & MILNER, A. D. 1992. Separate visual pathways for perception and action. *Trends in neurosciences,* 15**,** 20-25.

GORITZ, A. S., WOLFF, H. G. & GOLDSTEIN, D. G. 2008. Individual payments as a longer-term incentive in online panels. *Behavior Research Methods,* 40.

GOSLING, S. D., VAZIRE, S., SRIVASTAVA, S. & JOHN, O. P. 2004. Should we trust Web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist,* 59**,** 93–104.

GRAHAM, N., BECK, J. & SUTTER, A. 1992. Nonlinear processes in spatial-frequency channel models of perceived texture segregation: Effects of sign and amount of contrast. *Vision research,* 32**,** 719-743.

GRAHAM, N. & SUTTER, A. 1996. Effect of spatial scale and background luminance on the intensive and spatial nonlinearities in texture segregation. *Vision Research,* 36**,** 1371-1390.

GRAHAM, N. & SUTTER, A. 1998. Spatial summation in simple (Fourier) and complex (non-Fourier) texture channels. *Vision research,* 38**,** 231-257.

GRAHAM, N. & SUTTER, A. 2000. Normalization: Contrast-gain control in simple (Fourier) and complex (non-Fourier) pathways of pattern vision. *Vision Research,* 40**,** 2737-2761.

GRIFFITH, D. A. 2003. *Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization*, Springer Science & Business Media.

HAIDER, B., KRAUSE, M. R., DUQUE, A., YU, Y., TOURYAN, J., MAZER, J. A. & MCCORMICK, D. A. 2010. Synaptic and network mechanisms of sparse and reliable visual cortical activity during nonclassical receptive field stimulation. *Neuron,* 65**,** 107-21.

HANAZAWA, A. & KOMATSU, H. 2001. Influence of the direction of elemental luminance gradients on the responses of V4 cells to textured surfaces. *J Neurosci,* 21**,** 4490-7.

HARALICK, R. M. 1979. Statistical and structural approaches to texture. *Proceedings of the IEEE,* 67**,** 786-804.

HARRIS, K. D. & THIELE, A. 2011. Cortical state and attention. *Nature reviews neuroscience,* 12**,** 509-523.

HARTLINE, H. K. 1938. The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *American Journal of Physiology,* 121**,** 400-415.

HAYTON, J. C., ALLEN, D. G. & SCARPELLO, V. 2004. Factor Retention Decisions in Exploratory Factor Analysis: a Tutorial on Parallel Analysis. *Organizational Research Methods,* 7.

HEEGER, D. J. 1992a. Half-squaring in responses of cat striate cells. *Visual neuroscience,* 9**,** 427-443.

HEEGER, D. J. 1992b. Normalization of cell responses in cat striate cortex. *Visual neuroscience,* 9**,** 181-197.

HEER, J., BOSTOCK, M. & 2010. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. *ACM Human Factors in Computing Systems (CHI)***,** 203-212.

HEGDÉ, J. & VAN ESSEN, D. C. 2000. Selectivity for complex shapes in primate visual area V2. *J Neurosci,* 20**,** 61-66.

HEILBRONNER, R. & BARRETT, S. 2013. *Image analysis in earth sciences: microstructures and textures of earth materials*, Springer Science & Business Media.

HEINEMANN, E. G. 1955. Simultaneous brightness induction as a function of inducing-and test-field luminances. *Journal of experimental psychology,* 50**,** 89.

HEISENBERG, M. & WOLF, R. 1984. *Vision in Drosophila. Genetics of microbehaviour*, Springer Verlag.

HOF, P. R., VOGT, B. A., BOURAS, C. & MORRISON, J. H. 1997. Atypical form of Alzheimer's disease with prominent posterior cortical atrophy: a review of lesion distribution and circuit disconnection in cortical visual pathways. *Vision Res,* 37**,** 3609-25.

HORTON, J. J., RAND, D. G. & ZECKHAUSER, R. J. 2011. The online laboratory: conducting experiments in a real labor market. *Experimental Economics,* 14**,** 399-425.

HOTELLING, H. 1933. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology,* 24**,** 417.

HUBEL, D. H. 1988. *Eye, brain, and vision*, Scientific American Library New York.

HUBEL, D. H. & WIESEL, T. N. 1959. Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology,* 148**,** 574-591.

HUBEL, D. H. & WIESEL, T. N. 1961. Integrative action in the cat's lateral geniculate body. *The Journal of physiology,* 155**,** 385-98.

HUBEL, D. H. & WIESEL, T. N. 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology,* 160**,** 106-54.

HUBEL, D. H. & WIESEL, T. N. 1965. Receptive Fields and Functional Architecture in Two Nonstriate Visual Areas (18 and 19) of the Cat. *Journal of neurophysiology,* 28**,** 229-89.

HUBEL, D. H. & WIESEL, T. N. 1968. Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology,* 195**,** 215-43.

HURRI, J. 1997. Independent component analysis of image data. *Master's thesis, Dept. of Computer Science and Engineering, Helsinki University of Technology, Espoo, Finland*.

HURRI, J., HYVÄRINEN, A., KARHUNEN, J. & OJA, E. Image feature extraction using independent component analysis. In Proc. NORSIG'96, 1996. Citeseer.

HYVÄRINEN, A. & OJA, E. 2000. Independent component analysis: algorithms and applications. *Neural networks,* 13**,** 411-430.

IBBOTSON, M. R., PRICE, N. & CROWDER, N. A. 2005. On the division of cortical cells into simple and complex types: a comparative viewpoint. *Journal of neurophysiology,* 93**,** 3699-3702.

IPEIROTIS, P. 2010. Analyzing the Amazon Mechanical Turk Marketplace. *ACM XRDS (Crossroads),* 17.

JACOBSON, L. & WECHSLER, H. 1984. Invariant analogical image representation and pattern recognition. *Pattern recognition letters,* 2**,** 289-299.

JAIN, A., ROSS, A. & PRABHAKAR, S. Fingerprint matching using minutiae and texture features. Image Processing, 2001. Proceedings. 2001 International Conference on, 2001. IEEE, 282-285.

JEONG, D. H., ZIEMKIEWICZ, C., RIBARSKY, W., CHANG, R. & CENTER, C. V. 2009. Understanding principal component analysis using a visual analytics tool. *Charlotte Visualization Center, UNC Charlotte*.

JONES, J. P. & PALMER, L. A. 1987. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of neurophysiology,* 58**,** 1233-1258.

JULESZ, B. 1962. Visual pattern discrimination. *Information Theory, IRE Transactions on,* 8**,** 84-92.

JULESZ, B. 1975. Experiments in the visual perception of texture. *Scientific American,* 232**,** 34-43.

JULESZ, B., GILBERT, E. N. & VICTOR, J. D. 1978. Visual discrimination of textures with identical third-order statistics. *Biological cybernetics,* 31**,** 137-40.

KASTNER, S., DE WEERD, P. & UNGERLEIDER, L. G. 2000. Texture segregation in the human visual cortex: a functional MRI study. *Journal of Neurophysiology,* 83**,** 2453-2457.

KERSTEN, D. 1987. Predictability and redundancy of natural images. *JOSA A,* 4**,** 2395-2400.

KINGDOM, F. A. 2011. Lightness, brightness and transparency: a quarter century of new ideas, captivating demonstrations and unrelenting controversy. *Vision Res,* 51**,** 652-73.

KITTUR, A., CHI, E. H. & SUH, B. Crowdsourcing User Studies with Mechanical Turk. Conference on Human Factors in Computing Systems, 2008. 453-456.

KUFFLER, S. W. 1953. Discharge patterns and functional organization of mammalian retina. *Journal of neurophysiology,* 16**,** 37-68.

LANDY, M. S. & GRAHAM, N. 2004. Visual perception of texture. *The visual neurosciences,* 2**,** 1106-1118.

LANDY, M. S. & KOJIMA, H. 2001. Ideal cue combination for localizing texture-defined edges. *JOSA A,* 18**,** 2307-2320.

LANGLOIS, D., CHARTIER, S. & GOSSELIN, D. 2010. An introduction to independent component analysis: InfoMax and FastICA algorithms. *Tutorials in Quantitative Methods for Psychology,* 6**,** 31-38.

LAUGHLIN, S. 1981. A simple coding procedure enhances a neuron's information capacity. *Zeitschrift fur Naturforschung. Section C: Biosciences,* 36**,** 910-2.

LENNIE, P. 1998. Single units and visual cortical organization. *PERCEPTION-LONDON-,* 27**,** 889-936.

LEVITT, J. B. & LUND, J. S. 1997a. Contrast dependence of contextual effects in primate visual cortex. *Nature,* 387**,** 73-6.

LEVITT, J. B. & LUND, J. S. 1997b. Contrast dependence of contextual effects in primate visual cortex. *Nature,* 387**,** 73-76.

LU, C. S., CHUNG, P. C. & CHEN, C. F. 1997. Unsupervised texture segmentation via wavelet transform. *Pattern Recognition,* 30**,** 729-742.

MA, L., WANG, Y. & TAN, T. Iris recognition using circular symmetric filters. Pattern Recognition, 2002. Proceedings. 16th International Conference on, 2002. IEEE, 414-417.

MADDESS, T. 2015. Histogram of Primitive Analysis of Deterministic and Stochastic Ternary Textures. *In Preparation*.

MADDESS, T., DAVEY, M. & YANG, E. 1999. Discrimination of complex textures by bees. *Journal of Comparative Physiology A,* 184**,** 107-117.

MADDESS, T. & KULIKOWSKI, J. J. 1999. Apparent fineness of stationary compound gratings. *Vision research,* 39**,** 3404-16.

MADDESS, T., MCCOURT, M., BLAKESLEE, B. & CUNNINGHAM, R. 1988. Factors governing the adaptation of cells in area-17 of the cat visual cortex. *Biological cybernetics,* 59**,** 229-236.

MADDESS, T. & NAGAI, Y. 2001. Discriminating isotrigon textures [corrected]. *Vision research,* 41**,** 3837-60.

MADDESS, T., NAGAI, Y., JAMES, A. C. & ANKIEWCZ, A. 2004. Binary and ternary textures containing higher-order spatial correlations. *Vision research,* 44**,** 1093-113.

MADDESS, T., NAGAI, Y., VICTOR, J. D. & TAYLOR, R. R. 2007. Multilevel isotrigon textures. *Journal of the Optical Society of America. A, Optics, image science, and vision,* 24**,** 278-93.

MAFFEI, L. & FIORENTINI, A. 1973. The visual cortex as a spatial frequency analyser. *Vision research,* 13**,** 1255-1267.

MAHON, L. & DE VALOIS, R. 2001. Cartesian and non-Cartesian responses in LGN, V1, and V2 cells. *Visual neuroscience,* 18**,** 973-981.

MANTE, V. & CARANDINI, M. 2005. Mapping of stimulus energy in primary visual cortex. *Journal of neurophysiology,* 94**,** 788-798.

MARČELJA, S. 1980. Mathematical description of the responses of simple cortical cells*. *JOSA,* 70**,** 1297-1300.

MARESCHAL, I. & SHAPLEY, R. M. 2004. Effects of contrast and size on orientation discrimination. *Vision research,* 44**,** 57-67.

MARTIN, K. A. & WHITTERIDGE, D. 1984. Form, function and intracortical projections of spiny neurones in the striate visual cortex of the cat. *The Journal of physiology,* 353**,** 463-504.

MARTINEZ, L. M. & ALONSO, J.-M. 2001. Construction of complex receptive fields in cat primary visual cortex. *Neuron,* 32**,** 515-525.

MASON, W. & SURI, S. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behav Res Methods,* 44**,** 1-23.

MASON, W. A. & WATTS, D. J. 2009. Financial incentives and the performance of crowds. *Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP '10)***,** 77–85.

MATERKA, A. & STRZELECKI, M. 1998. Texture analysis methods–a review. *Technical university of lodz, institute of electronics, COST B11 report, Brussels***,** 9-11.

MAUNSELL, J. H. & NEWSOME, W. T. 1987. Visual processing in monkey extrastriate cortex. *Annual review of neuroscience,* 10**,** 363-401.

MCCREADIE, R. M. C., MACDONALD, C. & OUNIS, I. 2010. Crowdsourcing a news query classification data set. *Proceedings of the ACM SIGIR 2010 Workshop on Crowdsourcing for Search Evaluation (CSE 2010)***,** 31-38.

MECHLER, F. & RINGACH, D. L. 2002. On the classification of simple and complex cells. *Vision research,* 42**,** 1017-1033.

MEGLEN, R. R. 1992. Examining large databases: a chemometric approach using principal component analysis. *Marine Chemistry,* 39**,** 217-237.

MISHKIN, M., UNGERLEIDER, L. G. & MACKO, K. A. 1983. Object vision and spatial vision: two cortical pathways. *Trends in neurosciences,* 6**,** 414-417.

MOVSHON, J. A., THOMPSON, I. D. & TOLHURST, D. J. 1978a. Receptive field organization of complex cells in the cat's striate cortex. *The Journal of physiology,* 283**,** 79-99.

MOVSHON, J. A., THOMPSON, I. D. & TOLHURST, D. J. 1978b. Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *The Journal of physiology,* 283**,** 53-77.

MURPHY, P. C. & SILLITO, A. M. 1987. Corticofugal feedback influences the generation of length tuning in the visual pathway. *Nature,* 329**,** 727-9.

NORRIS, M. & LECAVALIER, L. 2010. Evaluating the use of exploratory Factor analysis in developmental disability psychological research. *J Autism Dev Disord,* 40**,** 8-20.

O'CARROLL, D. 1993. Feature-detecting neurons in dragonflies.

OLSHAUSEN, B. A. & FIELD, D. J. 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature,* 381**,** 607-9.

OLSHAUSEN, B. A. & FIELD, D. J. 1997. Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vision research,* 37**,** 3311-25.

OLSHAUSEN, B. A. & FIELD, D. J. 2004. Sparse coding of sensory inputs. *Current opinion in neurobiology,* 14**,** 481-7.

OPPENHEIM, A. V. & LIM, J. S. 1981. The importance of phase in signals. *Proceedings of the IEEE,* 69**,** 529-541.

OPPENHEIMER, D. M., MEYVIS, T. & DAVIDENKO, N. 2009. Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology,* 45**,** 867-872.

ORBAN, G. A. 2008. Higher order visual processing in macaque extrastriate cortex. *Physiological Reviews,* 88**,** 59-89.

OZEKI, H., FINN, I. M., SCHAFFER, E. S., MILLER, K. D. & FERSTER, D. 2009. Inhibitory stabilization of the cortical network underlies visual surround suppression. *Neuron,* 62**,** 578-92.

PALMER, L. A., JONES, J. P. & MULLIKIN, W. H. 1985. *Functional organization of simple cell receptive fields*, Wiley, New York.

PAOLACCI, G., CHANDLER, J. & IPEIROTIS, P. 2010. Running experiments on Amazon Mechanical Turk. *Judgment and decision making,* 5**,** 411-419.

PASUPATHY, A. 2006. Neural basis of shape representation in the primate brain. *Prog Brain Res,* 154**,** 293-313.

PEARSON, K. 1901. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science,* 2**,** 559-572.

POLAT, U. & SAGI, D. 1993. Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vision research,* 33**,** 993-999.

PONTIN, J. 2007. Artificial intelligence, with help from the humans. *New York Times*, 25 March.

PRESS, W. H. 2007. *Numerical recipes 3rd edition: The art of scientific computing*, Cambridge university press.

PURPURA, K. P., VICTOR, J. D. & KATZ, E. 1994a. Striate cortex extracts higher-order spatial correlations from visual textures. *Proceedings of the National Academy of Sciences,* 91**,** 8482-8486.

PURPURA, K. P., VICTOR, J. D. & KATZ, E. 1994b. Striate cortex extracts higher-order spatial correlations from visual textures. *Proceedings of the National Academy of Sciences of the United States of America,* 91**,** 8482-6.

PURVES, D., AUGUSTINE, G. J., FITZPATRICK, D., HALL, W. C., LAMANTIA, A.-S., MCNAMARA, J. O. & WHITE, L. E. 2001. Neuroscience. *Sunderland (MA): Sinauer Associates*.

PURVES, D., SHIMPI, A. & LOTTO, R. B. 1999. An empirical explanation of the cornsweet effect. *J Neurosci,* 19**,** 8542-51.

REDDY, K. R., BADAMI, S. & BALASUBRAMANIAN, V. 1994. *Oscillations and Waves*, Universities Press.

REGAN, D. & HONG, X. H. 1994. Recognition and detection of texture-defined letters. *Vision Res,* 34**,** 2403-7.

REGAN, D. & SIMPSON, T. 1995. Multiple sclerosis can cause visual processing deficits specific to texture-defined form. *Neurology,* 45**,** 809-15.

REICHARDT, W. & ROSENBLITH, W. Autocorrelation, a principle for evaluation of sensory information by the central nervous system. Symposium on Principles of Sensory Communication 1959, 1961. MIT Press, 303-317.

REYMENT, R. & JORESKOG, K. 1996. *Applied Factor Analysis in the Natural Sciences,* Cambridge, UK, Cambridge University Press.

REYMENT, R. A., JORESKOG, K.G. 1996. Applied Factor Analysis in the Natural Sciences. Cambridge University Press.

RIVEST, J. & CABANAGH, P. 1996. Localizing contours defined by more than one attribute. *Vision research,* 36**,** 53-66.

RIZZO, M., ANDERSON, S. W., DAWSON, J. & NAWROT, M. 2000. Vision and cognition in Alzheimer's disease. *Neuropsychologia,* 38**,** 1157-69.

ROSENBLATT, M. & SLEPIAN, D. 1962. N th Order Markov Chains with Every N Variables Independent. *Journal of the Society for Industrial & Applied Mathematics,* 10**,** 537-549.

ROSLI, Y., BEDFORD, S. M. & MADDESS, T. 2009. Low-spatial-frequency channels and the spatial frequency-doubling illusion. *Investigative ophthalmology & visual science,* 50**,** 1956-63.

ROSS, J., IRANI, I., SILBERMAN, M., ZALDIVAR, A. & TOMLINSON, B. 2010. Who are the CrowdWorkers? Shifting Demographics in Amazon Mechanical Turk. *CHI EA 2010***,** 2863-2872.

RUBIN, N. 2001. The role of junctions in surface completion and contour matching. *PERCEPTION-LONDON-,* 30**,** 339-366.

SANDLER, R. & LINDENBAUM, M. Gabor filter analysis for texture segmentation. Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on, 2006. IEEE, 178-178.

SCHWARTZ, E. L. 1980. Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to perceptual coding. *Vision research,* 20**,** 645-669.

SEAMONS, J. W., BARBOSA, M. S., BUBNA-LITIC, A. & MADDESS, T. 2015. A lower bound on the number of mechanisms for discriminating fourth and higher order spatial correlations. *Vision research,* 108**,** 41-48.

SEKULAR, R. & BLAKE, R. 1994. Perception (3rd edn). New York: McGraw-Hill.

SEKULER, R., WILSON, H. R. & OWSLEY, C. 1984. Structural modeling of spatial vision. *Vision research,* 24**,** 689-700.

SERENO, M. I., DALE, A., REPPAS, J., KWONG, K., BELLIVEAU, J., BRADY, T., ROSEN, B. & TOOTELL, R. 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science,* 268**,** 889-893.

SHANNON, C. E. 2001. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review,* 5**,** 3-55.

SHANNON, C. E. & WEAVER, W. 1948. *The Mathematical Theory of Communication*, University of Illinois Press.

SHAPLEY, R. & ENROTH-CUGELL, C. 1984. Visual adaptation and retinal gain controls. *Progress in retinal research,* 3**,** 263-346.

SHEPHERD, G. M. 1998. *The synaptic organization of the brain*, Oxford University Press New York.

SIMONCELLI, E. P. & OLSHAUSEN, B. A. 2001. Natural image statistics and neural representation. *Annual review of neuroscience,* 24**,** 1193-216.

SIMPSON, W. A. & MCFADDEN, S. M. 2005. Spatial frequency channels derived from individual differences. *Vision research,* 45**,** 2723-7.

SKITKA, L. J. & SARGIS, E. G. 2006. The internet as psychological laboratory. *Annu Rev Psychol,* 57**,** 529-55.

SNOWDEN, R., SNOWDEN, R. J., THOMPSON, P. & TROSCIANKO, T. 2012. *Basic vision: an introduction to visual perception*, Oxford University Press.

SOLOMON, S. G., LEE, B. B. & SUN, H. 2006. Suppressive surrounds and contrast gain in magnocellular-pathway retinal ganglion cells of macaque. *The Journal of neuroscience : the official journal of the Society for Neuroscience,* 26**,** 8715-26.

SRINIVASAN, G. & SHOBHA, G. Statistical texture analysis. Proceedings of world academy of science, engineering and technology, 2008. 1264-1269.

SRINIVASAN, M., ZHANG, S. & WITNEY, K. 1994. Visual discrimination of pattern orientation by honeybees: Performance and implications forcortical'processing. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 343**,** 199-210.

SRINIVASAN, M. V., LAUGHLIN, S. B. & DUBS, A. 1982. Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal*

Society of London. Series B, Containing papers of a Biological character. Royal Society, 216**,** 427-59.

STONE, J. V. 2004. *Independent component analysis*, Wiley Online Library.

STONE, M. & BARTRAM, L. 2008. Alpha, Contrast and the Perception of Visual Metadata. *16th Color and Imaging Conference.* Society for Imaging Science and Technology.

STRZELECKI, M. & MATERKA, A. Markov random fields as models of textured biomedical images. Proc. 20th National Conf. Circuit Theory and Electronic Networks KTOiUE'97, 1997. 493-498.

SUHR, D. D. 2005. Principal component analysis vs. exploratory Factor analysis. *SUGI 30 Proceedings***,** 203-230.

TAO, L., SHELLEY, M., MCLAUGHLIN, D. & SHAPLEY, R. 2004. An egalitarian network model for the emergence of simple and complex cells in visual cortex. *Proceedings of the National Academy of Sciences,* 101**,** 366-371.

TAYLOR, R. R. 2008. *Neural Computation of Statistical Structure.* PhD, The Australian National University.

TAYLOR, R. R., MADDESS, T. & NAGAI, Y. 2008. Spatial biases and computational constraints on the encoding of complex local image structure. *Journal of vision,* 8**,** 19 1-13.

TEICHERT, T., WACHTLER, T., MICHLER, F., GAIL, A. & ECKHORN, R. 2007. Scale-invariance of receptive field properties in primary visual cortex. *BMC neuroscience,* 8**,** 38.

TKACIK, G., PRENTICE, J. S., VICTOR, J. D. & BALASUBRAMANIAN, V. 2010. Local statistics in natural scenes predict the saliency of

synthetic textures. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 18149-54.

TODD, J. T., OOMES, A. H., KOENDERINK, J. J. & KAPPERS, A. M. 2004. The perception of doubly curved surfaces from anisotropic textures. *Psychol Sci,* 15**,** 40-6.

TOOTELL, R. B., SWITKES, E., SILVERMAN, M. S. & HAMILTON, S. L. 1988. Functional anatomy of macaque striate cortex. II. Retinotopic organization. *The Journal of Neuroscience,* 8**,** 1531-1568.

TOYAMA, K., KIMURA, M. & TANAKA, K. 1981a. Cross-Correlation Analysis of Interneuronal Connectivity in cat visual cortex. *Journal of neurophysiology,* 46**,** 191-201.

TOYAMA, K., KIMURA, M. & TANAKA, K. 1981b. Organization of cat visual cortex as investigated by cross-correlation technique. *Journal of neurophysiology,* 46**,** 202-14.

TURANO, K. & PANTLE, A. 1989. On the mechanism that encodes the movement of contrast variations: velocity discrimination. *Vision research,* 29**,** 207-221.

VAN ESSEN, D. C., ANDERSON, C. H. & FELLEMAN, D. J. 1992. Information processing in the primate visual system: an integrated systems perspective. *Science,* 255**,** 419-423.

VAN HATEREN, J. 1992a. Real and optimal neural images in early vision. *Nature*.

VAN HATEREN, J. 1992b. Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation. *Journal of Comparative Physiology A,* 171**,** 157-170.

VAN HATEREN, J. 1992c. A theory of maximizing sensory information. *Biological cybernetics,* 68**,** 23-29.

VAN HATEREN, J. H. & VAN DER SCHAAF, A. 1998. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society of London. Series B: Biological Sciences,* 265**,** 359-366.

VAN KLEEF, J. P., CLOHERTY, S. L. & IBBOTSON, M. R. 2010. Complex cell receptive fields: evidence for a hierarchical mechanism. *The Journal of physiology,* 588**,** 3457-3470.

VAN SANTEN, J. P. & SPERLING, G. 1985. Elaborated reichardt detectors. *JOSA A,* 2**,** 300-320.

VAZ, S., FALKMER, T., PASSMORE, A. E., PARSONS, R. & ANDREOU, P. 2013. The case for using the repeatability coefficient when calculating test-retest reliability. *PLoS One,* 8**,** e73990.

VICTOR, J. D. 1985. Complex visual textures as a tool for studying the VEP. *Vision research,* 25**,** 1811-27.

VICTOR, J. D. 1988. Models for preattentive texture discrimination: Fourier analysis and local feature processing in a unified framework. *Spatial vision*.

VICTOR, J. D. 1994. Images, Statistics, and Textures: Implications of Triple Correlation Uniqueness for Texture Statistics and the Julesz Conjecture. *Journal of the Optical Society of America A,* 11**,** 1680-1684.

VICTOR, J. D. 1995. Isodipole Textures: A Window on Cortical Mechanisms of Form Processing. *In:* PAPATHOMAS, T. V. C., C.; GOREA, A; KOWLER, E (ed.) *Early Vision and Beyond.* The MIT Press.

VICTOR, J. D. & BRODIE, S. E. 1978. Discriminable textures with identical Buffon needle statistics. *Biological Cybernetics,* 31**,** 231-234.

VICTOR, J. D., CHUBB, C. & CONTE, M. M. 2005. Interaction of luminance and higher-order statistics in texture discrimination. *Vision research,* 45**,** 311-328.

VICTOR, J. D. & CONTE, M. M. 1991. Spatial organization of nonlinear interactions in form perception. *Vision research,* 31**,** 1457-88.

VICTOR, J. D. & CONTE, M. M. 1996. The role of high-order phase correlations in texture processing. *Vision Res,* 36**,** 1615-31.

VICTOR, J. D. & CONTE, M. M. 2005. Local processes and spatial pooling in texture and symmetry detection. *Vision research,* 45**,** 1063-73.

VICTOR, J. D. & CONTE, M. M. 2012. Local image statistics: maximum-entropy constructions and perceptual salience. *Journal of the Optical Society of America. A, Optics, image science, and vision,* 29**,** 1313-45.

VICTOR, J. D. & KNIGHT, B. W. 2003. Simultaneously band and space limited functions in two dimensions, and receptive fields of visual neurons. *Perspectives and Problems in Nolinear Science.* Springer.

VICTOR, J. D., THENGONE, D. J. & CONTE, M. M. 2013. Perception of second- and third-order orientation signals and their interactions. *J Vis,* 13**,** 21.

VINJE, W. E. & GALLANT, J. L. 2000. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science,* 287**,** 1273-6.

VON GUNTEN, A., BOURAS, C., KOVARI, E., GIANNAKOPOULOS, P. & HOF, P. R. 2006. Neural substrates of cognitive and behavioral deficits in atypical Alzheimer's disease. *Brain Res Brain Res Rev,* 51**,** 176-211.

WACHTLER, T., LEE, T.-W. & SEJNOWSKI, T. J. 2001. Chromatic structure of natural scenes. *JOSA A,* 18**,** 65-77.

WATSON, A. B. 1983. *Detection and recognition of simple spatial forms*, Springer.

WERKHOVEN, P., SPERLING, G. & CHUBB, C. 1993. The dimensionality of texture-defined motion: a single channel theory. *Vision research,* 33**,** 463-485.

WERKHOVEN, P., SPERLING, G. & CHUBB, C. 1994. Perception of apparent motion between dissimilar gratings: spatiotemporal properties. *Vision research,* 34**,** 2741-2759.

WHITE, M. 1979. A new effect of pattern on perceived lightness. *Perception,* 8**,** 413-416.

WIGNER, E. 1932. On the quantum correction for thermodynamic equilibrium. *Physical Review,* 40**,** 749.

WILLIAMS, S. M., MCCOY, A. N. & PURVES, D. 1998. The influence of depicted illumination on brightness. *Proceedings of the National Academy of Sciences,* 95**,** 13296-13300.

WOLFSON, S. S. & LANDY, M. S. 1995. Discrimination of orientation-defined texture edges. *Vision research,* 35**,** 2863-2877.

YANG, E.-C. & MADDESS, T. 1997. Orientation-sensitive neurons in the brain of the honey bee (Apis mellifera). *Journal of Insect Physiology,* 43**,** 329-336.

YIN, Q., KIM, J.-N. & SHEN, L. 2009. Rotation-invariant texture classification using circular Gabor wavelets. *Optical Engineering,* 48**,** 017001-017001-13.

ZETZSCHE, C., KRIEGER, G. & WEGMANN, B. 1999. The atoms of vision: Cartesian or polar? *JOSA A,* 16**,** 1554-1565.

ZETZSCHE, C. & ROHRBEIN, F. 2001. Nonlinear and extra-classical receptive field properties and the statistics of natural scenes. *Network,* 12**,** 331-50.

ZWICK, W. R. & VELICER, W. F. 1986. Comparison of five rules for determining the number of components to retain. *Psychological Bulletin,* 99**,** 432-442.

Chapter 2: A Lower Bound on the Number of Mechanisms for discriminating Fourth- and Higher-order Spatial Correlations

2.1 Abstract

Research on single striate cortical neurons has often concentrated on their responses to stimuli defined by two-point correlations. Texture discrimination studies using a relatively small palette of isotrigon textures have indicated that humans are sensitive to third- and higher-order spatial correlations.

To further evaluate the underlying mechanisms of texture discrimination, subjects discriminated random binary noise patterns from ten new isotrigon texture types.

Factor analysis revealed that as few as three mechanisms may govern the detection of fourth- and higher-order image structure. This supports the findings of previous studies using different isotrigon textures.

The computation of higher-order correlations by the brain is neuro-physiologically plausible. The mechanisms identified in this study may represent some short range nonlinear combination of recursive and/or rectifying processes.

2.2 Introduction

Natural images contain large amounts of structural information characterised by higher-order spatial correlations (Franz and Schölkopf, 2005). Neurons have limited information capacities and energy budgets, so this volume of

information presents a considerable challenge. The human visual system must employ mechanisms to filter out redundant or non-salient information, but retain that which is behaviourally relevant (Barlow, 2001, Barlow, 1963).

The orientation and the spatial-frequency tuning of mammalian simple cells is suited to a *sparse* coding system which minimizes informational redundancy (Field, 1987). The antagonistic centre-surround receptive field architecture exploits the degree of second-order correlation in images to remove redundant information. The information removed in such *predictive encoding* processes is related to second-order correlations (Srinivasan et al., 1982). In this manner, the dynamic range of a neuron can be devoted to encoding a small range of non-redundant contrasts, and the visual system can encode spatial detail in a manner that minimizes the effects of intrinsic noise. Interestingly, the visual system also appears to filter out the predictable higher-order structural information from of natural scenes, whilst retaining and transmitting that which is not predictable (and therefore behaviourally relevant) (Tkacik et al., 2010). End-stopping may be particularly important in filtering out redundant higher-order structural information (Zetzsche and Nuding, 2005).

Taken together, these studies highlight the importance of investigating higher-order correlations in images and how they are processed. Such investigations may also tell us a great deal about cortical functioning (Victor, 1995, Purpura et al., 1994, Victor and Conte, 1996). Previous research into the responses of single striate cortical neurons has concentrated on two-point correlation properties, as captured by spatial frequency and orientation tuning. There has been less research into third- and higher-order

correlations, although the visual system is sensitive these relationships in natural images (Tkacik et al., 2010).

Using artificially generated textures we can probe the sensitivities and limitations of the human visual system to higher-order correlations. These artificial images can have very well controlled statistical properties, and an important class of such test images are the original set of i*sotrigon textures* (e.g. Figure 1). The word "isotrigon" derives from the fact that the average first to third order correlation functions of these textures is zero, as is the case for uniform noise. The obvious structure in these textures is therefore exclusively due to fourth- and higher-order spatial correlations (Maddess et al., 2004, Maddess et al., 2007, Julesz et al., 1978, Victor, 1994, Gilbert, 1980). Thus, in order to discriminate a particular isotrigon texture from noise, it is necessary to identify its complex, higher-order structure; isotrigon textures cannot be discriminated based on luminance or other lower-order properties. Our discrimination ability seems to be pre-attentive and changes little for long presentation times (Taylor et al., 2008).
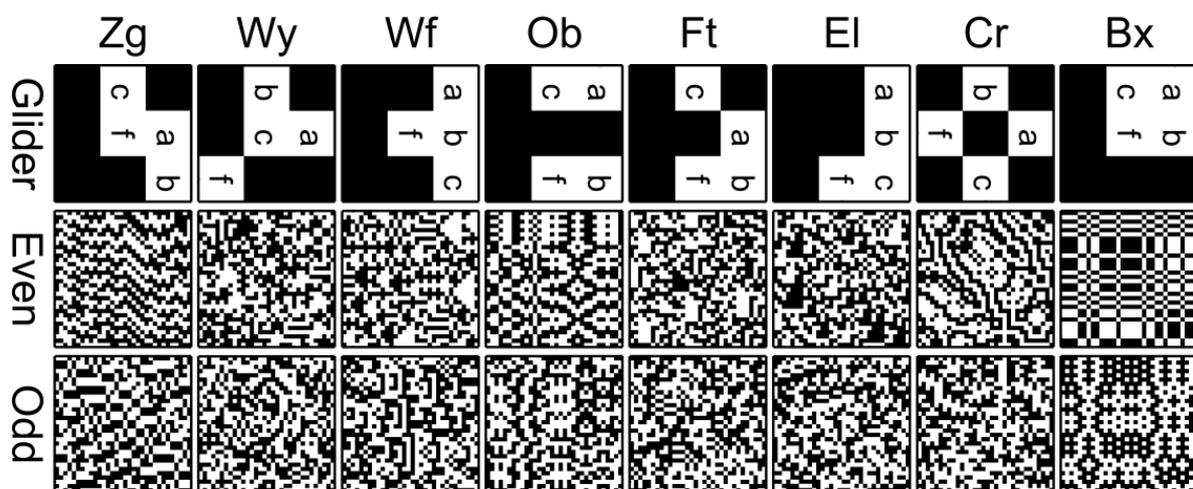

Figure 1. Examples of the first set of binary isotrigon textures to be published (middle and left column). The textures are created by a recursive

rule-based process where the rule is operated in concert with *glider* patterns (Victor and Conte, 1991). To begin the process a matrix is set randomly dark (-1) or bright (1) with equal probability, each entry representing a pixel. A 3x3 pixel glider is then moved over the matrix. The gliders are shown in the left column, where they are greatly magnified, their pixels actually match those of the textures. How the glider affects the underlying pixels is determined by the shape of the glider's active regions. Glider pixels marked *a,b,c* are the input pixels, and *f is* the position of the output pixel. Two rules termed *Even* and *Odd* were used to create all the textures. The Even and Odd rules are *f = abc* and *f = -abc*. The glider is moved in steps of one pixel across the matrix modifying the entries using the selected rule until the texture is complete. Since former outputs become inputs the process is recursive. In fact only the first few rows and columns need to be seeded with random -1 and 1 values and those will determine the outcome, hence these methods are said to be deterministic.

Humans and primates have been shown to be sensitive to the higher-order correlations in isotrigon textures in a number of psychophysical (Maddess and Nagai, 2001, Maddess et al., 2007, Victor and Conte, 2005), VEP (Victor and Conte, 1991), fMRI (Beason-Held et al., 1998a, Beason-Held et al., 2000), PET (Beason-Held et al., 1998b) studies, and also single cell recordings (Purpura et al., 1994). Although they are artificially generated, the same structural features that give isotrigon textures visual salience also create visual salience in natural images (Tkacik et al., 2010). Therefore, they are an ideal tool for evaluating human texture perception. This paper investigates a lower bound on how many mechanisms might operate to discriminate correlations at fourth and higher orders.

Most of the isotrigon textures used to date have been created using a recursion procedure in which a combinatorial rule operates on n pixels selected by a gliding template, or *glider*, determines the value of an output pixel (Victor and Conte, 1991). For details see the caption of Figure 1. Here we introduce the VnL2 isotrigon textures where n can be 4, 5 or 6 and the gliders are 4x4 pixels (Fig. 2), the L2 refers to the number of grey levels. This is the first paper to investigate these particular isotrigon textures.

Humans use a number of neurophysiological mechanisms to capture visually salient higher order structure. Evidence has been presented that the number of independent mechanisms is less than 10 (Taylor et al., 2008), and is more likely 3 to 4 (Maddess and Nagai, 2001, Maddess et al., 2007). One statistically well principled method is to use Factor analysis of human performance functions from many individuals, or many repeats for single individuals, to infer the number and form of underlying independent mechanisms (Rosli et al., 2009, Maddess and Kulikowski, 1999, Sekuler et al., 1984, Simpson and McFadden, 2005). For example, the technique has been used to evaluate the mechanisms which underpin contrast sensitivity (Peterzell and Teller, 1996, Peterzell et al., 1993). In this study, we apply that approach to performance functions for discriminating the new VnL2 textures from random textures. Additionally, given that many repeats were involved, we also examined learning effects.
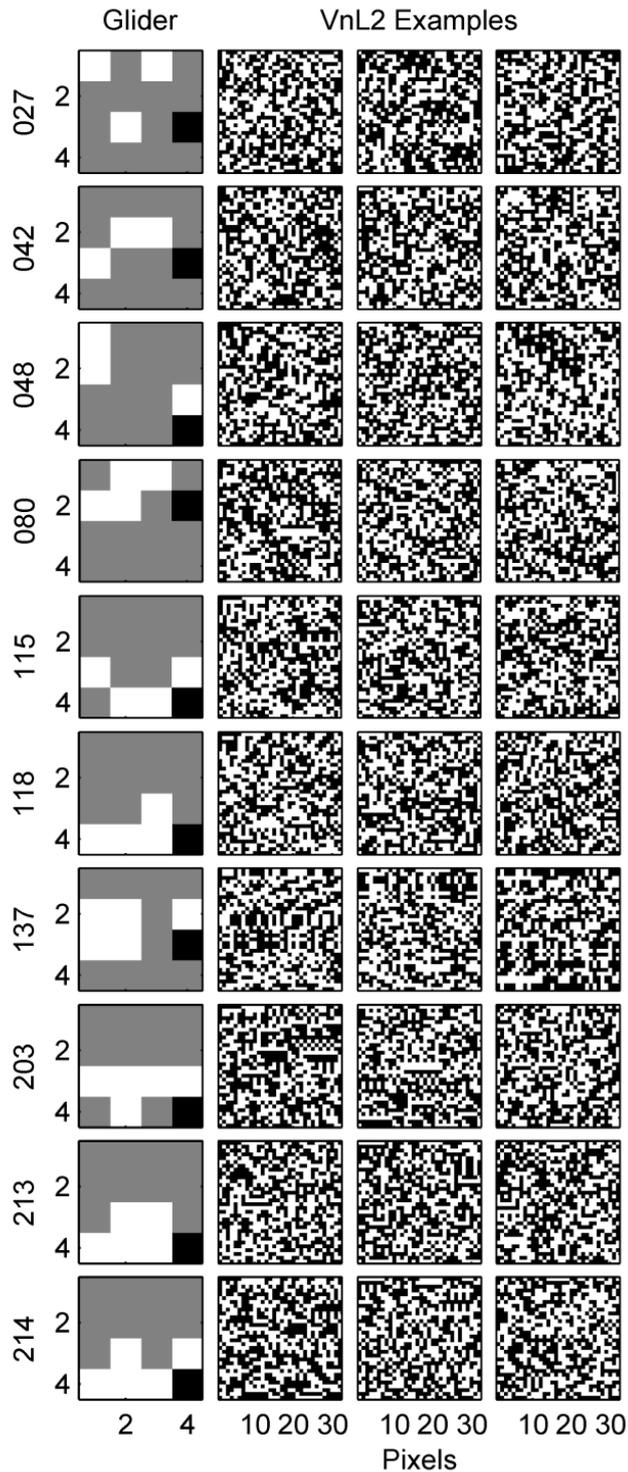
Figure 2. Examples of the newer VnL2 textures and their gliders. In this nomenclature, "V" refers to the number of input variables in the 4x4 pixel gliders, and "L" refers to the number of grey levels. The white glider pixels are the input pixels. The product of those pixels was then placed at the

location of the black glider pixel in the recursion process. The combinatorial rule is thus similar to the Even rule of Figure 1.

2.3 Materials and Methods

Two healthy male subjects (JWS and ALB) were recruited to the study (age 34 and 22). They had corrected to normal vision. Four other normal subjects aged 22, 33, 40, and 56 (one female) also participated to make a third, combined data set. In this data set, only early data from JWS and ALB was used, to equate the level of experience of the six participants. Informed consent was obtained under ANU Human Experimentation Ethics Committee protocol 238/04. All research adhered to the tenets of the Declaration of Helsinki.

The VnL2 textures (Fig. 2) were created by the procedure discussed above (Victor and Conte, 1991, Maddess and Nagai, 2001). Several hundred VnL2 texture types were created and screened for being isotrigon by examining their mean second- and third-order correlation functions (Maddess et al., 2004, Maddess et al., 2007). All texture production, stimulus presentation, and analysis was done using MatLab (The Mathworks, Natick, MA).

Ten VnL2 textures were selected because they exhibit a good variety of image structure, from quite salient to almost random in appearance. The set presumably modulates the human texture mechanisms well. Importantly, no texture was so salient that the performance came close to saturating at 100%: this would remove the variability that Factor analysis is based upon.

The achromatic (colour temperature 6500 K) texture patterns were displayed at mean luminance of 395 cd/m^2 on a Phillips MGD 403 ultra-high brightness monitor at a pixel resolution of 512×420 square pixels, and a refresh rate of 101 Hz. The ambient illumination was that provided by the monitor in an otherwise darkened room. Subjects viewed the patterns binocularly from a fixed distance of 46 cm, and head position of the subjects was maintained using a chin rest. The angular size of the stimuli is given in Figure 3 and they were 11.3 and 22.6 degrees square.

The experiment was designed as a series of classification tasks in which subjects attempted to discriminate between VnL2 textures and random binary noise. Two sizes of textures were employed (16x16 and 32x32 pixels), providing two difficulty levels. Subjects practiced using the larger size before being tested with the smaller stimuli (Fig. 3). The sizes were similar to previous studies (Maddess and Nagai, 2001, Victor and Conte, 1991), which had shown that the relevant range of salient correlations is less than ± 6 pixels.

The stimuli were presented for 297 ms within a smooth temporal window. During each texture presentation, the change in contrast of the textures incorporated a sigmoidal onset and offset as determined by a Blackman function and details of the exact function are given elsewhere (Maddess and Nagai, 2001). This was designed to mitigate the nonlinear effects associated with abrupt changes in contrast. The textures were at full contrast for 204 ms which constitutes pre-attentive viewing (Maddess and Nagai, 2001, Julesz et al., 1973). Subjects used a mouse button to indicate whether the texture was

137

random or non-random. To facilitate learning, incorrect choices were indicated by a tone (Maddess and Nagai, 2001).
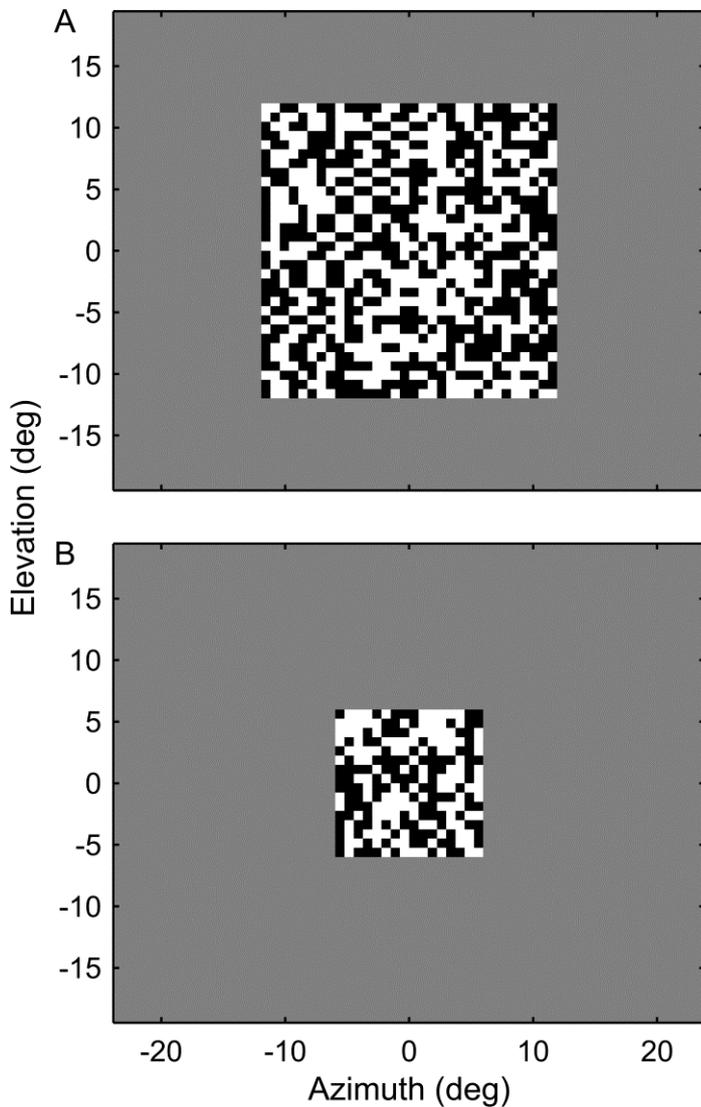

Figure 3. Two sizes of textures were employed (32x32 and 16x16 pixels), which thereby provided two difficulty levels. In the first phase of testing, subjects made 20 discriminations between 32x32 pixel textures. Then, in the second phase, 20 discriminations were made between 16x16 pixel textures. The same isotrigon texture was compared with random textures in each of these blocks of 40 discriminations. The order of these texture test blocks was randomised. For some subjects, this was repeated several times and in later blocks only the smaller textures were shown in further repeats. Each

texture pixel was actually 8 display pixels square. The textures were displayed on a neutral grey background at the middle luminance of the texture checks.

Test sessions proceeded as follows. One of the 10 VnL2 texture types was selected at random for testing. Before testing the subject inspected a sheet depicting 15 32x32 pixel examples of the texture type to be tested. For example, a subject was shown printed examples of texture 048. The subject was then presented with a sequence of 20 images: either an example of texture 048 or a random binary noise pattern. The subject was asked to determine whether the image they were presented was random or textured.

First twenty 32x32 pixel examples were tested, and then twenty 16x16 pixel examples, the larger textures providing some training for the smaller stimuli. We will refer of the testing of 40 examples as a test "block". After the last of the images was presented, a new texture was selected at random from the remaining 9 VnL2s. Once again, printed examples of the texture were provided before the trial commenced. This pattern was repeated until blocks for all 10 texture types had been tested. In later trials of subjects ABL and JWS the 32x32 examples were replaced by 16x16 meaning that 40 examples were shown in each block. For each texture size the mean performance, i.e. probability correct, did calculated providing a performance function comprise ten points, one for each texture type.

Subjects JWS and ABL each completed 16 blocks as described above (6,400 individual presentations). The test blocks were not carried out in one sitting, but were spread out over a period of days. Data from four other

subjects was combined (minimum of 40 presentations each). The data from the first few presentations to JWS and ABL were included to provide two further data sets from relatively inexperienced subjects for 16x16 and 32x32 textures. These data sets are labelled All16 and All32.

A further data set from a previous study was added (Maddess and Nagai, 2001). The spatial and temporal presentation aspects we identical except that the two subjects saw blocks of 12 texture/random alternates for each of 32x32, 16x16 and 8x8 pixels in block. The 18 V3L2 textures were the same as in Figure 1 except that Even and Odd versions of texture called Triangle were added. The two subjects each repeated 5 sets of blocks, and were very experience viewers (TM and YN).

We applied Factor analysis to estimate how many neural mechanisms might contribute to the performance functions, and their possible weightings for each texture type. This exploratory statistical method has been used by ourselves (Maddess and Kulikowski, 1999, Rosli et al., 2009) and others (Peterzell and Teller, 1996, Peterzell et al., 1993, Sekuler et al., 1984, Simpson and McFadden, 2005) to address similar questions. Factor analysis exploits covariance across performance functions to provide information about any underlying mechanisms. Simply stated, factor models can present simplified representations of the many data variables in terms of a small number of unobservable variables: in the present case putative texture discrimination mechanisms that determine the form of the measured performance functions. The basic equation of a Factor analysis is:

$$X = S\,L' + E \qquad \text{(Eq. 1)}$$

…where X is the data matrix, in the present case the rows were observations (performance functions) for the 10 textures (columns), one row for each repeat or subject depending on the data set. S is the matrix of factor *scores*, i.e. the estimated responses of the mechanisms to each texture over the repeats. L is the matrix of factor *loadings*: the loadings are regression weights that are used to combine the factor scores into the performance variables (X), here representing the influences of the independent putative mechanisms (factors) upon the discrimination of each texture type.

Importantly, if these loadings are similar for different subjects then the factors are more likely to be real. S has size N-repeats x k-factors (see below), and L is N-textures x k-factors. The loadings were computed from R, the correlation matrix of X by singular value decomposition (SVD), such that $R= F{\times}D{\times}G$ providing:

$$L=F{\times}\text{diag}(D)^{0.5} \qquad \text{(Eq. 2)}$$

…and the factor scores were given by:

$$S = XL(L'L)^{-1} \qquad \text{(Eq. 3)}$$

The columns of F (or G, they are equal in this case) are the eigenvectors of R and the elements of the diagonal matrix D are the eigenvalues. The eigenvalues provide the proportion of the variance in R provided by each eigenvector. Commonly the eigenvalues are sorted and plotted to form a so called *scree plot*. If some small number of factors determines the observed

variables then a sharp step can occur in this *scree* plot the number of eigenvalues before the step providing an indication of the number (k) of factors to consider, the remaining small components being discarded as noise. Note that Eqs. 2 and 3 are computed for the k columns of F and D, which provide the k highest eigenvalues to provide a k-factor model, the selected eigenvectors being the estimated *factors.* We will refer to this SVD based method as a Principal Components (PC) Factor analysis. The overall objective is to find the smallest set of factors that account for the original data well. Here this would provide a lower bound on the number of mechanisms for discriminating higher order structure.

One measure of the adequacy of these models is provided by the *communalities* (Reyment and Joreskog, 1996), which are akin to r^2-values and indicate what proportion of each of the original variables is accounted for by the k-factor model, hence *the communalities for each of the points in the performance functions examined should be uniformly high* if the reduced factor model accounts well for all the input variables (e.g. Fig. 7D,F). Before forming L (Eq. 2) the loadings were rotated according the Varimax method (Reyment and Joreskog, 1996). Through the Varimax method each rotated factor is a combination of only the highly interdependent input variables. That is the factors tend to be loaded onto (i.e. reflect) the input variables.

2.4 Results

2.4.1 Mean Performance

The mean performance functions for all subjects are shown in Figure 4. The upper two panels show the means for the first and last 10 repeats for JWS and ABL.
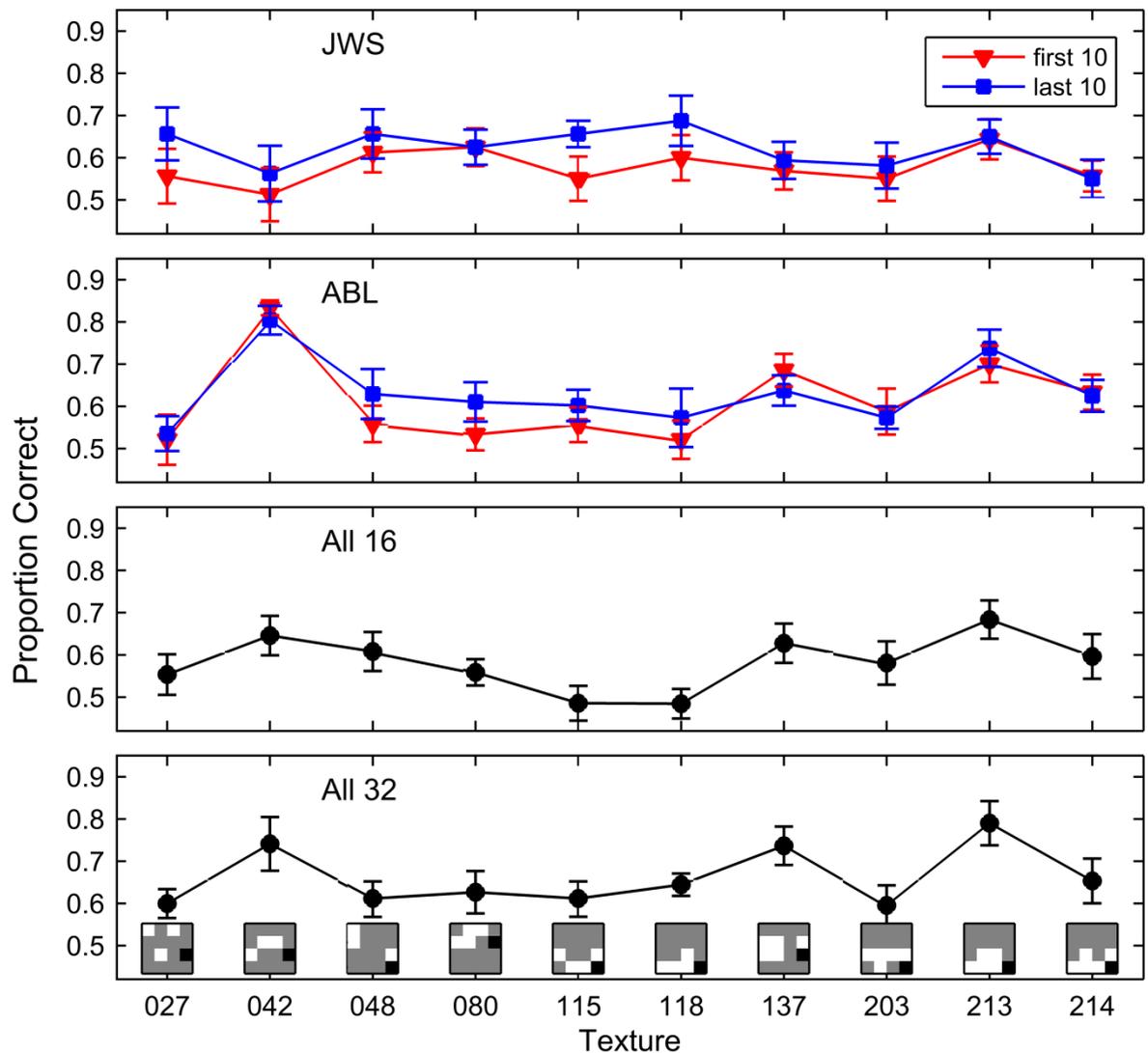

Figure 4. Mean *performance functions* for all subjects by texture type (glider name). The performance data (probability correct) is presented separately according to subjects and the texture size. For All16 (16x16 textures) and All32 (32x32 textures) the error bars are SE for N=6 subjects. The JWS and

ABL the data use the first and last 10 repeats of 26 and 21 total for 16 x16 textures. The glider shapes have been added in the bottom panel to improve the link between the performance functions and Figure 2.

For relatively naïve subjects, the performance data is split according to texture size (All16 and All32). Although ABL and JWS differ in their performance on texture 042, all subjects agree quite well on all other textures in terms of the form of their mean performance functions. The performance of last 10 repeats of JWS was on average significantly higher than his first 10 ($p = 0.04$), but that of ABL was not.

2.4.2 Factor Analysis

Figure 5 shows scree plots which display the proportion of variance accounted for the 10 possible factors from each of the 4 data sets (JWS, ABL, All16 and All32).
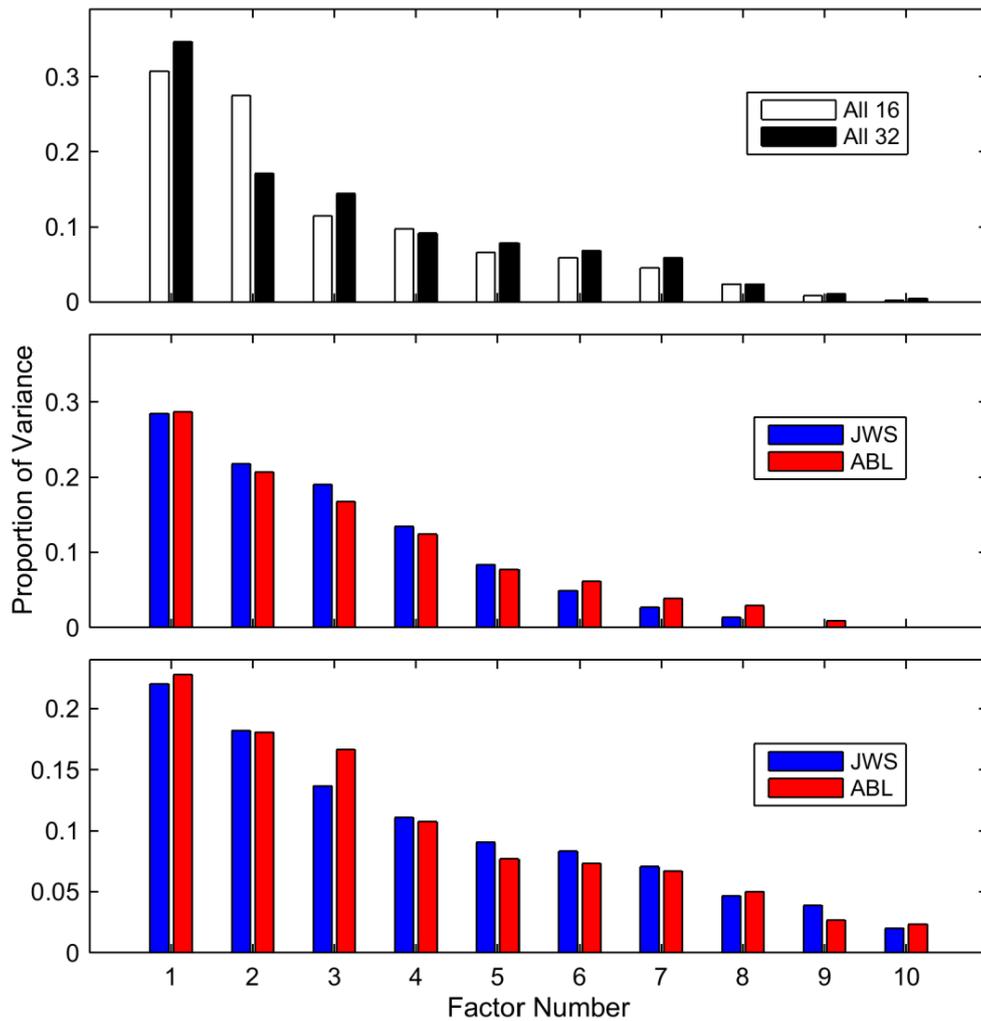
Figure 5. Top: Scree plots showing the proportion of variance accounted for in the data sets of Figure 4. The proportions of variance are plotted in order of their relative influence, from the largest to smallest. For All16, the scree plot shows a step between the 2nd and 3rd factors, possibly indicating the presence of 2 underlying mechanisms. First 10: the factors seem to drop off linearly rather than exponentially, as they would if they were noise. Last 10: ABL seems to show a marked step between the 3rd and 4th factors. JWS shows a more linear decline.

The proportions are plotted in order of their relative influence, from the largest to smallest. If the data being evaluated is governed by a small number of principal mechanisms, we may see a few large eigenvalues

145

followed by a significant drop-off: the lesser factors exert less influence on our data and trail off exponentially into noise.

There is a drop-off in the accounted proportion of variance after the first two factors for All16. The last 10 visits of ABL also seem to drop off after 3 factors. This may suggest that two, or perhaps three, independent neurophysiological mechanisms govern VnL2 texture discrimination. Better evidence for this idea comes from the communalities and the factor loadings.

2.4.3 Communalities

Figure 6 shows an evaluation of the factor model validity for N factors using communalities. The communalities indicate the proportion of variance accounted for within the data for each texture. A robust model will provide a balanced account of the data for each texture, i.e. communalities of even size. Communalities obtained for the largest 5 factors are shown for the All16 data set, which is the most representative of typical performance among the available data sets. It is clear that by 3 to 4 factors, a balanced account of the data is obtained.
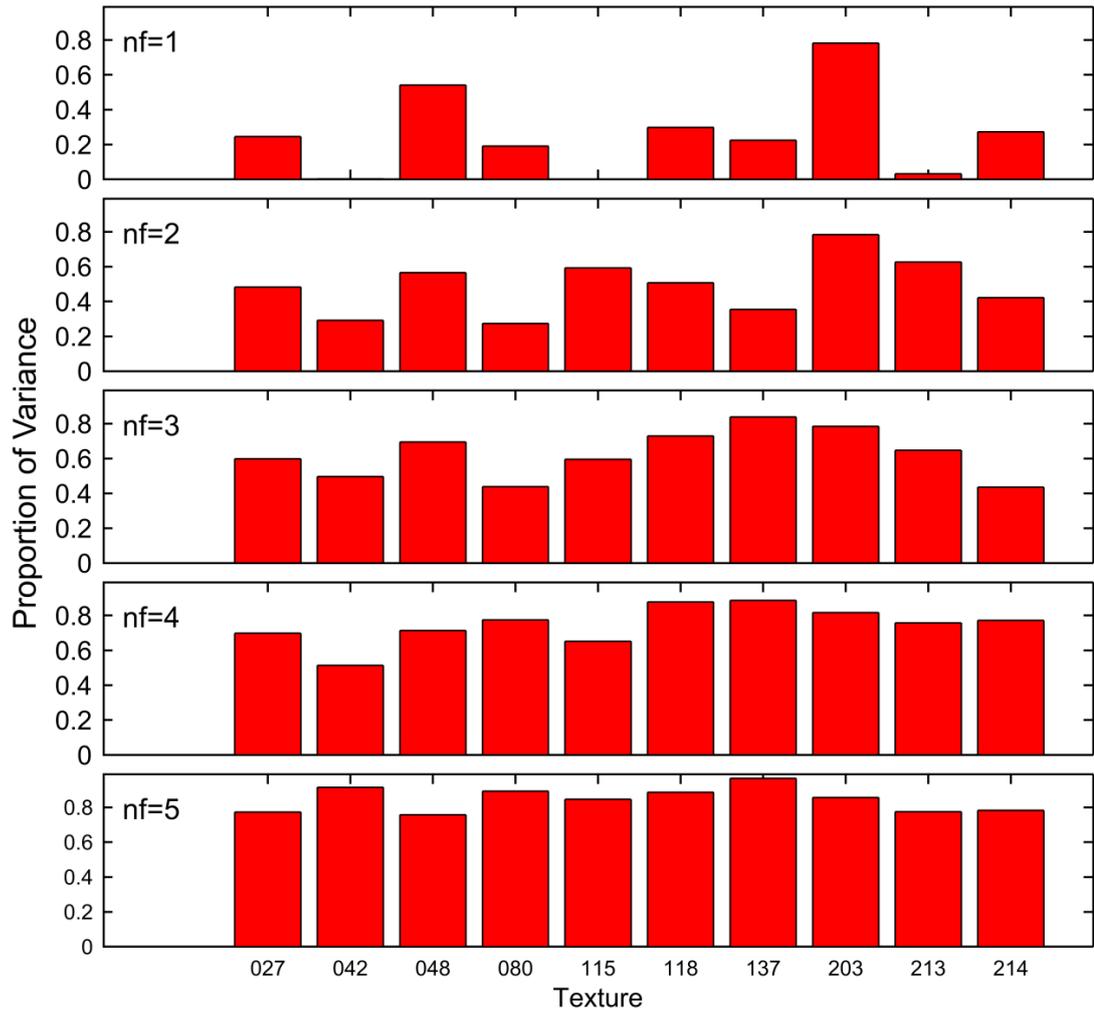
Figure 6. The communalities for five different factor models, beginning with a model including only the largest single factor (nf = 1), to a model containing the five largest factors (nf = 5) for the All16 data set. As the number of factors grows, the profile of bars becomes flatter, indicating that the models progressively account for the data in a more balanced way. The other three data sets were examined and showed many common features, such as texture 042 being less well described until 4 or more factors are included.

A three factor model is sufficient to describe the majority of the data points moderately well. Texture 042 was the least well described by the three factor model in Figure 6. It was also the least well described for 2 of the other 3 data sets (not shown). Across the 4 data sets, the mean (± SD) communality for 042 was 0.59 ± 0.17. For the other textures taken together, the three factor model provided an overall mean communality of 0.60 ± 0.09, the small SD indicating a rather even distribution of communalities. Including 042 in the four data sets, the 3-factor models accounted for 0.54 ± 0.19, 0.58 ± 0.18, 0.63 ± 0.14 and 0.66 ± 0.16 of the variance, for JWS, ABL, ALL16 and All32 respectively.

2.4.4 Factor Loadings

We now address the contributions of the top three factors to the performance of subjects on a given texture (Figure 7). This can be done by examining the so-called *factor loadings,* which are like regression weights between the factors and performance for each texture. Loadings with large non-zero weights indicate the possible form of the putative texture discrimination mechanisms. To highlight the differences and similarities between the four data sets, we compared their raw correlation coefficients. Given that the functions describing the loadings for each factor can be a 10-point function of any shape, significant correlation between those functions strongly indicates that they are real.
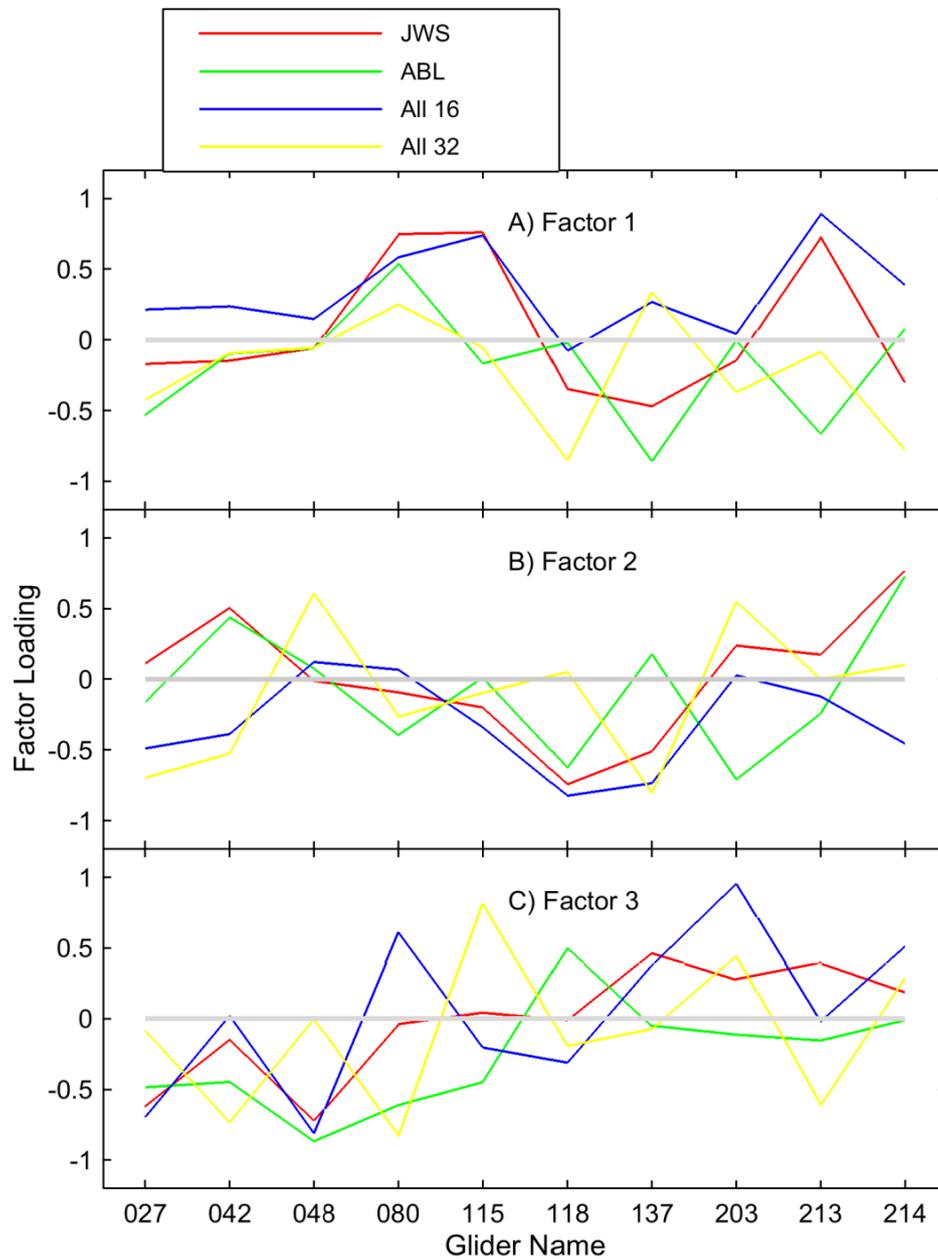
Figure 7. An analysis of the contributions of the top three factors to the performance of each subject. Factor loadings are regression weights between the factors and performance for each texture. These weights can be positive or negative, thus weights near 0 contribute little. This allows us to evaluate what proportion each factor contributes to performance functions (Figure 4) for each texture. The loadings are thus an indication of the relative influences of the putative texture discrimination mechanisms.

In this analysis, factor loadings FL1 to FL3 of data set 1 (JWS) were compared to those of data sets ABL, All16, and All32. A problem with Factor analysis is that the ordering of factors used here is based on the variance accounted for by each factor in each data set. Thus, a small amount of noise might cause the order of factors to be permuted. Furthermore, the sign of the factor loadings is arbitrary. Therefore, we used correlation analysis as a way of determining which factors were most similar between data sets. In Figure 7, we grouped the factor loadings from each data set that had the highest correlations. As expected, this only permuted a few factors by one step.

For Figure 7A the best correlation was between JWS and All16 ($r=0.84$, $p=0.002$). For Figure 7B, the best two correlations were between JWS and ABL or All16 ($r=0.54$ and 0.57; $p=0.11$ and 0.09). For Figure 7C, the best correlation was again between JWS and All16 ($r=0.75$, $p=0.01$), and JWS and ABL were marginally correlated ($r=0.60$, $p=0.06$).

2.4.5 Other textures

Data from a previous study (Maddess and Nagai, 2001) was added (as discussed in Section 2.3). These included the V3L2 textures of Figure 1, but the data for the Box-Even texture had low variance so it was dropped from the Factor analysis. Figure 8 shows the results for 16x16 pixel data from the 17 textures. Figure 8A shows a marked drop in the variance explained after 2 factors, and perhaps a second drop at 4 factors. The communalities for a three factor model were quite flat (Fig. 8B) and explained 0.85 ± 0.11 of the variance (mean ± SD). A four factor model accounted for 0.94 ± 0.06 of the variance for each texture. Results for 32x32 pixel versions were similar.
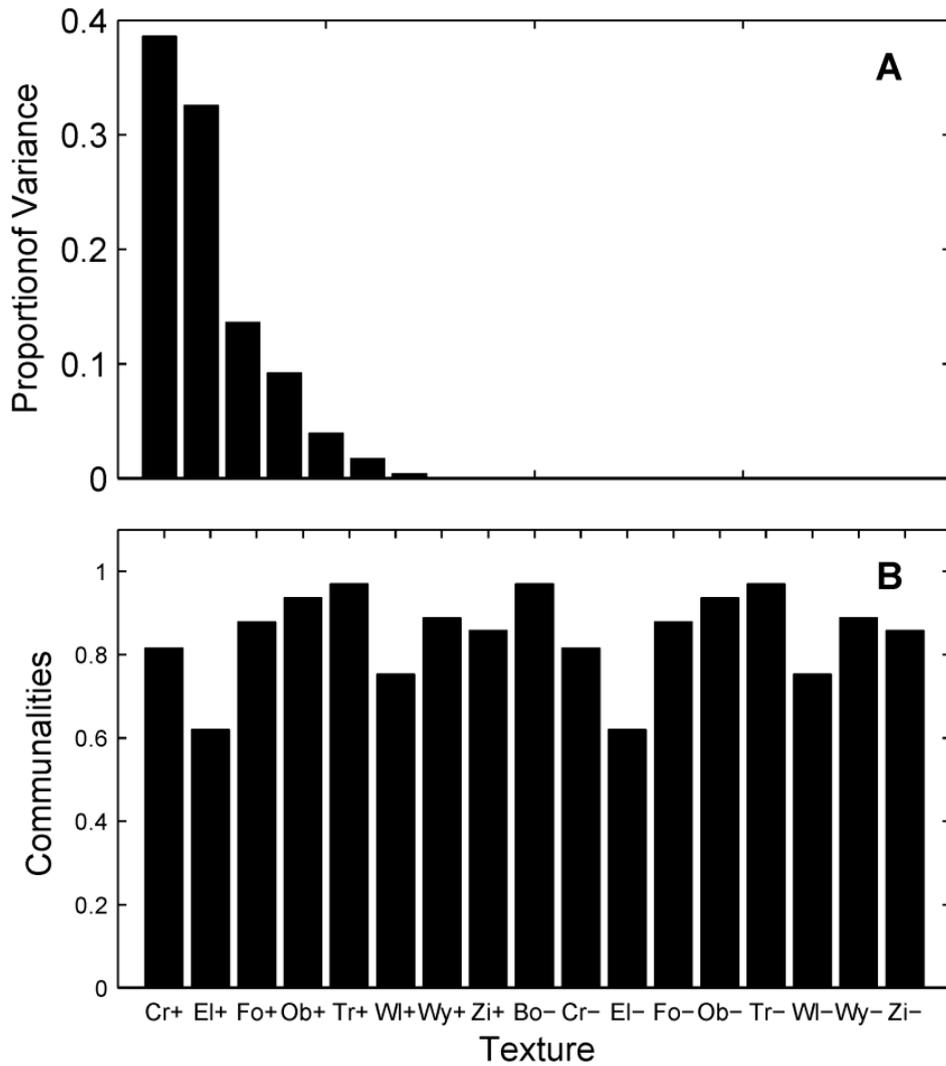
Figure 8. Results from the seventeen V3L2 textures. They are the same textures as those of Figure 1 plus 2 more, but not the Box-Even texture (see Section 2.3). A) Scree plot showing an abrupt drop between the 2nd and 3rd factors. B) the communalities for a 3 factor model account for an average of 0.85 ± 0.11 of the variance (mean ± SD) for discrimination of each texture. On the abscissa labels "+" indicates the Even rule (Figure 1 caption), and "−" the Odd rule; e.g. Cr+ is Cross-Even, and Cr− is Cross-Odd. The order of texture types thus follows Figure 1.

2.5 Discussion

Subjects differed somewhat in their ability to detect image structure defined by fourth- and higher-order spatial correlations (Figure 4). This study supports the possibility that approximately three mechanisms govern higher-order texture discrimination. This is similar to results reported for other isotrigon textures using different methods (Maddess and Nagai, 2001, Maddess et al., 2007). Three factor models gave a relatively balanced set of communalities for the VnL2 textures (Figure 6), and accounted for about 85% of the variance for a separate V3L2 17-texture data set from two subjects (Figure 8). For those V3L2 textures, the lowest communalities were for the Even and Odd EI textures, which interestingly are among the least visually salient of these textures (Maddess and Nagai, 2001, Tkacik et al., 2010, Victor and Conte, 1989, Victor and Conte, 1991); that type of fourth-order correlation rarely occurs in natural scenes (Tkacik et al., 2010).

Texture 042 was the least well explained by the 3 factor model, having the lowest communalities in 3 of the 4 data sets (e.g. Figure 6). As a way of possibly addressing this issue, the Minkowski functionals for the textures were also investigated. These are three combinations of low to 4th order correlations computed for 2x2 pixel blocks of the texture: the Area, Perimeter and Euler Number. On a square lattice, there are actually two Euler numbers based on 4- and 8-way connectedness, χ_4 and χ_8. Linear combinations of the functionals have been suggested to explain the performance data for discrimination of isotrigon textures (Barbosa et al., 2013). In this case, there are some notable differences in the Minkowski functional data across the

textures used here, but there was nothing that uniquely separated texture 042 from the others.

The study of Barbosa et al. (2013) asked if linear combinations of various moments of the Minkowski functionals could explain   human performance functions for 33 types of isotrigon texture, including 25 of those used here. The model with the lowest deviance contained combinations of some of the mean, variance, kurtosis and skew of χ_4 and χ_8. Thus, it is possible that the mechanisms for V2LN and VnL3 discrimination are related, but we have no evidence for that here.

An interesting feature of the Minkowski functionals is that they can be extended to describe 3D materials composed of two components, either two materials (like fibreglass) or a material and voids (like sponge), and in these cases they are related to the strength of the materials. Thus, the Minkowski functionals may be related to the surface properties of some real surface textures and how they inform us about the mechanical properties of the materials (Schroder-Turk et al., 2011, Schroder-Turk, 2010). This concept will be discussed further in the contest of material perception in Chapter 5 (Section 5.2).

In terms of the factor loadings (Figure 7), the consistent correlation between JWS and the 6-person All16 data set was encouraging in that they were two of the largest data sets, having 26 and 16 repeats respectively, and so were the less likely to be affected by noise. Overall the largest correlations between the loading functions were for the data sets for presentations of 16x16 pixel textures. Whether this implies something different is happening for the larger 32x32 pixel examples will require more work. Taken together,

153

the fact that a 3 factor model could produce quite evenly distributed communalities, and that the resulting factor loadings showed significant correlation, indicates that a lower bound of about 3 mechanisms is reasonable. Experiments on a larger set of different textures would be required to verify this preliminary result.

It should be noted that PCA analysis extracts orthogonal (uncorrelated) components. However, it is possible that the true mechanisms for extracting information about 4th and higher order textures may not be orthogonal, and therefore the number of mechanisms may be larger. That is, a larger number of correlated mechanisms could span the same 3D space. Alternatively our 10 textures may not probe all the dimensions of the true set of mechanisms.

The work of Victor et al. (2012) indicates that, for correlations of order 1 to 4, computed within 2x2 blocks of pixels, there is possibly a 10-dimensional space. This can be reduced to 5-dimensions or fewer if rotational symmetry is considered and some of the dimensions show some correlation between them (Victor et al., 2013, Victor and Conte, 2012). That set of correlations only includes a single measure for 4th order correlations, and larger patch sizes, e.g. 3x3 pixels, would be needed to be considered to survey more complex 4th order interactions, which could open up a large number of possible dimensions.

Certainly, there is evidence for salient and non-salient higher order correlations at such scales in isotrigon and natural textures (Tkacik et al., 2010). Previous studies have suggested that the range of salient correlations is probably less than 6 pixels, and a self-similar analysis occurs for a wide range of pixel sizes (Maddess and Nagai, 2001, Victor and Conte, 1989). We

have previously provided evidence that the number of independent mechanisms is less than 10 (Taylor et al., 2008), and is perhaps as small as 3 to 4 (Maddess and Nagai, 2001, Maddess et al., 2007) including different types of isotrigon textures than those used here.

Although we address mechanisms sensitive to 4th and higher order correlations here, it is worth mentioning that elegant studies of interactions between 1st and 4th order (Victor et al., 2005), and 2nd and 3rd order correlations (Victor et al., 2013), have been previously reported. These have implications for the total number of mechanisms involved in texture discrimination more generally. See also related work on the black shot mechanism (Chubb et al., 2004).

Other than JWS, none of the subjects showed significant evidence of learning. One possible feature of interest was a change in the scree plot for ABL with learning (Figure 5). In a previous study, some subjects markedly improved their ability to discriminate V3L2 textures; the increase in probability correct for 16x16 textures was 20% for one and 30% for another texture type in just 5 repeats (Maddess and Nagai, 2001). Learning has also been studied for isotrigon textures that had 3 grey levels (Taylor et al., 2008). The effects were moderate however, with discrimination performance improving little, either after experience or when the textures were presented for several seconds. A recent study by our group, which used a set of 17 V3L2 textures, found strong evidence of learning effects described by an exponential rise with a time constant of about 5 days (Coy et al., 2014). That data also showed that a three factor model was reasonable.

How might fourth-order correlations be computed by the brain? Maddess and Nagai (2001) have proposed cascading two motion computation-like mechanisms to yield the quadruple products required for fourth-order correlations. Nagai and Maddess have shown that simple models of cortical processing, based on recursion, can discriminate isotrigon textures (Nagai et al., 2009). They found that oscillator networks, containing as few as 3 oscillators, were able to emulate human discrimination performance for 53 binary and ternary isotrigon texture types.

The formation of recursively applied products is physiologically plausible and may occur via dendritic back-propagation (Stuart et al., 1997, Buzsaki and Kandel, 1998) or dendritic spiking (Mel, 1993, Stuart et al., 1997, Hausser et al., 2000). An unpublished modelling study from our group has also shown that dendritic back propagation in one pyramid cell is sufficient to discriminate some isotrigon textures from random ones (Taylor, 2013). Overall, the mechanisms identified in this study, and previous studies (Maddess and Nagai, 2001, Maddess et al., 2007), may represent some combination of recursive or rectifying processes.

In conclusion, in this study a novel set of isotrigon textures (VnL2) were used to evaluate the underlying mechanisms of texture discrimination. Subjects discriminated random binary noise patterns from the ten isotrigon textures. Factor analysis revealed that as few as three mechanisms govern the detection of fourth and higher order image structure. The mechanisms identified in this study may represent some nonlinear combination of recursive and/or rectifying processes computed over relatively short ranges.

2.7 Acknowledgments

2.8 References

BARBOSA, M. S., BUBNA-LITIC, A. & MADDESS, T. 2013. Locally countable properties and the perceptual salience of textures. *Journal of the Optical Society of America A,* 30**,** 1687-1697.

BARLOW, H. 2001. Redundancy reduction revisited. *Network,* 12**,** 241-53.

BARLOW, H. B. 1963. The Information Capacity of Nervous Transmission. *Kybernetik,* 2**,** 1.

BEASON-HELD, L. L., PURPURA, K. P., KRASUSKI, J. S., DESMOND, R. E., MANGOT, D. J., DALY, E. M., OPTICAN, L. M., RAPOPORT, S. I. & VANMETER, J. W. 2000. Striate cortex in humans demonstrates the relationship between activation and variations in visual form. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale,* 130**,** 221-6.

BEASON-HELD, L. L., PURPURA, K. P., KRASUSKI, J. S., MAISOG, J. M., DALY, E. M., MANGOT, D. J., DESMOND, R. E., OPTICAN, L. M., SCHAPIRO, M. B. & VANMETER, J. W. 1998a. Cortical regions involved in visual texture perception: a fMRI study. *Brain research. Cognitive brain research,* 7**,** 111-8.

BEASON-HELD, L. L., PURPURA, K. P., VAN METER, J. W., AZARI, N. P., MANGOT, D. J., OPTICAN, L. M., MENTIS, M. J., ALEXANDER, G.

E., GRADY, C. L., HORWITZ, B., RAPOPORT, S. I. & SCHAPIRO, M. B. 1998b. PET reveals occipitotemporal pathway activation during elementary form perception in humans. *Visual neuroscience,* 15**,** 503-10.

BUZSAKI, G. & KANDEL, A. 1998. Somadendritic backpropagation of action potentials in cortical pyramidal cells of the awake rat. *Journal of neurophysiology,* 79**,** 1587-91.

CHUBB, C., LANDY, M. S. & ECONOPOULY, J. 2004. A visual mechanism tuned to black. *Vision Res,* 44**,** 3223-32.

COY, D. G., BARBOSA, M., SEAMONS, J. W. G., & MADDESS, T. 2014. Learning Effects for Discrimination of Isotrigon Textures. (Honours Thesis; Journal Article In Preparation).

FIELD, D. J. 1987. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America. A, Optics and image science,* 4**,** 2379-94.

FRANZ, M. O. & SCHÖLKOPF, B. 2005. Implicit Volterra and Wiener series for higher-order image analysis. *In:* SAUL, L. K., WEISS, Y. & BOTTOU, L. (eds.) *Advances in neural information processing systems.* Cambridge MA: MIT Press.

FREEMAN, J., ZIEMBA, C. M., HEEGER, D. J., SIMONCELLI, E. P. & MOVSHON, J. A. 2013. A functional and perceptual signature of the second visual area in primates. *Nature neuroscience,* 16**,** 974-981.

GILBERT, E. N. 1980. Random Colorings of a Lattice of Squares in the Plane. *SIAM Journal on Algebraic and Discrete Methods,* 1**,** 152-159.

HAUSSER, M., SPRUSTON, N. & STUART, G. J. 2000. Diversity and dynamics of dendritic signaling. *Science,* 290**,** 739-44.

JULESZ, B., GILBERT, E. N., SHEPP, L. A. & FRISCH, H. L. 1973. Inability of humans to discriminate between visual textures that agree in second-order statistics-revisited. *Perception, 2,* 391-405.

JULESZ, B., GILBERT, E. N. & VICTOR, J. D. 1978. Visual discrimination of textures with identical third-order statistics. *Biological cybernetics,* 31, 137-40.

MADDESS, T. & KULIKOWSKI, J. J. 1999. Apparent fineness of stationary compound gratings. *Vision research,* 39, 3404-16.

MADDESS, T. & NAGAI, Y. 2001. Discriminating isotrigon textures [corrected]. *Vision research,* 41, 3837-60.

MADDESS, T., NAGAI, Y., JAMES, A. C. & ANKIEWCZ, A. 2004. Binary and ternary textures containing higher-order spatial correlations. *Vision research,* 44, 1093-113.

MADDESS, T., NAGAI, Y., VICTOR, J. D. & TAYLOR, R. R. 2007. Multilevel isotrigon textures. *Journal of the Optical Society of America. A, Optics, image science, and vision,* 24, 278-93.

MEL, B. W. 1993. Synaptic integration in an excitable dendritic tree. *Journal of neurophysiology,* 70, 1086-101.

NAGAI, Y., TAYLOR, R. R., LOH, Y. W. & MADDESS, T. 2009. Discrimination of complex form by simple oscillator networks. *Network,* 20, 233-52.

PETERZELL, D. H. & TELLER, D. Y. 1996. Individual differences in contrast sensitivity functions: the lowest spatial frequency channels. *Vision Res,* 36, 3077-85.

PETERZELL, D. H., WERNER, J. S. & KAPLAN, P. S. 1993. Individual differences in contrast sensitivity functions: the first four months of life in humans. *Vision Res,* 33**,** 381-96.

PURPURA, K. P., VICTOR, J. D. & KATZ, E. 1994. Striate cortex extracts higher-order spatial correlations from visual textures. *Proceedings of the National Academy of Sciences of the United States of America,* 91**,** 8482-6.

REYMENT, R. & JORESKOG, K. 1996. *Applied Factor Analysis in the Natural Sciences,* Cambridge, UK, Cambridge University Press.

ROSLI, Y., BEDFORD, S. M. & MADDESS, T. 2009. Low-spatial-frequency channels and the spatial frequency-doubling illusion. *Investigative ophthalmology & visual science,* 50**,** 1956-63.

SCHRODER-TURK, G. E., MICKEL, W., KAPFER, S. C., KLATT, M. A., SCHALLER, F. M., HOFFMANN, M. J., KLEPPMANN, N., ARMSTRONG, P., INAYAT, A., HUG, D., REICHELSDORFER, M., PEUKERT, W., SCHWIEGER, W. & MECKE, K. 2011. Minkowski tensor shape analysis of cellular, granular and porous structures. *Advanced materials,* 23**,** 2535-53.

SCHRODER-TURK, G. E., MICKEL, W., KAPFER, S. C. 2010. Minkowski Tensors of Anisotropic Spatial Structure. *arXiv,* 1009.

SEKULER, R., WILSON, H. R. & OWSLEY, C. 1984. Structural modeling of spatial vision. *Vision research,* 24**,** 689-700.

SIMPSON, W. A. & MCFADDEN, S. M. 2005. Spatial frequency channels derived from individual differences. *Vision research,* 45**,** 2723-7.

SRINIVASAN, M. V., LAUGHLIN, S. B. & DUBS, A. 1982. Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London B,* 216**,** 427-59.

STUART, G., SCHILLER, J. & SAKMANN, B. 1997. Action potential initiation and propagation in rat neocortical pyramidal neurons. *Journal of physiology,* 505 (Pt 3)**,** 617-32.

TAYLOR, R. R. 2013. *Neural Computation of Statistical Structure. Chapter 4: A Greater Purpose: Computation of Statistical Structure within Pyramidal Cells.* PhD Thesis, ANU.

TAYLOR, R. R., MADDESS, T. & NAGAI, Y. 2008. Spatial biases and computational constraints on the encoding of complex local image structure. *Journal of vision,* 8**,** 19 1-13.

TKACIK, G., PRENTICE, J. S., VICTOR, J. D. & BALASUBRAMANIAN, V. 2010. Local statistics in natural scenes predict the saliency of synthetic textures. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 18149-54.

VICTOR, J. D. 1994. Images, Statistics, and Textures: Implications of Triple Correlation Uniqueness for Texture Statistics and the Julesz Conjecture. *Journal of the Optical Society of America A,* 11**,** 1680-1684.

VICTOR, J. D. 1995. Isodipole Textures: A Window on Cortical Mechanisms of Form Processing. *In:* PAPATHOMAS, T. V. C., C.; GOREA, A; KOWLER, E (ed.) *Early Vision and Beyond.* The MIT Press.

VICTOR, J. D., CHUBB, C. & CONTE, M. M. 2005. Interaction of luminance and higher-order statistics in texture discrimination. *Vision Res,* 45**,** 311-28.

VICTOR, J. D. & CONTE, M. M. 1989. Cortical interactions in texture processing: scale and dynamics. *Vis Neurosci,* 2**,** 297-313.

VICTOR, J. D. & CONTE, M. M. 1991. Spatial organization of nonlinear interactions in form perception. *Vision research,* 31**,** 1457-88.

VICTOR, J. D. & CONTE, M. M. 1996. The role of high-order phase correlations in texture processing. *Vision Res,* 36**,** 1615-31.

VICTOR, J. D. & CONTE, M. M. 2005. Local processes and spatial pooling in texture and symmetry detection. *Vision research,* 45**,** 1063-73.

VICTOR, J. D. & CONTE, M. M. 2012. Local image statistics: maximum-entropy constructions and perceptual salience. *J Opt Soc Am A Opt Image Sci Vis,* 29**,** 1313-45.

VICTOR, J. D., THENGONE, D. J. & CONTE, M. M. 2013. Perception of second- and third-order orientation signals and their interactions. *J Vis,* 13**,** 21.

ZETZSCHE, C. & NUDING, U. 2005. Nonlinear and higher-order approaches to the encoding of natural scenes. *Network,* 16**,** 191-221.

Chapter 3: Developing and Validating an Isotrigon Texture Discrimination Task using Crowdsourcing

3.1 Abstract

Humans and primates are visually sensitive to the higher order correlations. One way to probe the sensitivity of the human visual system to higher order correlations is to use isotrigon textures. Isotrigon textures have the same structural features that create visual salience in natural images. Laboratory studies using isotrigon textures have found the lower bound of the independent neurological mechanisms which underpin texture discrimination to be approximately 3. However, these have utilised relatively few subjects.

We describe the development of a texture discrimination task for use with the crowdsourcing platform Amazon Mechanical Turk. We show the method to be robust across a variety of platforms. Texture discrimination data was then gathered from 121 subjects and compared to two independent laboratory data sets. Based on Pearson's correlation and coefficients of repeatability, mTurk is capable of producing laboratory quality data. Factor analysis was also performed and indicated the presence of 3 to 4 independent texture discrimination factors, consistent with previous published studies. This is encourging considering the variety of browsers, screen sizes, screen DPIs and OS recorded.

Due to perceived technical limitations, few visual psychometric studies using mTurk have been published. This study fulfils three important roles: it is an evaluation of crowdsourcing as a platform for visual psychometric research;

a guide for developing and testing such studies; and an analysis of binary isotrigon texture discrimination using a large number of unsupervised, naïve subjects. With the necessary safeguards, mTurk is a promising, and underutilized, platform for visual psychometric research at reduced cost and increased scale.

3.2 Introduction

Research into the responses of single striate cortical neurons has concentrated on two-point correlation properties, as captured by spatial frequency and orientation tuning. Nonetheless, humans and primates are sensitive to the higher order spatial correlations; this has been demonstrated in a number of psychophysical (Maddess and Nagai, 2001, Maddess et al., 2007, Victor and Conte, 2005, Seamons et al., 2015), VEP (Victor and Conte, 1991), fMRI (Beason-Held et al., 1998a, Beason-Held et al., 2000), PET (Beason-Held et al., 1998b) studies, and single cell recordings (Purpura et al., 1994, Freeman et al., 2013).

One way to probe the sensitivity of the human visual system is to use artificially generated textures, such as the i*sotrigon textures*. The word "isotrigon" derives from the fact that the average first- to third-order correlation functions of these textures is zero, as is the case for uniform noise. The salient structure in these textures is therefore exclusively due to fourth- and higher-order spatial correlations (Maddess et al., 2004, Maddess et al., 2007, Caelli et al., 1978, Victor, 1994, Gilbert, 1980). In order to discriminate a particular isotrigon texture from noise, it is necessary to

identify its complex, higher-order structure. Thus, isotrigon textures cannot be discriminated based on luminance or other lower-order properties. Although artificially generated, isotrigon textures have the same structural features that create visual salience in natural images (Tkacik et al., 2010). Therefore, they are an excellent tool for evaluating human texture perception.

Humans use a number of neurological mechanisms to capture visually salient higher order structure. Evidence has been presented that the number of independent mechanisms is certainly less than 10 (Taylor et al., 2008, Barbosa et al., 2013), and is possibly as low as 3 to 4 (Maddess and Nagai, 2001, Maddess et al., 2007, Seamons et al., 2015). One statistically well principled method is to use Factor analysis of human psychometric functions from many individuals, or many repeats for single individuals, to infer the number and form of underlying independent mechanisms (Rosli et al., 2009, Maddess and Kulikowski, 1999, Sekuler et al., 1984, Simpson and McFadden, 2005). In this study, we will use both modalities to develop and implement an isotrigon texture discrimination task using the crowdsourcing platform *Amazon Mechanical Turk* (mTurk). This is one of the few studies of human visual performance using mTurk.

Launched in 2005, mTurk has developed into the largest crowdsourcing platform (see (Paolacci et al., 2010, Mason and Suri, 2012) for reviews on using mTurk to conduct experiments). It now has in excess of 200,000 Workers registered from over 100 countries (Ross et al., 2010, Pontin, 2007). Anyone with internet access can register to use mTurk as a Requester or a Worker. Requesters create *Human Intelligence Tasks* (HITs) which Workers

are paid to complete (Paolacci et al., 2010, Mason and Suri, 2012). At the time of writing the number of active HITs on mTurk was 403,336 (AWS, 2014).

MTurk has several advantages over traditional methods of subject recruitment and study implementation. It provides access to a large, established workforce, which facilitates rapid recruiting (Ross et al., 2010, Paolacci et al., 2010) and provides an established technical infrastructure with an integrated payment system (Mason and Suri, 2012, Goritz et al., 2008). Quality control measures, such as catch trials and qualification tasks, can be incorporated into HITs (Paolacci et al., 2010, Kittur et al., 2008, Heer et al., 2010). MTurk Workers are slightly more demographically diverse than standard Internet samples and significantly more diverse than US college samples (Buhrmester et al., 2011, Gosling et al., 2004). Therefore, mTurk allows researchers to access subjects that would be difficult to access by other means.

One of the greatest advantages offered by mTurk is its low cost. Recent research found that Workers have a reservation wage (the minimum pay rate for which they would complete a HIT) of $1.38 USD per hour (Ipeirotis, 2010a, Mason and Suri, 2012, Buhrmester et al., 2011). Compensation rates appear to have less effect on data quality than on the *rate* of data collection (Buhrmester et al., 2011, Paolacci et al., 2010, Mason and Watts, 2009). In addition to cost savings and reduced recruiting effort, crowdsourcing via mTurk can scale to levels which would be prohibitive in a laboratory setting.

A concern with a novel platform such as mTurk is whether the data derived is of high quality. Issues such as demographic variations, subject motivation

166

and subject expertise apply to many mTurk studies and have been well studied (Ipeirotis, 2010b, Kittur et al., 2008, Mason and Suri, 2012, Ipeirotis, 2010a). Several authors have successfully used mTurk to replicate laboratory studies (Paolacci et al., 2010, Horton et al., 2011, Crump et al., 2013). However, there is a lack of published mTurk studies involving visual psychometric testing. Therefore, issues specific to visual studies, such display resolution and physical size, ambient lighting, subject viewing distance and angle, have not been studied extensively in the mTurk context. Crowdsourced visual perception experiments lack control over many of these conditions. Furthermore, instructional guidelines may not necessarily be followed (Heer et al., 2010, Buhrmester et al., 2011, Crump et al., 2013, Oppenheimer et al., 2009). It is therefore important to explore how variations in working conditions affect the data derived from crowdsourced visual studies.

At the time of writing only three papers were identified which used mTurk to administer tasks of a visual perceptual nature (Heer et al., 2010, Cole et al., 2009, Freeman et al., 2013). Cole et al. recruited 550 mTurk Workers to evaluate 3D line drawings and indicate surface normals. They gathered 275,000 observations and used them to rate rendering techniques. However, the data gathered was not compared to supervised laboratory data and collection statistics were not reported (Cole et al., 2009). Heer et al. (Heer et al., 2010) carried out a series of graphical perception experiments, including a replication of laboratory alpha contrast experiment (Stone and Bartram, 2008).

Freeman et al. analysed the single-unit responses of macaque V1 and V2 neurons to synthetic stimuli which replicated the higher-order structure of natural textures (Freeman et al., 2013). The process of generating these stimuli is interesting. Starting with photographs of textures taken from nature, two sets of stimuli were derived: spectrally matched noise and "naturalistic" textures. The latter were synthesised using an optimization process (gradient descent) which utilized the image statistics of the original photograph (Portilla and Simoncelli, 2000, Heeger and Bergen, 1995). The authors report that the resulting naturalistic textures have the same overall orientation and spatial frequency content as the original photographs, but lack their higher-order statistical dependencies (Freeman et al., 2013).

Freeman et al. carried out an mTurk study in which perceptual sensitivity functions to naturalistic textures were calculated (Freeman et al., 2013). They report gathering over "300 hour of behavioural data from thousands of human observers" (exact numbers not reported). Each Worker was paid $0.40 USD for ~5 minutes of work ($4.80 USD per hour). Neither demographic data nor platform data were reported, and the authors did not report using catch trials or qualification tasks (Freeman et al., 2013).

An innovative crowdsourced study was also recently conducted by Mitroff et al. (Mitroff et al., 2015). They reported using a mobile App called Airport Scanner to gain insights into visual search. Standalone Apps represent a viable alternative to crowdsourcing platforms in some instances, but at increased technical challenge and with a limited capacity to financially compensate users (Mitroff et al., 2015).

A previous laboratory study has explored the effect of contrast and pixel size on isotrigon discrimination. Maddess et al. carried out texture discrimination experiments using 18 isotrigon textures at several contrast levels and pixel sizes. They found the average performance to be consistent across all conditions (Maddess and Nagai, 2001). The isotrigon texture discrimination experiment is thus believed to be resistant to contrast and pixel size variations. Nonetheless, all HITs must be carefully tested before they are deployed and to allow this Amazon provides a Sandbox environment.

This study is divided into two parts. Firstly, we will discuss the development of the texture discrimination HIT, including a pilot study within the non-public Sandbox environment. These experiments used mTurk, but were conducted by laboratory members under more controlled conditions and are referred to as *Lab* experiments. We will discuss what was learned from the pilot study, how the HIT was adapted and prepared for use on the live mTurk site. Secondly, we will analyse the data derived from the live study (referred to as *Live* experiments). Therefore, this paper will fulfil three roles: a guide to developing, testing and implementing visual psychometric studies using mTurk; an evaluation of mTurk as a platform for visual psychometric research; an analysis of texture discrimination using a large number of unsupervised, naïve subjects. The mTurk experimental modality developed in this Chapter also lays the foundations for a larger series of mTurk experiments, using ternary textures, in Chapter 4.

3.3 Materials and Methods

3.3.1 Subjects

The ethical aspects of this research have been approved by the ANU Human Research Ethics Committee under two ANU Human Ethics protocols: one for the Lab experiments and one for the Live experiments (2010/194, 2014/237). As it was a joint experiment, a mirror protocol for the Live Experiments was also approved at Weill Cornell Medical College (0904010359-A003). The experiments were conducted in accordance with the Declaration of Helsinki.

All subjects were instructed to download and read a subject consent form before commencing the study. All subjects were instructed that, if they did not agree to the terms outlined in the subject consent form, they should not participate in the study. Subjects wishing to participate in the study then indicated their consent via a button click. We confirm that the named ethics committees detailed above specifically approved this study protocol.

3.3.2 Textures

This study utilized 16 common types of isotrigon textures (Figures 1 and 2) (Maddess et al., 2004, Maddess et al., 2007, Caelli et al., 1978, Victor, 1994, Gilbert, 1980).
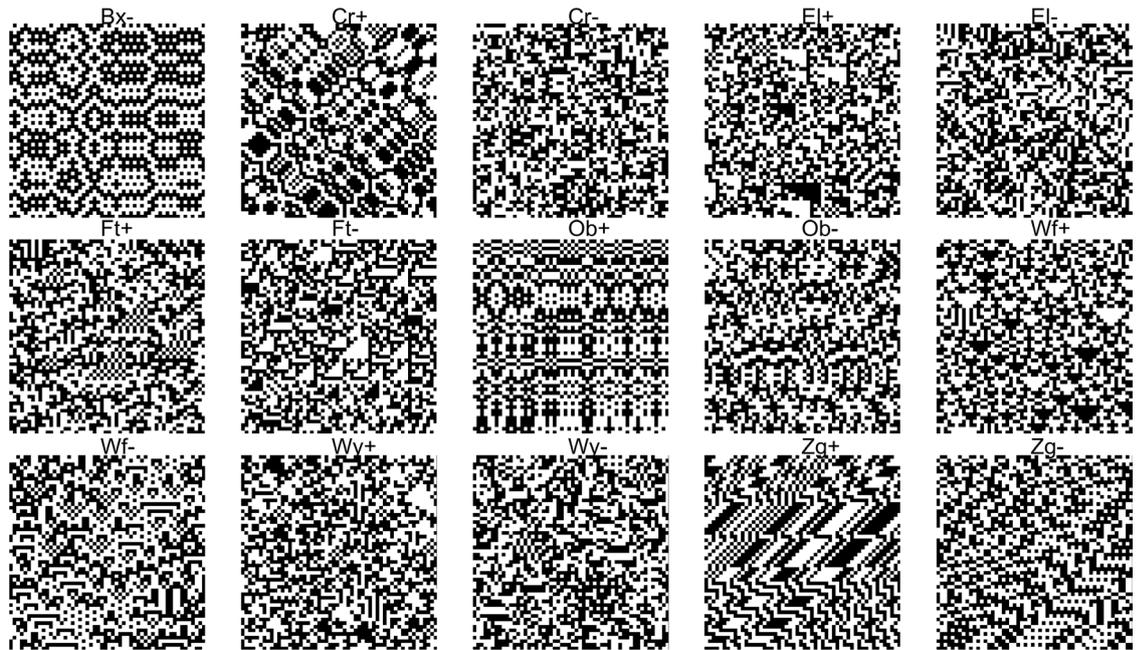
Figure 1. The 15 binary isotrigon textures used in the HIT (Box Even (Bx+) has been excluded, but is shown in Figure 2). The following abbreviations are used throughout the text: Box Odd (Bx-), B: Cross Even (Cr+), Cross Odd (Cr-), El Even (El+), El Odd (El-), Foot Even (Ft+), Foot Odd (Ft-), Oblong Even (Ob+), Oblong Odd (Ob-), Wolf Even (Wf+), Wolf Odd (Wf-), Wye Even (Wy+), Wye Odd (Wy-), Zigzag Even (Zg+), Zigzag Odd (Zg-).

3.3.3 Amazon Mechanical Turk HIT

An external HIT was developed, which included an isotrigon texture discrimination task. Whereas internal HITs are quick and easy to develop, use Amazon templates, and are hosted on Amazon servers, they lack the advanced features that we required. External HITs are much more flexible and may include text, images, movies, and even interactive Flash or Java applications (Heer et al., 2010, Paolacci et al., 2010, Mason and Suri, 2012).

171

Amazon S3 was used to host the binary isotrigon texture repository and other associated files.

The HIT itself begins with detailed instructions on how the isotrigon texture discrimination task proceeds (a screen capture of the HIT page is shown in Appendix 6.1). It includes a slideshow containing examples of the binary isotrigon textures for priming subjects. The HIT also includes a monitor calibration step which records the DPI of the Worker's screen and dictates the required viewing distance that should be used during the task. Workers are informed that they should complete the calibration at least once, assuming their work platform remains consistent. Other information such as the operating system, browser, and anonymous Worker ID are automatically captured.

On commencing the texture discrimination task, Workers are shown an image and are required to indicate the location of a distinct band. This can either be a background of texture with a foreground band of binary noise (this condition is referred to as "Foreground") or vice versa ("Background" condition) (Fig. 2). This method was developed by Victor et al. (Victor et al., 2013) and is intended to provide a more complex and naturalistic stimuli than those traditionally used in visual perception experiments using gratings. As in natural stimuli, local features have multiple orientations, and multiple orientations can occur even at the same location. Moreover, orientation cues can arise not only from pairwise spatial correlations, but from higher-order ones as well (Victor et al., 2013). There are four possible band positions, which were selected at random, as shown in Figure 2, i.e. all experiments were four-alternative forced choice. A secondary consequence of using two

conditions is it discourages fixation on single corner position, which could potentially result in performance of greater than 25% (greater than chance).
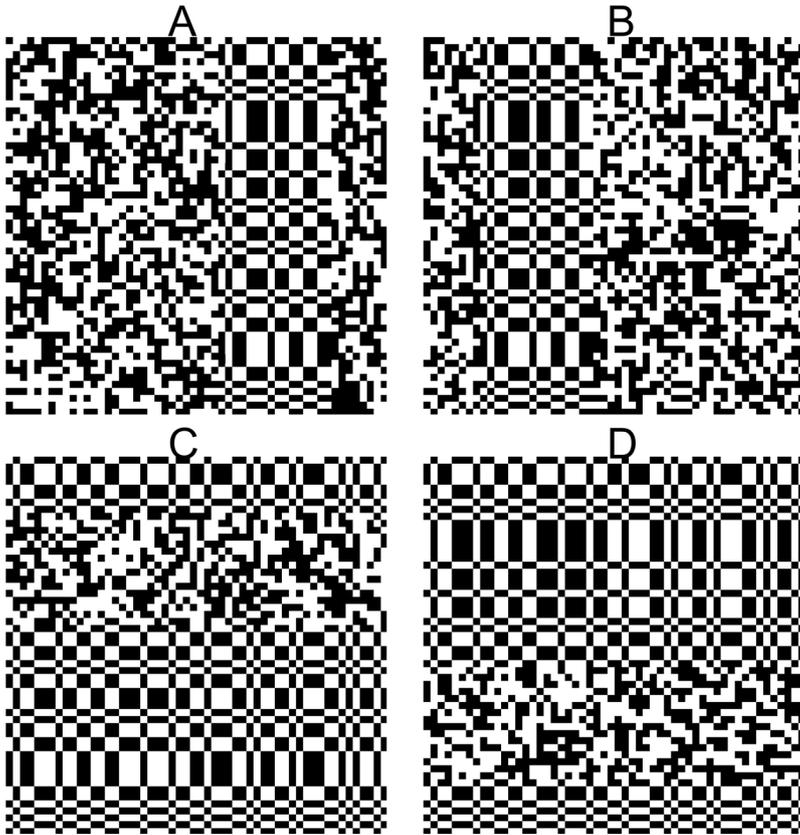

Figure 2. The four possible band positions and conditions used in the HIT. Top left: right band, foreground condition. Top right: left band, foreground condition. Bottom left: top band, background condition. Bottom right: bottom band, background condition. All examples are shown using the Box Even (Bx+) Texture.

All texture production and analysis was completed using MatLab (The Mathworks, Natick, MA). The isotrigon textures were created using a recursion procedure in which a combinatorial rule operated on *n* pixels selected by a gliding template, or *glider*, determines the value of an output pixel (Victor and Conte, 1991, Maddess and Nagai, 2001). During the HIT,

the textures were presented in a random order. The 16 isotrigon textures chosen had been previously used in supervised laboratory discrimination tasks (Maddess and Nagai, 2001, Seamons et al., 2015, Coy, 2014). Because the Box Even stimuli are the most salient, they were used as *catch trials* - salient texture discriminations which Workers were expected to pass, and therefore could be used to test their attentiveness and detect false negatives (Fig. 1 and 2).

The stimuli were presented for 2,000 ms. Subjects used a mouse to indicate the side containing the band and visual feedback was provided as to the veracity of their choice. In the initial Lab experiments and the first of two large crowdsourced experiments (Live1) the HITs contained a *Missed* button, which allowed Workers to skip a texture if they were distracted during its presentation (in which case the missed texture was reinserted at the end of the texture queue). The Missed button was removed from the second of two large crowdsourced experiments (Live2) to examine the effect of its absence. Each HIT utilised a single type of isotrigon texture of those discussed above. Each HIT lasted a maximum of 5 minutes and comprised 20 texture presentations. The four Box-Even catch trials were also included and distributed at random.

3.3.4 Testing on Different Platforms

In our initial Lab experiments, six different platforms, machines 1 to 6 (M1 to M6) were chosen on which to test the HIT. Monitor luminance and other physical characteristics of the platforms were measured. Luminance was

measured in candelas per m^2 (United Detector Technology S370) Luminance was measured in an otherwise darkened room and a white area of the Amazon Worker Sandbox page was used as the measurement target (AWS, 2014). An ambient luminance reading was taken from a neutral surface (a white wall) adjacent to the work area. In all cases, for a given machine, luminance readings were kept consistent for all repetitions. The characteristics of the six platforms are summarised in Table 1 (Appendix 6.2).

Six copies of the HIT were uploaded to the Amazon Sandbox environment for testing, one for each machine (M1-6). One subject (JWGS) completed 3 repetitions (rep0, rep1, rep2) of 15 HITs on each of the six different machines (for a total of 270 HITs). Therefore, each repetition included examples of all 15 isotrigon texture families on each platform, plus catch trials. Subject JWGS had corrected to normal vision and was assumed to be representative of the general population, albeit more experienced, in terms of texture discrimination.

Following analysis of that data, a number of minor modifications were made to the HIT before proceeding to the Live study. The Live1 and Live2 HITs were rolled out in two phases. The second phase incorporated minor modifications as detailed below (Section 3.3.6).

3.3.5 Statistical Techniques

Factor analysis was performed on the Lab and Live data sets. Factor analysis is a family of statistical methods whose goal is to identify the underlying relationships between variables within a data set (Reyment, 1996, Norris and Lecavalier, 2010). Before Factor analysis, all performance scores were transformed to d' units (d-Prime), which accounts for performance saturation for the easier textures (Macmillan and Creelman, 1991, Seamons et al., 2015); these methods have been published in detail (Seamons et al., 2015) and discussed in Chapter 2.

Once factors have been extracted, decisions must be made regarding the optimal number of factors to retain. In this study we primarily used the proportion of variance for each texture explained for a N-factor model (the communalities), which is accurate and reliable (Cattell, 1966, Hayton et al., 2004, DeVellis, 2012, Zwick and Velicer, 1986). After principal factors have been identified, they may be used to compute a reconstructed correlation matrix, factor loadings and factor scores (Reyment, 1996). The factor scores are regression weights indicating the contribution of each of the factors to the detection of each texture type. The factor loadings are similar weights, but indicate the contribution of each factor to each repeat or variables, such as the different platforms used. In our recent study using the same methods, but different isotrigon textures, we found evidence of 3 to 4 independent mechanisms (Seamons et al., 2015).

A central theme of this study is repeatability or reliability. Lab and Live performance data will be evaluated using relative and absolute test-retest

reliability measures. These include the relative technique, Pearson's product moment correlation coefficient (r) (Kirk, 2007). However, strong correlation does not necessarily mean that there is good agreement between test-retest scores. Two measures may be highly correlated, but vary considerably in terms of absolute values (Bland and Altman, 1986, Bland and Altman, 1999, Bland, 2000, Vaz et al., 2013). Therefore, we will also use the absolute reliability measure, coefficient of repeatability (CR). The CR is the value below which the absolute differences between two measurements lie within 0.95 probability (Bland and Altman, 1986, Bland and Altman, 1999, Bland, 2000, Vaz et al., 2013). Coefficients of reproducibility are commonly derived from Bland-Altman plots.

3.3.6 Modifications of the HIT

As discussed below in Section 3.4.3, the Lab study did not demonstrate a significant difference in texture discrimination performance between the Foreground and Background conditions (Victor et al., 2013). The incorporation of both conditions is believed for increase the difficulty of the task and prevent subjects from fixating on border regions (Victor et al., 2013). Therefore, a decision was made to use the Foreground condition only in the Live study phases. Simplifying the experimental protocol may be also beneficial when working with naïve subjects.

A small difference in discrimination performance was observed based on band position (Section 3.4.3). Therefore, the HIT was altered so that it

177

always contained equal proportions of each of the 4 band positions. A repeat analysis of this effect was performed using the Live1 data.

The HIT instructions were also shortened and clarified. It is extremely important that Workers read and understand the HIT instructions if they are to produce high quality data (Crump et al., 2013). To further clarify the user interface, a blank holding page was introduced between texture presentations (in the original HIT, a binary noise image was used). The visual feedback images were changed to a simple tick and cross, again to make them more readily understandable. Providing feedback to Workers is a factor in maintaining data quality (Dow et al., 2012).

The catch trials were changed to Box-Odd (Figure 2). Box-Even, as used in the Lab study, was considered too easy to encourage attentiveness in Workers. Therefore, it was exchanged for the slightly more challenging Box-Odd texture. The slideshow of texture examples was also enlarged to facilitate more effective priming of subjects.

A bonus payment system was implemented to encourage Workers to complete HITs for each of the 15 texture families. The bonus was set at $0.80 AUD. Each HIT takes approximately 1 minute to complete. Therefore, if the Worker completes all 15 HITs (one for each texture family), the Worker will earn $1.50 AUD plus the $0.80 AUD bonus ($2.30 AUD) in 15 minutes. This equates to an hourly rate of $9.20 AUD. Although recent research found that Workers have a reservation wage of $1.38 USD per hour, we took the position on the ethical grounds that the compensation rate should be consistent with that of conventional laboratory subjects (Ipeirotis, 2010a,

Mason and Suri, 2012). A link to a PDF of a subject consent form was implemented, as consistent with the ethical approval obtained.

An exploit which allowed Workers to rapidly cycle through the HIT was identified and removed. Cross platform testing was performed and a number of implementation faults removed, including those specific to Mac Firefox (32.03) and older versions of Internet Explorer. An image caching strategy was also applied to ensure that the times of texture presentation were consistent and not limited by network speed.

3.4 Results: Lab Phase

3.4.1 Lab Testing

3.4.2 Texture Discrimination Performance by Machine

The performance data for the machines M1 to M6 (all three replications) was grouped and plotted in Figure 3. Based on the combined data set, the textures with highest median performance were Oblong-Even (85.7%), Zigzag-Even (95.2%), and Box-Odd (100%). The textures with the lowest median performance were Wye-Even (50.0%), Foot-Odd (47.6%), and Wye-Odd (50.0%). The remaining textures had performance scores which clustered around of 60-80%. Across all 15 textures, the average median performance was 68.4%.
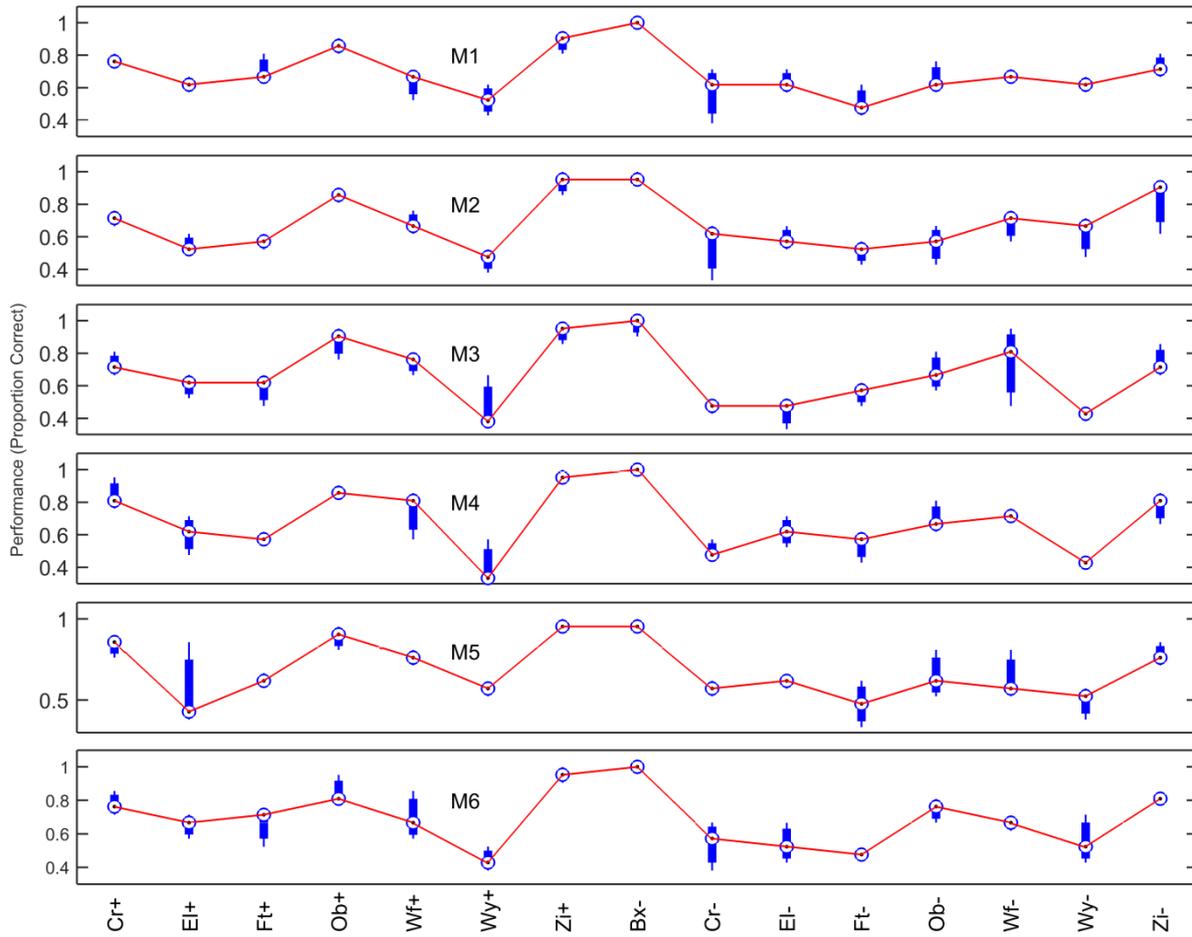
Figure 3. Texture discrimination performance (probability correct) by machine type. Median performance is shown in red. The blue boxes indicate the 25th and 75th percentiles.

To begin to analyse the differences in performance between machines, the median performance for machines 1 to 6 were co-plotted (figure not shown). The trends of texture discrimination performance across all 6 machines were superficially very similar. The largest differences in performance occurred for the textures El-Even (M6 > M5 by 23.8%), Wye-Even (M5 > M4 by 23.8%), Wolf-Odd (M3 > M5 by 23.8%), Wye-Odd (M2 > M4 by 23.8%). The smallest differences in performance occurred for the textures Zigzag-Even (M6 > M5 by 4.8%) and Box-Odd (M5 > M4 by 4.8%). Across the 15 textures, the mean

difference between the highest and lowest performance for any given machine was 15.6%. To further evaluate the differences in texture discrimination performance between machines, colorrmaps were produced (Figure 4).
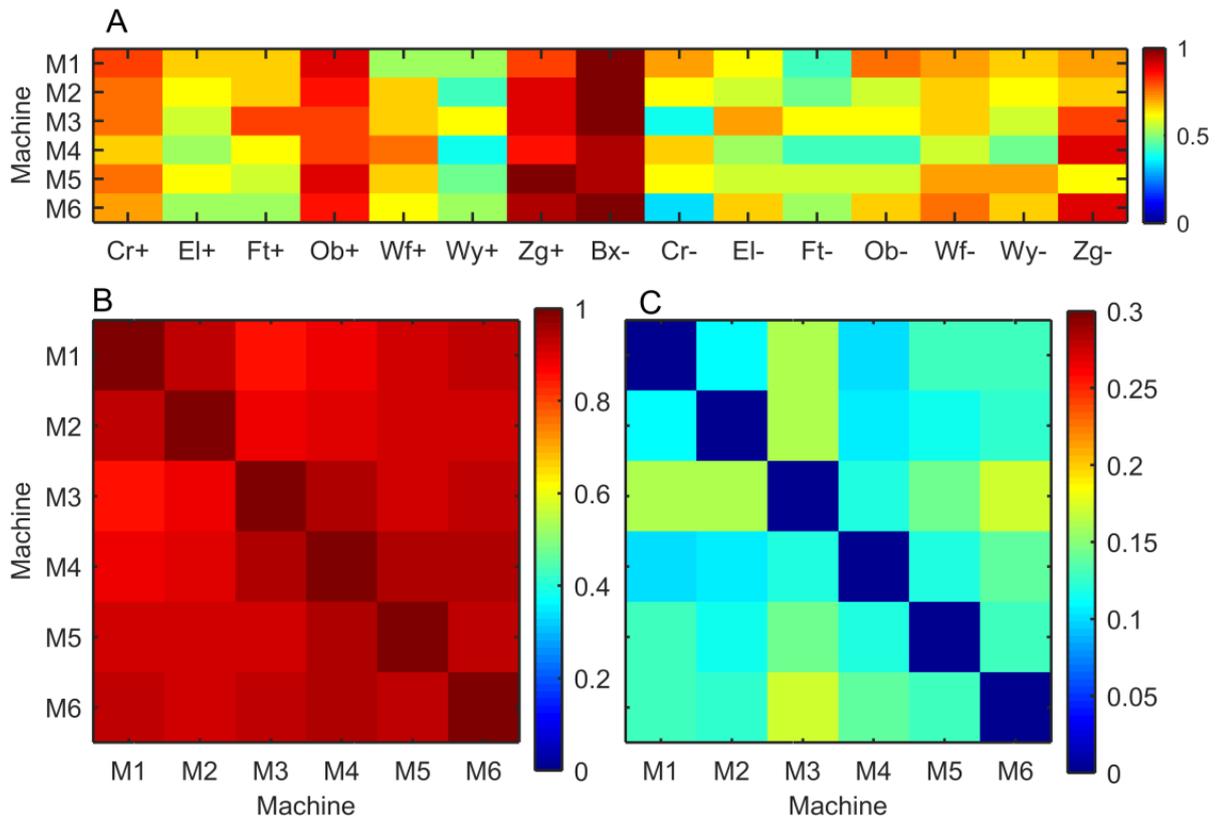

Figure 4. Colourmaps derived from the M1-M6 data set. A: Colourmap showing median texture discrimination performance (probability correct) by machine type. Colourmap showing the correlations (B) and the coefficients of repeatability (C) from Bland-Altman plots between median texture discrimination performances for each machine.

Machines 2, 3 and 4 had the lowest mean performance of the six platforms (67.0%, 66.2%, and 68.9% respectively). Machine 2 was a netbook, machine 3 was a Dell Desktop (768x1024 monitor) and machine 4 was an iPad Mini

(1920x1200 monitor). See Table 1 (Appendix 6.2) for detailed information on the properties of the six machines tested. The effect of physical screen size is most noticeable for textures Cr-, El- and Wy-. These machines represent the *extremes* of screen size evaluated in this study. Log median performance data was also analysed (figure not shown). If the difference in performance between machines was multiplicative, the log mean performance data would exhibit a constant shift. This did not appear to be the case.

Machines 1, 2 and 3 had the lowest inter-machine correlations of the 6 platforms tested. Nonetheless, the inter-machine correlations were universally high; the mean inter-machine correlation for machine 1 was 90.1%, machine 2 90.9%, and machine 3, 90.3%. The highest mean inter-machine correlation was for machine 6 (93.3%). The lowest were between machine 3 and machine 1 (85.1%), machine 3 and machine 2 (87.7%), machine 4 and machine 1 (89.0%), and between machine 4 and machine 2 (90.5%). The inter-machine correlations of log median performance were also analysed, but were not significant different, suggesting that a multiplicative relationship between monitor size and level of performance was not present.

All combinations of machines were analysed using Bland-Altman plots (an example Bland-Altman plot is included in Appendix 6.3) and coefficients of repeatability were calculated. Across all 6 machines, the coefficients of repeatability (CR) were *low*. The mean CR across all 6 machine types was 13.1%, which indicates *good repeatability*. The lowest CR between individual machines was 10.1% (machines 1 and 4). The highest was 17.3% (machines 3 and 6). The machine with the lowest mean CR when paired with the

remaining 4 machine types was machine 3 (9.1%); the highest was 12.6% for machine 1. Factor analysis was also performed using the Lab data and Factor scores were plotted (see Appendix 6.4). This confirmed that machine-specific features had little impact on binary isotrigon discrimination performance.

3.4.3 Analysis of Foreground versus Background Conditions and Band Positions

As shown in Figure 2, two texture conditions were used in the testing phase of this study: Foreground and Background. To evaluate the significance of these two conditions on texture discrimination performance, the mean performance data for the two conditions was separated and the median performance and interquartile ranges (IQRs) plotted in Figure 5.

Across all 15 textures, the mean difference in median performance between the Foreground and Background conditions was 0.077. The largest differences in median performance between the Foreground and Background conditions were observed for textures Cr+ (difference in median performance of 0.301) and Zi- (0.221). The smallest differences in median performance were observed for textures Bx- (0) and Fo- (0). Apart from the Cr- and Zi+ textures, the trends in median performance were very similar between the two conditions across the 15 textures.
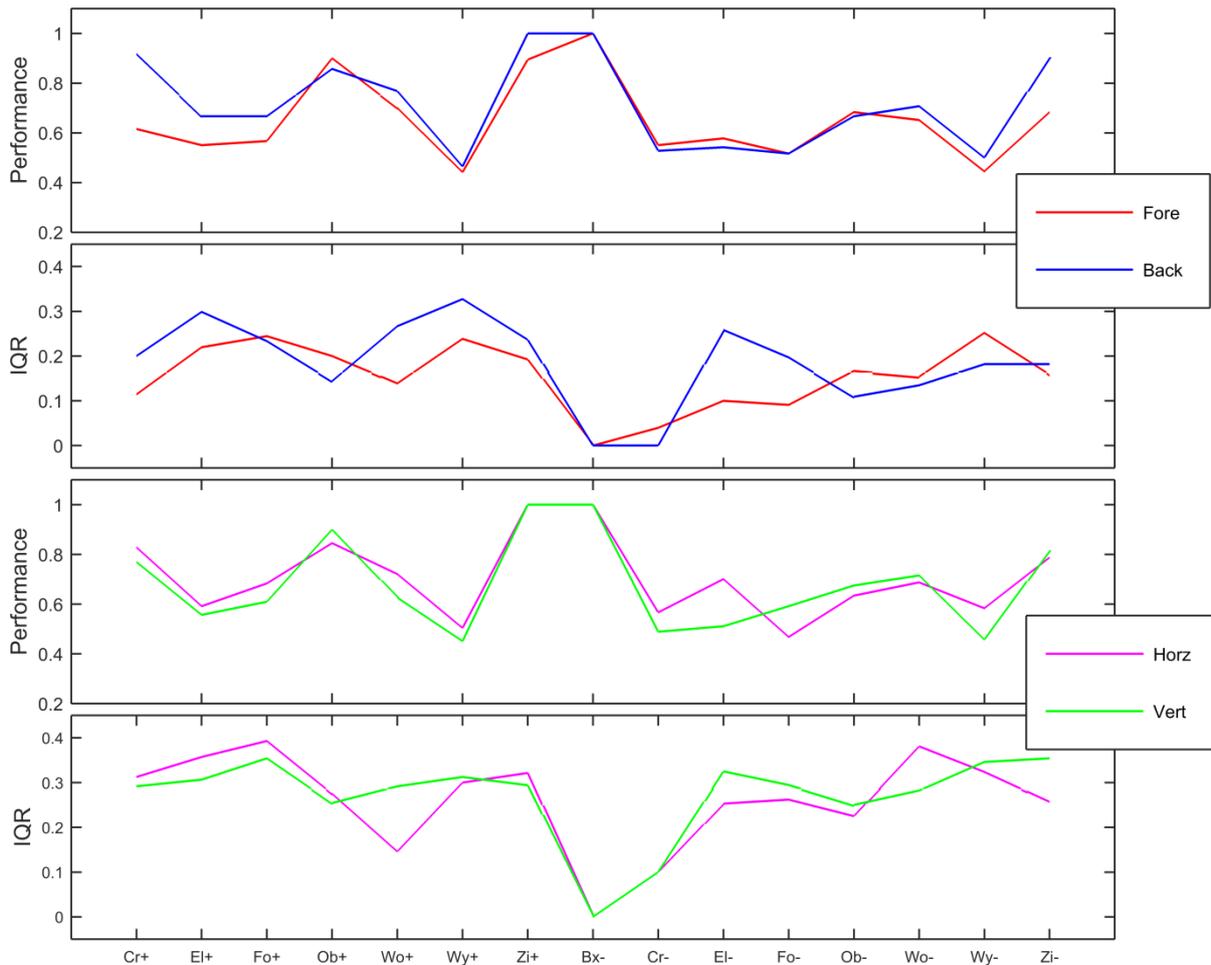
Figure 5. Plots of median texture discrimination performance and interquartile range (IQR) according to texture condition (top 2) and band position (bottom 2). The texture condition plots show performance data from trials using Foreground (Fore) versus Background (Back) conditions (as defined in Fig. 2). The band position plots show performance data from trials using horizontal (Horz) versus vertical (Vert) bands.

The interquartile ranges (IQRs) for the Foreground versus Background conditions were also very similar. The mean difference in the Foreground and Background variances was 0.064 across all textures. The largest differences in variance were observed for textures El- (difference in IQR of

0.158) and Wo+ (0.128). The smallest differences in variance were observed for textures Bx- (0) and Fo+ (0.011). Based on median performance and IQR, the differences between the Foreground and Background conditions are not significant.

The effect of band position was also analysed. A random number of each band position condition was presented during each HIT. Therefore, because some Workers may receive a disproportionate number of each condition, it was important to evaluate their significance. The performance data was separated for the four conditions and the median performance for vertical and horizontal bands were co-plotted, along with their IQRs (Fig. 5).

The absolute difference in median performance between the horizontal and vertical conditions was 6.57%. The largest differences in median performance between the two conditions were observed for textures El-Odd (absolute difference in median performance of 19.0%), Foot-Odd (12.4%), and Wye-Odd (12.6%). The smallest differences in median performance were observed for textures Zigzag-Even and Box-Odd (0%). Based on this analysis, there is a small but measurable difference in performance for horizontal and vertical bands for certain textures. This will be revisited in our analysis of the Live data below (Section 3.5.2).

3.5 Results: Live Phase

3.5.1 Summary Statistics for Live1 and Live2 Data

Live1 and Live2 both contained 32 assignments. Each assignment was composed of 15 texture discrimination tasks (one for each texture family), making a total of 960 texture discrimination tasks. Each unique Worker had access to a maximum of 15 HITs, one of each texture type. Live1 was uploaded to mTurk on 21 Oct 2014 at 11:24:04 EST and Live2 on 7 November 2014 at 12:43:08 EST.

Live1 took 48.72 hours to complete. The mean HIT completion rate was therefore 0.164 HITs per minute, 9.84 HITs per hour, or 1 HIT every 6 minutes 5.4 seconds. Live2 took only 20.42 hours to complete. The average HIT completion rate was therefore 0.392 HITs per minute, 23.50 HITs per hour, or 1 HIT every 2 minutes 56.3 seconds (Fig. 6A). This is approximately double the rate observed in Live1.

Live1 included 69 unique Workers; 18 (26.1%) completed all 15 texture families, thereby earning the bonus payment. The mean number of HITs completed was 6.96 and the mode was 1 HIT (21 Workers, 30.4%). Live2 included 58 unique Workers and 15 (25.9%) completed all 15 texture families, earning the bonus. The mean number of HITs completed was 8.28 and the mode was 15 HITs. In Live2, 6 Workers (10.3% of Workers) were identified who participated in the original study. Therefore, 17.2% of the HITs were completed by Workers who were not strictly naïve subjects.
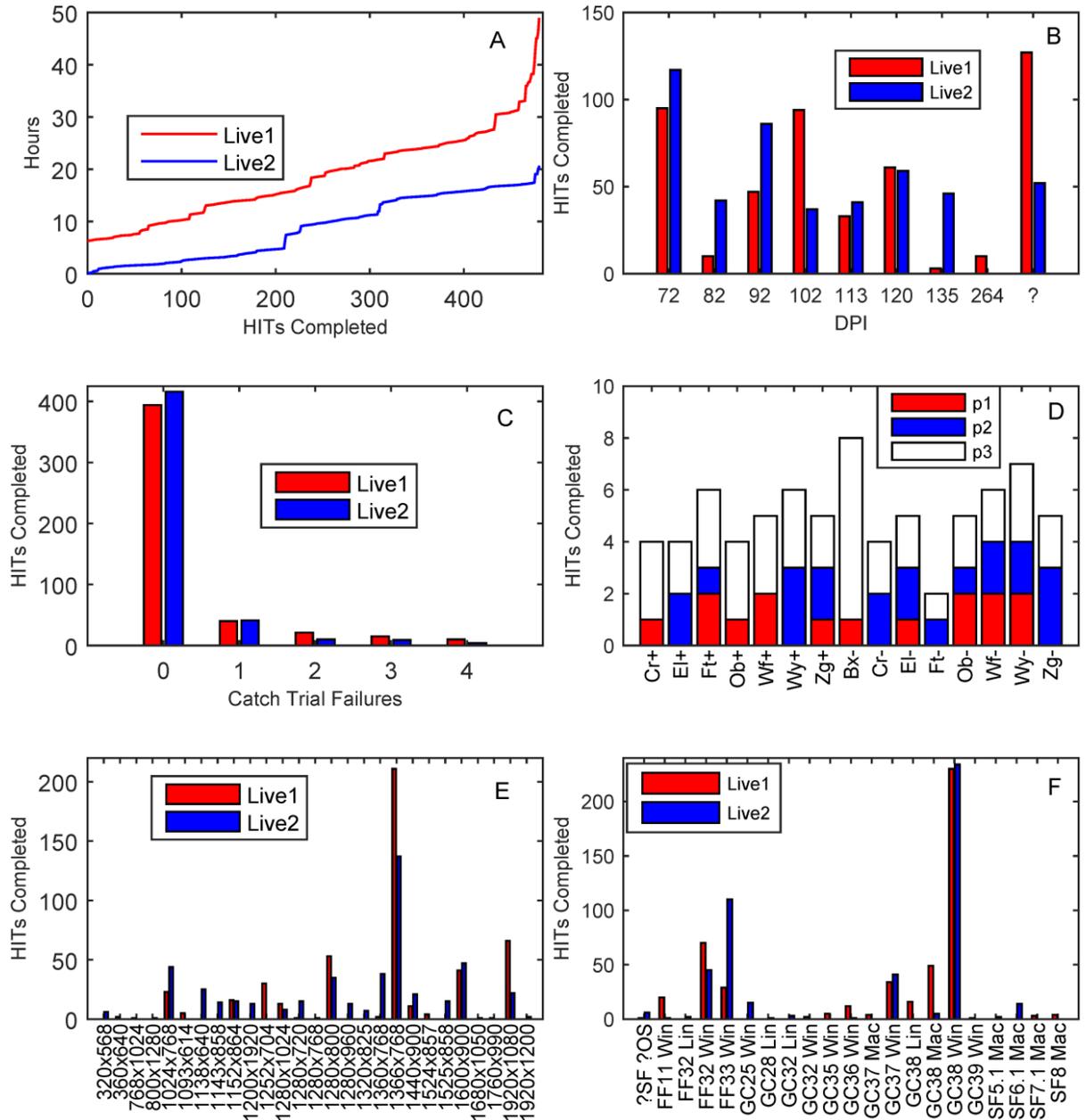
Figure 6. Summary statistics for the Live 1 and Live 2 studies. A: rate at which HITs were completed (in hours). B: DPI of the monitors used to complete the HITs. C: catch trial failures per HIT. D: catch trial passes by texture type (4 passes excluded). E: screen types (monitor size) used to complete the HITs. F: browser and operating system combinations used to complete the HITs.

In Live1, 9 different monitor DPIs were recorded (including "unknown") and in Live2, 8 different DPIs. The most significant difference concerned the number of Workers that skipped the monitor calibration step. In Live1, "unknown" was the most common DPI recorded, with 127 HITs (26.4%). After adding an explicit warning to the HIT page, this was reduced to 52 HITs (10.8%). In Live2, 72 DPI was the most common with 117 HITs (24.4%) (Fig. 6B).

Figure 6D shows catch trial passes per HIT for Live1 and Live2. In Live1, 394 (82.1%) Workers passed all 4 catch trials. HITs where Workers failed 2 or more catch trials were rejected on data quality grounds. Therefore, 90.4% (434) of HITs were of acceptable quality. In Live1, the total cost for the HITs was 434 * $0.10 AUD, plus 17 * $0.80 bonus payments (equals $57.00 AUD). For further analysis, Live1 HITs were also excluded where Workers used the Missed button more than once. Based on the 417 retained HITs, the cost per HIT was $0.137 AUD, or an estimated hourly cost of $8.20 AUD.

In Live2, 416 (86.7%) Workers passed all 4 catch trials and the number of HITs of acceptable quality was 457 (95.2%). Therefore, 13.1% more HITs were retained in Live2 versus Live1. The total cost for the HITs was 457 * $0.10 AUD, plus 15 * $0.80 bonus payments (equals $57.70 AUD). The cost per HIT was $0.126 AUD, which gives an estimated hourly cost of $7.57 AUD.

It was important to confirm that the effects of catch trials were not favouring particular texture types. Figure 6C shows the catch trial passes by texture type (4 catch trial passes not shown due to the difference in magnitude).

188

There is a reasonably even distribution of catch trial failures across textures, indicating that the catch trials did not bias any particular texture type.

Statistics were also gathered on Missed button usage in Live1. In 417 (86.9%) of the 480 HITs, Workers did not use the Missed button. Due to the frequent use of the Missed button by some Workers, the missed button was disabled in Live2.

Additional statistics were gathered on screen type, browser, and operating systems used in Live1 and Live2. 16 different screen types were recorded in Live1 and 21 in Live2. Note that the pixel dimensions are given in horizontal by vertical format. Thus, 800x600 (landscape) and 600x800 (portrait) are considered *different* screen sizes. In both Live1 and Live2, 1366x768 was by far the most common monitor type (Fig. 6E). 15 different browser and operating system (OS) combinations were recorded in Live1, and 14 in Live2. In both studies, Google Chrome 38 on Windows was the most common browser/OS combination (Fig.6F). In both studies, the most popular OS was Windows (86.9% of HITs in Live1, 93.1% in Live2) (Fig. 6F).

3.5.2 Revisiting the Effect of Screen Size and Band Position

In the Lab study, it was observed that two monitor types, which represented the extremes of monitor size, had an effect on texture discrimination performance (Section 3.4.1). The same analysis was carried out for a combined Live1 and Live2 data set. The performance data was stratified by monitor type. After removing catch trial failures, there was sufficient data to analyse the monitor types: 1024x768 (64 HITs), 1252x704 (29 HITs),

1280x800 (84 HITs), 1366x768 (322 HITs) and 1920x1080 (86 HITs). The median performance data for these screen resolutions was co-plotted (figure not shown) and all monitors produced very similar results. Some minor variation was observable between monitor types, but this probably results from noise; there was no substantial evidence that monitor types produce significant variations in performance. Colourmaps were then produced, showing the median performance by monitor type and the corresponding correlations between monitor types (Fig. 7).
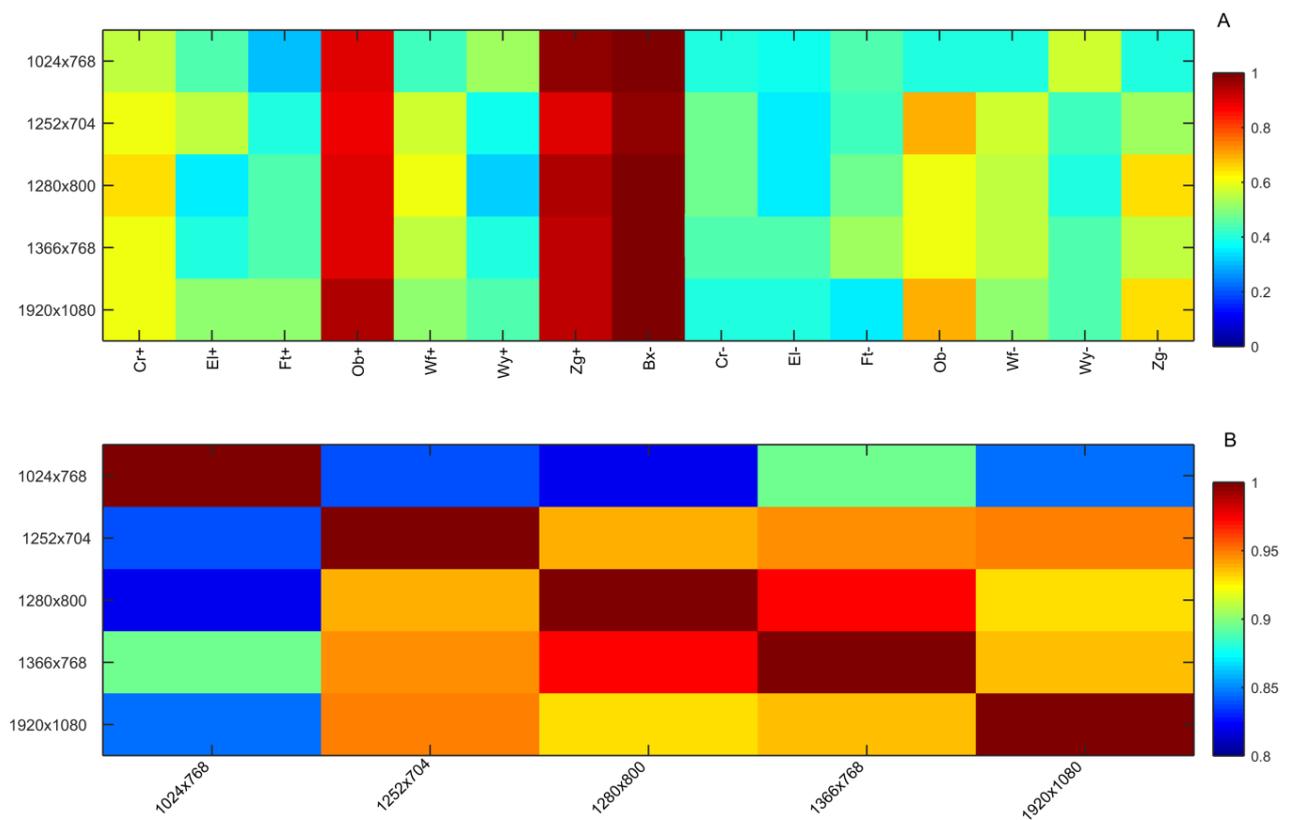
Figure 7. Colourmap derived from the combined Live1 and Live2 data sets. A: Colourmap showing median texture discrimination performance (probability correct) by monitor resolution. B: Colourmap showing the correlations between median texture discrimination performance for each monitor size.

190

Figure 7B shows that the performance data for all monitor types are highly correlated. One monitor produced slightly lower correlations across the other monitor types (1024x768). However, the magnitude of the difference is small and the mean correlation for that monitor is still high across all other platforms at 84.9%. Therefore, screen resolution does not appear to be a significant contributor to the variation in texture discrimination performance observed.

Physical screen sizes were also calculated, using the DPI and screen resolution data. The combined Live performance data was then stratified into four groups according to physical screen size (diagonal in inches): 10-13 in, 14-16 in, 17-19 in, and 20-22 in. The median performance scores of the four groups were highly correlated (85.8% to 96.2%) (see Appendix 6.5), although the smallest displays showed the lowest correlation with the others.

The Lab testing phase of this study found that band position had a small effect on discrimination performance (Section 3.4.3). This analysis was repeated using the combined Live1 and Live2 data sets (figure not shown) and only a very small difference in performance was observed based on band position. The largest differences in median performance between horizontal minus vertical conditions were observed for textures Foot-Even (4.3%), Cross-Odd (5.9%) and Zigzag-Odd (4.6%). These textures may contain directional elements which alter how difficult they are to discriminate in certain orientations. However, the effect is small.

3.5.3 Live Performance Data versus Two Comparison Data Sets

The Live data sets were compared to 2 other data sets which used the same texture types and were also administered using mTurk. The first comparator data set (which we have called M1-M6) was derived from the Lab testing phase discussed above. It used the same mTurk HITs, but in the Sandbox environment (Fig. 8A).
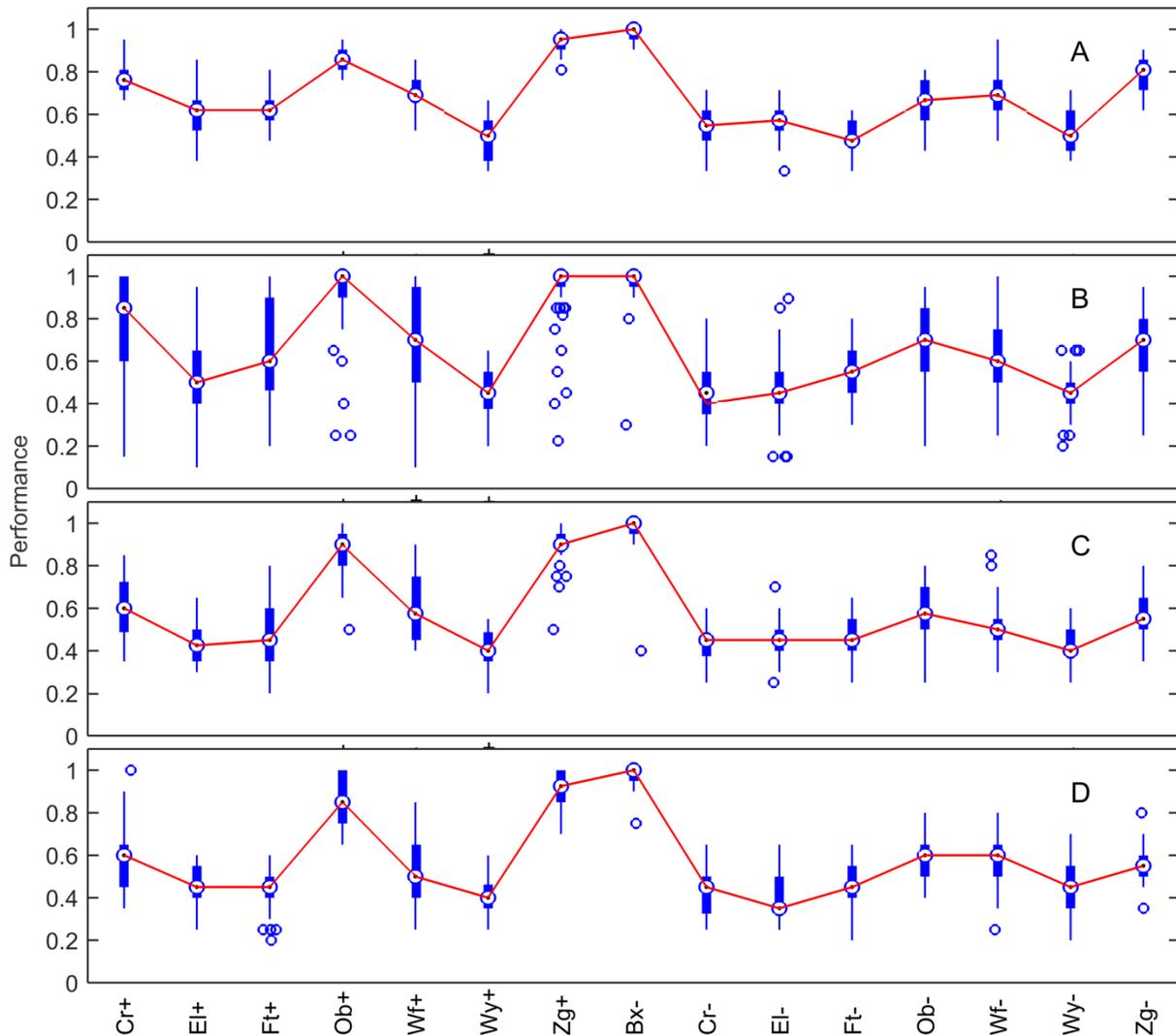

Figure 8. Texture discrimination performance by data set. Median performance is shown in red. Boxes are 25th to 75th percentiles and whiskers are the 10th and 90th percentiles, small circles are outliers. A:

the M1-M6 data set (machines M1-M6 from the Lab data set combined). B: the DC lab data set. C: the Live 1 data set. D: the Live 2 data set.

In the M1-M6 data set, one subject (JWGS) completed 3 repetitions (rep0, rep1, rep2) of 15 HITs on each of six different platforms (for a total of 270 HITs). The textures used were the same as those in the Live method. The distribution of the median performance scores for this data set is strikingly similar to that of the Live mTurk data. A superficial inspection suggests that the all data points appear to have been shifted upwards 10-15%, and this is perhaps unsurprising given the experience of the subject with the visual psychometric task.

A second comparator data set was provided by Dominique Coy (DC) of the Maddess group (Figure 8B) (Coy, 2014). It constitutes texture discrimination data from naïve subjects under supervised laboratory conditions and it used same Sandbox environment as the M1-M6 data set. In this case, 6 naïve subjects were recruited to complete the task under supervised laboratory conditions. They included five females and two males whose ages ranged from 20 to 24 years. A qualified optometrist completed standard eye examinations on all 6 subjects. Each of the subjects completed HITs using 17 texture families (2 of which are excluded here, to match those used in this study), producing 84 HITs for each of the 15 texture families examined.

To facilitate comparison of the four data sets, the median performance and IQRs for each data set were plotted on the same axis (Fig. 9). The trends of all four data sets were very similar. The DC and M1-M6 data sets have

slightly higher performance scores than the Live data sets and this is particularly noticeable for the textures which are more difficult to discriminate. The DC data set has the highest IQRs and the most outliers (Fig. 9B). Live1 and Live2 are very similar in terms of median performance and IQRs.
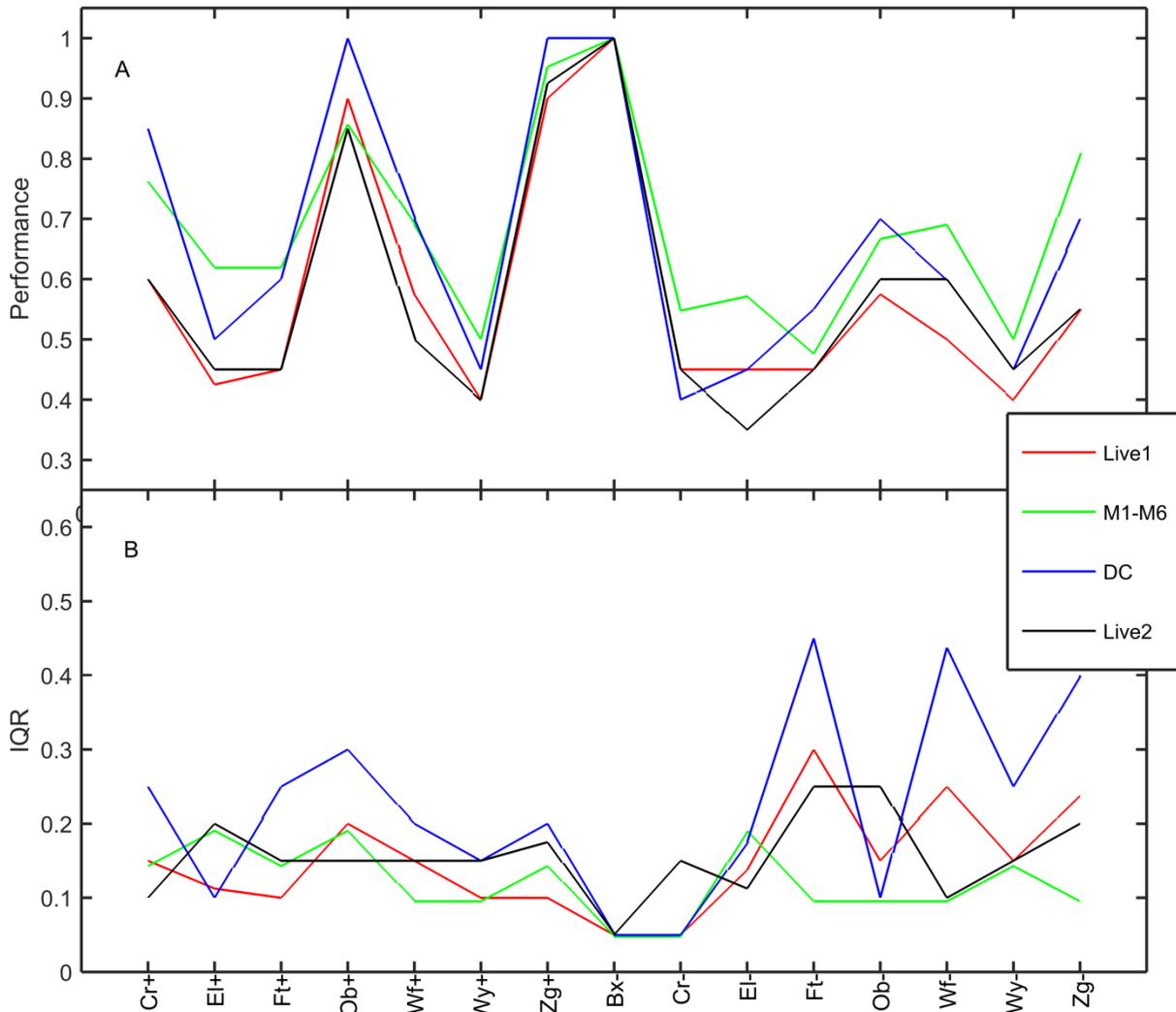

Figure 9. Median (A) and interquartile range (B) plots by data set.

The M1-M6 data set has the lowest IQRs across the 15 textures, with a mean IQR of 0.12. This was expected as the subject was experienced in the testing modality. It is immediately clear how consistently low the IQRs of Live1 and Live2 are, with mean IQRs of 0.149 and 0.156 across the 15

textures. Indeed, the mTurk Live IQRs are very similar to the M1-M6 IQRs for the first 10 textures. The DC data set is significantly more variable, with a mean IQR of 0.22. This is most noticeable for textures Cross-Odd and Oblong-Odd.

3.5.4 Relative and Absolute Repeatability of Live1 and Live2 Data

The relative and abolsoute test-retsest reliability for the data sets were evaluated. Firstly, Pearson's correlation coefficients (*r*) were compluted and displayed in a colourmap (Fig. 10B) alongside the median performance data (Fig. 10A). There is a strong linear correlation between both comparator data sets and the Live mTurk data sets. The M1-M6 and DC data sets generally produce slightly higher performance scores than the Live data sets. The correlation between the M1-M6 data set and Live1 was 91.9% and Live2, 91.4%. The correlation between the DC data set and Live1 was 93.9% and Live2, 92.1%. The correlation between Live1 and Live2 was 97.0%. The correlations of log mean performance were also computed, but they were not markedly different, suggesting the absence of a multiplicative relationship.

Bland-Altman plots were also examined (see Appendix 6.3) and the coefficient of repeatability (CR) range from X to Y and was smallest between Live1 and Live 2 at 9.4% (Bland and Altman, 1986, Bland and Altman, 1999, Bland, 2000, Vaz et al., 2013) . The CR for the DC versus the Live1 data was 14.3% and M1-M6 versus Live1 data was 15.6%. These scores are quite low and they suggest that the Live1 mTurk study has accurately reproduced the texture discrimination results that derived from supervised laboratory trials.

The coefficient of repeatability for the DC versus the Live2 data was 16.2% and M1-M6 versus Live2 data was 16.0%. These scores are slightly higher, but once again they are *low*, suggesting that the Live2 mTurk study has accurately reproduced the supervised laboratory trials. The CR between Live1 and Live2 was only 9.4% and this very low score indicates that the Live study itself has a good repeatability.
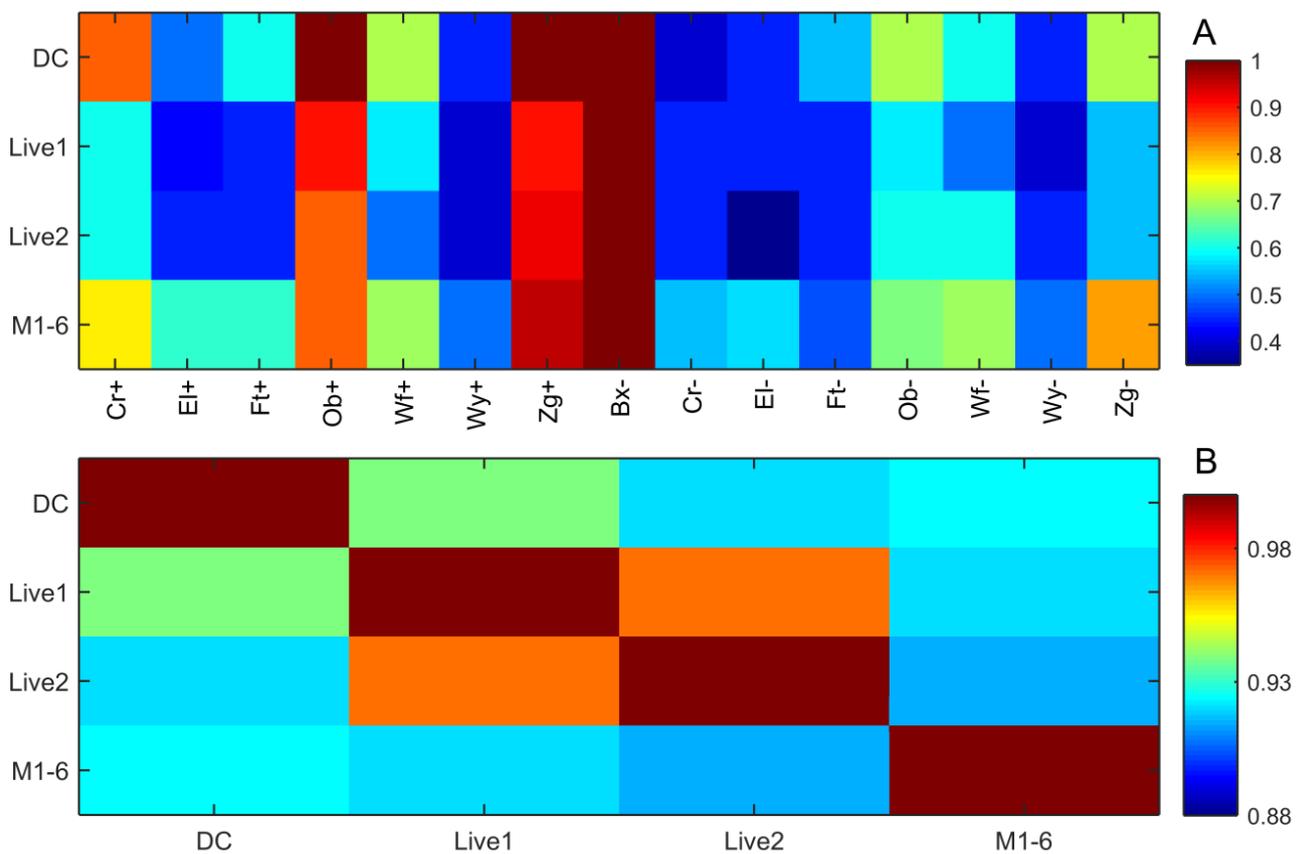

Figure 10. Colourmap derived from the Live and Lab data sets. A: Colourmap showing median texture discrimination performance by data set. B: Colourmap showing the correlations between median texture discrimination performance for each data set.

3.5.5 Factor Analysis of the Combined Lab and Live Datasets

The Live1, Live2, and both Lab data sets (M1-M6 and DC) were combined and the probabilities correct were transformed into d-prime values (Macmillan and Creelman, 1991). The adjusted data was then analysed using Factor analysis. Two forms of Factor analysis were performed, based on varimax rotated principal components (PCs) (Reyment, 1996, Norris and Lecavalier, 2010). Details of all these methods are available from our recent publication (Seamons et al., 2015) and outlined in Chapter 2.

Figure 11 shows the resulting PC Scree plot, with the proportion of variance accounted for by the possible factors from the pooled data set. The proportions are plotted in order of their relative influence, from largest to smallest. If the data being evaluated is governed by a small number of principal mechanisms, we will see a small number of large eigenvalues, followed by a significant drop-off; the lesser factors should exert little influence on our data and therefore are expected to trail off exponentially into noise (Seamons et al., 2015). There is clear a drop-off in the accounted proportion of variance after the first 2 factors. This suggests that, as a first estimate, 2 is the lower bound of the independent neurological mechanisms that govern texture discrimination. The lesser factors do not fall away as markedly as expected.
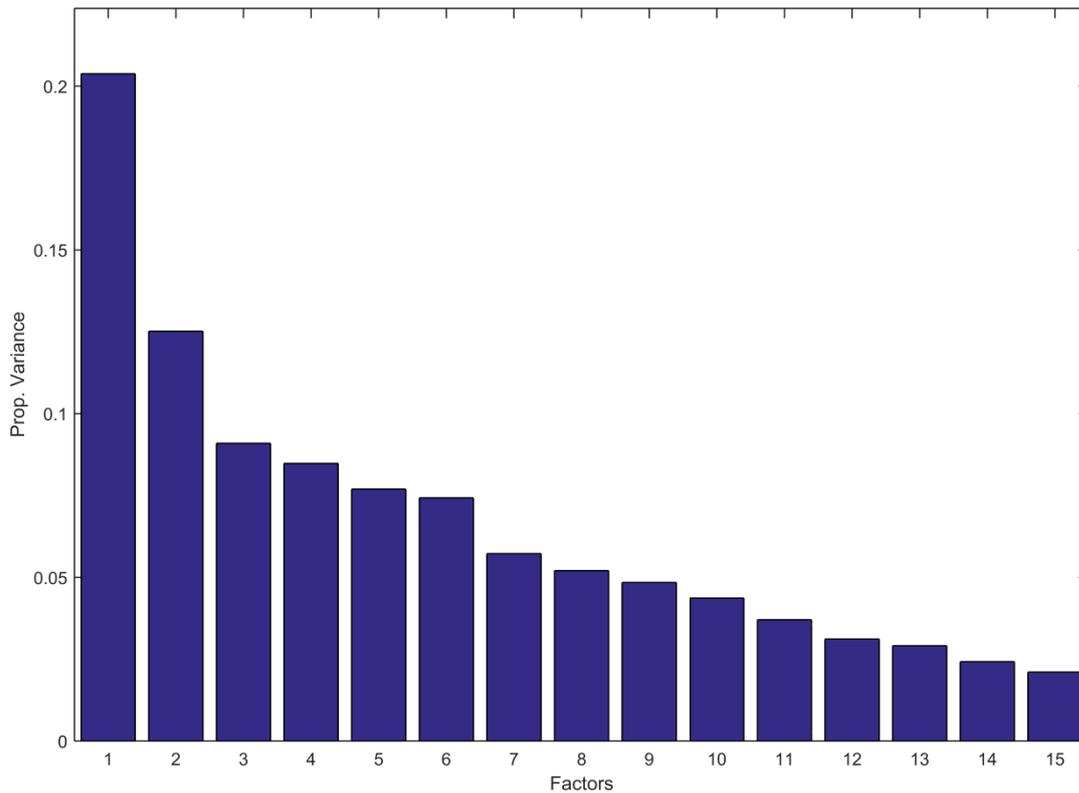
Figure 11: Scree plot showing the proportion of variance accounted for by 15 possible factors from the combined Live and Lab data sets (Live1, Live2, M1-M6, and DC). The proportions of variance accounted for by each factor are plotted in order of their relative influence, from largest to smallest.

Further analysis was carried out to support the presence of the principal factors. Figure 12 shows the proportion of variance accounted for by N factors for each texture, often called *communalities* (Reyment, 1996). The first graph shows that the *per texture* proportion of variance explained by the single largest factor is poor. As additional factors are added (in descending order from the Scree plot of Fig. 12), the reconstruction error declines significantly. It can be seen that 3 to 4 factors are required to account for the majority of the data evenly across all textures. As subsequent factors are added above 4, there was little further improvement in the reconstruction.

198
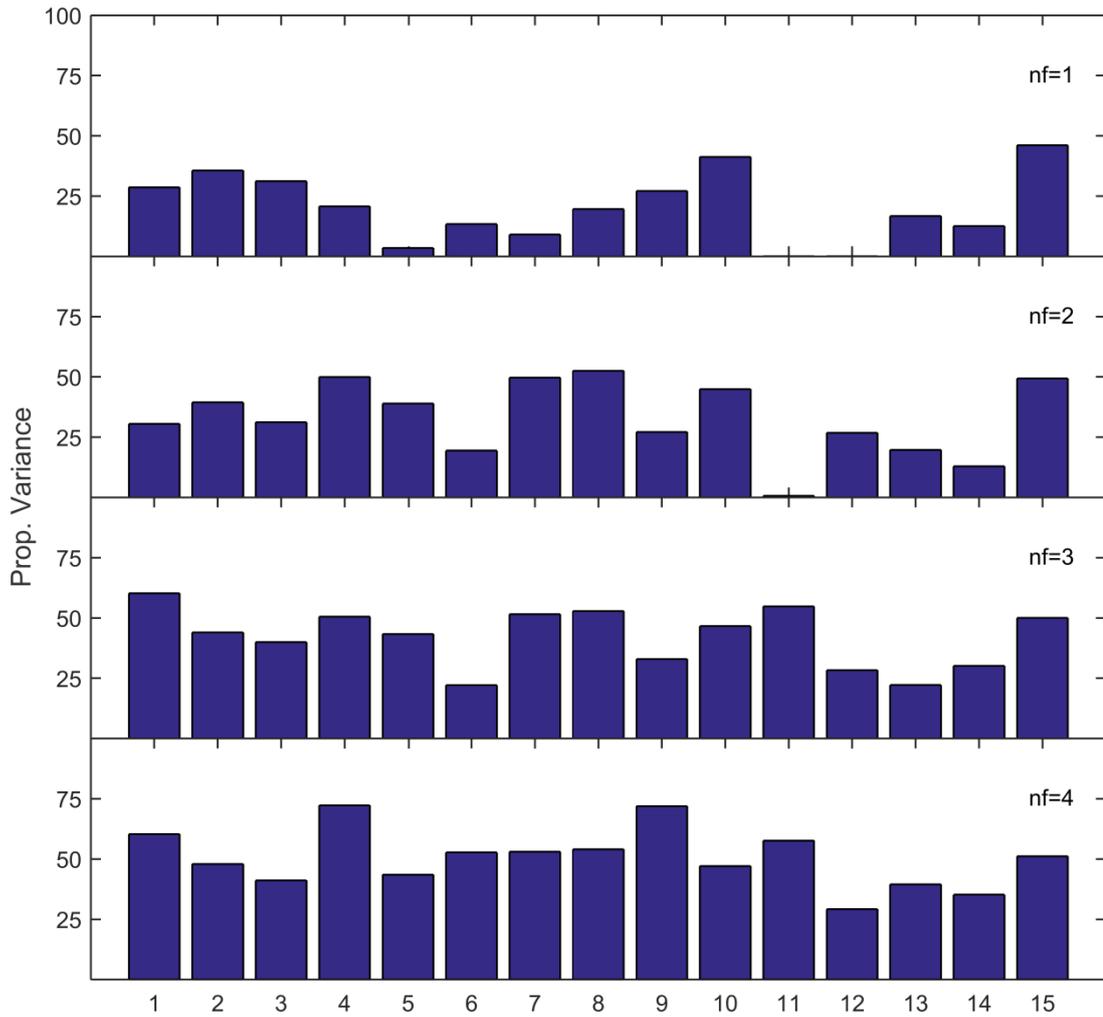
Figure 12: A sequence of plots showing the proportion of variance explained by reconstructing the performance data using *nf* factors .The first graph shows that the proportion of variance explained for each texture by just the single largest factor is poor. As additional factors are added, the reconstruction error declines significantly. It can clearly be seen that 3 to 4 factors are sufficient to account for the majority of the proportion of variance in the data. Including additional factors above this threshold generates only a minor improvement in the reconstruction.

3.6 Discussion

A concern with novel platforms such as mTurk is whether the data derived is of the requisite quality. The primary purpose of this study was to address this question with regard to a visual psychometric task: discrimination of binary isotrigon textures. This was accomplished by comparing two sets of Live mTurk data (Live1 and Live2) to data derived from supervised laboratory tests (DC and M1-M6).

The rate of HITs completed in this study was broadly consistent with previous findings (Heer et al., 2010, Buhrmester et al., 2011). HIT participation rates are affected by the level of compensation and how long the HIT takes to complete (Buhrmester et al., 2011). For example, Buhrmester et al. offered $0.02 USD for a 30 minute task, but still accumulated 25 completed HITs within 5 hours of posting. When the compensation was increased to $0.50 USD, they obtained the same number of completed HITs in less than 2 hours of posting (Buhrmester et al., 2011). For Live2, the average HIT completion rate was approximately double the rate observed in Live1. This may be attributed to the Requester reputation gained from having previously completed an mTurk study. Live1 was the first mTurk study executed on the Requester account, which could have dissuaded some Workers from joining the study.

In terms of catch trial failures, the quality of the data derived was excellent (90.4% HITs retained in Live1 and 95.2% in Live2, a 13.1% improvement between the studies). This improvement may have resulted from the presence of repeat Workers, who completed 17.2% of the total HITs in Live2. Considering that the HITs used a slightly more challenging catch trial (Box-

Odd), the rates of HIT retention based on catch trials are surprising. There was an even distribution of catch trial failures across textures, indicating that the catch trials did not bias any particular texture type.

A relatively high proportion of Workers skipped the calibration test for monitor DPI. Despite adding a warning to the Live2 study, 52 HITs (10.8%) were recorded where Workers did not complete the calibration test at least once. There were also problems with Workers not using the latest version of their browser and overusing the Missed button in Live1, despite explicit instructions otherwise. These findings suggest that clear directions alone are *not sufficient* to guarantee Worker behaviour. In future studies, implementation of a qualification task and programmatic enforcement of calibration steps (for example, locking the HIT Start button until calibration has been completed) should be implemented where possible. That being said, the high contrast of the stimuli minimised the effects of display differences and the self-similar analysis of texture by the brain minimised the effect of display size and resolution (Maddess and Nagai, 2001, Victor et al., 2005).

The percentage of HITs rejected is comparable with previous findings and may be improved by adding a qualification task in addition to the catch trials. For example, Heer et al. repeated one experiment after removing qualification tasks and found that over 10% of the responses were unusable (Kittur et al., 2008, Heer et al., 2010). After incorporating a qualification task, only 0.75% of responses were rejected. Qualification tasks help to ensure that Workers understand the task and are suitably engaged (Kittur et al., 2008, Heer et al., 2010).

The Live1 and Live2 data was compared to two Lab data sets from supervised laboratory tests (DC and M1-M6) which used similar protocols. The distribution of median performance scores for the Live data sets was superficially very similar to the Lab results. The performance scores of the DC data set were slightly higher than the Live data sets across the textures examined. The performance scores of the M1-M6 data set were shifted upwards 10-15% with respect to the Live data sets. These findings may be attributed to the fact that the subjects were under supervised laboratory conditions and more experienced than mTurk subjects (in the case of the DC data set, subjects started as naïve subjects, but became more experienced with repeated visits). This is consistent with previous findings that the discrimination of these textures can be improved by learning, although many repeats appear to be required (Maddess and Nagai, 2001, Taylor et al., 2008, Coy, 2014).

A strong linear correlation was observed between both Lab data sets and the Live mTurk data, based on median performance. The correlations of log mean performance were also computed, but they were not markedly different suggesting the absence of a multiplicative relationship.

Absolute repeatability was evaluated using coefficient of repeatability (CR), derived from the Bland–Altman plots (see Appendix 6.3). The CRs between Lab and Live data sets were low, suggesting that the Live mTurk study had accurately reproduced the texture discrimination results that were produced by supervised laboratory trials. The CR between Live1 and Live2 was only 9.4% and this very low score indicates that the Live study itself has a *good repeatability*. Based on these metrics, it appears that mTurk can produce

comparable quality data to that derived from laboratory studies in the case of an isotrigon texture discrimination task.

The strong agreement between Live and Lab data sets is interesting considering the considerable variations in the browser, physical screen size, screen resolution and operating systems observed. During the Lab phase of this study, we investigated whether factors such as screen size, pixel size. and luminance would affect texture discrimination. In the assessment of repeatability, the CRs and Pearson's correlation coefficients were consistently low across the machine types tested. Factor scores were also plotted from a Factor analysis of the M1-M6 data (see Appendix 6.4). These are estimates of the contributions of the four largest factors, showing the relative contributions of each factor for each machine. Machine-specific features appeared to have little impact on discrimination performance. Small differences were observed for two screen sizes (the extremes of those tested). This finding was supported by the subsequent analysis of the Live data (see Section 3.5.2).

Factor analysis of the combined Lab and Live datasets suggested that 3 to 4 is the lower bound of the independent neurological mechanisms that govern texture discrimination. This finding is similar to results reported for other isotrigon textures using different methods. These studies found that the number of independent mechanisms is less than 10 (Taylor et al., 2008, Barbosa et al., 2013), and is more likely 3-4 (Maddess and Nagai, 2001, Maddess et al., 2007, Coy, 2014, Seamons et al., 2015). In our previous study using different isotrigon textures, we found evidence of as few as 3 independent mechanisms (Seamons et al., 2015). It is notable that there is a

strong congruence between Factor analyses performed from multiple laboratory-derived samples, and that derived from *unsupervised naïve subjects using mTurk*. In interpreting the results of this Factor analysis, it should be noted that the HITs are designed such that each Worker are naïve subjects and they have access to a maximum of one discrimination task per texture type. Data derived from naïve subjects is expected to be noisier than that derived from more experienced subjects, who have been through a learning process. Previous studies suggest that learning effects require many repeats to develop (Maddess and Nagai, 2001, Taylor et al., 2008, Coy, 2014).

The findings of this study with regard to data quality are consistent with previous studies where mTurk has been used to replicate laboratory studies (Heer et al., 2010, Freeman et al., 2013). For example, Freeman et al. observed a 92% correlation between mTurk and laboratory data 15 textures tested (Freeman et al., 2013). Paolacci et al. replicated a series of classic judgment and decision-making experiments on mTurk at a cost of just $1.71 USD per hour per subject; quantitatively, there were only very minor differences between the mTurk and laboratory data sets (Paolacci et al., 2010). Similarly, Horton et al. and Crump et al. replicated some classic behavioural psychology experiments using mTurk , including the Prisoners' Dilemma, and found agreement with the laboratory data (Horton et al., 2011, Crump et al., 2013).

Considering that mTurk Workers are unsupervised, what accounts for the high data quality observed? Although compensation is a relevant issue for most Workers, it is not the sole driver; the number of Workers that rely on

mTurk for their primary income is quite low: 12% of US and 27% of Indian Workers (Ipeirotis, 2010a). Several authors have found that compensation rates appear to have less effect on data quality than on the *rate* of data collection (Buhrmester et al., 2011, Paolacci et al., 2010, Mason and Watts, 2009, Heer et al., 2010). Workers are not believed to be primarily financially motivated, but also derive secondary benefits from mTurk, such as entertainment and a sense of being productive (Buhrmester et al., 2011, Ipeirotis, 2010a, Paolacci et al., 2010, Ross et al., 2010). Indeed, a recent study found that the data quality declined in parallel with declining meaningfulness of the task (Chandler et al., 2013). It also suggests that tasks which are designed to have at least some entertainment value may be more successful at recruiting Workers (Crump et al., 2013).

At the time of writing only three studies were identified which used mTurk to administer tasks of a visual perceptual nature (Heer et al., 2010, Cole et al., 2009, Freeman et al., 2013). The study by Freeman et al. is particularly interesting as they correlated mTurk performance with single-unit and fMRI data (Freeman et al., 2013). They used synthetic stimuli which replicated the higher-order structure of natural textures. Starting with photographs taken from nature, two sets of stimuli were derived: spectrally matched noise and "naturalistic" textures. The latter were synthesised using an optimization process (gradient descent) based on the image statistics of the original photograph (Portilla and Simoncelli, 2000, Heeger and Bergen, 1995). The authors report that the resulting naturalistic textures had the same overall orientation and spatial frequency content as the original photographs, but lacked their higher-order statistical dependencies (Freeman et al., 2013).

The naturalistic stimuli induced a unique response in V2, but not V1, during macaque single-unit and human fMRI recordings (Freeman et al., 2013).

Freeman et al. then used mTurk to carry out 3 alternative forced choice trials where subjects were asked to identify the image which "looks different" from the other two. From this data, the authors calculated perceptual sensitivity functions. The mTurk perceptual sensitivity functions were significantly correlated with single-unit modulation in V2 (r = 0.74, P < 0.001, 16 cells) and fMRI modulation in V2 (r = 0.77, P < 0.0001, 2 subjects) (Freeman et al., 2013). The Freeman et al. study (Freeman et al., 2013) is an elegant demonstration of the potential of mTurk in combination with laboratory techniques.

In this study, we have shown that correlation, repeatability and Factor analysis support the use of mTurk for visual psychometric research. Indeed, mTurk appears capable of producing robust and repeatable results, results which are comparable with supervised laboratory studies. Isotrigon discrimination has been previously found to be resistant to contrast and pixel size variations (Maddess and Nagai, 2001) and it is likely that some experimental protocols will be less suited to mTurk. In order to maximise data quality derived from mTurk, a number of safeguards should be implemented. HITs should have clear instructions, screen calibration measures where appropriate, use destructive prior testing to identify bugs and exploits, and capture browser, DPI, and OS information. Cross-browser testing is particularly important as subtle bugs and performance variations are common. Catch trials and/or qualification tasks are crucial for maintaining data quality (Heer et al., 2010, Kittur et al., 2008). This is also true during the

data analysis phase, where it is essential that the data is sufficiently rich to allow sufficient stratification and filtering based on individual Worker performance. This study found that warning messages are not sufficient to ensure that Workers comply with the HIT instructions and therefore measures may be needed to enforce them (such as disabling the HIT until calibration has been completed).

For some studies, HITs may be technically complex to develop and require considerable time and effort to validate. This is particularly true of visual psychometric HITs, which perhaps offsets some of the gains derived from the mTurk system. Nonetheless, as in this study, many technical challenges can be addressed and overcome by careful implementation strategies. This study has found mTurk to be a promising, and underutilized, platform for visual psychometric research which can produce data of comparable quality to laboratory samples at greatly reduced cost and greatly increased scale.

3.7 Acknowledgments

3.8 References

AWS, A. W. S. 2014. *Amazon Mechanical Turk (mTurk)* [Online]. Available: https://www.mturk.com/mturk/welcome [Accessed 4 November 2014.

BARBOSA, M. S., BUBNA-LITIC, A. & MADDESS, T. 2013. Locally countable properties and the perceptual salience of textures. *JOSA A,* 30, 1687-1697.

BEASON-HELD, L. L., PURPURA, K. P., KRASUSKI, J. S., DESMOND, R. E., MANGOT, D. J., DALY, E. M., OPTICAN, L. M., RAPOPORT, S. I. & VANMETER, J. W. 2000. Striate cortex in humans demonstrates the relationship between activation and variations in visual form. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale,* 130, 221-6.

BEASON-HELD, L. L., PURPURA, K. P., KRASUSKI, J. S., MAISOG, J. M., DALY, E. M., MANGOT, D. J., DESMOND, R. E., OPTICAN, L. M., SCHAPIRO, M. B. & VANMETER, J. W. 1998a. Cortical regions involved in visual texture perception: a fMRI study. *Brain research. Cognitive brain research,* 7, 111-8.

BEASON-HELD, L. L., PURPURA, K. P., VAN METER, J. W., AZARI, N. P., MANGOT, D. J., OPTICAN, L. M., MENTIS, M. J., ALEXANDER, G. E., GRADY, C. L., HORWITZ, B., RAPOPORT, S. I. & SCHAPIRO, M. B. 1998b. PET reveals occipitotemporal pathway activation during elementary form perception in humans. *Visual neuroscience,* 15, 503-10.

BLAND, J. M. 2000. *An Introduction into Medical Statistics,* Oxford, Oxford University Press.

BLAND, J. M. & ALTMAN, D. G. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet,* 1, 307-10.

BLAND, J. M. & ALTMAN, D. G. 1999. Measuring agreement in method comparison studies. *Stat Methods Med Res,* 8, 135-60.

BUHRMESTER, M., KWANG, T. & GOSLING, S. D. 2011. Amazon's Mechanical Turk : A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science,* 6, 3-5.

CAELLI, T., JULESZ, B. & GILBERT, E. 1978. On perceptual analyzers underlying visual texture discrimination: Part II. *Biological Cybernetics,* 29, 201-214.

CATTELL, R. B. 1966. The Scree Test For The Number Of Factors. *Multivariate Behavioral Research,* 1, 245-276.

CHANDLER, M., DE LA CRUZ, F., DYDA, F., HICKMAN, A. B., MONCALIAN, G. & TON-HOANG, B. 2013. Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nat Rev Microbiol,* 11, 525-38.

COLE, F., SANIK, K., DECARLO, D., FINKELSTEIN, A., FUNKHOUSER, T., RUSINKIEWICZ, S. & SINGH, M. How Well do Line Drawings Depict Shape? ACM Transactions on Graphics, 2009.

COY, D. G., BARBOSA, M., SEAMONS, J. W. G. & MADDESS, T. 2014. Learning effects for Discrimination of Isotrigon Textures (Honours Thesis; Journal Article In Preparation).

CRUMP, M. J., MCDONNELL, J. V. & GURECKIS, T. M. 2013. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS One,* 8, e57410.

DEVELLIS, R. F. 2012. Factor Analysis. *Scale Development Theory and Applications.* 3rd ed. University of North Carolina, Chapel Hill, USA: SAGE Publications, Inc.

DOW, S., KULKARNI, A., KLEMMER, S. & HARTMANN, B. Shepherding the Crowd Yields Better Work. ACM 2012 conference on Computer Supported Cooperative Work 2012 New York. ACM, 1013-1022.

FREEMAN, J., ZIEMBA, C. M., HEEGER, D. J., SIMONCELLI, E. P. & MOVSHON, J. A. 2013. A functional and perceptual signature of the second visual area in primates. *Nature neuroscience,* 16, 974-981.

GILBERT, E. N. 1980. Random Colorings of a Lattice of Squares in the Plane. *SIAM Journal on Algebraic and Discrete Methods,* 1, 152-159.

GORITZ, A. S., WOLFF, H. G. & GOLDSTEIN, D. G. 2008. Individual payments as a longer-term incentive in online panels. *Behavior Research Methods,* 40.

GOSLING, S. D., VAZIRE, S., SRIVASTAVA, S. & JOHN, O. P. 2004. Should we trust Web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist,* 59, 93–104.

HAYTON, J. C., ALLEN, D. G. & SCARPELLO, V. 2004. Factor Retention Decisions in Exploratory Factor Analysis: a Tutorial on Parallel Analysis. *Organizational Research Methods, 7*.

HEEGER, D. J. & BERGEN, J. R. Pyramid-based texture analysis/synthesis. Proceedings of the 22nd annual conference on Computer graphics and interactive techniques, 1995. ACM, 229-238.

HEER, J., BOSTOCK, M. & 2010. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. *ACM Human Factors in Computing Systems (CHI)*, 203-212.

HORTON, J. J., RAND, D. G. & ZECKHAUSER, R. J. 2011. The online laboratory: conducting experiments in a real labor market. *Experimental Economics,* 14, 399-425.

IPEIROTIS, P. 2010a. Analyzing the Amazon Mechanical Turk Marketplace. *ACM XRDS (Crossroads),* 17.

IPEIROTIS, P. 2010b. Demographics of Mechanical Turk. *Working Paper CeDER-10-01.* NYU Center for Digital Economy Research.

KIRK, R. E. 2007. *Statistics: An Introduction*, Cengage Learning.

KITTUR, A., CHI, E. H. & SUH, B. Crowdsourcing User Studies with Mechanical Turk. Conference on Human Factors in Computing Systems, 2008. 453-456.

MACMILLAN, N. A. & CREELMAN, C. D. 1991. *Detection theory: A User's Guide,* Cambridge, Cambridge University Press.

MADDESS, T. & KULIKOWSKI, J. J. 1999. Apparent fineness of stationary compound gratings. *Vision research,* 39, 3404-16.

MADDESS, T. & NAGAI, Y. 2001. Discriminating isotrigon textures [corrected]. *Vision research,* 41, 3837-60.

MADDESS, T., NAGAI, Y., JAMES, A. C. & ANKIEWCZ, A. 2004. Binary and ternary textures containing higher-order spatial correlations. *Vision research,* 44, 1093-113.

MADDESS, T., NAGAI, Y., VICTOR, J. D. & TAYLOR, R. R. 2007. Multilevel isotrigon textures. *Journal of the Optical Society of America. A, Optics, image science, and vision,* 24, 278-93.

MASON, W. & SURI, S. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behav Res Methods,* 44, 1-23.

MASON, W. A. & WATTS, D. J. 2009. Financial incentives and the performance of crowds. *Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP '10),* 77–85.

MITROFF, S. R., BIGGS, A. T., ADAMO, S. H., DOWD, E. W., WINKLE, J. & CLARK, K. 2015. What can 1 billion trials tell us about visual search? *Journal of experimental psychology: human perception and performance,* 41, 1.

NORRIS, M. & LECAVALIER, L. 2010. Evaluating the use of exploratory Factor analysis in developmental disability psychological research. *J Autism Dev Disord,* 40, 8-20.

OPPENHEIMER, D. M., MEYVIS, T. & DAVIDENKO, N. 2009. Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology,* 45, 867-872.

PAOLACCI, G., CHANDLER, J. & IPEIROTIS, P. 2010. Running experiments on Amazon Mechanical Turk. *Judgment and decision making,* 5, 411-419.

PONTIN, J. 2007. Artificial intelligence, with help from the humans. *New York Times*, 25 March.

PORTILLA, J. & SIMONCELLI, E. P. 2000. A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision,* 40, 49-70.

PURPURA, K. P., VICTOR, J. D. & KATZ, E. 1994. Striate cortex extracts higher-order spatial correlations from visual textures. *Proceedings of the National Academy of Sciences of the United States of America,* 91, 8482-6.

REYMENT, R. A., JORESKOG, K.G. 1996. Applied Factor Analysis in the Natural Sciences. Cambridge University Press.

ROSLI, Y., BEDFORD, S. M. & MADDESS, T. 2009. Low-spatial-frequency channels and the spatial frequency-doubling illusion. *Investigative ophthalmology & visual science,* 50, 1956-63.

ROSS, J., IRANI, I., SILBERMAN, M., ZALDIVAR, A. & TOMLINSON, B. 2010. Who are the CrowdWorkers? Shifting Demographics in Amazon Mechanical Turk. *CHI EA 2010*, 2863-2872.

SEAMONS, J. W., BARBOSA, M. S., BUBNA-LITIC, A. & MADDESS, T. 2015. A lower bound on the number of mechanisms for discriminating fourth and higher order spatial correlations. *Vision research,* 108**,** 41-48.

SEKULER, R., WILSON, H. R. & OWSLEY, C. 1984. Structural modeling of spatial vision. *Vision research,* 24, 689-700.

SIMPSON, W. A. & MCFADDEN, S. M. 2005. Spatial frequency channels derived from individual differences. *Vision research,* 45, 2723-7.

STONE, M. & BARTRAM, L. 2008. Alpha, Contrast and the Perception of Visual Metadata. *16th Color and Imaging Conference.* Society for Imaging Science and Technology.

TAYLOR, R. R., MADDESS, T. & NAGAI, Y. 2008. Spatial biases and computational constraints on the encoding of complex local image structure. *Journal of vision,* 8, 19 1-13.

TKACIK, G., PRENTICE, J. S., VICTOR, J. D. & BALASUBRAMANIAN, V. 2010. Local statistics in natural scenes predict the saliency of synthetic textures. *Proceedings of the National Academy of Sciences of the United States of America,* 107, 18149-54.

VAZ, S., FALKMER, T., PASSMORE, A. E., PARSONS, R. & ANDREOU, P. 2013. The case for using the repeatability coefficient when calculating test-retest reliability. *PLoS One,* 8, e73990.

VICTOR, J. D. 1994. Images, Statistics, and Textures: Implications of Triple Correlation Uniqueness for Texture Statistics and the Julesz Conjecture. *Journal of the Optical Society of America A,* 11, 1680-1684.

VICTOR, J. D., CHUBB, C. & CONTE, M. M. 2005. Interaction of luminance and higher-order statistics in texture discrimination. *Vision research,* 45, 311-328.

VICTOR, J. D. & CONTE, M. M. 1991. Spatial organization of nonlinear interactions in form perception. *Vision research,* 31, 1457-88.

VICTOR, J. D. & CONTE, M. M. 2005. Local processes and spatial pooling in texture and symmetry detection. *Vision research,* 45, 1063-73.

VICTOR, J. D., THENGONE, D. J. & CONTE, M. M. 2013. Perception of second- and third-order orientation signals and their interactions. *J Vis,* 13, 21.

ZWICK, W. R. & VELICER, W. F. 1986. Comparison of five rules for determining the number of components to retain. *Psychological Bulletin,* 99, 432-442.

Chapter 4: Using Crowdsourcing to Identify Maximally Informative Directions within a Ternary Texture Feature Space

4.1 Abstract

This study employs a set of ternary textures with well-defined spatial correlation structure of different orders between three grey levels. The different textures were constrained in different orders of spatial correlations from 1st to 4th order and are referred to as gamma, beta_hv, beta_diag, theta and alpha respectively. The textures are stochastically defined allowing salience to be independently controlled by adjusting the level of decorrelation. The ternary textures can be projected onto so-called *trico* planes; in so-doing, three-level stimuli can be displayed coherently in two dimensions. The origin of these planes represents ternary noise, i.e. the point of zero salience. Particular directions outward from the origin on the plane point to mixtures of the canonical textures.

To the ideal observer, all textures defined by a given distance out from the origin of the trico plane are equally detectable. However, humans are *not* ideal observers; in accordance with the efficient coding hypothesis, the informational resources of the human visual system are constrained. The objective of this study is to determine which dimensions of the trico planes are independently processed.

216

In this study, we examined a subset of the possible ternary texture space. It consisted of 13 texture planes (of a possible 66). There were 30 textures drawn from each plane, defined by 6 vectors (rays) centred at the origin; each ray had 6 predefined, equal levels of decorrelation. The total number of textures evaluated was therefore 390. Because of the large number of textures available for study, a logical way to gather performance functions was to use a crowdsourcing service. We developed a crowdsourcing method for evaluating ternary texture discrimination, based on that previously described for binary isotrigons (Seamons et al., 2015a). The laboratory testing phase was composed of 468 replications. The live (crowdsourced) studies were implemented in two phases containing 532 ("Live1") and 427 unique Workers ("Live2"). In each live study, 32 replications of each texture type were tested, making a total of 12,480 per live study.

Perceptual salience was found to vary for each image statistic examined. We observed the textures to have a salience rank order of gamma > beta_hv > beta_diag > alpha > theta, which is congruent with that previously reported for related binary stochastic textures (Victor and Conte, 2012). The most salient statistic (gamma) is defined by the pixel-by-pixel mean and variance. Second-order statistics affect the spatial-frequency content of the stimulus and are thus readily detectable by linear filters. However, that fourth-order correlations (alphas) are more salient than third-order correlations (thetas) is less readily explicable.

We also examined performance along vectors between ternary noise and 50:50 mixtures of the three canonical textures per trico-plane: white:black, black:grey, and white:grey. For the gammas and betas, the white:black and

grey-bias directions were consistently the least salient. For the thetas and alphas, the black:grey and grey:white directions were consistently the least salient. The observed differences reflect the sensitivities and limitations of neural processing and are therefore a manifestation of efficient coding.

We hypothesised that the grey tokens may confer non-salience. Indeed, for the gammas and betas, the grey-bias was consistently the second least salient. However, this relationship did not hold for thetas or alphas.

The effect of texture presentation *order* was also examined. In Live1, textures were presented in a pseudo-random order (random steps from within a single trico plane). In Live2, the textures were presented in step order along each ray, starting with the most salient example. Counter-intuitively, the order of presentation *did not* significantly affect texture discrimination performance. An analysis of 31 repeat Workers found evidence of learning for the beta textures, whereas performance for the gammas, thetas and alphas appeared to be already maximal.

4.2 Introduction

The natural world is rich with texture and the surface of any visible object is textured at some scale. We have an intuitive sense of what texture means, but it is harder to define in precise terms. According to one definition, texture is an *organised area phenomenon* which can be decomposed into *primitives* with specific spatial distributions (Haralick, 1979). However, textures are often highly variable and demonstrate contradictory properties that challenge strict definition (Landy and Graham, 2004, Bergen and Adelson, 1991,

Adelson, 2001). As Victor points out, "…basic visual judgments are fundamentally statistical in nature…" and, in this study, our examination of human texture perception will concentrate on image statistics (Victor and Conte, 2012).

When studying human texture perception, the selection of appropriate visual stimuli is challenging. The space of signals that the visual system encounters is rich and varied; thus, many different image statistics could potentially drive visual judgments. Mathematically-convenient stimuli, such as those built from white noise, may fail to sample features that are present in natural scenes (David and Gallant, 2005). Conversely, naturalistic stimuli encapsulate the required image features, but make subsequent analysis intractable as their image statistics are complex and interwoven.

What is required is a statistically well-principled set of stimuli in which multiple image statistics can be *independently* controlled. As we will discuss in detail below (Section 4.3.3), the stochastic ternary textures developed by Victor et al. meet this criterion (Victor and Conte, 2012). The number of ternary textures available for study is large (Maddess et al., 2007, Victor et al., 2013, Victor and Conte, 2012). One sensible way to gather performance functions for such a large number of textures is to use a crowdsourcing service to carry out distributed visual psychometric experiments (Paolacci et al., 2010).

Crowdsourcing websites coordinate the supply and demand of tasks that require human input. Launched in 2005, Amazon Mechanical Turk (mTurk) has developed into the largest crowdsourcing platform (Ross et al., 2010, Pontin, 2007). MTurk provides the elements required to conduct research

studies: a large, persistent pool of research subjects, an integrated payment system, a streamlined process for study design, recruitment, and data collection (Paolacci et al., 2010). Therefore, mTurk has developed into a labour market for tasks, termed Human Intelligence Tasks (HITs), which vary from surveys and language translations, to psychometric experiments (Mason and Suri, 2012). We have discussed the use of mTurk for visual psychometric research in detail in Chapter 1 (Section 1.6) and in Chapter 3 (Seamons et al., 2015a).

A concern with a novel platform such as mTurk is whether the data produced is of high quality. Due to perceived technical limitations, few visual psychometric studies using mTurk have been published (Heer et al., 2010, Cole et al., 2009, Freeman et al., 2013, Seamons et al., 2015a). In our previous study (Chapter 3), we discussed the development and rigorous evaluation of a binary isotrigon discrimination task using crowdsourcing (Seamons et al., 2015a). We showed that a binary isotrigon discrimination test based on mTurk was robust across a variety of platforms. Texture discrimination data was gathered from 121 subjects and compared to two independent laboratory data sets. Based on Pearson's correlation and coefficients of repeatability, mTurk was shown to be capable of producing laboratory quality data. Factor analysis was also performed and indicated a lower bound of 3-4 independent texture discrimination factors, consistent with previous published studies (Seamons et al., 2015b, Maddess and Nagai, 2001, Maddess et al., 2007), in addition to an unpublished study by our group (Coy, 2014). This is encourging considering the variety of browsers, screen sizes, screen resolutions and operating system recorded

(Seamons et al., 2015a). In this study, we will describe the development, testing and implementation of a modified version of the HIT previously described (Seamons et al., 2015a).

The objective of this study is to determine which dimensions of the ternary texture space are independently processed. The efficient coding hypothesis states that neurons have limited energy budgets and information capacities; therefore, they should not waste energy nor bandwidth by transmitting *redundant* information (Barlow, 1961, Barlow, 2001). To put it another way, in order to maximise available computing resources, neurons should encode as much *functional* information as is structurally and energetically sustainable. Reducing informational redundancy potentially increases the amount of information that can be transferred through the channels of the visual system.

As we will discuss below in Section 4.3.3, the ternary textures that will be used in this study can be visualised as *trico axes* (Figure 3); within the trico axes, to an ideal observer, symmetrical positions in all directions should be equally salient (Victor and Conte, 2012). The human visual system, however, has limited resources and therefore must employ mechanisms to filter out redundant or non-salient information, but retain that which is behaviourally relevant (Barlow, 2001, Barlow, 1963). If efficient coding holds, we would therefore expect some directions within the trico axes to be more informative than others. By using crowdsourcing to gather texture discrimination functions from a large number of subjects, we may be able to identify the maximally informative directions and gain an insight into the underlying mechanisms. In the case of binary isotrigon textures, a similar approach

demonstrated that salient textures with higher-order correlations echo features that confer salience in natural textures (Tkacik et al., 2010).

The ternary textures used here differ from their binary counterparts previously studied (Seamons et al., 2015a, Seamons et al., 2015b, Maddess and Nagai, 2001, Maddess et al., 2004) by the presence of a third token, grey. Therefore, a natural question is whether the grey element of ternary textures has special properties with regard to texture perception. Indications that grey is perceived differently come from perceptual errors, such as brightness induction and brightness assimilation, as discussed in Chapter 1 (Section 1.5.3) (Cornsweet, 1970, White, 1979). Cornsweet championed the idea that the simultaneous brightness contrast illusion resulted from reciprocal interactions between retinal neurons (Cornsweet, 1970). Although appealingly simple, these types of "top-down" approaches to brightness perception have now been supplanted by models based on low-level spatial filtering mechanisms combined with neural gain control (Dakin and Bex, 2003, Blakeslee and McCourt, 2008, Blakeslee and McCourt, 2004, Blakeslee and McCourt, 2012).

Several previous studies are directly relevant (Maddess et al., 2007, Victor and Conte, 2012, Victor et al., 2013, Maddess et al., 2004). Maddess et al. described a set of deterministic ternary textures (Maddess et al., 2004) and found that ternary isotrigon discrimination performance can be simulated using a bi-quadratic model (Maddess et al., 2007). Victor et al. investigated *binary* stochastic texture performance in their 2012 paper (Victor and Conte, 2012) using 6 normal subjects ages 25 to 51, 5 of which were experienced in visual psychometric tasks (Victor and Conte, 2012). This work was followed

up by Victor et al. in a 2013 study which focused on the beta and theta *binary* stochastic textures (Victor et al., 2013). Victor et al. also completed a study in 2005 which utilized *combinations* of binary stochastic textures (such as gamma-alphas) (Victor et al., 2005). Where possible, we will make direct comparisons with these studies in the Discussion below.

4.3 Materials and Methods

4.3.1 Methods Overview

This study can be divided into four phases. A summary of the study methods is shown below (Figure 1). In the Lab phase, the ternary HIT was developed and tested using the Amazon Mechanical Turk Sandbox environment, i.e. not using crowdsourced subjects. The development procedure has been described in detail in Chapter 3 and in (Seamons et al., 2015a) and ternary texture-specific differences are described in Section 4.3.4 below. Three subjects then completed a series of 468 different test HITs in order to determine the texture step (decorrelation) levels to be used in the Live study phases (described in Section 4.3.5). For each of the 13 textures being studied, 6 initial signal levels (steps) and 6 predefined rays were used.

Lab Phase
- Ternary HIT developed
- Destructive testing
- Step levels calculated using 3 subjects (468 HITs)

Live 1
- 13 texture families
- 12,480 HITs
- 532 unique workers
- Steps presented in random order within a trico plane
- Levels as calculated in Lab Phase

Live 2
- 13 texture families
- 12,480 HIT-equivalents
- 427 unique workers
- Steps presented in Ray order within a trico plane
- Levels adjusted from Live1 to improve signal to noise ratio

Analysis
- Live1 summary statistics
- Live2 summary statistics
- Live1 performance functions
- Live2 performance functions
- Comparison of Live1/2 performance
- Performance of repeat workers (learning?)
- Live2 Weibull functions
- Live 2 isocontours

Figure 1. Method summary for this study. The study can be divided into four phases: Lab, Live1, Live2 and Analysis.

The Live1 and Live2 phases followed the Lab phase (and are described in Section 4.3.6) in which the HIT was posted on mTurk and performance functions were gathered from a large number of naïve subjects. In Live1, the order of texture presentation was pseudo-random (random within a single trico plane). In Live2, textures were presented in step-order along each ray (see Sections 4.3.4 and 4.3.6). Based on an early analysis of the Live1 data, the step levels were adjusted in Live2 to improve the reliablity of the results.

Following the Live study phases, a detailed analysis of the performance functions was undertaken, i.e. the proablity correct as a function of the signal

(decorrealtion) level. This included a comparison of Live1 and Live2 data, which informed us about the significance of sequential texture presentation (Section 4.3.5); an analysis of repeat Workers and thus learning, which has been previously described for binary isotrigon textures (Maddess and Nagai, 2001, Taylor et al., 2008, Coy, 2014) (Section 4.3.6); and the derivation of iso-performance contours ("isocontours") from the Live2 data (Sections 4.3.7 and 4.3.8), which is the primary focus of this study.

4.3.2 Subjects

During the Lab phase, testing was conducted using 3 normal subjects (two male, one female). All subjects had normal or corrected-normal visual acuity. The subjects had some prior experience of texture discrimination tasks.

The Live1 and Live2 phases were implemented using the Amazon Mechanical Turk crowdsourcing platform. On the HIT page, all subjects were explicitly instructed to download and read a subject consent form before participating in the study. All subjects were instructed that, if they did not agree to the terms outlined in the subject consent form, they should not participate. Subjects wishing to participate in the study then indicated their consent via a button click. We confirm that the named ethics committees, detailed below, specifically approved this study protocol.

The ethical aspects of this research were approved by the ANU Human Research Ethics Committee under two ANU Human Ethics protocols: one for the Lab experiments and one for the Live1 and Live2 experiments (2010/194, 2014/237). As it was a joint experiment, a mirror protocol for the Live1/Live2

experiments was also approved at Weill Cornell Medical College (0904010359-A003). The experiments were conducted in accordance with the Declaration of Helsinki.

4.3.3 Ternary Textures

A large number of statistically constrained ternary textures have been generated (Maddess et al., 2004, Maddess et al., 2007, Taylor et al., 2008). The most common methods for creating them include a deterministic method (Maddess et al., 2004, Maddess et al., 2007) and a stochastic method, the latter of which has previously been described for the binary textures (Victor and Conte, 2012). Whereas the binary textures previously discussed (in Chapters 2 and 3) are composed of black and white pixels, the ternary textures have three levels: black, white and the mean luminance grey (contrasts -1, 1 and 0). The full range of ternary textures is larger than that of the binary textures.

Various types of statistically constrained textures, including isotrigon textures, can be produced using the stochastic method, which was developed by Victor et al. and has been previously described for binary textures (Victor and Conte, 2012). This process is readily adaptable for producing ternary textures. The statistics of the stochastic textures can be controlled in more sophisticated ways, but they share the same fundamental statistics as their deterministic counterparts. This was demonstrated by comparing the *histograms of primitives* (HOPs) of the two texture families examined(Maddess, 2015). HOPs count the number of small, unique texture samples (primitives, typically 3x3 pixel samples) within a texture example

and thereby uniquely identify a texture type (Maddess et al., 2007). The deterministic process for producing ternary textures, and its comparison to the stochastic process using HOPs, is discussed further in Appendix 6.6.

Palettes of stochastic ternary textures are created in a space referred to as a *trico-plane*, where the centre location contains ternary noise; the three corners of the space are texture archetypes and the remaining space is filled with quantitatively derived mixtures (Victor and Conte, 2012). The ternary textures are classified according to the number of pixels that are constrained within a 2x2 pixel clique. If a single pixel is constrained, the textures are termed "gamma". If 2 pixels are constrained, they are termed "beta_hv" or "beta_diag" (depending on the orientation of the constrained pixels). If triplets of pixels are constrained they are termed "theta" and quadruplets of pixels constrained are "alpha" (Victor et al., 2013, Victor and Conte, 2012). These pixel constraints are illustrated in Figure 2 below. Texture mixtures can also be created (for example, gamma-theta).

By generating the stimuli in this manner, local image statistics can be specified, whilst *long-range statistics* are maintained as random as possible (Victor et al., 2013). Note that of the 10 stimuli types specified in Figure 2, 8 potentially carry information about orientation (the betas and thetas). The gamma and alpha textures are *not oriented*.

Gamma_A

A	0
0	0

Beta_hv_AB

A	B
0	0

Beta_hv_AC

A	0
C	0

Beta_diag_AD

A	0
0	D

Beta_diag_BC

0	B
C	0

Theta_BCD

0	B
C	D

Theta_ACD

A	0
C	D

Theta_ABC

A	B
C	0

Theta_ABD

A	B
0	D

Alpha_ABCD

A	B
C	D
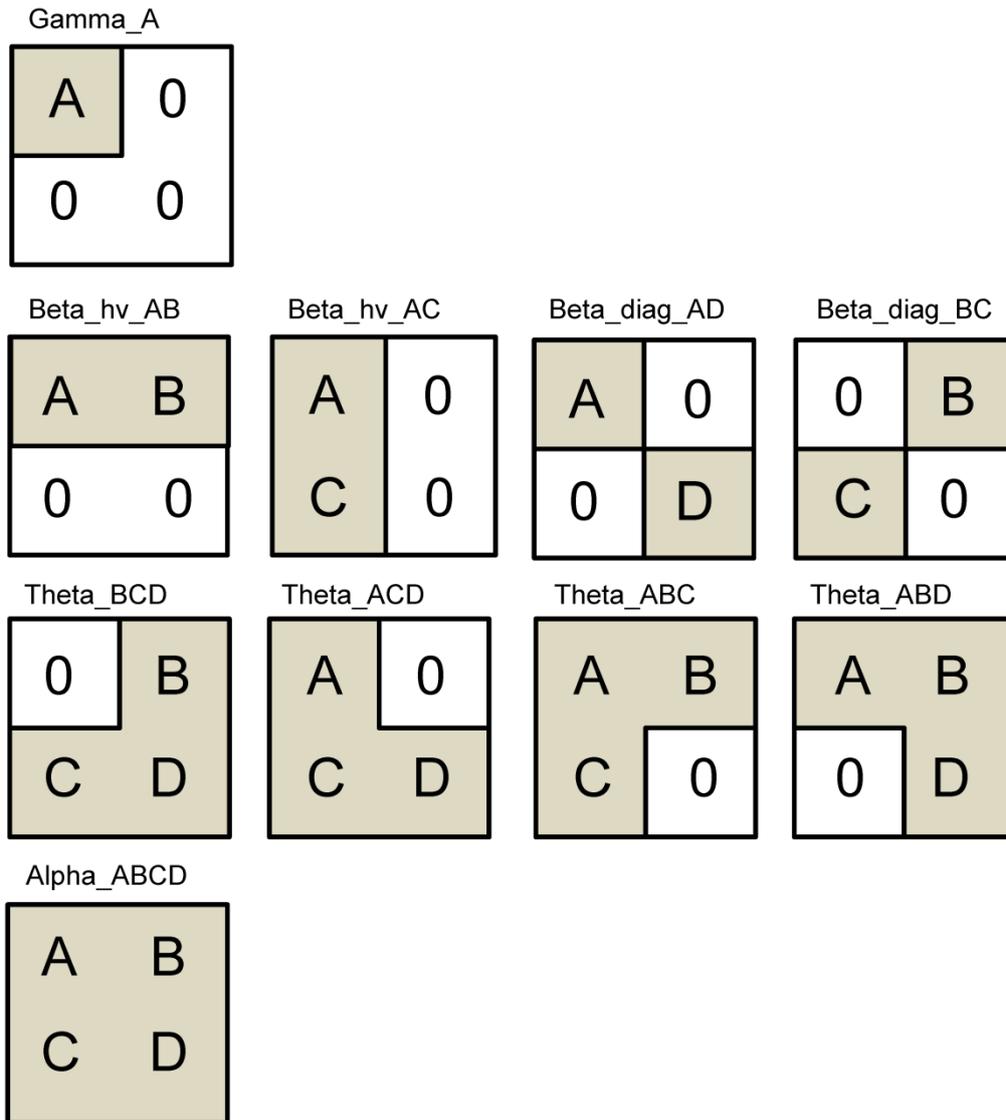
Figure 2. Illustration of the pixels constrained within the ternary texture templates. From top to bottom, these correspond to constraints on first, second, third, and fourth order spatial correlations. The specific texture types are labelled.

The stimulus is further specified by different *linear combinations* of pixels. The linear combination defines how the constrained pixels from the template interact. For example, alpha_ABCD1111 indicates that pixels ABCD are constrained and interact with the combination (rule): modulo 3 of (1*A +1*B +

1*C + 1*D). The pixel values are actually {0, 1, 2}, and later are mapped to the contrasts {-1, 0, 1}. The total number of texture planes (based on combinations of constrained pixels and linear combinations) is 66; the texture planes are shown in Table 1 below. In fact, just 33 texture planes are really unique given contrast inversion and in this study, we tested a reasonable subset of the texture types (13), which are highlighted in yellow; taking into consideration rotational and reflectional symmetry, these 13 represent the smallest set of clearly distinct textures.

Gamma	Beta_hv	Beta_diag	Theta	Alpha
Gamma_A1	Beta_hv_AB11	Beta_diag_AD11	Theta_BCD111	Alpha_ABCD1111
Gamma_A2	Beta_hv_AB12	Beta_diag_AD12	Theta_BCD112	Alpha_ABCD1112
	Beta_hv_AB21	Beta_diag_AD21	Theta_BCD121	Alpha_ABCD1122
	Beta_hv_AB22	Beta_diag_AD22	Theta_BCD122	Alpha_ABCD1221
	Beta_hv_AC11	Beta_diag_BC11	Theta_BCD211	Alpha_ABCD1121
	Beta_hv_AC12	Beta_diag_BC12	Theta_BCD221	Alpha_ABCD1211
	Beta_hv_AC21	Beta_diag_BC21	Theta_BCD212	Alpha_ABCD2111
	Beta_hv_AC22	Beta_diag_BC22	Theta_BCD222	Alpha_ABCD2211
			Theta_ACD111	Alpha_ABCD1212
			Theta_ACD112	Alpha_ABCD2112
			Theta_ACD121	Alpha_ABCD2121
			Theta_ACD211	Alpha_ABCD1222
			Theta_ACD122	Alpha_ABCD2221
			Theta_ACD221	Alpha_ABCD2212
			Theta_ACD212	Alpha_ABCD2122
			Theta_ACD222	Alpha_ABCD2222
			Theta_ABC111	
			Theta_ABC112	
			Theta_ABC121	
			Theta_ABC211	
			Theta_ABC122	
			Theta_ABC221	
			Theta_ABC212	
			Theta_ABC222	
			Theta_ABD111	
			Theta_ABD112	
			Theta_ABD121	
			Theta_ABD211	
			Theta_ABD122	
			Theta_ABD221	
			Theta_ABD212	
			Theta_ABD222	

Table 1. The total number of texture planes available for study (66). Taking into to consideration rotational and reflectional symmetry, the number of clearly distinct textures is 13; it is these that we selected for further study and these textures are highlighted in yellow.

The simplest stochastic manipulation is illustrated in Figure 3: controlling the statistics of single pixels without reference to any other pixels (gamma) (Victor and Conte, 2012), i.e. controlling the first order statistics. Here the canonical "textures" are purely black, white and grey images with different amounts of ternary noise, i.e. random pixels of equal probability of being black white or grey.

As a way of representing this set of stimuli, the 3 orthogonal colour axes (black, grey and white) are projected onto a triangular plane via use of the three complex cube root of -1 as the axis directions; this forms the axes of an equilateral triangle where the length of each axis vector is 1 (Victor and Conte, 2012). We refer to this arrangement as a *trico plane* and the 3 principle axes are the *trico axes*. The trico plane is a contrivance which allows what is a three-dimensional image space to be displayed in two-dimensions.

We refer to lines in the trico plane that traverse from the centre point to the borders as "rays". In order to explore the trico space, we have selected 6 rays across each trico plane: the 3 principle axes and the perpendicular bisectors of the borders. Along each ray that is a trico axis, the frequency of one of the three pixel levels (white, grey, or black) increases to the mid-point, where it becomes the dominant colouring. At the centre of the trico plane, the origin, all three colourings of pixels are equally likely; the central texture is therefore uniform ternary noise (Figure 3).
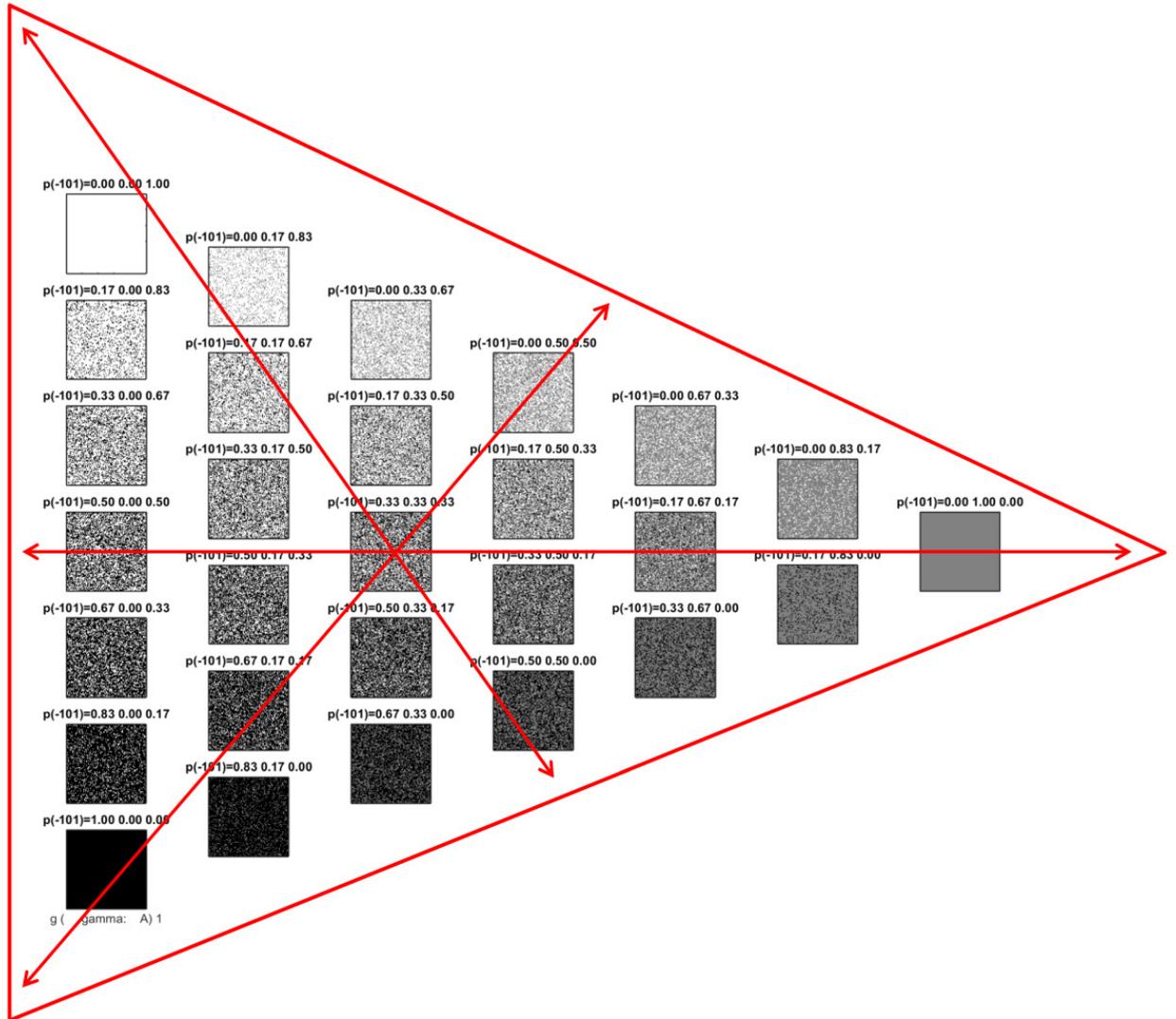
Figure 3. Example trico axis for the gamma ternary textures produced using the stochastic method, as described by Victor et al. for binary textures (Victor and Conte, 2012). The 3 orthogonal colour axes (black, grey and white) are projected onto a triangular plane. Along each outside edge, the frequency of one of the three pixel levels increases to the mid-point. The central space contains ternary noise. The six rays are shown in red; note that they extend from the origin. The captions above each ternary texture indicate the pixel

231

probabilities for black, grey, and white respectively. The titles above each texture example give the probabilities of each of the 3 contrasts.

Along each ray, the number of subdivisions (and hence the number of textures), are referred to as "steps" of signal level. In broad terms, the distance of the step from the centre of the trico axis affects its salience. To put it another way, the step is a measure of decorrelation between 1 and 0, with 0 being ternary noise (the central axis position) and 1 indicating no decorrelation (positions at the ends of each ray). In this context, it is clear that the rays and steps we use in this study, although numerous, are but a subset of those possible within the ternary texture space.

An important feature of the trico axes is that symmetrical changes in all directions should be equally detectable to the ideal observer (Victor and Conte, 2012). The human visual system, however, has limited resources as discussed by other authors in the context information theory (see Chapter 1, Section 1.4) (Barlow, 2001, Barlow, 1963). The visual system must therefore employ mechanisms to filter out redundant or non-salient information, but retain that which is behaviourally relevant (Barlow, 2001, Barlow, 1963). In informational terms, we might therefore expect some directions within the trico axes to be more informative than others.

In order to identify a given ternary texture, a specific nomenclature has been adopted. For example, "alpha_ABCD1111_B_0.0433" indicates that the ternary texture family is "alpha"; the arrangement of the pixels constrained are "ABCD1111", where "1111" indicates the linear combination; the pixel level (colour-bias) is "B" (black); and the step is 0.0433, which indicates a

232

high level of decorrelation, i.e. the pattern occurs 0.0433 of the distance from the centre along the principle axis (ray), pointing to the black-bias corner position.

An alternative way to visualize the ternary textures, which emphasizes the effect of the step (decorrelation), is shown in Figure 4. For each texture family (in this case gamma) we can display a grid of textures for the tested rays; these rays are shown in Figure 3 above. The textures are displayed with increasing levels of decorrelation (steps). So, in the example shown in Figure 4, 10 levels of decorrelation are displayed, from left (1, no decorrelation) to right (0, full decorrelation). This corresponds to gamma textures that are derived from starting from an outer position on the trico axis and then working in towards the middle in 10 steps, along each ray.

This method of display reinforces the strength of the stochastic method: by adjusting the level of decorrelation, the salience of each ternary texture can be independently controlled. When used in psychometric studies, this allows for the calculation of performance functions, whilst avoiding performance *saturation*. That is, by adjusting the level decorrelation, we can identify the noise levels that yield a criterion probability correct (e.g. a limen of 70%). This, and the ability to mix different types of ternary texture, is the primary motivation for using the stochastic method. After gathering performance functions for the reasonable dimensions within the trico axes, they can be used to evaluate perceptual sensitivity within ternary texture space (Victor et al., 2005, Victor et al., 2013, Victor and Conte, 2012).
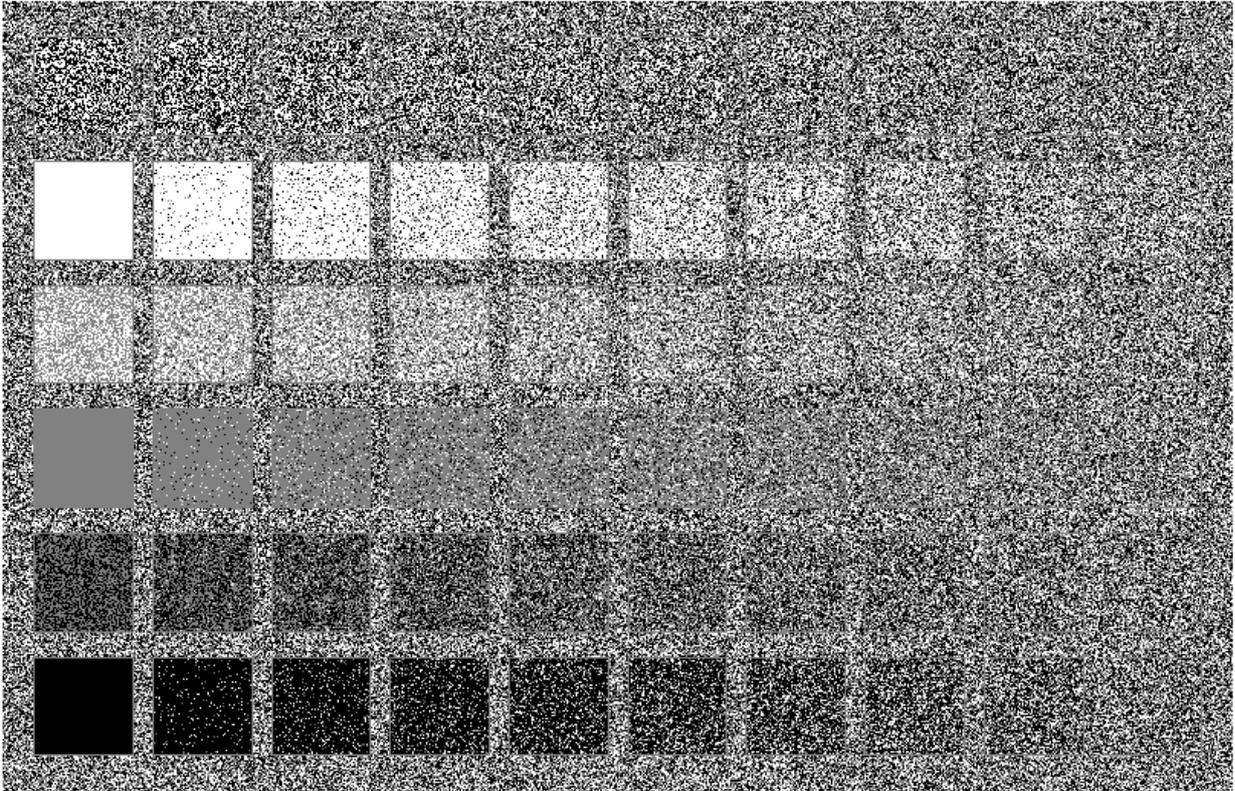
Figure 4. Examples of ternary gamma textures, displayed to emphasise the effect of decorrelation on texture salience. Six rays are represented by the Y axis, with *binary* noise on the top row, i.e. the perpendicular bisector between all white and all black. The rays correspond to those shown in Figure 3 above. Each step is represented by displacement along the X axis and in this case 10 equal steps are shown. Step 10 is equivalent to zero decorrelation (the leftmost position) and step 1 is equivalent to full decorrelation (the rightmost position). Compare the visual effect of increasing decorrelation with the ternary noise background. Most obviously humans have difficulty discriminating *binary* noise (the perpendicular bisector of black and white dominated textures in Fig. 3) from ternary noise. Grey-black and Grey-white textures are far more discriminable

from ternary noise. Apparently, the associated luminance cues are highly salient compared to the equiluminant black and white noise case.

The ternary textures that were selected for this study are shown in Figures 5 and 6 below. For each selected texture family, 6 rays were predefined. Along each ray, five steps were defined (six in the preliminary Lab experiments), where the step size was defined to maximise the signal to noise ratio based on Lab phase of testing (see Section 4.4.1). In total, 390 different ternary textures were used (13 trico axes, 6 rays on each axis, 5 steps on each ray).

Gamma_A1   **Beta_hv_AB11**   **Beta_hv_AB12**

Beta_diag_AD11   **Beta_diag_AD12**
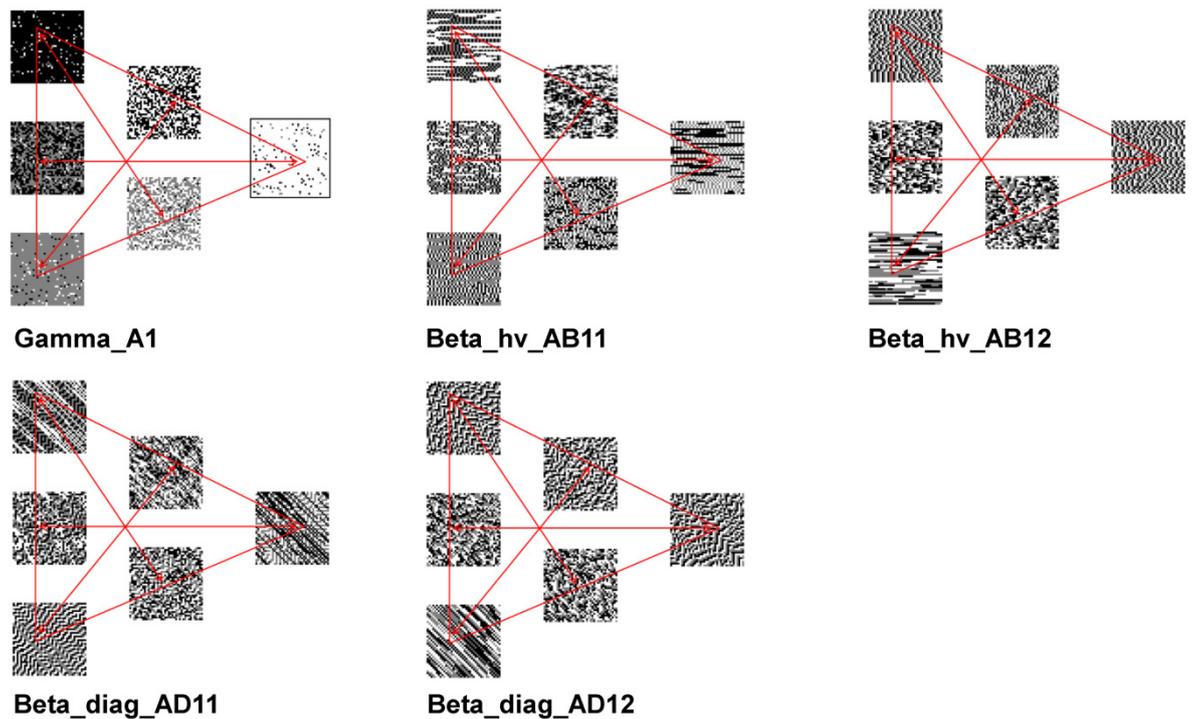
Figure 5. Reduced trico axes which illustrate 5 of the 13 ternary texture families used in this study. Note that except for Gamma_A1, only the fully correlated textures are shown. In the case of Gamma_A1, 10% decorrelated textures are shown to illustrate the effect of adding a small amount of ternary noise. The 6 rays are

shown as red vectors emanating from the origin. The origin is ternary noise, but is not shown.

Theta_BCD111   **Theta_BCD112**   **Theta_BCD121**

Theta_BCD122   **Alpha_ABCD1111**   **Alpha_ABCD1112**

Alpha_ABCD1122   **Alpha_ABCD1221**
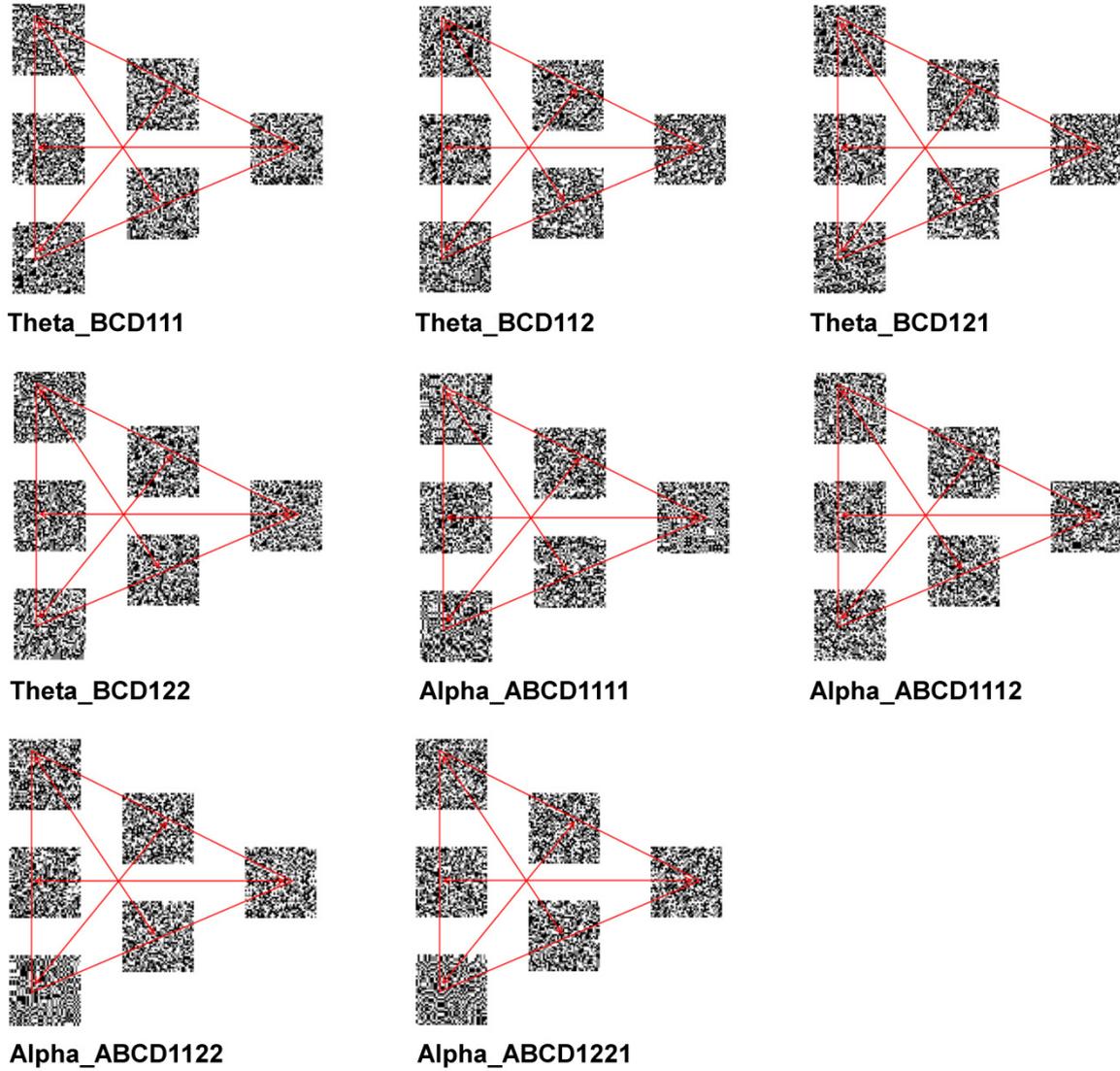
Figure 6. Reduced trico axes which illustrate the remaining 8 of the 13 ternary texture families used in this study. Note that only the fully correlated textures are shown.

4.3.4 Amazon Mechanical Turk HIT

An external HIT was developed, which included a ternary texture discrimination task, based on that previously described for binary isotrigon textures (see Chapter 3) (Seamons et al., 2015a). Whereas internal HITs are quick and easy to develop using Amazon templates, and are hosted on Amazon servers, they lack the advanced features that we required. External HITs are more flexible and may include text, images, movies, and Flash or Java applications (Heer et al., 2010, Paolacci et al., 2010, Mason and Suri, 2012). Amazon S3 was used to host the texture repository and other associated files. All development and testing of the HIT was performed using the Amazon Mechanical Turk Sandbox environment.

The structure of the HIT was as follows (a screen capture of the HIT page is included in Appendix 6.7). In the first instance, Workers were required to complete a subject consent form. This was followed by detailed instructions on how to conduct the texture discrimination task, along with a slideshow containing examples of the ternary textures to be presented. There is a strong association between the clarity of the HIT instructions and data quality (Crump et al., 2013).

As with the binary isotrigon HIT previously reported (Seamons et al., 2015a), the HIT included a monitor calibration step. This captured the resolution in dots per inch (DPI) of the Worker's screen, as well as dictating the viewing distance that should be used during the task. Workers were informed that they should complete the calibration at least once (and each time their work platform changed). Other information, such as the operating system,

237

browser, and Worker ID, were automatically captured for subsequent analysis. Some basic demographic information was also recorded including gender and age group.

The compulsory calibration step allowed us to control the effective monitor gamma, independently from local settings. Before starting the HIT, Workers were asked look at the image below (Figure 7) and pick the grey rectangle in the image that was closest to the average brightness of the black and white textured surround, i.e. the mid-grey. A noise background was chosen because we found that it was highly resistant to spatial aliasing effects (Yellott, 1983). The checks of the background were also lightly spatially low-pass filtered to further reduce its high spatial frequency content. Thus, the calibration pattern could be resized over a very broad range without causing Moire patterns (that can create false grey levels in the background). Checking of the corresponding tick-box caused a Cascading Style Sheets (CSS) filter to be applied to the HIT page, which normalised the gamma of the Worker's monitor. The calibration step was enforced by locking the HIT Start button until this step was completed. Note that the use of CSS filters required extensive cross-browser testing as they are not universally supported.
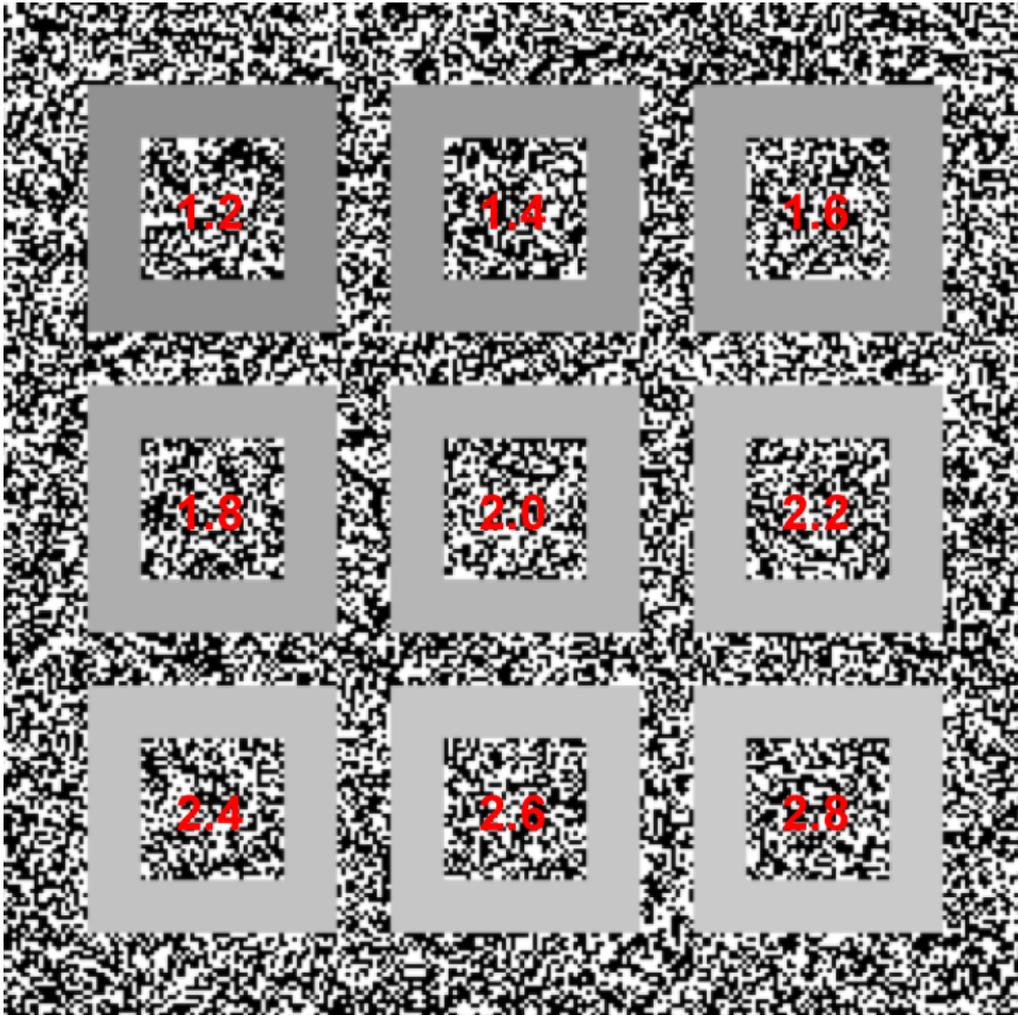
Figure 7. Gamma calibration task developed for the ternary texture HIT. Workers selected the grey rectangle that is closest to the average brightness of the black and white surround, i.e. the mid-grey. This effectively allowed us to control the relative gamma of the HIT presented texture, independently from local monitor settings. In the live studies reported below, the minimum gamma recorded was 1.45 and the maximum was 2.01 (see Section 4.4.2 below).

Directly above the texture discrimination task, a priming strip was displayed. The strip was implemented to change *dynamically* to illustrate the texture family being tested. Each priming strip constituted one ray and 5 examples,

corresponding to the 5 steps being tested. The steps were arranged in step order, from left to right (as in Figure 4 above). i.e.: the least decorrelated step (step 5) is shown in the leftmost position. The specific step being tested is highlighted in blue (Figure 8).
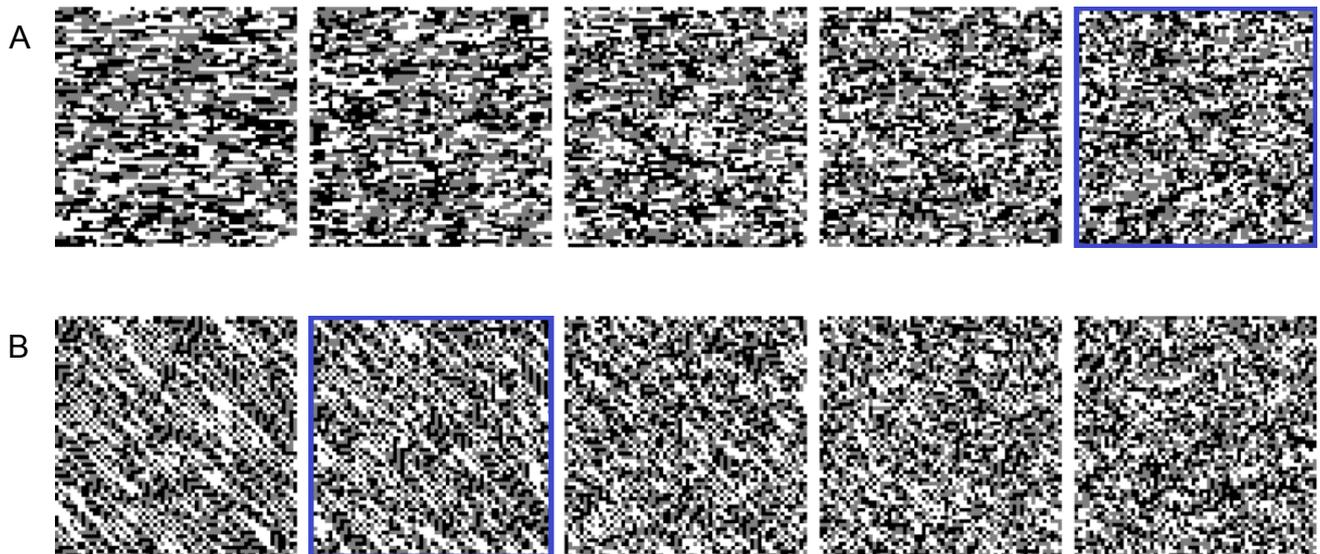
Figure 8. Examples of priming strips used during the ternary texture HIT. The textures in these examples are A: beta_hv_AB12 and B: beta_diag_AD11. Textures are displayed in step order (step 5 is leftmost). The texture type being tested is highlighted in blue. One priming strip occurs per HIT.

During the texture discrimination task, ternary textures were presented to each Worker in a semi-random order. Specifically, each Worker was allocated a randomly determined trico plane to work on, from the 13 planes being studied. Then, within this trico plane, they were presented with textures to discriminate in a *random order from within that plane* (Lab and Live1) or *in step order* (Live2). Thus, in Live2, the first ternary texture presented was the least decorrelated (step 5), followed by steps 4 to 1 *in order*; upon

completion of the ray, the next ray within the same trico plane was selected at random from those remaining. This difference between Live1 and Live2 allowed us to analyse *the effect of presentation order on learning*; the Live2 presentation schema might be reasonably expected to promote learning (and thus improve performance).

During the task, Workers were required to identify the position of a band of the test texture within a field of ternary noise (the Foreground condition), or the opposite, with the texture as the background and a band of noise (Background condition). There were four possible band positions, as shown in Figure 9. The band positions occurred in equal quantities and were presented in random order during the HIT. Thus, all experiments were four-alternative forced choice texture segregation tasks.

Subjects were told that the target was equally likely to appear in any one of four locations (top, right, bottom, left), and were instructed to maintain central fixation, rather than to attempt to scan the stimulus. This method was developed by Victor et al. (Victor et al., 2013) and was intended to provide a more complex and naturalistic stimuli than those traditionally used in visual perception experiments using gratings. As in natural stimuli, local features have multiple orientations, and multiple orientations can occur even at the same location (Victor et al., 2013). A secondary consequence of using Foreground and Background conditions is that it discouraged fixation on a single corner position, which could potentially result in performance of greater than chance (>25%).

Visual feedback for correct and incorrect responses was given during the task. Providing feedback to Workers promotes learning and has previously been identified as a factor in maintaining data quality (Dow et al., 2012).
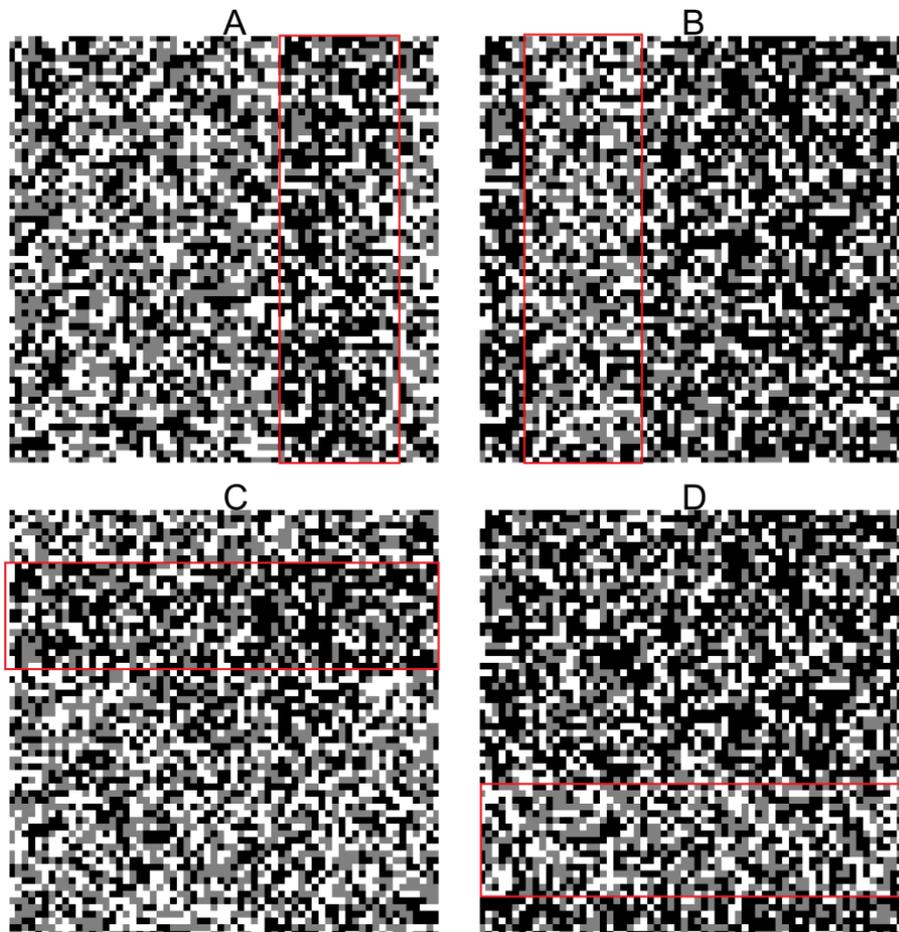
Figure 9. The four possible band positions in the ternary HIT. In this example, the ternary texture used is a highly salient example of a gamma texture (black-bias). The bands and conditions shown are as follows. A: right band Foreground. B: left band Background. C: top band Foreground. D: bottom band Background. For clarity, the bands have been outlined in red (bands were *not* highlighted during the HITs proper).

Each HIT was composed of 40 texture presentations. The first 20 had a presentation time of 2,000 ms (referred to as the "long" condition); the second 20 presentations lasted for 200 ms each ("short" condition). Victor et al. used presentation times of 120-160 ms (Victor et al., 2013, Victor and Conte, 2012). Previous studies have found texture discrimination of related textures to be pre-attentive and changes little for longer presentation times (Taylor et al., 2008). An unexpected consequence of the large number of texture presentations was that image pre-caching was required on the Worker's device; if textures were instead loaded on demand, presentation times would be dependent on network latency. Many tablets, such as iPads, have limited memory available for caching and were therefore excluded from the study.

In Lab and Live1 phases, one HIT corresponded to one step (40 texture presentations). In Live2, in order to implement sequential presentation, all 5 steps from one ray were incorporated into a single HIT, and then presented in step order; thus, each HIT in Live2 corresponded to 5 steps and 200 texture presentations. During analysis, Live2 HITs were simply divided into 5 "HIT-equivalents".

As with the binary isotrigon task previously reported (Seamons et al., 2015a), catch trials were incorporated to encourage Worker attentiveness and allow an assessment of data quality. Catch trials and/or qualification tasks have been found to be important for maintaining data quality (Heer et al., 2010, Kittur et al., 2008). In this study, highly salient examples of the gamma (black-bias) texture were used as catch trials. Examples of these textures are shown in Figure 9 above.

4.3.5 Lab Testing: Weibull Functions

After the implementation of the ternary HIT, the texture step (decorrelation) levels to be used in the Live1 study were determined. Our goal was to prevent performance saturation and maximise the signal to noise ratio during the Live study phases. If the texture discrimination task was too easy, subject performance will saturate. Conversely, if the textures were too difficult, noise will dominate performance.

Under laboratory conditions, 3 normal subjects (two male, one female) completed 468 HITs between them: one replication each of 13 texture families, 6 rays, and 6 steps, where the initial step levels were approximated. Note that 6 steps were used (as opposed to 5 in the Live phases) to improve the fitting of performance functions. All subjects had experience of isotrigon discrimination tasks and the HIT protocol used in this study.

The performance data was used to fit Weibull functions to the mean performance separately along each ray, using the procedure outlined in Victor et al. 2005 and subsequent studies (Victor et al., 2005, Victor et al., 2013, Victor and Conte, 2012). In order to fit the Weibull function, the initial Weibull parameters were produced based on a linear approximation; the parameters arising from that prediction were given to an iterative method (gradient descent) to achieve the final Weibull fit. Victor et al. previously reported Weibull shape parameters (the exponent) in the range 2.2 to 2.6 (Victor and Conte, 2012). Then, the inverse Weibull function was used to find the degree of decorrelation corresponding to a selected limen.

The cumulative distribution function of the Weibull distribution is defined as follows in Equation 1:

$$F(x; k, \lambda) = 1 - e^{-(x/\lambda)^k} \qquad \text{(Eq. 1)}$$

…for $x \geq 0$, and $F(x; k; \lambda) = 0$ for $x < 0$. The shape of the function is determined by the free parameters k, the shape (or steepness) parameter and λ, the scale (or centring) parameter. Figure 10 shows how various values of these free parameters affect the shape of the Weibull function.
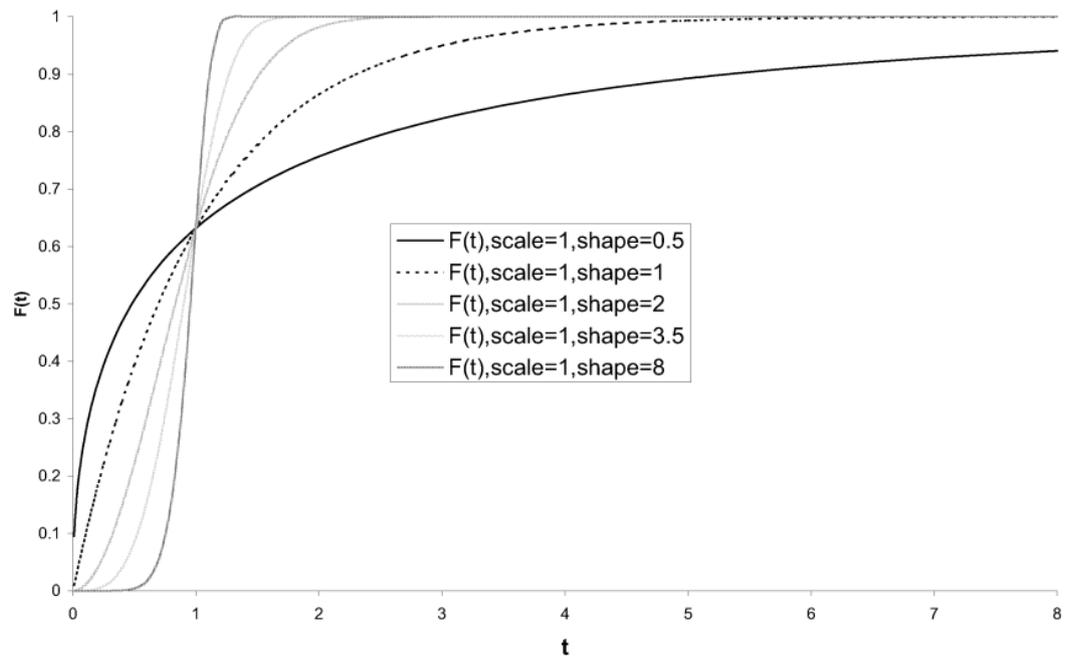
Figure 10. Example of how the cumulative distribution of the Weibull function varies according to the free parameters k (shape parameter) and λ (scale parameter) (Verma et al., 2010).

245

The value of the image statistic that yields performance at some defined point between floor and ceiling was then calculated for each texture and each ray. The threshold was set to predict a fraction correct (limen) of 0.625. Based on these predictions, the steps (decorrelation levels) for the Live1 ternary textures were established. Effectively, the step levels allow us to independently set the *difficulty* of the texture discrimination task for a given ray. After this testing phase, a repository of ternary textures for live testing was produced that should yield reasonable performance levels. The textures were uploaded to Amazon S3 and linked to the mTurk HIT implementation.

4.3.6 Live1 and Live2 Phases

Live1 was uploaded to Amazon mTurk on 5 December 2014 at 14:10:53 EST. In total, 32 HITs of each texture type were uploaded, each based on a different random number seed. As discussed above, 13 ternary texture families were selected, each with 5 steps (established on the basis of Weibull functions in the Lab phase) and 6 rays; this made for a total of 390 HIT types. 32 HITs of each type were posted, making a total of 12,480 HITs.

Live2 was uploaded to mTurk on 3 June 2015 at 14:38:33 EST. In total, 32 HITs of each texture type were uploaded: again, the same 13 ternary texture families were selected, each with 6 rays and 5 steps, established on the basis of Weibull functions, but *modified* based on Live1 results to improve Worker performance (discussed below in Section 4.4.5). To allow sequential presentation, the 5 steps were combined into a single HIT; that made a total of 78 HIT types. 32 HITs of each type were posted, making a total of 2,496

HITs. In post processing, each HIT was divided up into 5 HIT-equivalents (1 for each step), producing the same number of replications as in Live1 (12,480).

In Live1, 8 catch trials were randomly distributed throughout the HITs. Thus, the short and long presentation times could contain 0-8 catch trials. In Live2, the distribution of catch trials was adjusted so that each presentation condition contained 4 catch trials.

The compensation rate for each HIT in Live1 was set at $0.10 AUD. As each HIT in Live2 contained 5 steps, this was increased to $0.50 AUD in Live2. Based on the high quality data obtained from the binary isotrigon HIT previously described (Seamons et al., 2015a), a bonus payment system was implemented to encourage Workers to complete HITs for each of the texture families. Although recent research found that Workers have a reservation wage of $1.38 USD per hour, on ethical grounds the compensation rates were made consistent with that of conventional laboratory subjects (Ipeirotis, 2010a, Mason and Suri, 2012). Consistent with good experimental procedure, a feedback website was produced so study participants could track the progress of the study.

4.3.7 Isodiscrimination Contours

After gathering performance functions for the reasonable dimensions within the trico axes in the Live phases, this data was used to evaluate perceptual salience within the three-dimensional ternary texture space, projected onto

the trico plane. The procedure for doing this is outlined by Victor et al. (Victor et al., 2005, Victor et al., 2013, Victor and Conte, 2012). In the first instance, performance functions were fitted to Weibull functions in the process outlined in Section 4.3.5. The Weibull function parameters were then used to predict the step (decorrelation) level that was predicted to yield some criterion performance (termed "a_r"). This was taken as the threshold for a given fraction correct, in the corresponding direction of the trico plane. This analysis is carried out for the Live2 data in Sections 4.4.7 and 4.4.8 below.

The reciprocal, $1/a_r$, indicates perceptual sensitivity to changes in the direction of a single ray; when plotted in an "isodiscrimination contour" (or "isocontour ") this forms a visual representation of perceptual sensitivity for each trico plane (Victor and Conte, 2012). An example isocontour is shown in Figure 11. The axis of elongation (highest threshold) corresponds to the direction in which changes are *least salient*.

Where Weibull functions cannot be fitted, due to insufficient discrimination performance levels, the observed performance at zero decorrelation may instead be plotted; in this case, the axis of elongation corresponds to the direction in which changes are *most salient*. In these cases, performance functions were nowhere near saturation. Therefore, we obtained the unsaturated performance to a fixed signal level: 1.0. For clarity, these two forms of plot are presented and analysed separately (see Section 4.4.8 below).
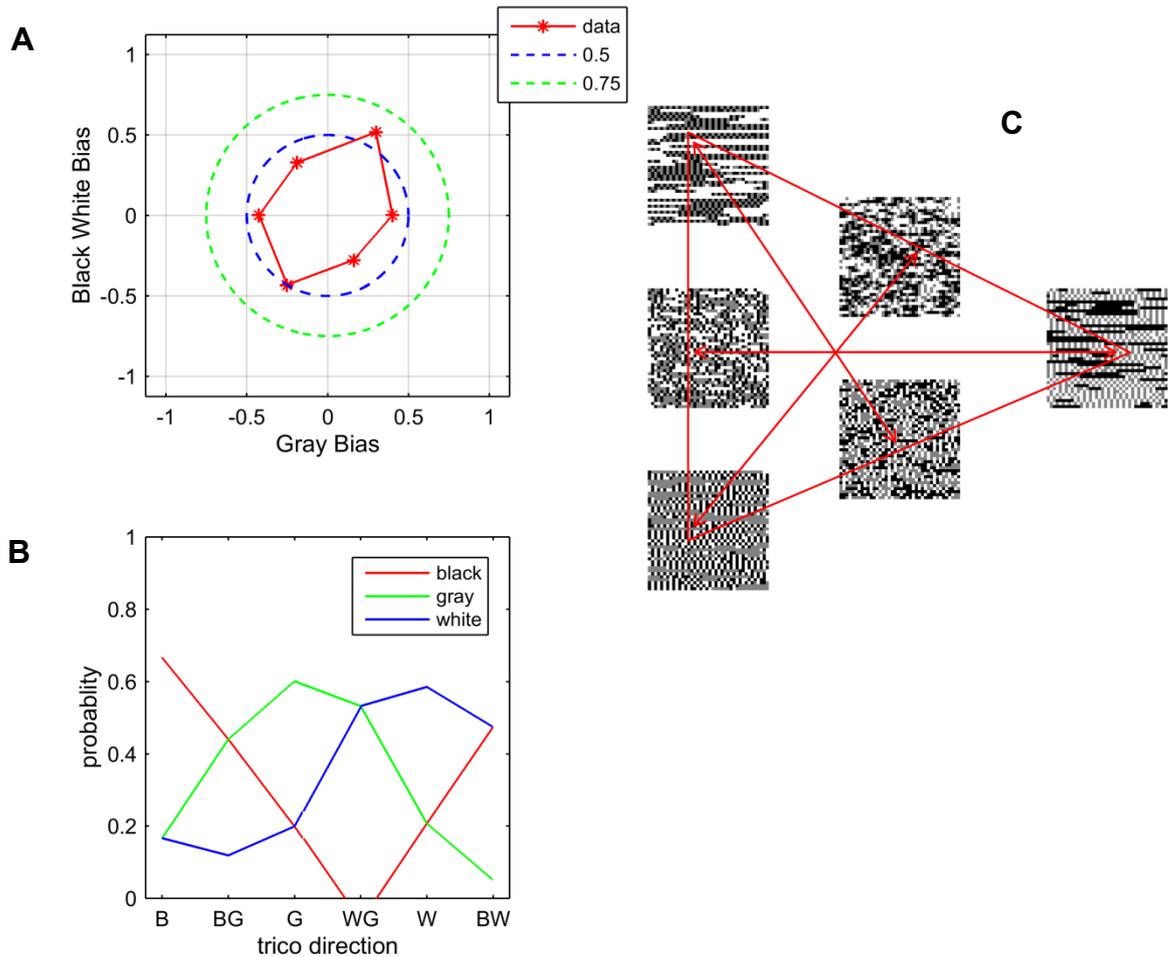
Figure 11. Example of an isodiscrimination contour for the beta_hv_AB11 textures. A: For each ray, the step levels have been plotted that yield a predicted performance of 60%. Thus, the axis of elongation corresponds to the direction in which changes are *least salient*. B: Corresponding plot of the pixel probabilities for each ray at the step levels that predict 60% performance. To assist the reader, the corresponding reduced trico plane is shown in C.

4.4 Results

4.4.1 Lab Phase: Estimation of Initial step Levels Using Weibull Functions

As described above (in Section 4.3.5), under laboratory conditions 3 normal subjects (two male, one female) completed 468 HITs between them using comparable platforms. All subjects had experience of isotrigon discrimination tasks and the HIT protocol used in this study. Weibull functions were plotted and step levels established which could be used to confidently predict 0.625 performance limen.

An example of the Weibull functions derived in Lab testing is shown in Figure 12. After Lab testing, a repository of ternary textures was produced using the step levels obtained. The textures were then uploaded to Amazon S3 and used in Live1. The actual step levels used in Live1 are summarised in Table 2 (in Section 4.4.5).
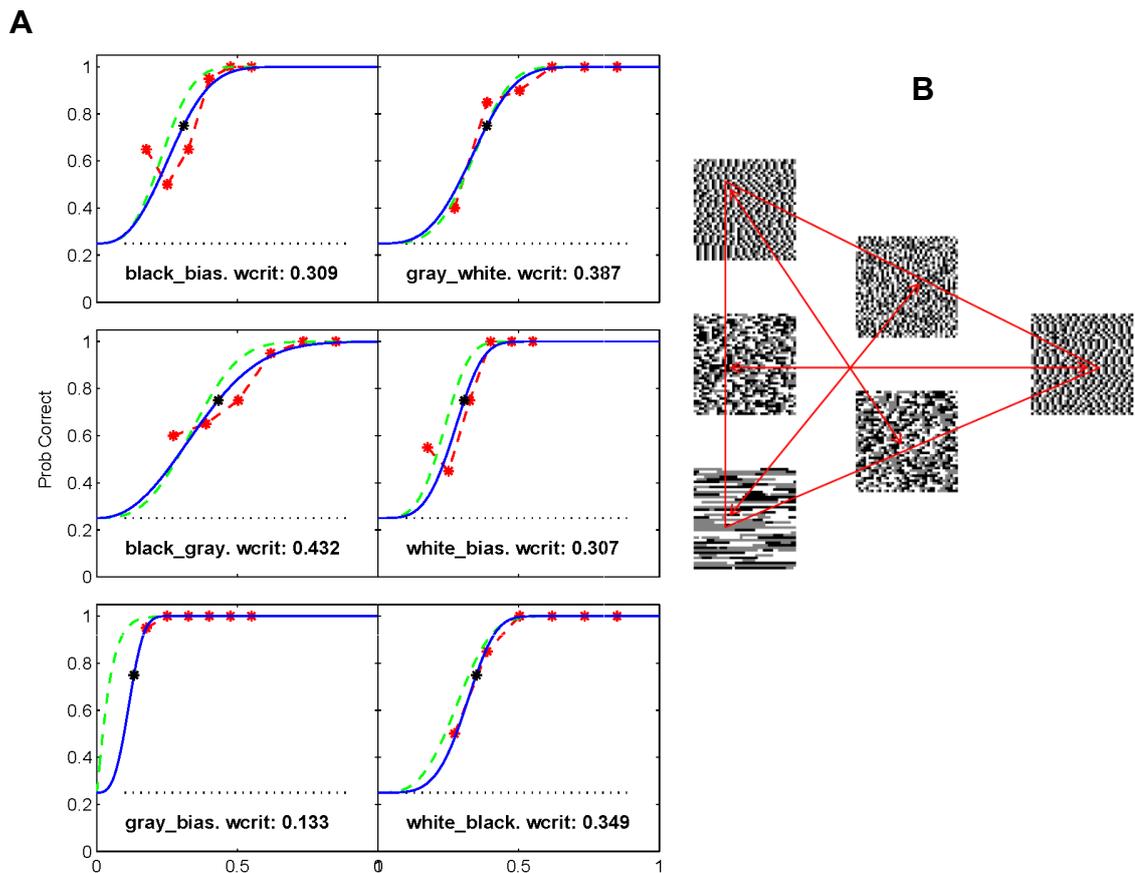
Figure 12A. Example Weibull functions derived from Lab testing for the beta_hv_AB12 textures. Using functions such as these, the step levels, "wcrit" (Weibull function critical value), were identified which predicted 0.625 performance. The wcrit values are indicated under each Weibull plot. The green dashed line indicates the initial prediction to the Weibull fit, given by a linear approximation; the parameters arising from that prediction were given to an iterative method to achieve the final blue fitted line. B: the corresponding reduced trico plane for the beta_hv_AB12 texture.

4.4.2 Live1 and Live2: Summary Statistics

Summary statistics for the Live1 and Live2 studies are shown in Figure 13. We will briefly consider some of the most interesting summary stats. Such reporting is significant, considering the dearth of published visual psychometric mTurk studies.

The first HIT of Live1 was completed on 5 December 2014 at 22:27:11 EST and the last completed on 31 December 2014 at 07:34:38 EST. Therefore, the total time to completion was +617.4 hours; that is one HIT per 2.97 minutes. A slightly reduced rate of HIT completion was observed between 24 December 2014 07:17:09 and 26 December 2014 06:42:44, presumably due to the Christmas holiday period. This is consistent with previous data on seasonal fluctuations in the mTurk workforce (Mason and Suri, 2012, Ipeirotis, 2010b) (Figure 13A). The maximum HIT completion time was 298.0 seconds, minimum 76.0 seconds, and mean 169.8.

The first HIT of Live2 was completed on 3 June 2015 at 14:49:47 EST and the last completed on 12 June 2015 at 18:48:50 EST. Therefore, the total time to completion was +220.0 hours; after taking into consideration that each HIT contained 5 steps, one "HIT-equivalent" was completed every 1.06 minutes (Figure 13A). The maximum HIT-equivalent completion time was 179.2 seconds, minimum 62.4 seconds, and mean 104.7.

In Live1, 532 unique Workers participated in the study. In Live2, 427 unique Workers participated. Across the 12,480 HITs of Live1, 9 different DPIs were recorded (including unknown). A DPI of 264 was most common with 2,577

HITs (20.7%). In Live2, 9 different DPIs were recorded and a DPI of 72 was most common with 2,555 HITs (20.5%) (Figure 13B).
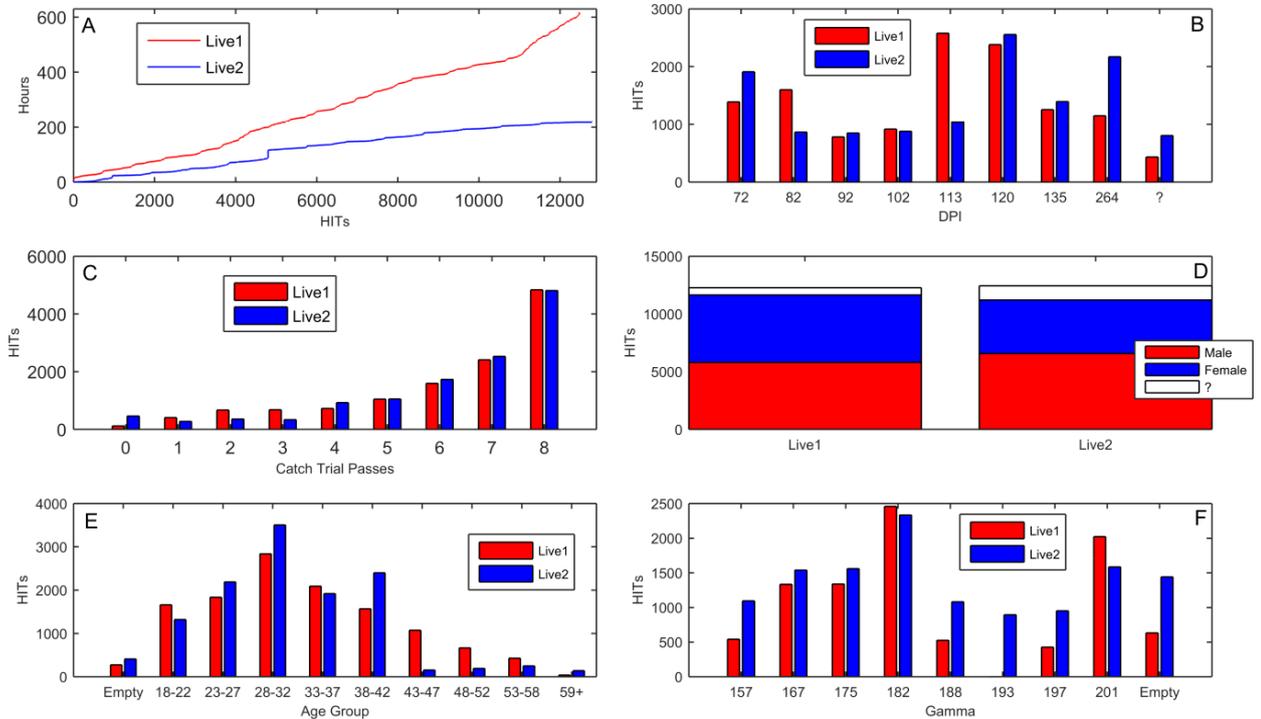
Figure 13. Summary statistics for Live1 and Live2. A: time course for Live1 and Live2 (HITs completed per hour). B: screen DPIs recorded. C: catch trials passed. D: gender distribution. E: age group distribution. F: screen gammas x10.

In total, 9 different gamma settings were recorded in Live1 and 10 in Live2 (Figure 13F). In Live1, the most common gamma was 201 with 3,197 HITs (25.6%) and the least common was Unknown with 2 HITs (0.02%). In Live2, the most common gamma was 182 with 2,335 HITs (18.7%) and the least common was 193 with 895 HITs (7.1%) (Figure 13F).

In Live1, 54 browser and operating system (OS) combinations were recorded. The most common was Google Chrome 39 for Windows with 6,070 HITs (48.6%). The most common OS was Windows with 11,843 HITs

(94.9%). In Live2, 39 browser and operating system (OS) combinations were recorded. The most common was Google Chrome 43 for Windows with 6,050 HITs (48.5%), followed by Mozilla Firefox 38 for Windows with 4,440 HITs (35.6%). The most common OS was Windows with 11,890 (95.3%).

In Live1, the number of Workers in each gender were approximately equal with 5,810 HITs male (46.6%) and 5,837 HITs female (46.8%). In Live2, the number of Workers in each gender was also approximately equal with 6,570 HITs male (52.6%) and 4,640 HITs female (37.2%) (Figure 13D). In both cases, gender distribution across the texture families was approximately equal.

The most common (self-reported) age group in Live1 was 28-32 years with 2,839 HITs (22.7%). The least common age group was 59+ years with 38 HITs (0.3%). HITs were approximately normally distributed across the age groups. The most common age group in Live2 was 28-32 years with 3,505 HITs (28.1%). The least common age group was 59+ years with 140 HITs (1.1%) (Figure 13E).

The total number catch trial passes per HIT in Live1 were evaluated (Figure 13C). Each HIT contained 8 catch trials in total, randomly distributed between the long and short conditions. The most frequent was 8 catch trial passes with 4,837 HITs (38.8%). HITs were retained if the total catch trial passes were >= 4 (11,288 HITs, 90.4% retained). In Live2, each HIT contained 8 catch trials in total, 4 in the long presentation condition and 4 in the short presentation condition. The most frequent was 8 catch trial passes with 4,814 HITs (38.6%). Again, HITs were retained if total catch trial passes

were >= 4 (11,052 HITs, 88.6% retained). The number of catch trial passes was reasonably consistent across texture types.

Every Worker that submitted HITs in Live1 was paid $0.10 AUD per HIT, regardless of catch trial passes. Those that completed 28 or more HITs received a $0.80 bonus (there were 41 in total and some Workers earned more than one bonus). Therefore, the total paid for 12,480 HITs was $1280.80 AUD. The cost per HIT was $0.113 AUD ($6.81 AUD per hour).

Every Worker that submitted HITs in Live2 was paid $0.50 AUD per HIT, regardless of catch trial passes. Those that completed all 6 rays of a trico plane received a $1.00 bonus (there were 309 bonuses in total). Therefore, the total paid for 12,480 HIT-equivalents was $1,557.00 AUD. Based on the 11,052 HITs retained with catch trial passes >= 4, the cost per HIT was $0.1409 AUD ($8.45 AUD per hour).

Lastly, the number of repeat Workers was analysed (Figure 14). The performance of the repeat Workers is analysed below in Section 4.4.6. In total, 31 Workers *participated in both* Live1 and Live2. In the figure below, the number of HITs completed per Worker has been sorted for Live1, and then plotted over the number of HITs completed in Live2.
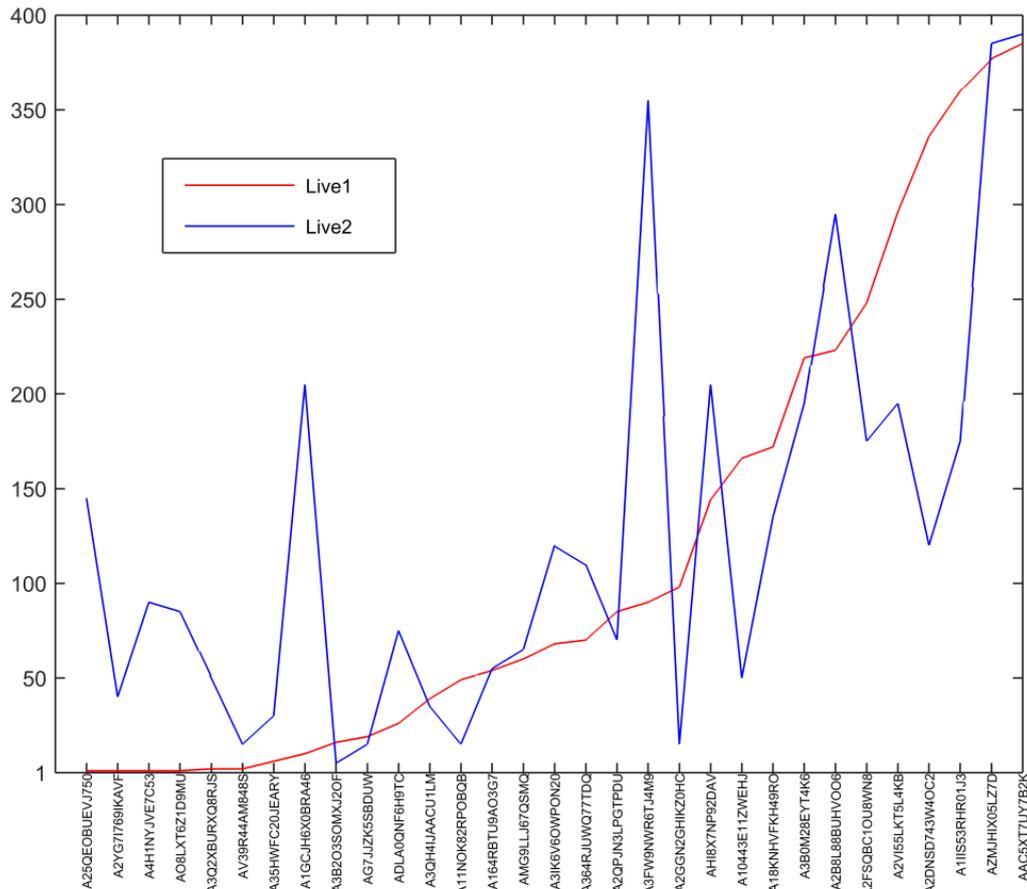
Figure 14. Repeat Workers in Live1 and Live2. In total, 532 Workers participated in Live1 and 427 in Live2; 31 Workers participated in *both* studies and the number of HITs they completed per study is plotted here.

4.4.3 Live1 Performance Data

All Live1 performance data was analysed *after* filtering the data by catch trial passes. i.e.: data from Workers that failed >= 4 catch trials was excluded. In the first instance, boxplots were produced for each ternary texture type (one example is shown in Figure 15; for completeness, the other plots are included in Appendix 6.8).
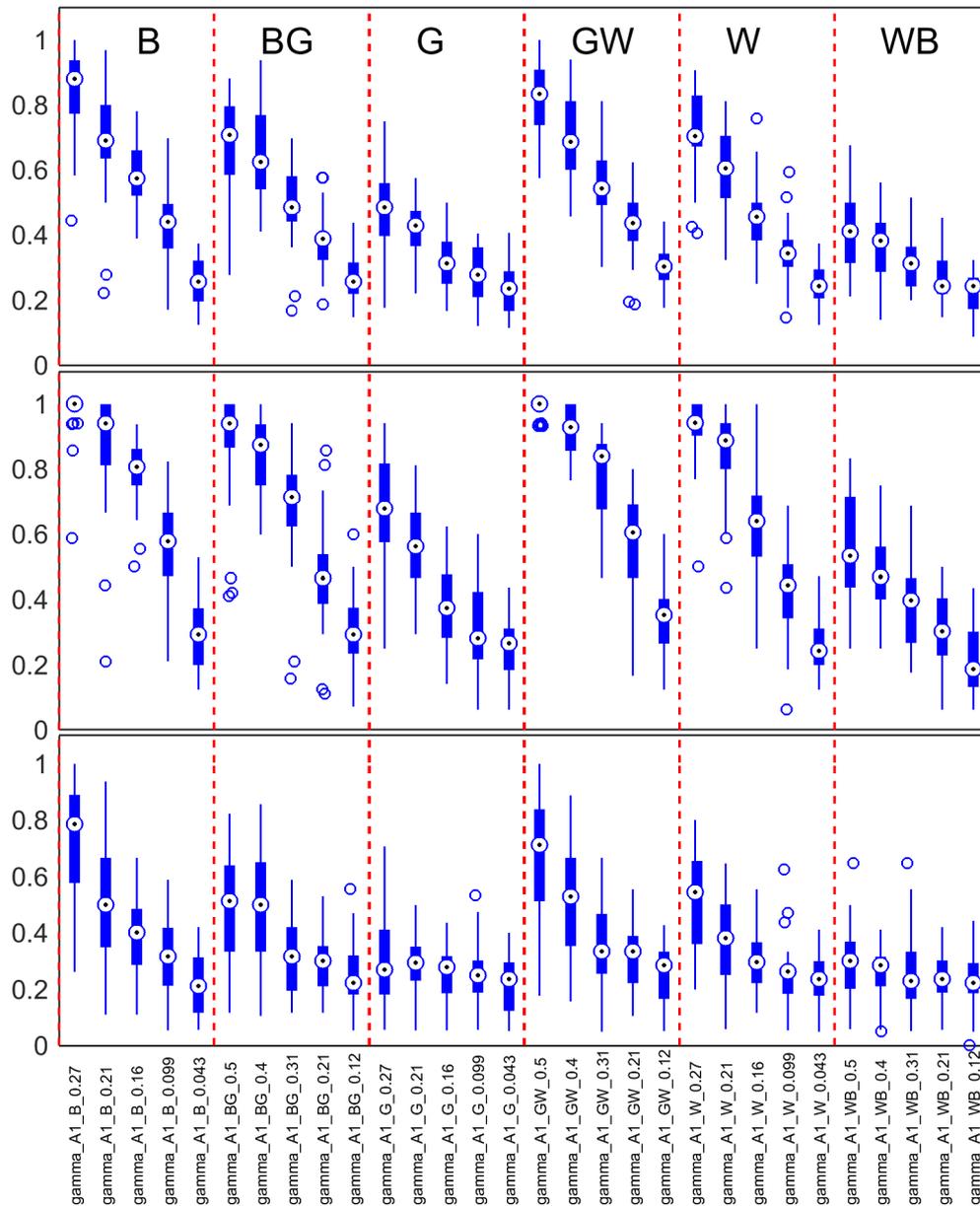
Figure 15. Performance data for gamma_A1 ternary textures in Live1.

HITs where Workers failed >= 4 catch trials were excluded. The three

plots refer to the combined (top), long (middle) and short (bottom)

presentation times. The dashed red lines divide the plots into the 6

rays with the following colour-biases: B (Black); BG (Black-grey); G

(Grey); GW (Grey-white); W (White); WB (White-black). The step and

ray values are indicated on the X axis, using the Victor nomenclature discussed in Section 4.3.3.

Based on these performance functions, Weibull plots were produced from the Live1 data. Due to the *low discrimination performance* levels for some texture types, Weibull plots could not always be fitted to the data. Therefore, the step levels were adjusted to make the textures more salient in Live2. The main change was to increase the minimum signal level tested, given that in most cases the maximum signal level of 1.0, or very close to it, was already being tested. The Live1 data is further analysed in Sections 4.4.5 and 4.4.6. The actual step levels used in Live1 and Live2 are summarised in Table 2 in Section 4.4.5.

4.4.4 Live2 Performance Data

All Live2 performance data was analysed *after* filtering the data by catch trial passes (again, Workers that failed >= 4 catch trials were excluded). Boxplots were produced for each ternary texture type (one example is shown in Figure 16, the other plots are included in Appendix 6.9).
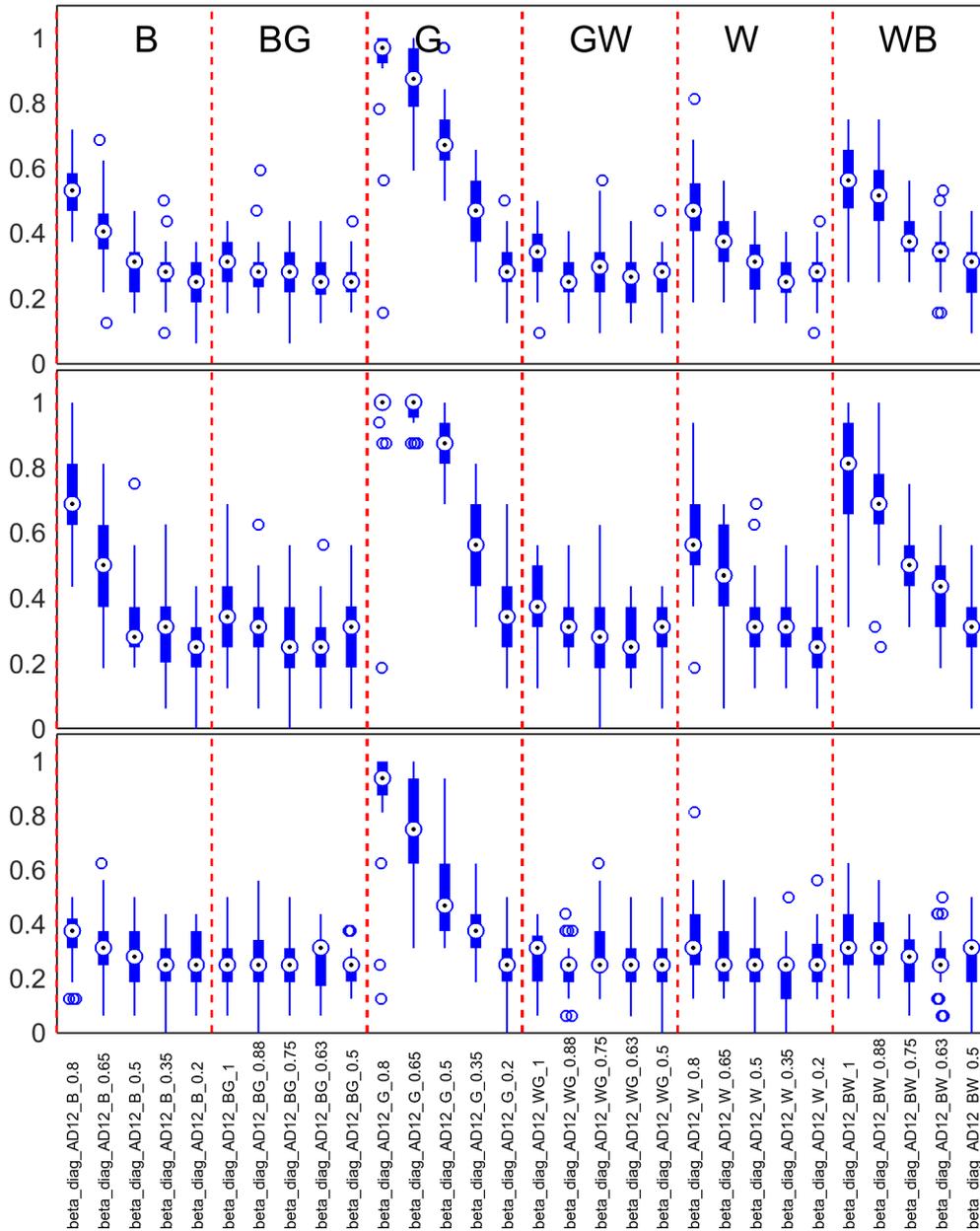
Figure 16. Performance data for beta_diag_AD12 ternary textures in Live2. HITs where Workers failed >= 4 catch trials were excluded. The three plots indicate the combined (top), long (middle) and short (bottom) presentation times. The dashed red lines divide the plots into the 6 rays with the following colour-biases: B (Black); BG (Black-grey); G (Grey); GW (Grey-white); W (White); WB (White-black). The step and ray values are indicated on the X axis.

As will be shown below, Live2 should have improved discrimination performance, so Live2 was analysed further in terms of Weibull plots and isodiscrimination contours (Sections 4.4.7 and 4.4.8). First though, we will analyse the differences between the Live1 and Live2 data sets, and therefore analyse the effect of sequential versus pseudo-random (random within a trico plane) presentation.

4.4.5 Comparison of Live1 and Live2 Performance

Colourmaps were produced to compare the Live1 and Live2 performance functions for the short, long and combined presentation times. The comparison of the Live1 and Live2 performance data informs us of the effect of pseudo-random, versus sequential presentation *by step*, within the trico plane (Figures 17 and 18).
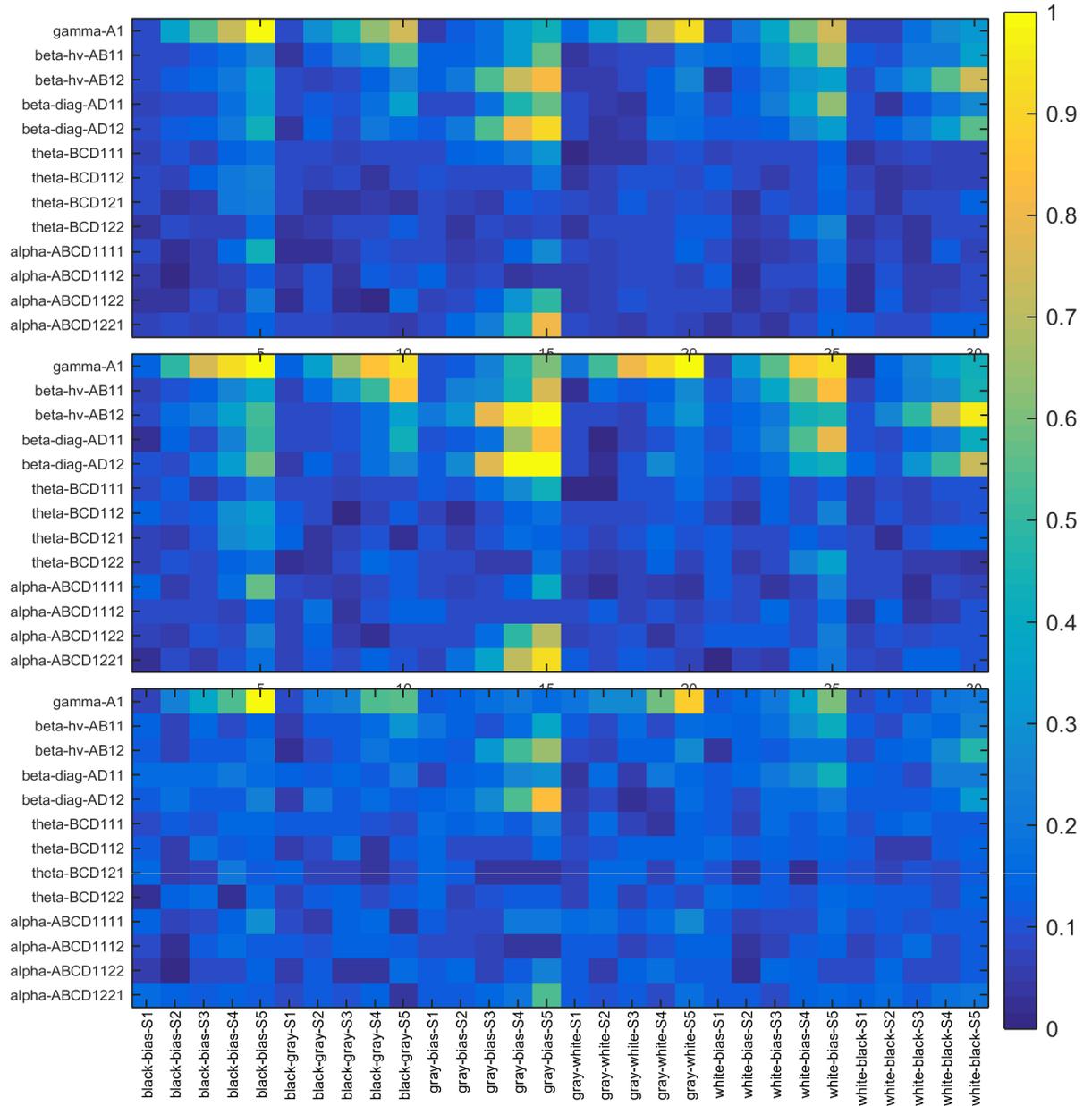
Figure 17. Colourmap of median performance data for Live1. HITs where Workers failed >= 4 catch trials were excluded. The three plots indicate the combined (top), long (middle) and short (bottom) presentation times. The steps and rays are indicated on the X and Y axes respectively. Thus each row represents a trico-plane (Figs. 6 and 7) and the columns represent steps of correlation, from lowest to highest going left to right in blocks of 5 steps.
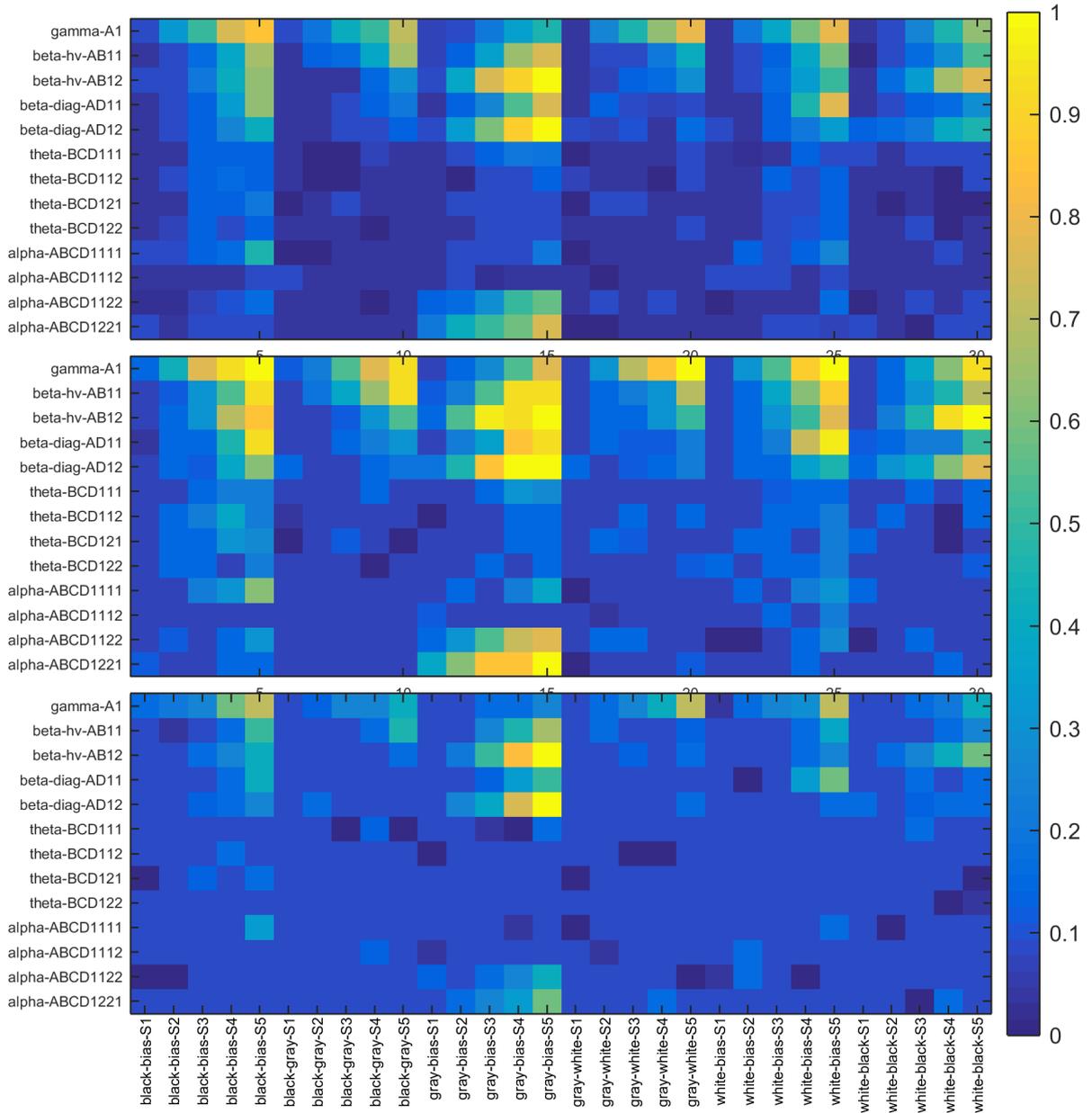
Figure 18. Colourmap of median performance data for Live2. HITs where Workers have failed >= 4 catch trials were excluded. The three plots indicate the combined (top), long (middle) and short (bottom) presentation times. The steps and rays are shown as for Fig. 17.

When making a direct comparison between Live1 and Live2 performance data, it is essential that we take into consideration the changes in the step levels implemented in Live2. One way to account for this is to only examine

the step 5 performance values. These values are very similar between the Live1 and Live2 studies, as summarised in the following table, in which the largest differences are highlighted (Table 2).

Live1	B	BG	G	GW	W	WB
gamma_A1	0.27	0.5	0.27	0.5	0.27	0.5
beta_hv_AB11	0.47	0.78	0.47	0.78	0.47	0.78
beta_hv_AB12	0.47	0.78	0.47	0.78	0.47	0.78
beta_diag_AD11	0.67	1	0.67	1	0.67	1
beta_diag_AD12	0.67	1	0.67	1	0.67	1
theta_BCD111	0.97	1	0.97	1	0.97	1
theta_BCD112	0.97	1	0.97	1	0.97	1
theta_BCD121	0.97	1	0.97	1	0.97	1
theta_BCD122	0.97	1	0.97	1	0.97	1
alpha_ABCD1111	1	1	1	1	1	1
alpha_ABCD1112	1	1	1	1	1	1
alpha_ABCD1122	1	1	1	1	1	1
alpha_ABCD1221	1	1	1	1	1	1

Live2	B	BG	G	GW	W	WB
gamma_A1	0.3	0.5	0.3	0.5	0.3	0.8
beta_hv_AB11	0.7	1	0.7	1	0.6	1
beta_hv_AB12	0.7	1	0.7	1	0.6	1
beta_diag_AD11	0.8	1	0.8	1	0.8	1
beta_diag_AD12	0.8	1	0.8	1	0.8	1
theta_BCD111	1	1	1	1	1	1
theta_BCD112	1	1	1	1	1	1
theta_BCD121	1	1	1	1	1	1
theta_BCD122	1	1	1	1	1	1
alpha_ABCD1111	1	1	1	1	1	1
alpha_ABCD1112	1	1	1	1	1	1
alpha_ABCD1122	1	1	1	1	1	1
alpha_ABCD1221	1	1	1	1	1	1

Table 2. Summary table of the decorrelation levels used in Live1 and Live2 (step 5 only, the minimum decorrelation level). The most distinct step levels are highlighted. The texture names are shown in the leftmost column. The colour-biases are shown in the top row as follows: B (Black); BG (Black-grey); G (Grey); GW (Grey-white); W (White); WB (White-black).

263

With these differences and similarities in mind, we can now repeat the colourmap above, but only consider the step 5 values, i.e. the textures for each trico axis with the higher degree of correlation. Colourmaps for the long and short presentation times are shown below in Figures 19 and 20.
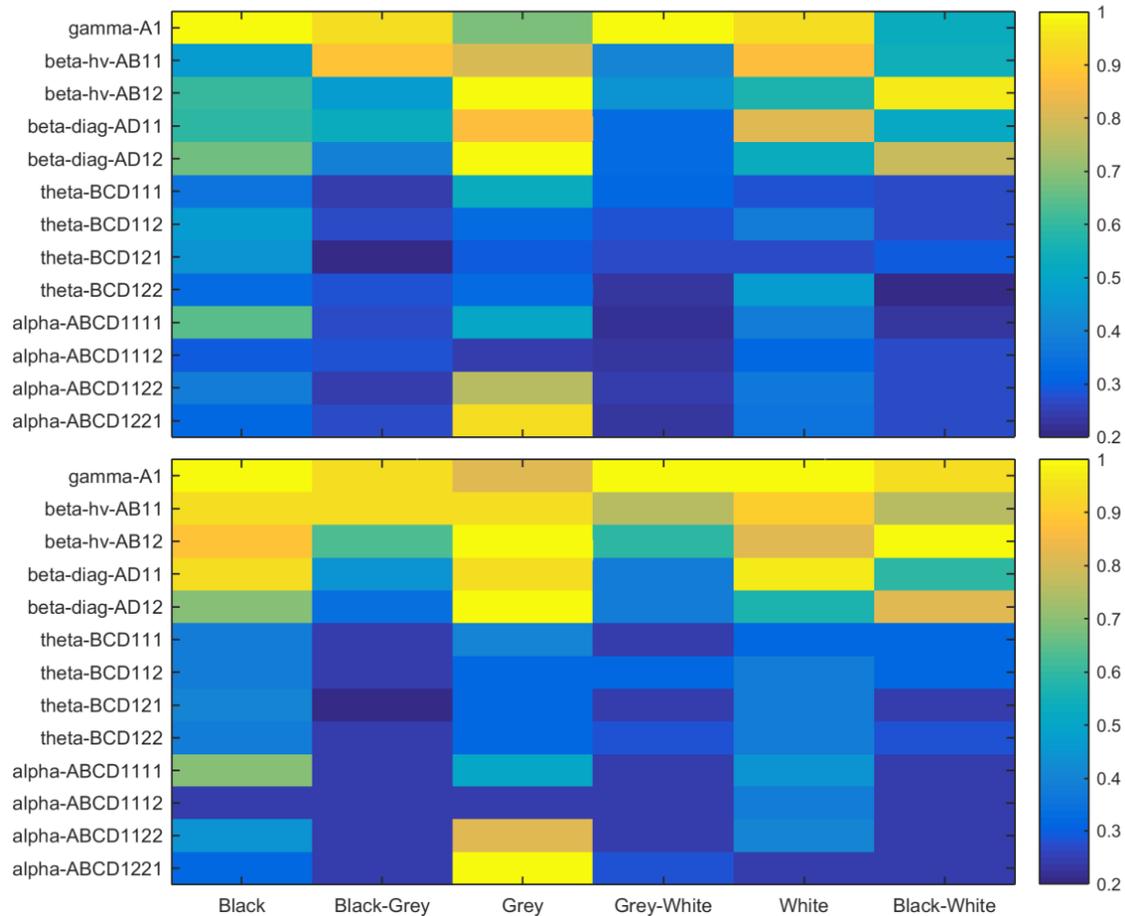

Figure 19. Colourmap of median performance data for Live1 (top) and Live2 (bottom) for the *long* presentation condition. HITs where Workers failed >= 4 catch trials were excluded. Only step 5 values are shown, which allows for direct comparison between the Live1 and Live2 phases (see Table 2).
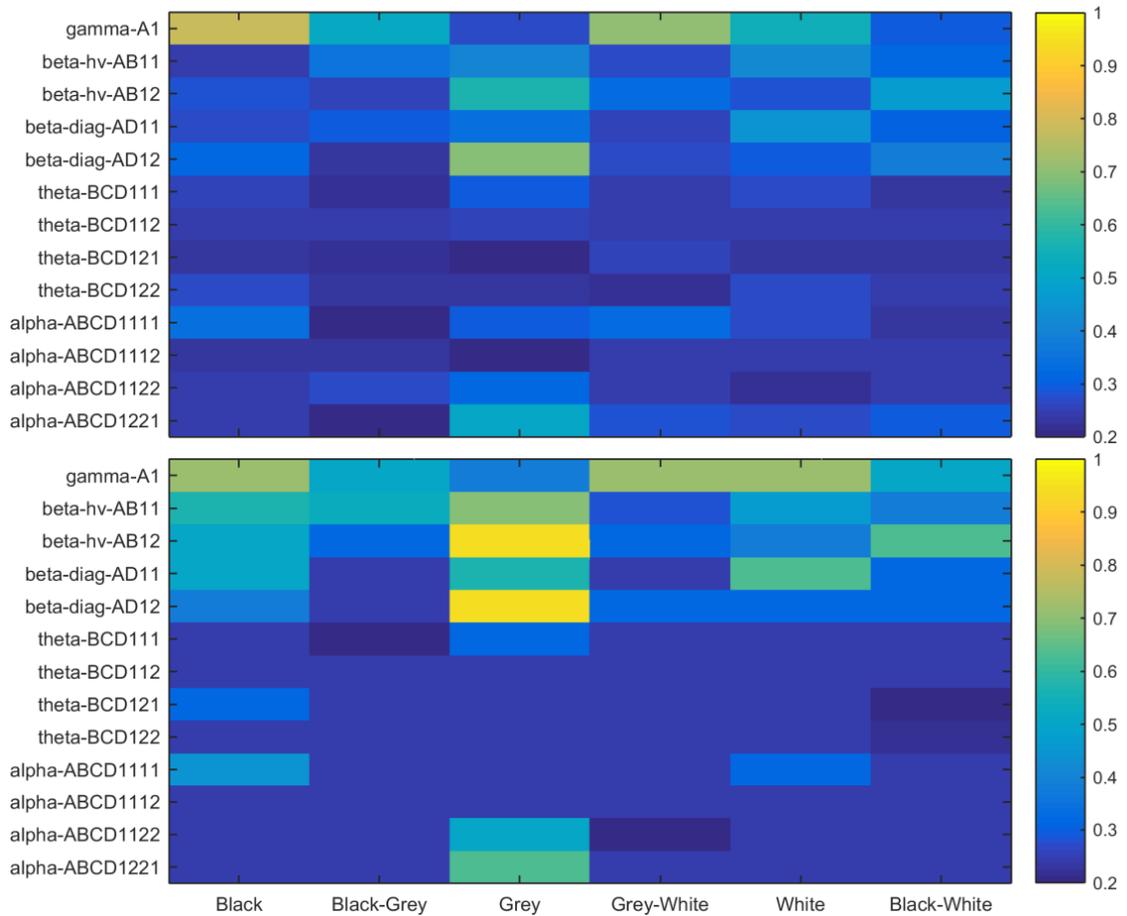
Figure 20. Colourmap of median performance data for Live1 (top) and Live2 (bottom) for the *short* presentation time. HITs where Workers failed >= 4 catch trials were excluded. Only step 5 values are shown. Again, refer to Table 2 for actual step levels.

Previous studies, including some using the current alpha textures, have found our discrimination ability to be pre-attentive and relatively unaffected by presentation times (Taylor et al., 2008). In contrast, in the current study, performance functions for the short condition are clearly lower for all textures examined. Therefore, further analysis was performed using the long condition data.

265

4.4.6 An Evaluation of Repeat Workers

In total, 31 Workers participated in both Live1 and Live2 (see Figure 14 in Section 4.4.2). Where possible, the repeat performance of these Workers was analysed; 11 Workers were identified as having sufficient repeat performance data available. Again, only step 5 values were compared (see Table 2 in Section 4.4.5). One example of this analysis is shown below in Figure 21.
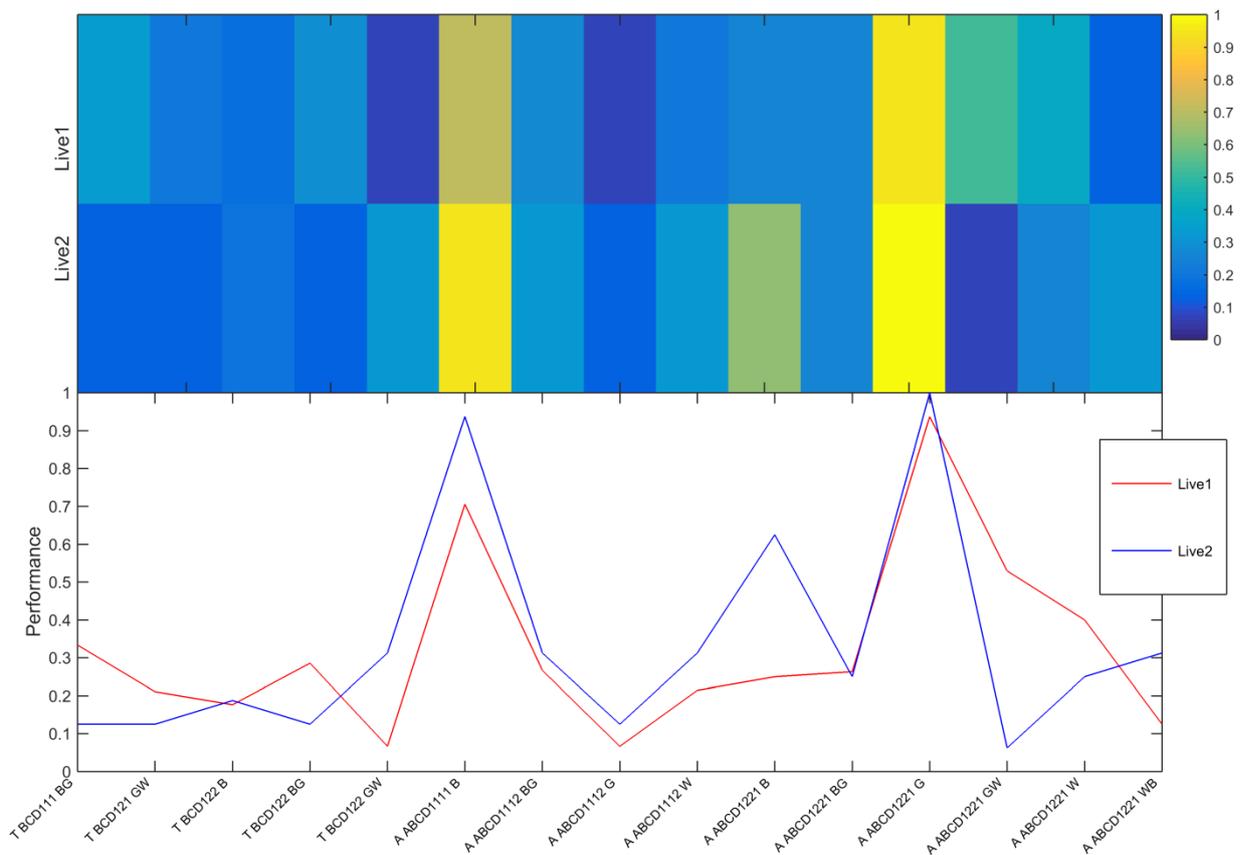
Figure 21. Plots of median performance for repeat Worker ID AAC5XT7UY7B2K. Top: Colourmap of median performance for Live1 and Live2. Bottom: plots of Live1 and Live2 median performance. This analysis is based on the *long* presentation time and HITs where Workers failed >= 4 catch trials were excluded. Only mean

performance levels for step 5 textures are plotted. Recall that for the 4AFC test, chance performance is 0.25.

For all 11 Workers, there appeared to be a significant improvement in performance from Live1 to Live2 for the second order ternary textures (betas), suggesting a learning effect. This data is summarised in Figure 22, where Live2 performance has been subtracted from Live1 performance; thus, positive values indicate improvement. The black areas indicate where comparison data was not available. i.e.: the given Worker did not complete HITs for the given ternary texture in *both* Live1 and Live2.
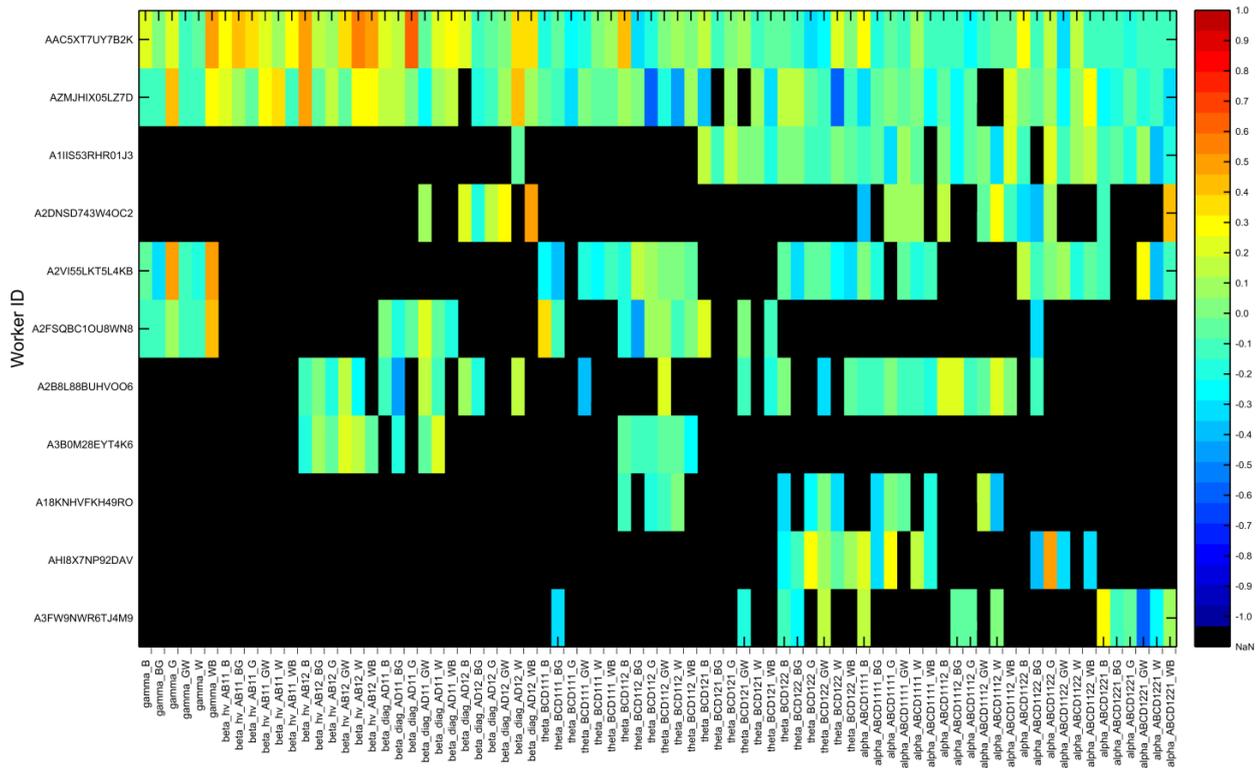
Figure 22. Colourmap of median performance data for 11 repeat Workers where Live1 has been subtracted from Live2. Positive values indicate improvement in performance between Live1 and Live2 (and

suggest learning). Learning is most evident for the beta textures. This analysis is based on the *long* presentation time and HITs where Workers failed >= 4 catch trials were excluded. Only step 5 values are considered. Black areas indicate insufficient data for analysis.

To further evaluate possible learning effects in repeat Workers, a series of Bland-Altman plots were produced for the second-order textures (Figures 23 to 25). Note that the numbers of Workers sampled is show in parentheses next to the texture name; refer to Figure 22 above, which indicates where data was available to analyse a given Worker/texture pair. For example, two Workers completed Live1 and Live2 HITs for texture beta_hv_AB11 (top plot in Figure 23). As above, only step 5 values are considered (to allow direct comparison). Therefore, each Worker produced 6 performance scores in Live1 and Live2 for beta_hv_AB11 (one for each colour bias).
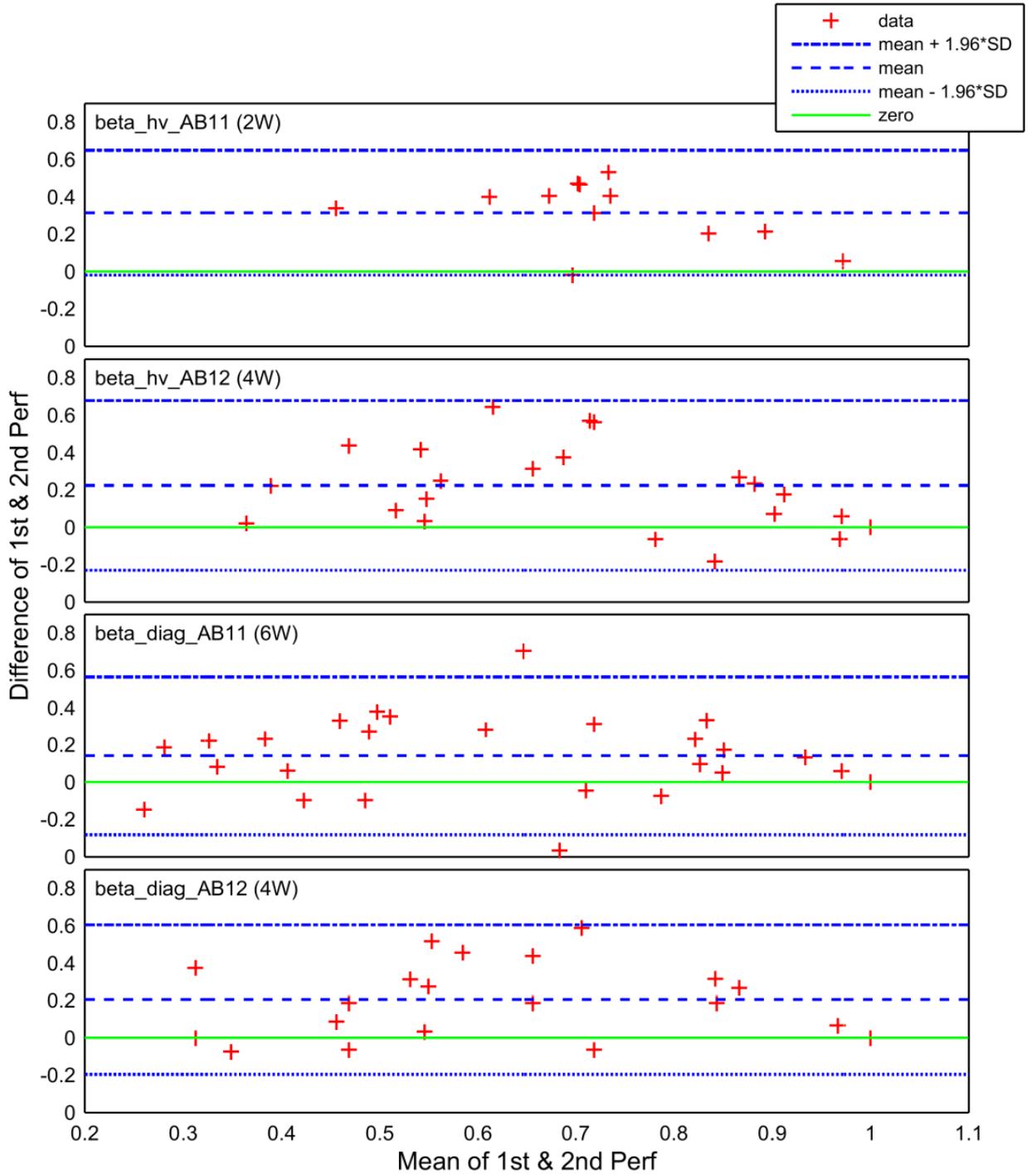
Figure 23. Bland-Altman plots for the repeat Workers for the beta texture types. The numbers of Workers sampled is indicated in parentheses next to the texture name.
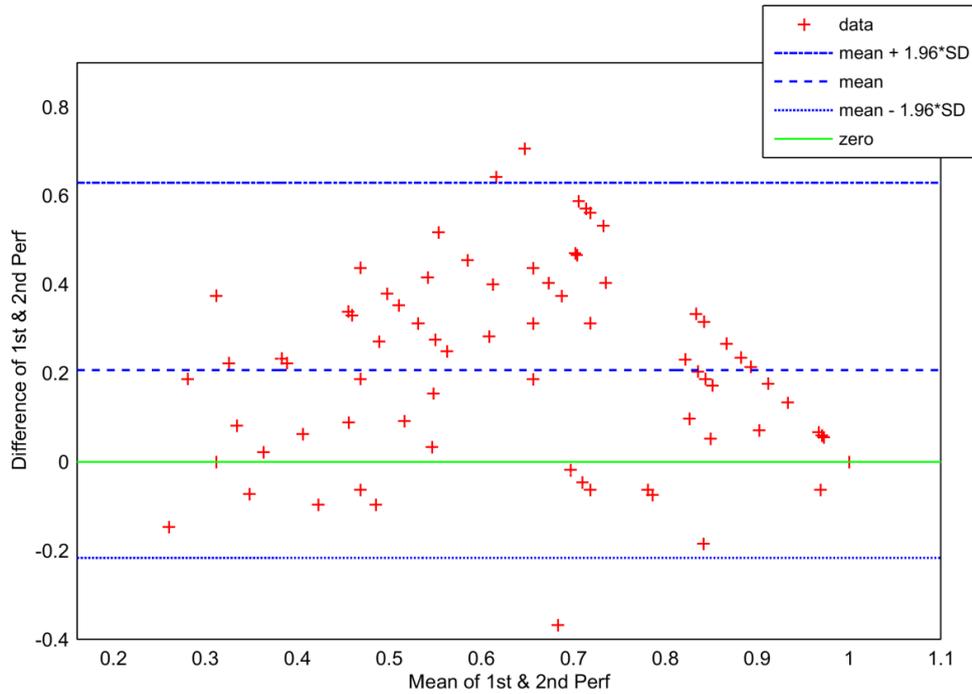
Figure 24. Bland-Altman plot for the repeat Workers for all 4 beta textures pooled. In total, data from 6 repeat Workers was sampled.
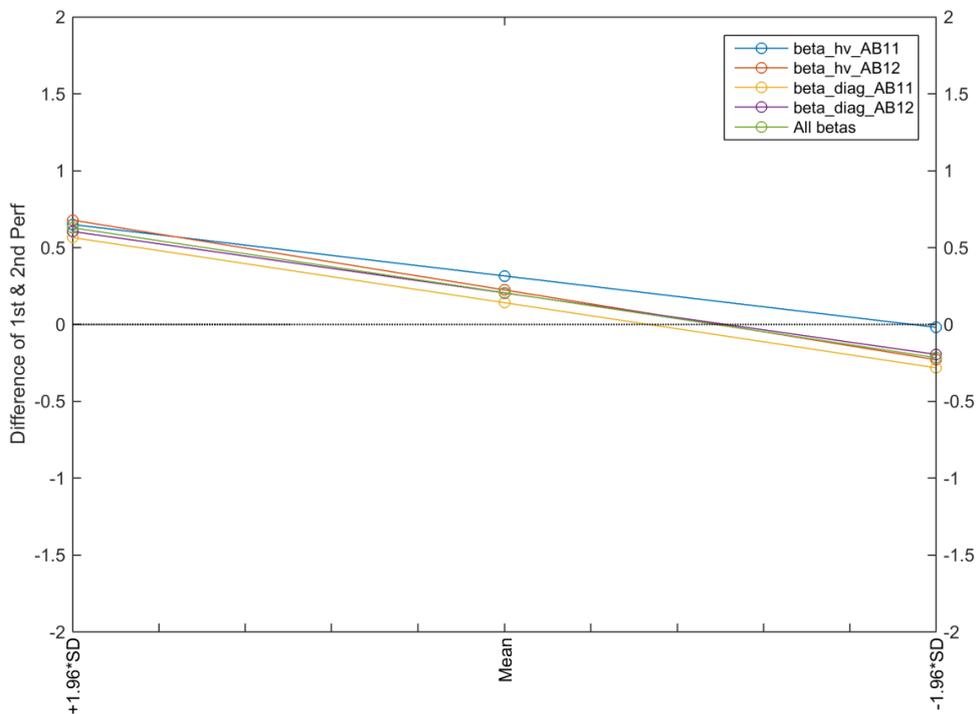
Figure 25. Summary of mean and mean+/-1.96*SD data derived from Bland-Altman plots. The legend indicates which beta texture family was sampled.

4.4.7 Weibull Functions from Live2 Data

Weibull functions were derived from the Live2 performance data; the analysis proceeded as follows. In the first instance, in order to try to obtain the best signal to noise ratio, performance functions were plotted using various subsets of the data; this included the number of catch trial passes, the minimum number of HITs completed per Worker, and the subset of performance functions (50th percentile, 75th percentile, RMS). Based on a comparison of these conditions, >=20 HITs completed and 75th percentile performance were chosen, as shown in Figure 26A below. Note also the arrangement of the plots in the context of the trico plane. The plots contain inset examples of the ternary textures examined; these correspond to the reduced trico plane shown in Figure 26B.

Based on an examination of these plots, it appears that performance functions for the *corners* of the trico plane (i.e. for the canonical black, white and grey dominated correlations) are higher than those of the quantitatively derived 50:50 mixtures for the thetas and alphas. A clear example of this is shown for the alpha_ABCD1111 in Figure 27 below.

Weibull plots were then produced using the data where Workers completed >=20 HITs, 7 catch trial passes, and 75th percentile. An example Weibull plot is shown in Figure 28. Where possible, the step levels were identified which predicted 0.60 performance. Where Weibull functions could not be fitted, the observed performance at zero decorrelation was instead recorded. These values were then used to form isodiscrimination contours (See Section 4.4.8).
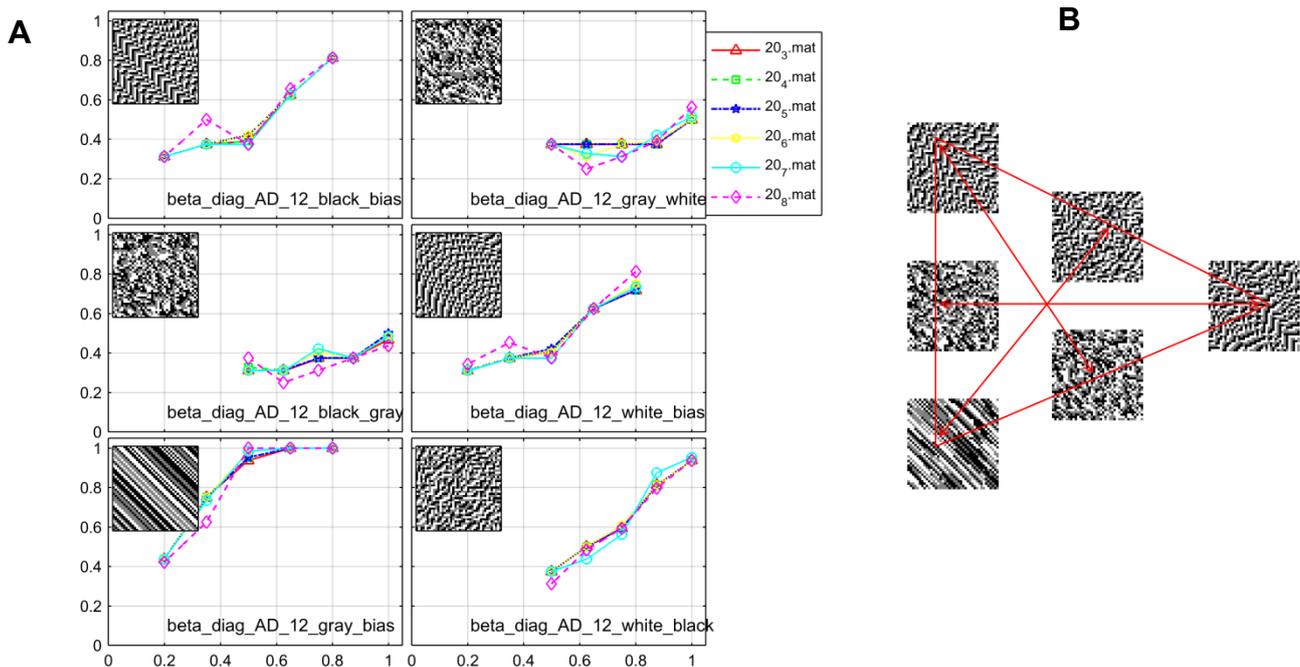
Figure 26. A: Plots of 75th percentile performance for beta_diag_AD12, where Workers completed >=20 HITs. Examples of the textures examined are inset. Different levels of catch trial passes are compared, as indicated by the different coloured plots. For example, the red plot indicates performance by Workers that completed >=20 HITs and passed >=3 catch trials ("20_3"). B: a miniature representation of the beta_diag_AD12 trico axis, which puts the performance plots in the context of the trico plane.
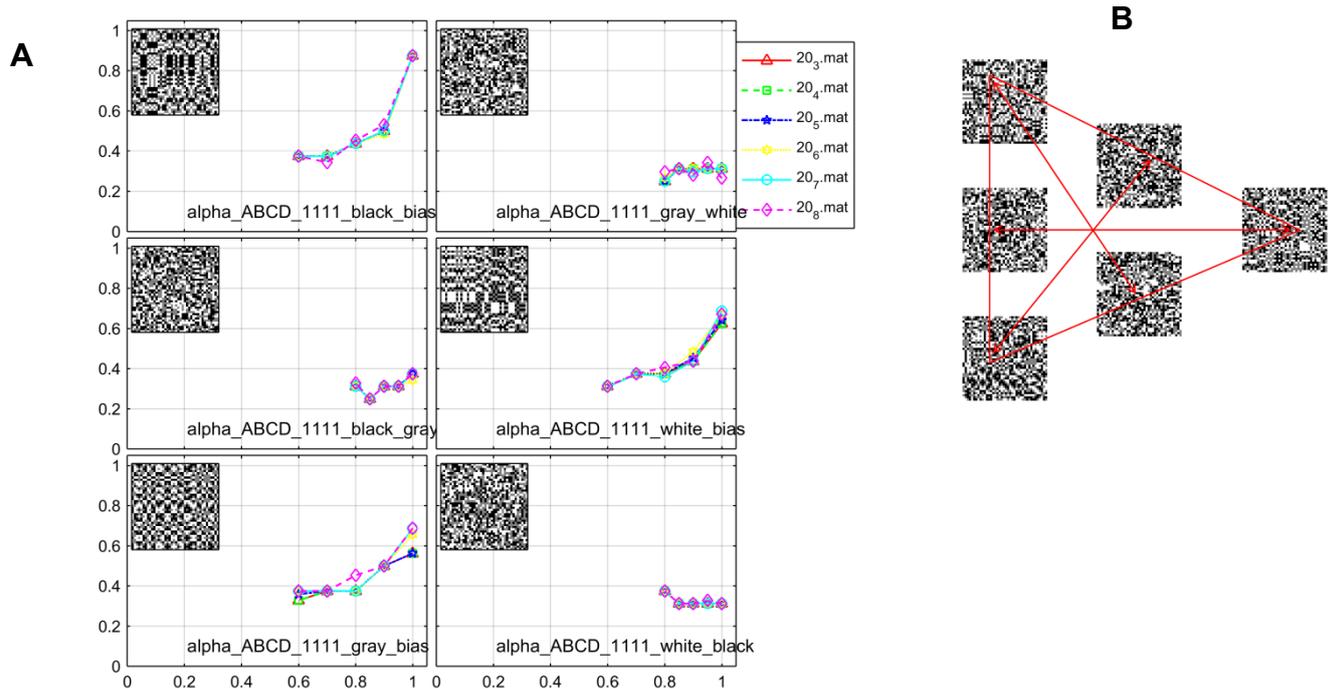
Figure 27. A: Plots of 75th percentile performance for alpha_ABCD1111, where Workers completed >=20 HITs. Examples of the textures examined are inset. Different levels of catch trial passes are compared. Note the *markedly higher* performance for textures at the corners of the trico axis, versus the derived mixtures. B: a miniature representation of the alpha_ABCD1111 trico axis.
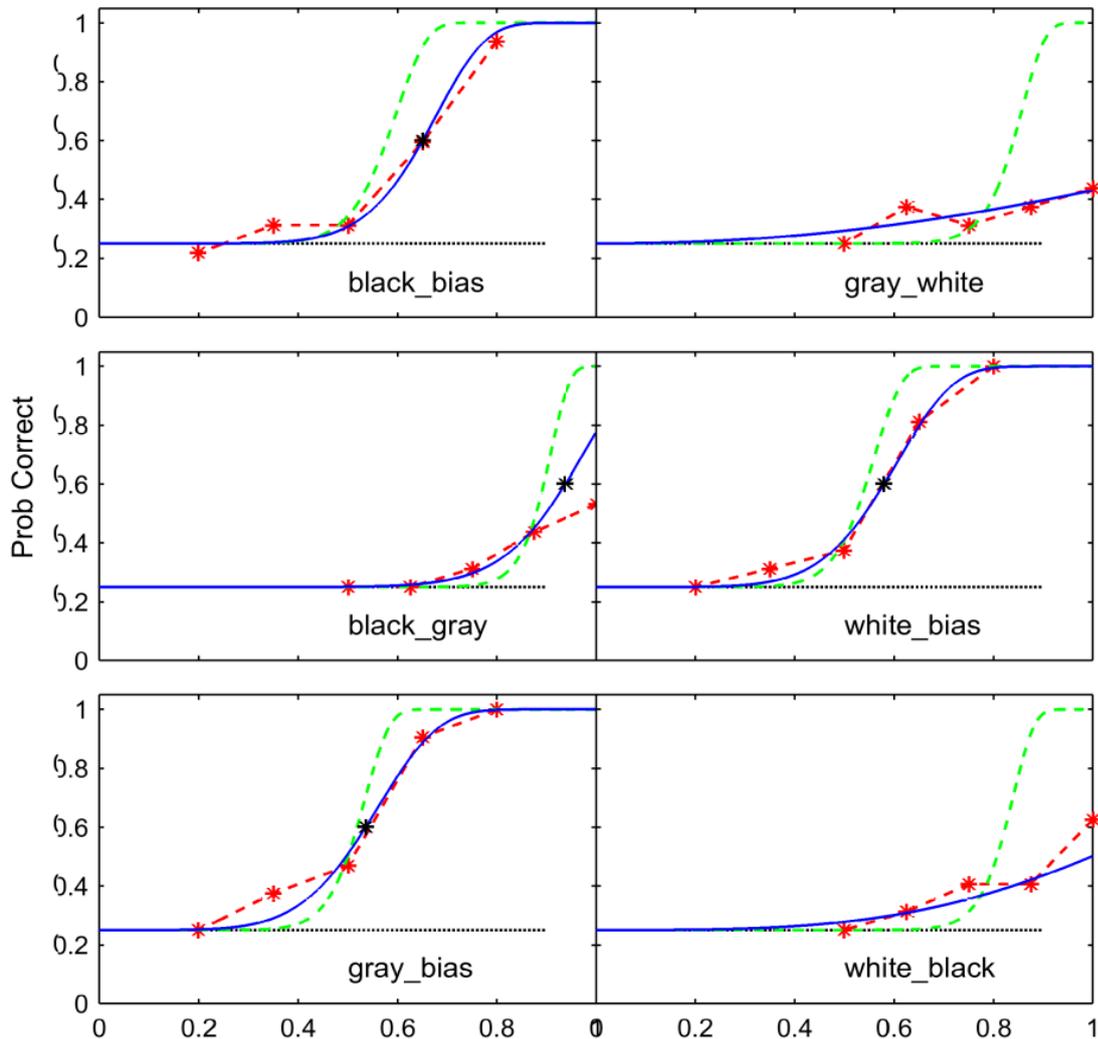
Figure 28. Weibull fit derived from Live2 data for the beta_diag_AD11 textures. Using Weibull functions such as these, the step levels were identified which predicted a 0.6 performance limen.

4.4.8 Isocontours from Live2

Isodiscrimination contours were produced for each ternary texture family, based on the values extracted from Weibull functions using Live2 data (Figures 29 and 30; also see Section 4.3.7 for details).

274

In some cases, for the theta and alpha textures, Weibull functions could not be successfully fitted; in this case, the performance level was as recorded at a step level of 1.0. This is reasonable because the low performance meant performance was not close to saturation. Therefore, we obtained the unsaturated performance to a fixed signal level: 1.0.
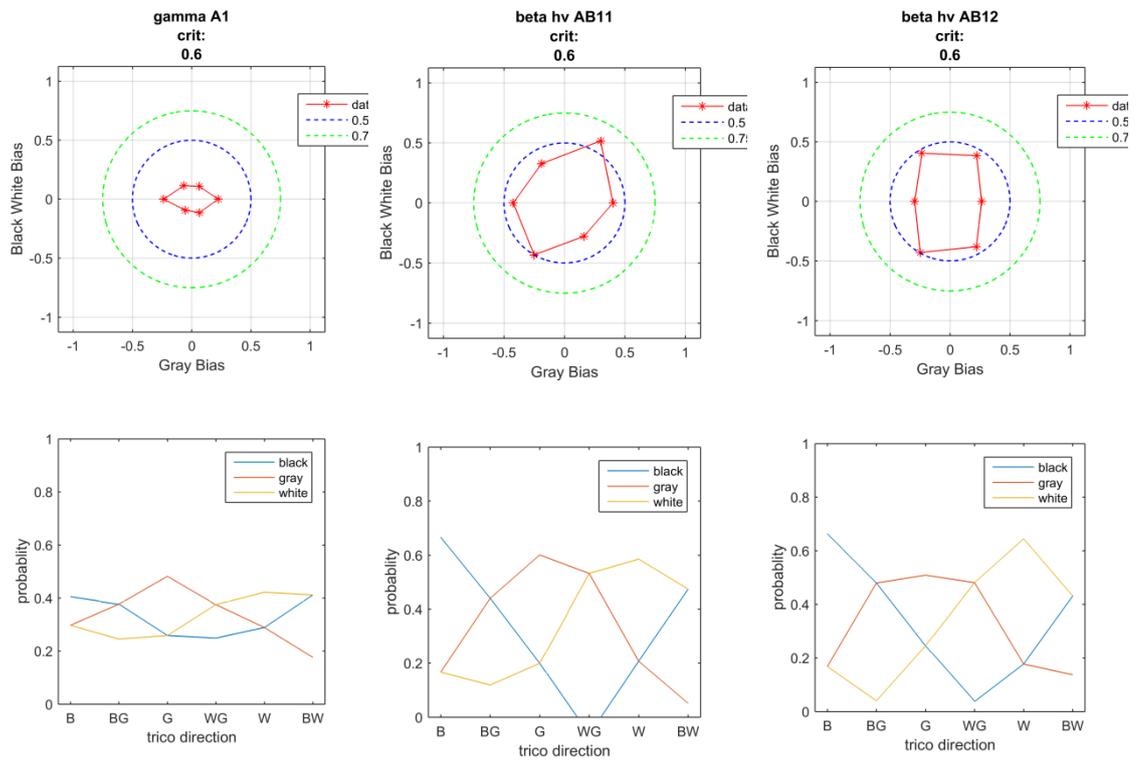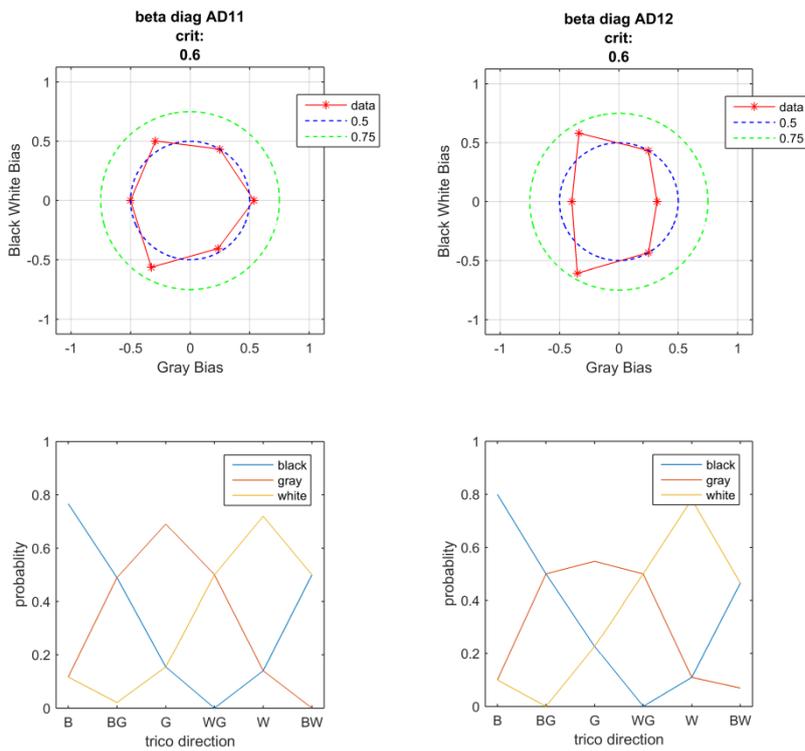
Figure 29. Isodiscrimination contours for 5 of the 13 ternary texture families. The isocontour parameters were derived from Weibull plots using 75th percentile performance data from Live2 (long presentation condition), where Workers completed >=20 HITs and >=7 catch trial passes. Upper plots: gamma and beta textures; step levels for each ray have been plotted which yield a predicted performance of 60% (wcrit). Thus, the axes of elongation correspond to the directions in which changes are *least salient*. Lower plots: corresponding plots of the pixel probabilities for each ray. Refer to Figures 5 and 6 for the corresponding reduced trico plots.
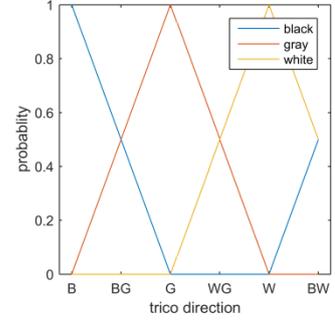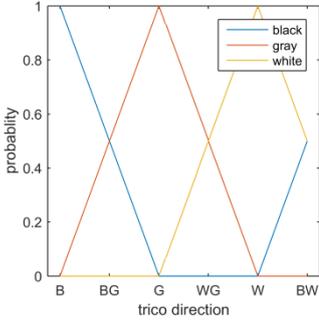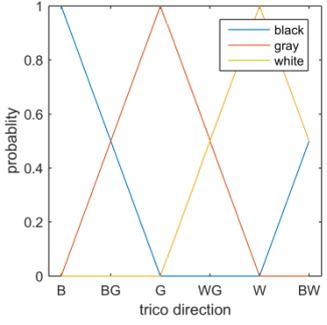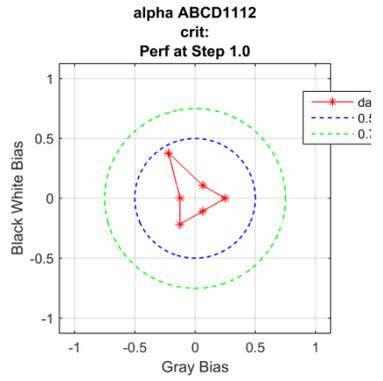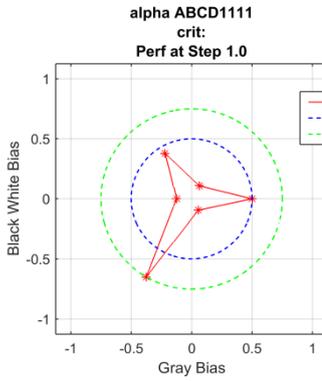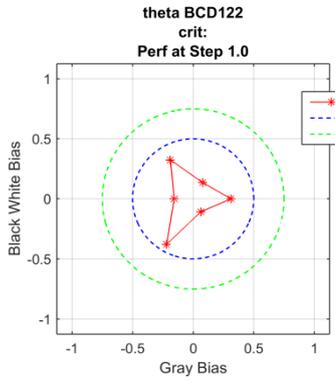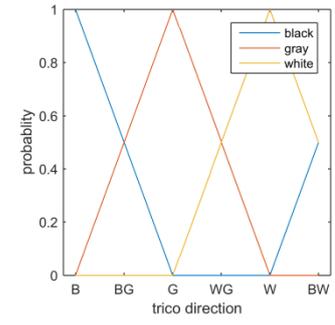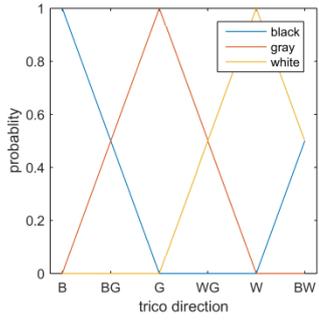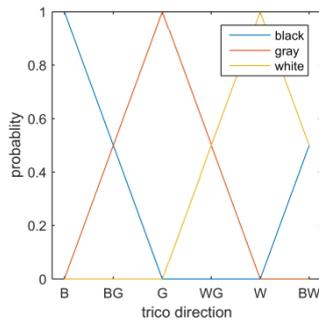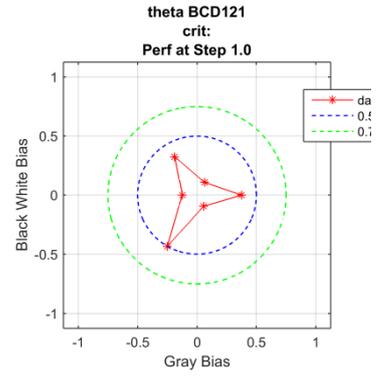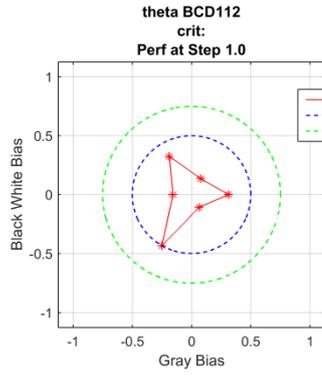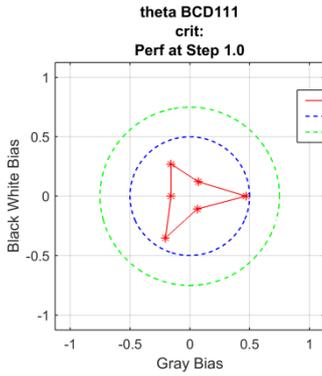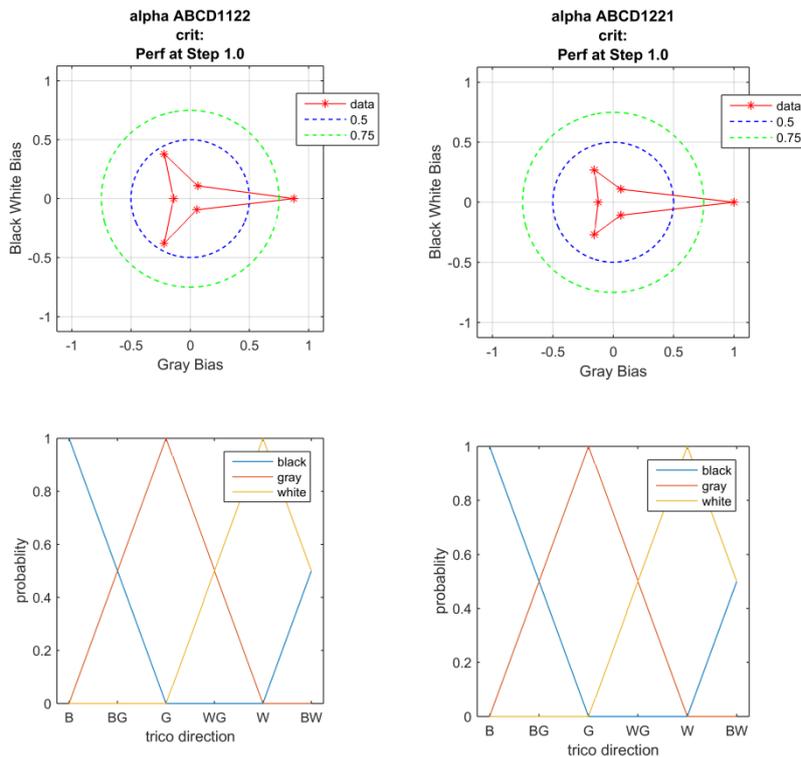
Figure 30. Isodiscrimination contours for 8 of the 13 ternary texture families. Isocontour parameters were derived from Weibull plots using 75th percentile performance data from Live2 (long presentation time), where Workers completed >=20 HITs and >=7 catch trial passes. Upper plots: theta and alpha textures; performance at step 5 (1.0 decorrelation, Perf at Step 1.0) is plotted; thus, the axis of elongation corresponds to the direction in which changes are *most salient*. Lower plots: corresponding plots of the pixel probabilities for each ray. Refer to Figures 5 and 6 for the corresponding reduced trico plots.

4.5 Discussion

We found that perceptual salience varied for each image statistic examined. For the gamma and beta textures, criterion step levels were obtained from Weibull plots, which predicted 0.60 performance (Figure 29). The resulting

mean thresholds were gamma 0.12, beta_hv 0.26, and beta_diag 0.34. This corresponds to a sensitivity rank-order of gamma > beta_hv > beta_diag.

Weibull functions could not be fitted for the thetas and alphas. Therefore, the observed performance at zero decorrelation was instead plotted on isocontours (Figure 30). These values are not directly comparable to the Victor data (although we can evaluate *relative salience*).

The mean performance values at step level 1.0 (zero decorrelation) were theta 0.18 and alpha 0.22. Note that, contrary to the threshold values discussed above for the gammas and betas, larger mean performance values indicate *more salient* textures. Adding these values to our sensitivity rank-order gives: gamma > beta_hv > beta_diag > alpha > theta.

The most salient statistic (gamma) is defined by the pixel-by-pixel mean and variance. Second-order statistics affect the spatial-frequency content of the stimulus and are thus readily detectable by linear filters; therefore, it is logical that the beta textures were more salient than the higher-order textures (thetas and alphas). However, the current study *also found* fourth-order correlations (alphas) to be more salient than third-order correlations (thetas), which is less readily explicable.

Using the isocontours (Figures 29 and 30 above), values were extracted for each texture type and ray and summarised in Figure 31 below.
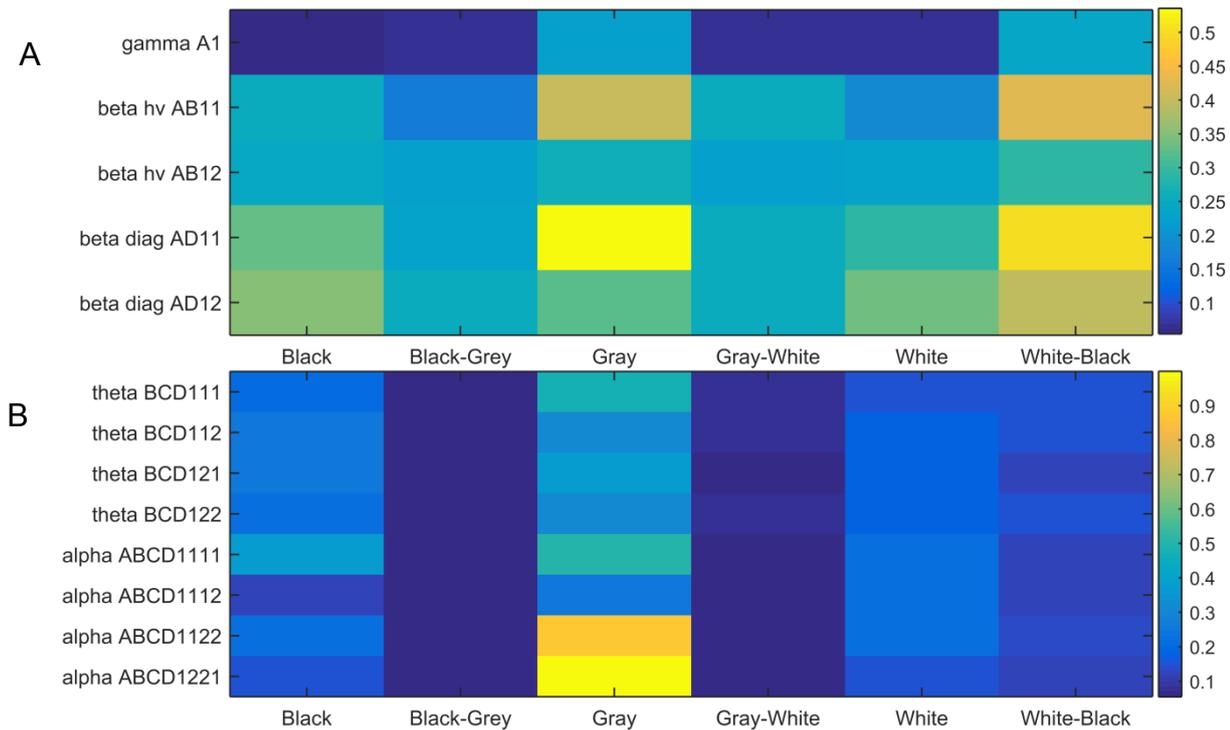
Figure 31. A: Threshold values for predicted performance of 0.60 (extracted from isocontours, Live2 data). Higher values correspond to the direction in which changes are *less salient*. B: Performance values at a step level of 1.0 (i.e.: no decorrelation). Higher values correspond to the direction in which changes are *more salient*.

The extracted values themselves are summarised in Table 3 below. Based on this data, it is clear that for the first- and second-order textures, the white-black and grey-bias rays were *consistently* the least salient, followed by white- or black-bias. For the third- and fourth-order textures, the black-grey and grey-white textures were *consistently* the least salient, followed by white-black. In both cases, for the first/second order and the third/fourth order textures, it is notable that the pattern of salience is *so consistent* across all textures considered.

Texture	Least Salient	2nd Least Salient	3rd Least Salient
Gamma_A1	0.24 (WB)	0.22 (G)	0.07 (W)
Beta_hv_AB11	0.42 (WB)	0.40 (G)	0.25 (GW)
Beta_hv_AB12	0.29 (WB)	0.26 (G)	0.25 (B)
Beta_diag_AD11	0.54 (G)	0.50 (WB)	0.33 (B)
Beta_diag_AD12	0.40 (WB)	0.35 (B)	0.34 (W)
Theta_BCD111	0.06 (BG)	0.07 (GW)	0.16 (W)
Theta_BCD112	0.06 (BG)	0.08 (GW)	0.16 (WB)
Theta_BCD121	0.05 (BG)	0.06 (GW)	0.12 (WB)
Theta_BCD122	0.06 (BG)	0.08 (GW)	0.16 (WB)
Alpha_ABCD1111	0.05 (BG)	0.06 (GW)	0.13 (WB)
Alpha_ABCD1112	0.06 (BG)	0.06 (GW)	0.13 (B/WB)
Alpha_ABCD1121	0.05 (BG)	0.06 (GW)	0.14 (WB)
Alpha_ABCD1221	0.06 (BG)	0.06 (GW)	0.13 (WB)

Table 3. Summary of the first, second and third least salient rays. Data isolated from isocontours (see Figure 31 above). The identity of the ray is indicated in parentheses as follows: White (W); White-Black (WB); Grey (G); Grey-White (GW); Black (B); Black-Grey (BG).

In Section 4.4.7, we also noted that performance functions for the *corners* of the trico plane are higher than those for the quantitatively derived mixtures. This is supported by the isocontour analysis of the thetas and alphas. How do these findings compare to those of previous studies by Victor et al. (Victor and Conte, 2012, Victor et al., 2013)?

Victor et al. investigated *binary* stochastic texture performance in their 2012 paper (Victor and Conte, 2012) using 6 normal subjects ages 25 to 51, 5 of which were practiced psychophysical observers in tasks involving texture discrimination. A four alternative choice task was used, the same protocol as previously described in our binary HIT (Victor and Conte, 2012, Seamons et al., 2015a) and used here (Section 4.3.4). They used pairs of image statistics and determined the salience of each coordinate in isolation, as well as

analysing how they interacted (Victor and Conte, 2012). After a 2 hour period over which performance functions stabilized, performance functions were calculated and fit to Weibull functions along each axis (Victor et al., 2005, Victor and Conte, 2012).

Victor et al. found the performance variation across observers to be small. In terms of the thresholds corresponding to a fraction correct of 0.625, there was ~10% scatter for gamma, beta_hv, and beta_diag and ~20% scatter for theta and alpha. Across all observers, the thresholds for a fraction correct of 0.625 were gamma 0.157, beta_hv 0.286, beta_diag 0.415, theta 0.824 and alpha 0.648 (Victor and Conte, 2012).

In both our study and that of Victor et al. (Victor and Conte, 2012) perceptual salience varied for of each image statistic examined. For the gammas and betas, where thresholds were calculated, our findings were very similar. We also derived the same rank order: gamma > beta_hv > beta_diag > alpha > theta. This is notable as Victor's study utilized experienced visual psychometric participants, whereas we sampled nearly 1,000 naïve, unsupervised subjects via a crowdsourcing platform. Again, note that Victor et al. used binary textures whereas in the current study we used ternary textures (although the white-black rays contain only binary textures).

In a follow-up study in 2013, Victor et al. focused on the *binary* beta and theta ternary textures (Victor et al., 2013). Again, human sensitivity was shown to be selective, with sensitivity to betas (second-order) being 2-3 times higher than that for thetas (third-order). In the current study, because threshold values were not calculated for the theta textures, we can instead

examine the step 5 (zero decorrelation) performance values for the betas versus the thetas (Live2 data, long condition, catch trial passes >=4). This comparison is shown in Figure 20 of Section 4.4.5 above. The mean performance for the Live2 step 5 betas was 0.757; the mean performance for the Live2 step 5 thetas was 0.313. Therefore, in the current study, human sensitivity to *ternary* betas is 2.42 times higher that than for thetas, which is consistent with the Victor study cited above for binary *textures* (Victor et al., 2013).

Victor et al. also found that sensitivity to positive and negative deviations of image statistics (black- and white-bias) was not significantly different for the betas. For thetas, sensitivity to positive correlations (white-bias) was 10% greater than sensitivity to negative correlations (black-bias) (Victor et al., 2013). In the current study, the black-bias and white-bias rays were analysed. Using Live2 data across the 4 subtypes of beta textures, the median black-bias threshold for Live2 was 0.287 (95% confidence interval (CI) of 0.218 to 0.357) and median white-bias was 0.262 (95% CI of 0.182 to 0.341). Because of the large confidence interval, we cannot conclude that there was a significant difference in black- and white-bias sensitivity.

For the theta textures, the median black-bias performance level at step 5 (zero decorrelation) was 0.234 (95% CI of 0.204 to 0.265) and median white-bias was 0.187 (95% CI of 0.175 to 0.200). Based on the median data, the sensitivity to positive correlations (white-bias) was about 20.1% less than sensitivity to negative correlations (black-bias), roughly double that found by Victor (Victor et al., 2013). However, according to the 95% confidence

interval, this sensitivity difference could be as little as 2.0% or as high as 33.9%.

To an ideal observer, each image statistic surveyed is independent and should be equally informative. The variations in salience that we have observed between rays within a given trico axis, *and* between different texture families, must reflect the sensitivities and limitations of neural processing. They are consequently a manifestation of efficient coding (Barlow, 2001, Barlow, 1961). In keeping with efficient coding, humans have restricted information budgets; therefore humans have evolved to be sensitive for certain, maximally informative directions within the texture space (see Chapter 1, Section 1.4). The evolution of the human visual system has been driven by the tasks that the organism must perform, environmental constraints, and the computational limitations of neurons (both structural and metabolic) (Simoncelli and Olshausen, 2001). Therefore, human sensitivity to certain image statistics is intrinsically linked to historical ecological and behavioural imperatives. Presumably then, these maximally informative directions correspond to those within natural images.

The Live1 and Live2 studies differed in their mode of texture presentation. In Live1, the textures were presented in a pseudo-random fashion: random steps from any of the 6 rays within a trico plane. In Live2, the textures were presented in step order along each ray, starting with the most salient example, giving the subject the best chance of learning the texture types without distraction by other texture types within the trico plane. In Section 4.4.5 we analysed the effect differing presentation order. The analysis used only step 5 values, which were directly comparable in the majority of cases.

An improvement from Live1 to Live2 was observed, but the magnitude of the improvement was *small*. This suggests that the order of presentation of ternary textures within a trico plane *does not* significantly increase texture discrimination learning. This finding is interesting, as it is counter-intuitive. We might expect that presenting the most salient example of a texture to a naïve Worker first, and then subsequently presenting increasingly more difficult examples in step order, might promote learning and therefore boost Live2 performance. This does not appear to be the case.

The performance of the 31 repeat Workers was analysed Section 4.4.6. This data appears to show that gamma performance is already maximal, or close to maximal, as the gammas are highly salient. Theta and alpha performance also appears to be close to maximal, even at low levels of decorrelation, but in this case the textures are highly non-salient. In contrast, performance on the beta textures is *not maximal*; thus, for naïve subjects, there is room for performance improvements after repeated presentations (and thus learning).

In the introduction and in Chapter 1 (Section 1.5.3) we considered whether grey is perceived differently. Is there evidence of this from our isocontours and extracted values (Figures 29, 30 and 31)? For the second-order ternary textures (gammas and betas), the grey-bias rays were consistently the *second least* salient. For the third- and fourth-order ternary textures (thetas and alphas), the black-grey and grey-white rays were consistently the least salient, followed by white-black; however, grey-bias was one of the more salient textures. Therefore, we did not find a consistent pattern which indicated that grey tokens confer non-salience, as we might have expected.

A more consistent pattern emerged which showed that for the thetas and alphas, performance functions for the *corners of the trico plane* were higher than those for the quantitatively derived mixtures. To put it another way, the canonical theta and alpha textures at the corners of the trico plane (white-, grey- and black-bias) were consistently more salient than the 50:50 mixtures (see Figure 31). Furthermore, contrast adaptation, the variance of the black-white mixture, is higher than any other ray.

4.6 Acknowledgements

4.7 References

ADELSON, E. H. On seeing stuff: the perception of materials by humans and machines. Photonics West 2001-Electronic Imaging, 2001. International Society for Optics and Photonics, 1-12.

BARLOW, H. 2001. Redundancy reduction revisited. *Network,* 12**,** 241-53.

BARLOW, H. B. 1961. Possible principles underlying the transformations of sensory messages.

BARLOW, H. B. 1963. The Information Capacity of Nervous Transmission. *Kybernetik,* 2**,** 1.

BERGEN, J. R. & ADELSON, E. 1991. Theories of visual texture perception. *Spatial vision,* 10**,** 114-134.

BLAKE, A., BULTHOFF, H. H. & SHEINBERG, D. 1993. Shape from texture: ideal observers and human psychophysics. *Vision Res,* 33**,** 1723-37.

BLAKESLEE, B. & MCCOURT, M. E. 2004. A unified theory of brightness contrast and assimilation incorporating oriented multiscale spatial filtering and contrast normalization. *Vision research,* 44**,** 2483-2503.

BLAKESLEE, B. & MCCOURT, M. E. 2008. Nearly instantaneous brightness induction. *Journal of Vision,* 8**,** 15.

BLAKESLEE, B. & MCCOURT, M. E. 2012. When is spatial filtering enough? Investigation of brightness and lightness perception in stimuli containing a visible illumination component. *Vision research,* 60**,** 40-50.

BUHRMESTER, M., KWANG, T. & GOSLING, S. D. 2011. Amazon's Mechanical Turk : A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science,* 6**,** 3-5.

CAVANAGH, P. & LECLERC, Y. G. 1989. Shape from shadows. *J Exp Psychol Hum Percept Perform,* 15**,** 3-27.

COLE, F., SANIK, K., DECARLO, D., FINKELSTEIN, A., FUNKHOUSER, T., RUSINKIEWICZ, S. & SINGH, M. How Well do Line Drawings Depict Shape? ACM Transactions on Graphics, 2009.

CORNSWEET, T. 1970. *Visual Perception,* New York, Academic Press.

COY, D. G., BARBOSA, M., SEAMONS, J. W. G., & MADDESS, T. 2014. Learning Effects for Discrimination of Isotrigon Textures. (Honours Thesis; Journal Article In Preparation).

CRUMP, M. J., MCDONNELL, J. V. & GURECKIS, T. M. 2013. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS One,* 8**,** e57410.

DAKIN, S. C. & BEX, P. J. 2003. Natural image statistics mediate brightness 'filling in'. *Proc Biol Sci,* 270**,** 2341-8.

DAVID, S. V. & GALLANT, J. L. 2005. Predicting neuronal responses during natural vision. *Network: Computation in Neural Systems,* 16**,** 239-260.

DOW, S., KULKARNI, A., KLEMMER, S. & HARTMANN, B. Shepherding the Crowd Yields Better Work. ACM 2012 conference on Computer Supported Cooperative Work 2012 New York. ACM, 1013-1022.

FREEMAN, J., ZIEMBA, C. M., HEEGER, D. J., SIMONCELLI, E. P. & MOVSHON, J. A. 2013. A functional and perceptual signature of the second visual area in primates. *Nature neuroscience,* 16**,** 974-981.

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. *Cereb Cortex,* 18**,** 2416-38.

HARALICK, R. M. 1979. Statistical and structural approaches to texture. *Proceedings of the IEEE,* 67**,** 786-804.

HEER, J., BOSTOCK, M. & 2010. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. *ACM Human Factors in Computing Systems (CHI)***,** 203-212.

IPEIROTIS, P. 2010a. Analyzing the Amazon Mechanical Turk Marketplace. *ACM XRDS (Crossroads),* 17.

IPEIROTIS, P. G. 2010b. Analyzing the amazon mechanical turk marketplace. *XRDS: Crossroads, The ACM Magazine for Students,* 17**,** 16-21.

KITTUR, A., CHI, E. H. & SUH, B. Crowdsourcing User Studies with Mechanical Turk.   Conference on Human Factors in Computing Systems, 2008. 453-456.

LANDY, M. S. & GRAHAM, N. 2004. Visual perception of texture. *The visual neurosciences,* 2**,** 1106-1118.

MADDESS, T. & NAGAI, Y. 2001. Discriminating isotrigon textures [corrected]. *Vision research,* 41**,** 3837-60.

MADDESS, T., NAGAI, Y., JAMES, A. C. & ANKIEWCZ, A. 2004. Binary and ternary textures containing higher-order spatial correlations. *Vision research,* 44**,** 1093-113.

MADDESS, T., NAGAI, Y., VICTOR, J. D. & TAYLOR, R. R. 2007. Multilevel isotrigon textures. *Journal of the Optical Society of America. A, Optics, image science, and vision,* 24**,** 278-93.

MASON, W. & SURI, S. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behav Res Methods,* 44**,** 1-23.

MASON, W. A. & WATTS, D. J. 2009. Financial incentives and the performance of crowds. *Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP '10)***,** 77–85.

OPPENHEIMER, D. M., MEYVIS, T. & DAVIDENKO, N. 2009. Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology,* 45**,** 867-872.

PAOLACCI, G., CHANDLER, J. & IPEIROTIS, P. 2010. Running experiments on Amazon Mechanical Turk. *Judgment and decision making,* 5**,** 411-419.

PONTIN, J. 2007. Artificial intelligence, with help from the humans. *New York Times*, 25 March.

ROSS, J., IRANI, I., SILBERMAN, M., ZALDIVAR, A. & TOMLINSON, B. 2010. Who are the CrowdWorkers? Shifting Demographics in Amazon Mechanical Turk. *CHI EA 2010***,** 2863-2872.

SEAMONS, J. W. G., BARBOSA, M. S., COY, D. & MADDESS, T. 2015a. Developing and Validating an Isotrigon Texture Discrimination Task using Crowdsourcing. Australian National University (In Preparation).

SEAMONS, J. W., BARBOSA, M. S., BUBNA-LITIC, A. & MADDESS, T. 2015b. A lower bound on the number of mechanisms for discriminating fourth and higher order spatial correlations. *Vision research,* 108**,** 41-48.

SIMONCELLI, E. P. & OLSHAUSEN, B. A. 2001. Natural image statistics and neural representation. *Annual review of neuroscience,* 24**,** 1193-216.

TAYLOR, R. R., MADDESS, T. & NAGAI, Y. 2008. Spatial-biases and computational constraints on the encoding of complex local image structure. *Journal of vision,* 8**,** 19 1-13.

TKACIK, G., PRENTICE, J. S., VICTOR, J. D. & BALASUBRAMANIAN, V. 2010. Local statistics in natural scenes predict the saliency of synthetic textures. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 18149-54.

TODD, J. T. 2004. The visual perception of 3D shape. *Trends Cogn Sci,* 8**,** 115-21.

TODD, J. T., OOMES, A. H., KOENDERINK, J. J. & KAPPERS, A. M. 2004. The perception of doubly curved surfaces from anisotropic textures. *Psychol Sci,* 15**,** 40-6.

VERMA, A. K., SRIVIDYA, A. & KARANKI, D. R. 2010. Basic Reliability Mathematics. *Reliability and Safety Engineering***,** 15-70.

VICTOR, J. D., CHUBB, C. & CONTE, M. M. 2005. Interaction of luminance and higher-order statistics in texture discrimination. *Vision research,* 45**,** 311-328.

VICTOR, J. D. & CONTE, M. M. 2012. Local image statistics: maximum-entropy constructions and perceptual salience. *Journal of the Optical Society of America. A, Optics, image science, and vision,* 29**,** 1313-45.

VICTOR, J. D., THENGONE, D. J. & CONTE, M. M. 2013. Perception of second- and third-order orientation signals and their interactions. *J Vis,* 13**,** 21.

WHITE, M. 1979. A new effect of pattern on perceived lightness. *Perception,* 8**,** 413-416.

YELLOTT, J. I. 1983. Spectral consequences of photorecptor sampling in the rhesus retina. *Science,* 221**,** 382-385.

Chapter 5: Thesis Summary and Future Directions

At the heart of this thesis is the evaluation of human texture perception using texture analysis algorithms. Basic visual judgments are statistical in nature and we took a primarily statistical approach to evaluating texture perception.

5.1 Thesis Overview

5.1.1 Chapter 2

In Chapter 2, we used binary isotrigon textures to evaluate the underlying mechanisms of higher-order texture discrimination; this chapter has been subsequently published (Seamons et al., 2015a).

The visually salient structure of binary isotrigon textures is exclusively due to fourth- and higher-order spatial correlations (Maddess et al., 2004, Maddess et al., 2007, Julesz et al., 1978, Victor, 1994, Gilbert, 1980). Thus, in order to discriminate a particular isotrigon texture from noise patterns, it is necessary to identify its average complex, higher-order structure: isotrigon textures cannot be discriminated based on luminance or other lower-order properties. Whereas some textures are clearly structured and readily recognisable, others cannot be easily differentiated from noise (Maddess and Nagai, 2001).

Humans use a number of neurophysiological mechanisms to capture visually salient higher order structure. One statistically well principled method to infer the number and form of the underlying independent mechanisms is to use

Factor analysis of multiple human performance functions using a texture discrimination task; ideally, about as many functions should be determined as texture types examined. The performance functions can be derived from many individuals, or many repeats from single individuals (Rosli et al., 2009, Maddess and Kulikowski, 1999, Sekuler et al., 1984, Simpson and McFadden, 2005). Here we reported using the latter methodology, *many repeats from single individuals*, but also comparing the derived factors across different individuals.

In Chapter 2, we described binary isotrigon discrimination experiments using ten novel VnL2 isotrigon textures in addition to 17 standard V3L2 isotrigon textures. Subjects were found to differ somewhat in their ability to detect image structure defined by fourth- and higher-order spatial correlations. After calculating performance functions for each subject (probability correct across the different textures) and their d-prime values, two forms of Factor analysis were performed based on rotated principal components and maximum likelihood estimates (Reyment, 1996, Norris and Lecavalier, 2010). We then analysed the number of neurological mechanisms which govern the detection of fourth- and higher-order image structure in two ways: using a Scree plot and by analysing communalities (Cattell, 1966, Hayton et al., 2004, DeVellis, 2012, Zwick and Velicer, 1986). Communalities indicated the proportion of variance accounted for within the data for each texture based on an *N factor model*; we sought the models that produced the most balanced communalities.

We found that three factor models gave relatively balanced communalities for the VnL2 textures, and accounted for about 85% of the variance for a

separate V3L2 17-texture data set from two subjects. For those V3L2 textures, the lowest communalities were for the Even- and Odd- EI textures: these are among the least visually salient binary isotrigon textures (Maddess and Nagai, 2001, Tkacik et al., 2010, Victor and Conte, 1989, Victor and Conte, 1991) and possess fourth-order correlations that rarely occur in natural scenes (Tkacik et al., 2010). This study supports the possibility that three is the *lower bound* of the number of mechanisms underlying higher order texture discrimination. The Maddess group has previously provided evidence that the number of independent mechanisms is less than 10 (Taylor et al., 2008, Barbosa et al., 2013), and is perhaps as small as 3 to 4 (Maddess and Nagai, 2001, Maddess et al., 2007), using different types of isotrigon texture and means of determining the estimated number of mechanisms.

A previous study by Barbosa et al. asked if linear combinations of various moments of the Minkowski functionals (MFs) could explain human performance functions for 33 types of isotrigon texture, 25 of which were used here (Barbosa et al., 2013). Texture 042 was the least well explained by the 3 factor model, showing the lowest communalities in 3 of the 4 data sets. In this case however, 042 was not distinct from the other textures based on MF content.

Other than JWS, none of the subjects showed significant evidence of learning. In a previous study, some subjects markedly improved their ability to discriminate V3L2 textures; the increase in probability correct for 16x16 textures was 20% for one and 30% for another texture type in just 5 repeats (Maddess and Nagai, 2001). A recent study by our group, which used a set

of 17 V3L2 textures, found strong evidence of learning effects described by an exponential rise with a time constant of about 5 days (Coy et al., 2014). That data also showed that a three factor model was reasonable.

The formation of recursively applied products is physiologically plausible and may occur via dendritic back-propagation (Stuart et al., 1997, Buzsaki and Kandel, 1998) or dendritic spiking (Mel, 1993, Stuart et al., 1997, Hausser et al., 2000). An unpublished modelling study from our group has also shown that dendritic back propagation in one pyramid cell is sufficient to discriminate some isotrigon textures from random ones (Taylor, 2013). Overall, the mechanisms identified in this study, and previous studies (Maddess and Nagai, 2001, Maddess et al., 2007), may represent some combination of recursive or rectifying processes.

5.1.2 Chapter 3

In Chapter 3, we continued our exploration of higher-order texture discrimination using a complementary approach to that employed in Chapter 2, i.e. analysing isotrigon discrimination performance functions *from many individuals*. An efficient way to implement such a study was to use a crowdsourcing platform. For this study, we utilised the crowdsourcing platform Amazon Mechanical Turk (mTurk) (Ross et al., 2010, Pontin, 2007).

A concern with a novel platform such as mTurk is whether the data derived is of high quality. Issues such as demographic variations, subject motivation and subject expertise apply to many mTurk studies and have been well studied (Ipeirotis, 2010b, Kittur et al., 2008, Mason and Suri, 2012, Ipeirotis,

2010a). Several authors have successfully used mTurk to replicate laboratory studies (Paolacci et al., 2010, Horton et al., 2011, Crump et al., 2013). However, there is a lack of published mTurk studies involving *visual* psychometric testing (Heer et al., 2010, Cole et al., 2009, Freeman et al., 2013). Therefore, this Chapter fulfilled the secondary role of evaluating crowdsourcing as a platform for visual psychometric research; in so-doing it also laid the foundations for a second, larger crowdsourced study (described in Chapter 4).

In the first part of Chapter 3, we discussed the development and implementation of a binary isotrigon discrimination task (HIT) using mTurk. The HIT was based on a four alternative forced choice task developed by Victor et al. (Victor et al., 2013). It was intended to provide a more complex and naturalistic stimuli than those traditionally used in visual perception experiments based on gratings. For testing purposes, a pilot study was completed within the non-public Sandbox environment. During this "Lab" phase, laboratory members completed 270 HITs on 6 different platforms. The effects of variations between platforms were then evaluated, using correlational analysis, Factor analysis, and coefficients of repeatability (Bland and Altman, 1986, Bland and Altman, 1999, Bland, 2000, Vaz et al., 2013).

Coefficients of repeatability and Pearson's correlation coefficients were consistently low across the machine types tested. Factor scores were also plotted from a Factor analysis. Machine-specific features appeared to have little impact on discrimination performance. Small differences were observed for two screen sizes (the extremes of those tested). In conclusion, the HIT was found to be robust to platform variations such as gamma, DPI, and

physical screen size; this is consistent with a previous study by our group which evaluated the effects of contrast and pixel size on isotrigon discrimination (Maddess and Nagai, 2001).

In the second part of Chapter 3, the HIT was uploaded to the live mTurk website and crowdsourced performance data was collected in two phases (Live 1 and Live2). We found the rate of HIT completion to be broadly consistent with previous reports (Heer et al., 2010, Buhrmester et al., 2011). In terms of catch trial failures, the quality of the data was excellent (90.4% HITs retained in Live1 and 95.2% in Live2). There was an even distribution of catch trial failure across textures, indicating that the catch trials did not bias any particular texture type.

To further evaluate its quality, the Live1 and Live2 data sets were compared to the Lab (M1-M6) data set and a second data set ("DC"), which was gathered under supervised laboratory conditions using naïve subjects (Coy et al., 2014). The distribution of median performance scores for the Live1/2 data sets was superficially very similar to the M1-M6/DC results. The performance scores of the DC data set were slightly higher than Live1/2 across the textures examined. The performance scores of the M1-M6 data set were shifted upwards 10-15% with respect to the Live1/2 data sets, which may indicate the experience of that subject to the study protocol. In the case of the DC data set, subjects started as naïve, but became more experienced with repeated visits; this is consistent with previous findings that the discrimination of these textures can be improved by learning, although many repeats appear to be required (Maddess and Nagai, 2001, Taylor et al., 2008, Coy, 2014).

A strong correlation was observed between M1-M6/DC data sets and Live1/2 based on median performance. Absolute repeatability was evaluated using coefficient of repeatability (CR), derived from the Bland-Altman plots. The CR between M1-M6/DC and Live1/2 were low and suggested that the Live mTurk study had accurately reproduced the texture discrimination results that were produced by supervised laboratory trials. The CR between Live1 and Live2 was only 9.4% and this very low score indicates that the live study itself has a good repeatability.

Based on these metrics, we concluded that, in the case of an isotrigon texture discrimination task, mTurk can produce comparable quality data to that derived from laboratory studies. The strong agreement between Live1/2 and M1-M6/DC data sets is interesting considering the considerable variations in the browser, screen size, screen DPI and OS reported. One of the few other visual psychometric mTurk studies observed a 92% correlation between mTurk and laboratory data for 15 textures tested (Freeman et al., 2013). Other non-visual studies have reported gathering high quality mTurk data based on judgment and decision-making (Paolacci et al., 2010) and other behavioural psychology experiments including the Prisoners' Dilemma (Horton et al., 2011, Crump et al., 2013).

Factor analysis of the Live1/2 data found that 3 to 4 is the lower bound of the independent neurological mechanisms that govern texture discrimination in this study. This finding is consistent with the results reported in Chapter 2 (Seamons et al., 2015a) and for other isotrigon textures using different methods (Maddess and Nagai, 2001, Maddess et al., 2007, Coy, 2014). In

light of the different study modalities, this strong congruence in notable and lends further support to the 3-factor model.

5.1.3 Chapter 4

In Chapter 4, we focused on *ternary* textures generated via the stochastic method (Victor and Conte, 2012). When studying human texture perception, the selection of appropriate visual stimuli is challenging. The space of signals that the visual system encounters is rich and varied; thus, many different image statistics could potentially drive visual judgments. The stochastic textures are statistically well-principled and allow multiple image statistics are be *independently* controlled. The stochastic method has previously been described for the binary textures (Victor and Conte, 2012).

Palettes of stochastic ternary textures are created in a space, where the centre location contains ternary noise; the three corners of the space are canonical texture types and the remaining space is filled with quantitatively derived mixtures (Victor and Conte, 2012). The canonical ternary textures have well-defined spatial correlation structure of different orders between the three grey levels of the textures. They are classified according to the number of pixels that are constrained within a 2x2 pixel grid. If a single pixel is constrained, the textures are termed "gamma". If 2 pixels are constrained, they are termed "beta_hv" or "beta_diag" (depending on the orientation of the constrained pixels). If triplets of pixels are constrained they are termed "theta" and quadruplets of pixels constrained are "alpha" (Victor et al., 2013, Victor and Conte, 2012). The ternary textures can be projected onto so-

called *trico* planes; in so-doing, three-level stimuli can be displayed coherently in two dimensions.

To the ideal observer, changes in all directions of the trico plane are equally detectable. However, humans are *not* ideal observers; in accordance with the efficient coding hypothesis, the informational resources of the human visual system are constrained. The efficient coding hypothesis states that neurons have limited energy budgets and information capacities; therefore, they should not waste energy nor bandwidth by transmitting *redundant* information (Barlow, 1961, Barlow, 2001). To put it another way, in order to maximise available computing resources, neurons should encode as much *functional* information as is structurally and energetically sustainable.

If efficient coding holds, we would expect some directions within the trico axes to be more informative than others. In light of the large number of stochastic ternary textures available, a logical way to proceed was to use crowdsourcing to gather texture discrimination functions from a large number of subjects. The ternary textures differ from their binary counterparts previously studied (Seamons et al., 2015a, Seamons et al., 2015b, Maddess and Nagai, 2001, Maddess et al., 2004) by the presence of a third token, grey. Therefore, a natural question is whether the grey element of ternary textures has special properties with regard to texture perception.

In Chapter 4, we examined a subset of the possible ternary texture space consisting of 13 texture planes (of a possible 66), each with 6 pre-defined rays with 5 noise levels. We developed a ternary texture testing crowdsourcing method based on that previously described for binary

isotrigons (Seamons et al., 2015b). We found that perceptual salience varied for each image statistic examined. The textures had a salience rank order of gamma > beta_hv > beta_diag > alpha > theta, which is congruent with that previously reported for related binary stochastic textures (Victor and Conte, 2012). The most salient statistic (gamma) is defined by the pixel-by-pixel mean and variance. Second-order statistics affect the spatial-frequency content of the stimulus and are thus readily detectable by linear filters. However, that fourth-order correlations (alphas) are more salient than third-order correlations (thetas) is less readily explicable.

We also examined performance along specific directions on the trico planes. The six axes explored different ad-mixtures of noise and texture. For the gammas and betas, the white-black and grey-bias directions were consistently least salient. For the thetas and alphas, the black-grey and grey-white directions were consistently least salient. To the ideal observer, each image statistic should be equally informative. Therefore, the observed differences reflect the sensitivities and limitations of neural processing and are therefore a manifestation of efficient coding. We hypothesised that the grey token of ternary texture may confer non-salience. Indeed, for the gammas and betas, the grey-bias was consistently the second least salient. However, this relationship did not hold for thetas, or alphas.

The effect of texture presentation *order* was also examined. In Live1, textures were presented in a pseudo-random fashion (random steps from within a single trico plane). In Live2, the textures were presented in step order along each ray, starting with the most salient example. Counter-intuitively, the order of presentation *did not* significantly affect texture

301

discrimination performance. An analysis of 31 repeat Workers found evidence of learning for the beta textures, whereas performance for the gammas, thetas and alphas appeared to be already maximal.

5.1.4 Summary

1. In Chapter 2, we evaluated human texture discrimination using a novel set of 10 isotrigon textures (VnL2) and also 17 standard V3L2 isotrigon textures

2. About 3 principal neurophysiological mechanisms may govern the detection of complex image structure; this is the lower bound. The Maddess group has previously provided evidence that the number of independent mechanisms is less than 10 (Taylor et al., 2008, Barbosa et al., 2013), and is perhaps as small as 3 to 4 (Maddess and Nagai, 2001, Maddess et al., 2007, using different types of isotrigon textures and means of determining the estimated number of mechanisms.

3. There was limited evidence of improved discrimination by learning, although it has been observed in other studies by our group. A recent study by our group, which used a set of 17 V3L2 textures, found strong evidence of learning effects described by an exponential rise with a time constant of about 5 days (Coy et al., 2014).

4. In Chapter 3, we described the development of a binary isotrigon texture discrimination task for use with the crowdsourcing system Amazon Mechanical Turk (mTurk).

5. Under laboratory conditions, we demonstrated that this experimental modality was robust across a variety of different platforms, which encompassed a range of browsers, operating systems, resolutions, physical screen sizes and contrasts.

6. Using mTurk, texture discrimination data was then gathered from 121 subjects and compared to two independent laboratory data sets. Based on Pearson's correlation, coefficients of repeatability, and Factor analysis, mTurk is capable of producing visual psychometric data which is of comparable quality to that derived in laboratory studies using the same testing modality. This validation of crowdsourced visual psychometric data is significant because, due to perceived technical limitations, very few such studies have been published to date (Heer et al., 2010, Cole et al., 2009, Freeman et al., 2013).

7. Factor analysis was performed and indicated the presence of 3-4 independent texture discrimination factors, again consistent with previous published studies (discussed above in point 4). This is

encourging considering the variety of browsers, screen sizes, screen resolutions and operating systems/browsers recorded.

8. In Chapter 4, we analysed the perceptual salience of a range of ternary textures using a crowdsourcing platform. We began by discussing the development of a ternary texture HIT, based on the binary HIT previously described in Chapter 3. The trico planes allow what are defined on three sensory dimensions to be displayed coherently in two dimensions. The objective of this study was to determine which dimensions of the trico plane are independently processed.

9. Perceptual salience was found to vary for each image statistic examined. We observed the textures to have a salience rank order of gamma > beta_hv > beta_diag > alpha > theta, which is congruent with that previously reported for related binary stochastic textures (Victor and Conte, 2012).

10. We also examined performance along vectors between ternary noise and 50:50 mixtures of the three canonical textures per trico-plane: white:black, black:grey, and white:grey. For the gammas and betas, the white:black and grey bias directions were consistently the least salient. For the thetas and alphas, the black:grey and grey:white directions were consistently the least

salient. The observed differences reflect the sensitivities and limitations of neural processing and are therefore a manifestation of efficient coding.

11. We hypothesised that the grey token may confer non-salience. Indeed, for the gammas and betas, the grey bias was consistently the second least salient. However, this relationship did not hold for thetas or alphas.

12. The effect of texture presentation *order* was also examined. In Live1, textures were presented in a pseudo-random order (random steps from within a single trico plane). In Live2, the textures were presented in step order along each ray, starting with the most salient example. Counter-intuitively, the order of presentation *did not* significantly affect texture discrimination performance. An analysis of 31 repeat Workers found evidence of learning for the beta textures, whereas performance for the gammas, thetas and alphas appeared to be already maximal.

We will now consider future research directions which follow naturally from the findings outlined herein. Four such research directions will be considered: a study of material perception using 3D stimuli derived from binary isotrigon textures; a study utilizing isotrigon discrimination as a clinical diagnostic aid in neurological disease; an investigation of 3D illusions in

ternary textures; and the application of the ternary texture testing modality to convolutional neural networks.

5.2 Future Directions

5.2.1 "Bones" and Material Perception

5.2.1.1 Material Perception

Material perception is the study of how we infer material properties from visual information. Based on everyday experience, we can distinguish numerous different categories of material and can recognize many specific materials within each class (Fleming, 2014). In some cases, this can be considered an extension of texture perception, in that we are recognising textures and associating them with learned associations about previously encountered materials (Landy and Graham, 2004). This must be the case where material properties cannot be directly perceived. However, judgements about material properties are often extended to *unfamiliar* materials and there is growing evidence that some of this information is visually encoded. As Fleming puts it: "…there is almost certainly more to material perception than our ability to categorize or recognize familiar materials" (Fleming, 2014). Research into material perception aims to understand what that visual information is, the neurological processes involved in extracting it, and how it is used to guide our behaviour (Anderson, 2011, Fleming, 2014, Gibson, 2013). This challenge has been aided by advances in computer graphics which have made it possible to simulate different materials in realistic illumination fields (Anderson, 2011).

What is the evidence that some material perception is visually encoded? Sharan et al. presented evidence that humans can identify a wide range of materials, even with presentations as short as 40 ms (Sharan et al., 2009). In another study, Fleming et al. showed subjects photographs of materials and asked them to rate qualities such as hardness and gloss. The ratings of the individual samples were systematically clustered into categories, suggesting that subjects could classify materials through visual judgments of their properties (Fleming et al., 2013).

Subjects also appear to be adept at distinguishing between photographs of *real* and *fake* materials, again with short (40 ms) presentation times (Sharan et al., 2008). This is notable as the differences between the two conditions were often subtle and hard to define. The neurophysiological basis for such judgements is unclear. However, a recent study by Freeman et al. indicates a particular functional role of V2 in the discrimination of natural versus synthetic image structure (Freeman et al., 2013). Much of the work on material perception has focussed on specific properties such as surface roughness, gloss and transparency (Anderson, 2011, Thompson et al., 2011, Zaidi, 2011, Padilla et al., 2008, Pont and Koenderink, 2008). Taken together, these findings seem to support the hypothesis that - at least in certain circumstances - the human visual system can estimate the properties of materials from retinal images.

Despite its subjective ease, material perception poses the visual system with significant challenges; the image of a given material can vary dramatically depending on variations in lighting, viewpoint and shape. This means that the visual system cannot recognize materials by simply matching the image

against a stored *template*. In order to recover the intrinsic properties of a material, the visual system must disentangle these various contributions (Fleming, 2014). The field of material perception is in its infancy, but three theoretical models have emerged (as reviewed in (Fleming, 2014)). It should be noted that, although developed through the study of gloss, surface roughness, and other parameters, models of material perception can be naturally extended to any material properties.

The theory of inverse optics proposes that the visual system attempts to accurately model a scene. In so-doing, it operates like a computer graphics program, reconstructing the positions of light sources, surface geometry and reflectance (Pizlo, 2001, Poggio et al., 1985). For example, to evaluate the reflectance of a sphere, the visual system would infer that the surface is spherical, model the scene surrounding the sphere, and use that information to dissect the various contributions of lighting and geometry to the retinal image. Once these factors are removed, the intrinsic reflectance of the object can be established (Fleming, 2014).

Von Helmholtz was one of the earliest proponents of this theory. He proposed that the visual system uses inverse optics to recover albedo (the proportion of incident light that is reflected by a surface) (von Helmholtz, 1962). However, it is now generally agreed that the brain does not perform literal inverse optics computations. Based on the information available to the visual system, such computations would be intractable (Anderson, 2011, Fleming, 2014). Modified inverse optics hypotheses have been developed which assume that the brain imposes constraints to reduce ambiguities. For example, it may be assumed that a single light source is present or that the

reflectance of a surface is uniform. Such assumptions can make inverse optics problems tractable for a limited set of viewing conditions and material properties (Romeiro and Zickler, 2010).

An alternative hypothesis proposes that the brain identifies image statistics that are *diagnostic* of material properties, but which remain reasonably *invariant* across different viewing conditions (Fleming and Bülthoff, 2005, Motoyoshi et al., 2007, Nishida and Shinya, 1998). In the natural world, viewing conditions are not completely arbitrary, so retinal images of materials might demonstrate statistical regularities. Detecting these *signature appearances* could allow the visual system to identify material properties without having to accurately model the entire scene (Fleming, 2014). Several investigators have provided experimental support for the *invariant image statistics* hypothesis (Motoyoshi et al., 2007, Fleming et al., 2003). For example, in a study of perceptions of gloss and albedo, Motoyoshi et al. concluded that human observers use the skew of the luminance histogram as the basis for their judgements (Motoyoshi et al., 2007). This was true whether the degree of skew was naturally occurring or introduced via digital manipulation.

Invariant image statistics have the advantage of being able to cope with arbitrary material properties, as long as they exhibit distinctive image features. A disadvantage of this model is that the visual system may be fooled when the assumed statistics of the natural world are infringed. For example, Fleming et al. found that human gloss constancy was good under naturalistic illuminations, but performance decreased significantly when using artificial sources of illumination (Fleming et al., 2003). The authors suggest

that the visual system relies on specular reflection *signatures* which, when infringed, cause gloss constancy to fail. However, like failures in luminance constancy (Kingdom, 2011), these failures may be rare under natural conditions.

A third hypothesis has emerged from gloss perception studies using conglomerate materials with different reliefs (Ho et al., 2008). Surfaces of identical reflectance are perceived as more glossy when they have shallow relief and frontal illumination. This observation does not appear to fit with a model where physical properties are being estimated by the visual system; why should relief affect gloss perception? In a follow-up study, Marlow et al. proposed that subjects predict surface gloss based on the "relative salience" of highlights (highlight size, contrast, etc.) (Marlow et al., 2012). This led Fleming to propose that the brain performs material perception by identifying *statistical regularities that are predictive of material properties* (Fleming, 2014). In the case of gloss constancy, subjects use the characteristics of reflections, such as highlights, to estimate reflectance. Thus, the goal of material perception may be to "…identify and measure statistically informative appearance attributes…", and not to estimate intrinsic physical properties; Fleming refers to these as *statistical appearance models* (Fleming, 2014).

An advantage of statistical appearance models is that they are much easier to compute and can be readily adapted to unfamiliar materials. This hypothesis may also explain phenomena that do not appear to fit with other models of material perception, such as the association between relief and gloss perception (Ho et al., 2008, Marlow et al., 2012, Fleming et al., 2011).

310

5.2.1.2 Minkowski Functionals and "Bones"

Minkowski functionals (MFs) are combinations of 1st to 4th order correlations computed for 2x2 pixel blocks of a binary texture. For 2D binary textures, the functionals are referred to as the area, perimeter and Euler number. Informally, area refers to the total area of the *holes* in the texture; perimeter refers to the total distance surrounding the holes; and the Euler number indicates the degree of connectedness within the texture, and can be thought of as a measure of porosity (Michielsen and De Raedt, 2001). On a square lattice of pixels, the Euler number can be based on 4- or 8-way connectedness (χ_4 and χ_8) (Michielsen and De Raedt, 2001).

As discussed in Chapter 2, the relationship between isotrigon texture discrimination performance and MFs has been investigated by Barbosa et al. (Barbosa, 2013). They considered whether linear combinations of MFs could model human performance functions for 33 types of binary isotrigon texture. The regression model with the lowest deviance contained combinations of some of the mean, variance, kurtosis and skew of χ_4 and χ_8 (Barbosa, 2013). This provided a more parsimonious account of texture perception than some other posited basis functions.

An interesting feature of the Minkowski functionals is that they can be extended to describe two-component 3D materials (Michielsen and De Raedt, 2001). Two component materials may either be composed of two different materials (like fibreglass), or one material and voids (like sponge). It has been shown that the MFs of 3D objects are predictive of mechanical properties of such materials (Schroder-Turk et al., 2011, Schroder-Turk, 2010, Michielsen and De Raedt, 2001). For higher dimensions, there is one

more MF than the dimensionality. If MFs can be related to the mechanical properties of the materials, then by extension the MFs of the surface texture might also provide clues to these mechanical properties (Barbosa, 2013, Michielsen and De Raedt, 2001). In this manner, *MFs may form a bridge between the areas of research discussed above: material perception and the perception of surface texture*.

Analogous with the early investigations of image statistics in 2D textures, a limiting factor in the investigation of material perception is the availability of suitable 3D stimuli with controlled and readily interpretable statistics. In this regard, we speculated whether the isotrigon textures described above could be extrapolated to 3D models. According to the inverse image statistics model (Pizlo, 2001, Poggio et al., 1985), it is conceivable that the brain computes MFs from a retinal image of a 3D object. This computation would occur on the basis of observation of surface material structure (for example, the voids and occlusions visible on the surface of a sponge). Recall that the MFs can be computed from local 2x2 blocks of the surface texture (Michielsen and De Raedt, 2001), i.e. highly local image properties. Because internal structure is not visible, an implicit assumption would be required that the material is *structurally consistent*. Based on inferred MF data, the physical properties of a material could then be computed.

Alternatively, MFs are candidate *invariant image statistics* (Fleming and Bülthoff, 2005, Motoyoshi et al., 2007, Nishida and Shinya, 1998) which might be used to infer physical properties of materials. This seems more likely, as characteristics defined by MFs would not be affected by changing environmental factors. It should be noted that the *perception* of MFs could be

affected by environmental factors however. For example, in the case of surface roughness, one study found that subject evaluations were biased by angle of illumination (Ho et al., 2006). It is therefore feasible that evaluations of different angles of illumination could enhance or detract from the appearance of holes, making them appear more or less deep. This has not been evaluated experimentally, although other studies of the perception of surface roughness may be relevant (Padilla et al., 2008, Pont and Koenderink, 2008).

Lastly, it could be that MFs are indirectly related to material perception, in accordance with the *statistical appearance model* (Fleming, 2014). i.e.: MFs describe the inherent properties of a material; the inherent properties of a material dictate its statistical appearance model; the brain uses the statistical appearance model to infer material properties. In this is the case, we might expect *relative judgements* of material properties to be reasonably consistent between different observers, but the *veracity* of those judgements in absolute terms to be poor. This might would indicate that direct computation of material properties is not occurring, rather an inference is being made based on statistical appearance (Fleming, 2014).

5.2.1.3 Study Design and Implementation

One way to test hypotheses of material perception is to develop novel 3D stimuli with defined, predicted MFs and calculable physical properties. These 3D stimuli could be displayed to subjects under consistent lighting conditions. Subjects would then be asked to rate the stimulus in terms of a physical

metric, such as compressive of tensile strength. The material perception data of the subjects would then be analysed in the context of known MF data and predicted physical properties.

The first step in producing 3D isotrigon texture-derived structures, which have been colloquially termed "Bones", was to decide how to produce the structures. Binary and ternary isotrigon textures are produced using deterministic and stochastic processes (Victor and Conte, 2012, Maddess et al., 2007, Maddess and Nagai, 2001). In the binary isotrigon case, these methods produce a 2D matrix of 1s and -1s representing black and white pixels. One way to produce 3D matrices is to combine two binary isotrigon 2D matrices by taking their outer products. i.e.: we take the *n outer products* of the two binary isotrigon matrices, where *n* is the length of the Bone (in this case 32). Before taking the outer products, the matrix is converted to 1s and 0s. This produces a 3D matrix of size 32 along each axis. This process is illustrated in Figure 1 below. To accommodate a wide variety of physical properties, it was decided that the Bones should be made with dimensions 32x32x96. This was achieved by combining 3 copies of the same 3D bone matrix, end to end.
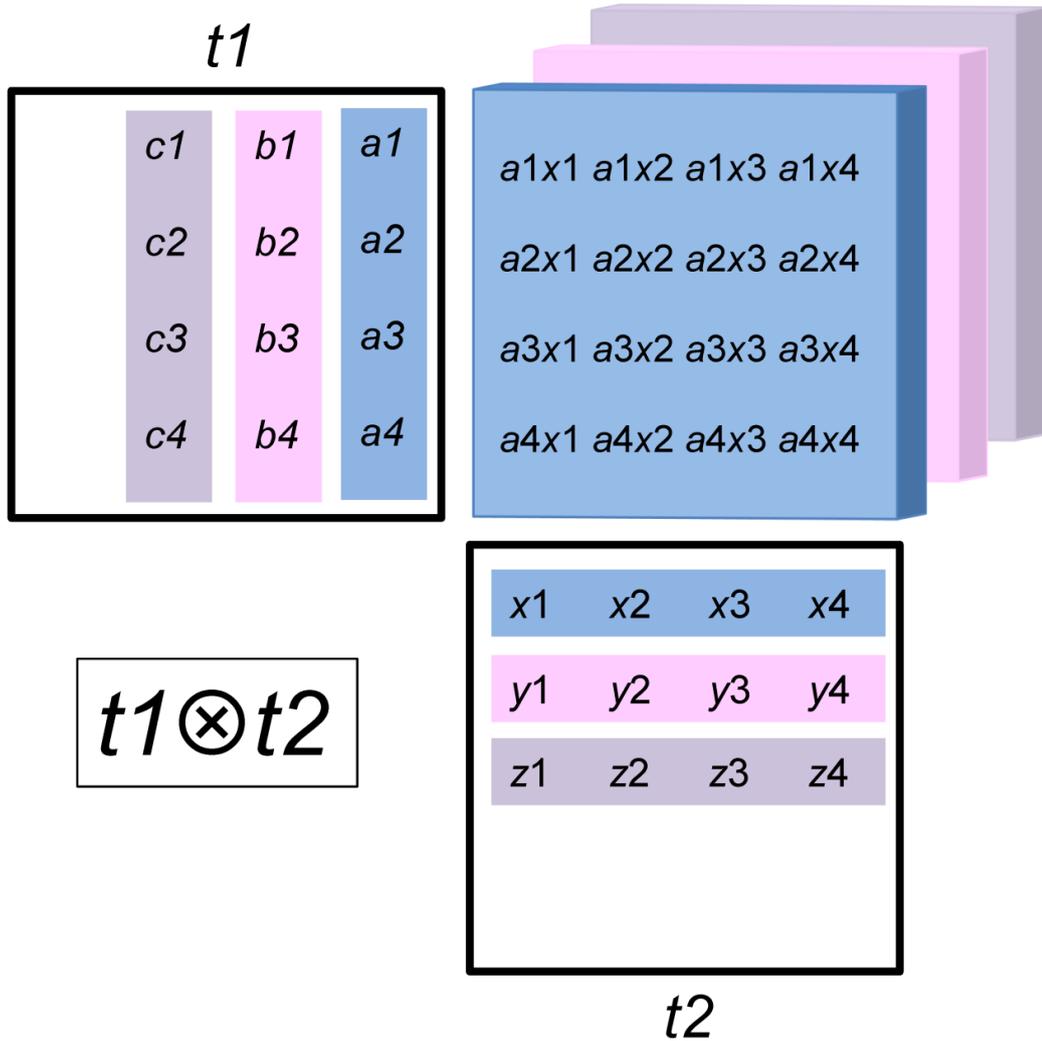
Figure 1. Generating the 3D Bone matrix by taking a series of outer products. *t1* and *t2* represent two binary isotrigon textures. Each of these textures is composed of a 2D matrix of 1s and 0s (of size 32x32). To form the first slice of the Bone matrix, the outer product is taken between row *x* and column *a* (highlighted in blue in the Figure). The resulting slice is a 32x32 matrix, in which each 1 represents a 1x1x1 cube and each 0 a 1x1x1 void. The next slice is formed by taking the outer product of row *y* and column *b* (pink), and so on. The series of outer products results in the formation of 32, 32 x32x1 slices,

and therefore a 32x32x32 cube. Three such cubes are stacked to form one Bone (of size 32x32x96).

Bones were systematically produced from 14 types of binary isotrigon texture, each of which had Even and Odd variants: Bar, Box, Corners, Cross, Diag, El, Foot, Oblong, Rod, Tee, Triangle, Wolf, Wye, Zigzag. The product of glider *a* and *b* is compositionally the same as the product of *b* and *a*, except the 3D matrix produced is a 90 degree rotation. Therefore, the number of possible combinations of 2 texture types chosen from the 28 texture types, where order is not important and repetition is not allowed, is 378. To indicate the taking of the outer products of two binary isotrigon matrices, we use the nomenclature "a⊗b", where *a* and *b* are texture types.

One interesting consequence of the outer product process is that the resulting Bone matrices always have 2 ends which are BoxEven. Examples of Bones produced in this manner are shown below in Figure 2.

The next step was to select a subset of the possible Bones which had reasonably *distinct* MFs. For each Bone type, there is some variation in MFs between examples (Figure 3).

A   B   C
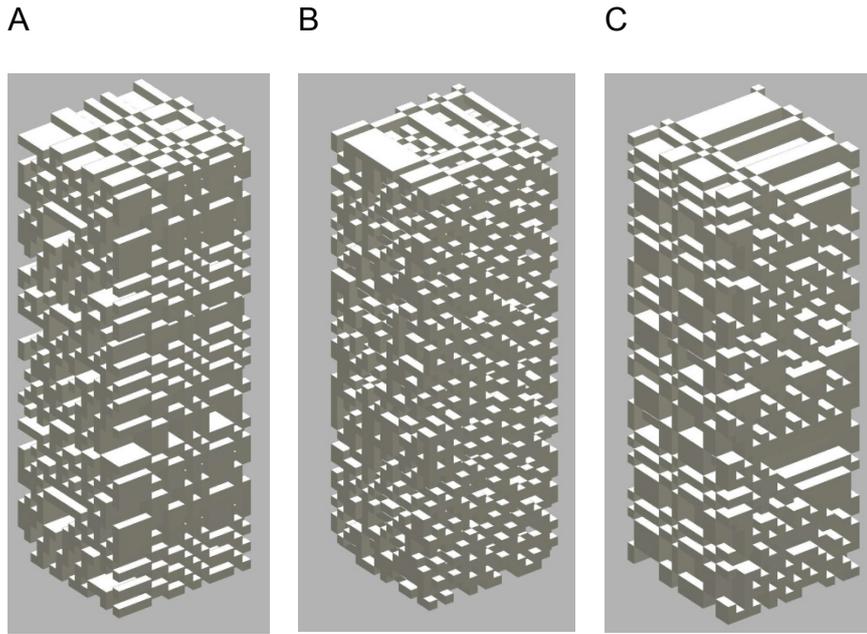
Figure 2. Bones from three families. A: BoxEven⊗TriangleEven. B: BoxOdd⊗CrossOdd. C: TriangleOdd⊗BoxEven.
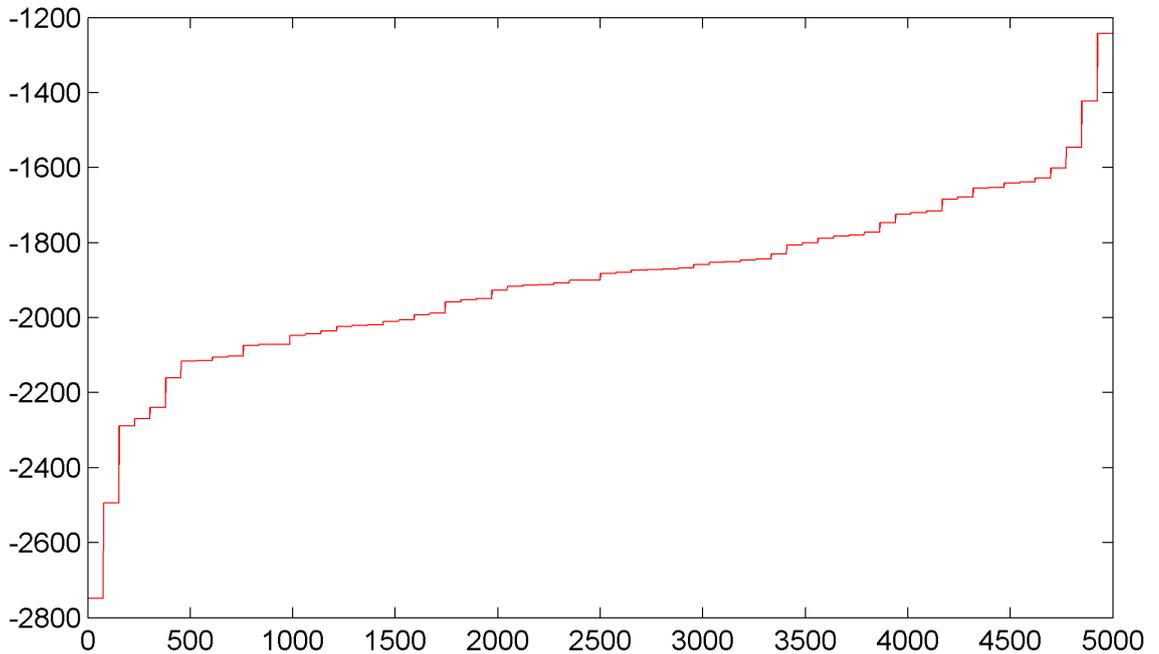
**Figure 3**. Plot of Euler number versus example number for 5,000 examples of the TeeOdd⊗CrossEven Bone. Note the variation in the magnitude of the Euler number, from -2,749 to -1,243.

In light of this sample variation, the MF data for 5,000 examples of each Bone type were plotted (Figures 4 and 5).
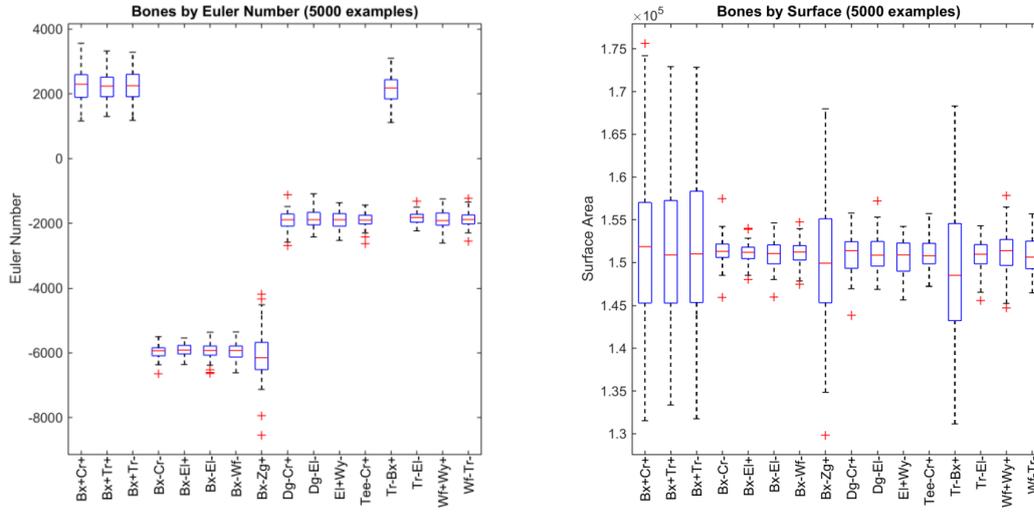

Figure 4. MFs (Euler number and Surface) for 5,000 examples of Bones from 16 Bone families.
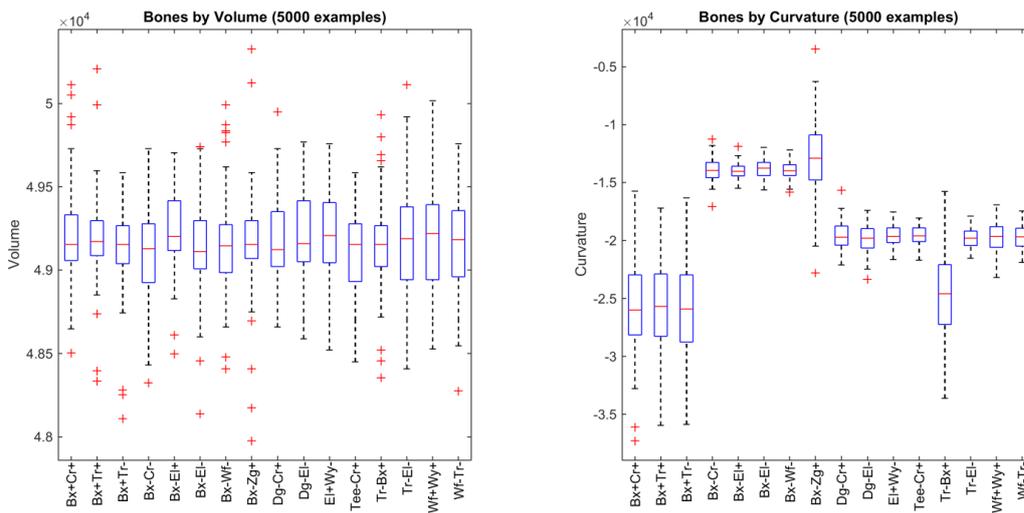

Figure 5. MFs (Volume and Curvature) for 5,000 examples of Bones from 16 Bone families.

Based on this analysis, the 11 Bone families with the most distinct MFs were selected to test. The Euler numbers for these Bone families are plotted below in Figure 6, based on 10 examples of each Bone.
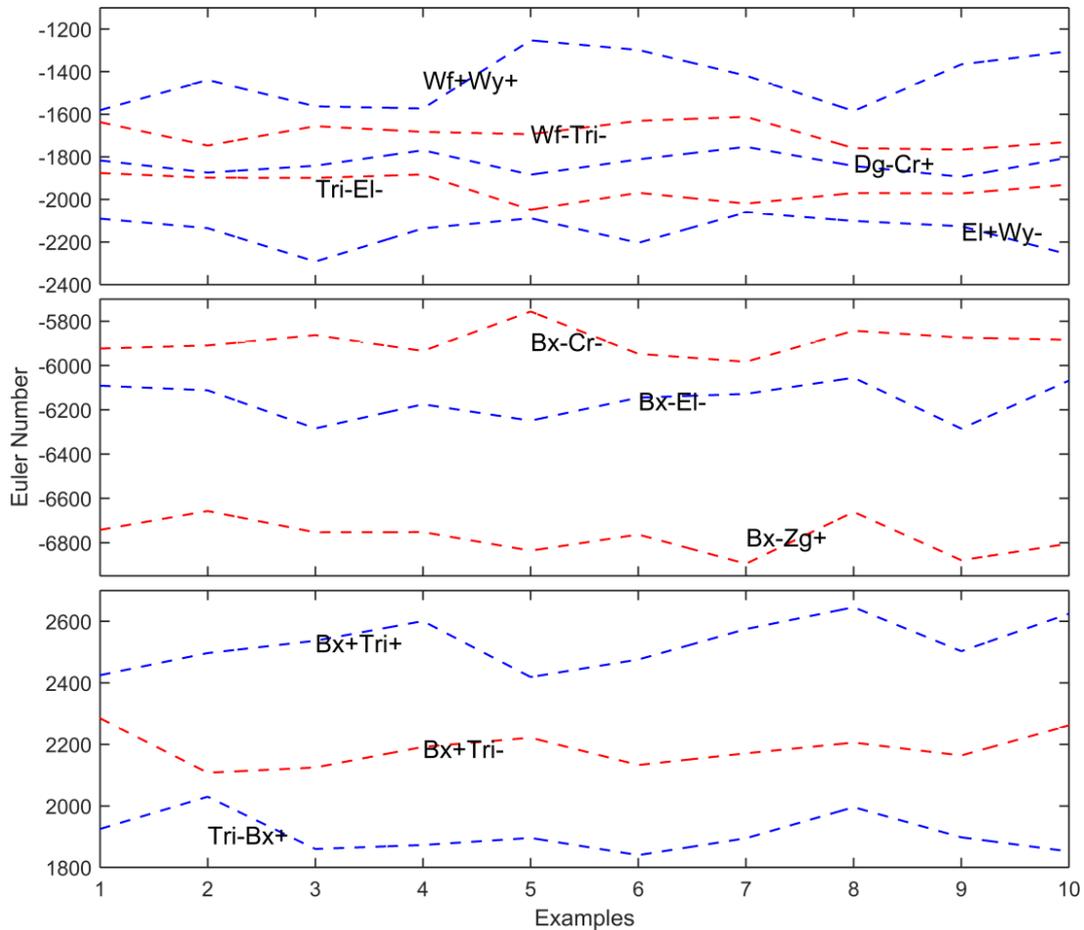

Figure 6. Selected Bone families, based on Euler number. The 11 most distinct Bone families were selected, based on Euler numbers for 5,000 examples.

In total, 11 Bone families were selected for maximum variability in MFs. Ten examples of each Bone was produced. In addition to the 3D surface models produced in Matlab, static images were produced for priming stimuli, and STL files were produced for use with COMSOL Multiphysics software

(COMSOL, 2014). COMSOL Multiphysics can be used to calculate the predicted material properties of an STL model.

For the purposes of a pilot study, a Matlab graphical user interface (GUI) was developed and implemented, which allowed Bone examples of be displayed in a random order and subject material perception judgements recorded (Figure 7). Within the GUI, the Bones are rendered as simple "surf" objects in consistent light fields. The Bone examples rotate around the Y axis to show the different faces.
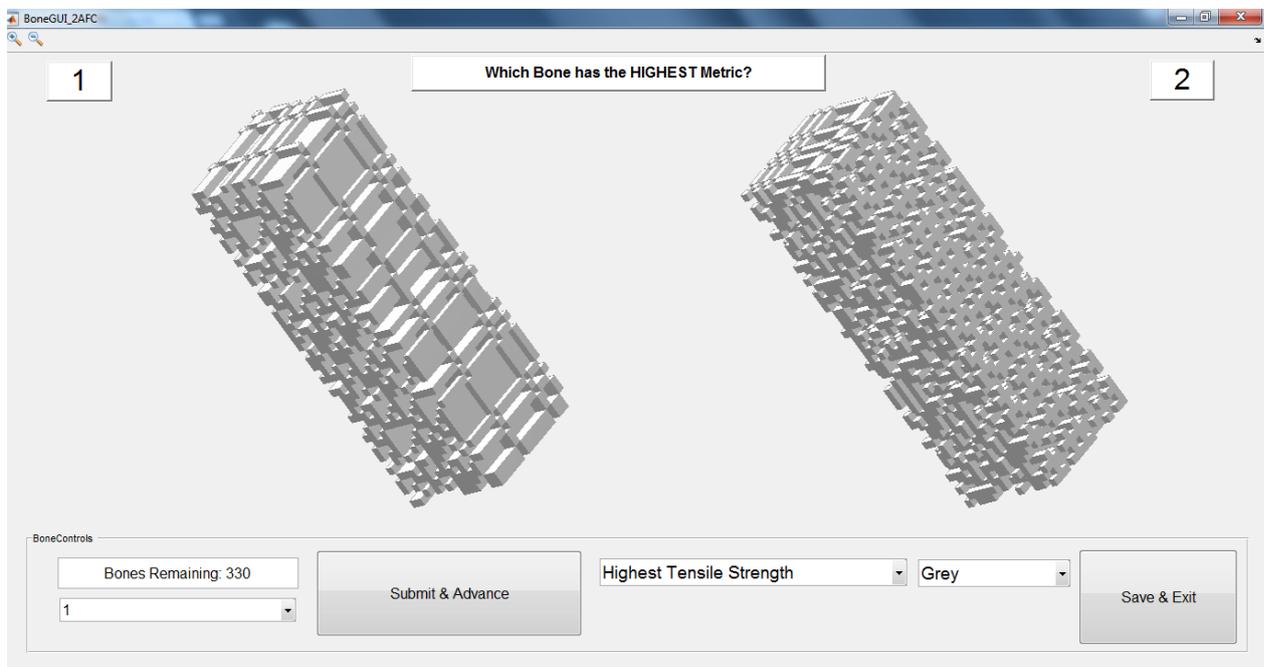

Figure 7. Matlab GUI for presenting Bone examples and recording material perception judgements. Bones are presented in a random order. On presentation, the Bones rotate around the Y axis, to display the distinct faces.

In the first instance, one subject (JWGS) completed 1,320 replications for the four metrics discussed above.

5.2.1.4 Possible Study Extensions

Based on the results of the pilot study, a number of study extensions might be considered.

Matlab provides basic support for rendering and lighting 3D objects, but it is not its primary purpose. Because we are asking subjects to make material perception judgements about the stimuli, it is important that they look *realistic*. Therefore, the use of sophisticated 3D modelling software may be warranted, such as the open source 3D graphics and animation program Blender. Blender can produce photo-realistic object models with complex light fields and a variety of different surface properties. According to our hypothesis, material perceptual judgements should be unaffected by cosmetic changes to the surface and this could also be explored. Some examples of Blender objects, using a variety of different surface materials, are shown in Figure 8.

A related study extension might be to 3D print Bone examples using the STL files already produced. This is relatively cheap and there are many commercial outfits which offer small scale 3D printing in a variety of materials. The 3D printed Bone examples might fulfil two purposes. Firstly, they could be used to *prime* subjects before testing. Subjects could view them before performing the material perception task, to help them bridge the conceptual gap between the visual presentation of simulated materials and the perception of "real materials". Secondly, the physical properties of the 3D printed examples could be destructively tested, using methods such as uniaxial tensile testing; this would enable us to evaluate the veracity of COMSOL predictions.
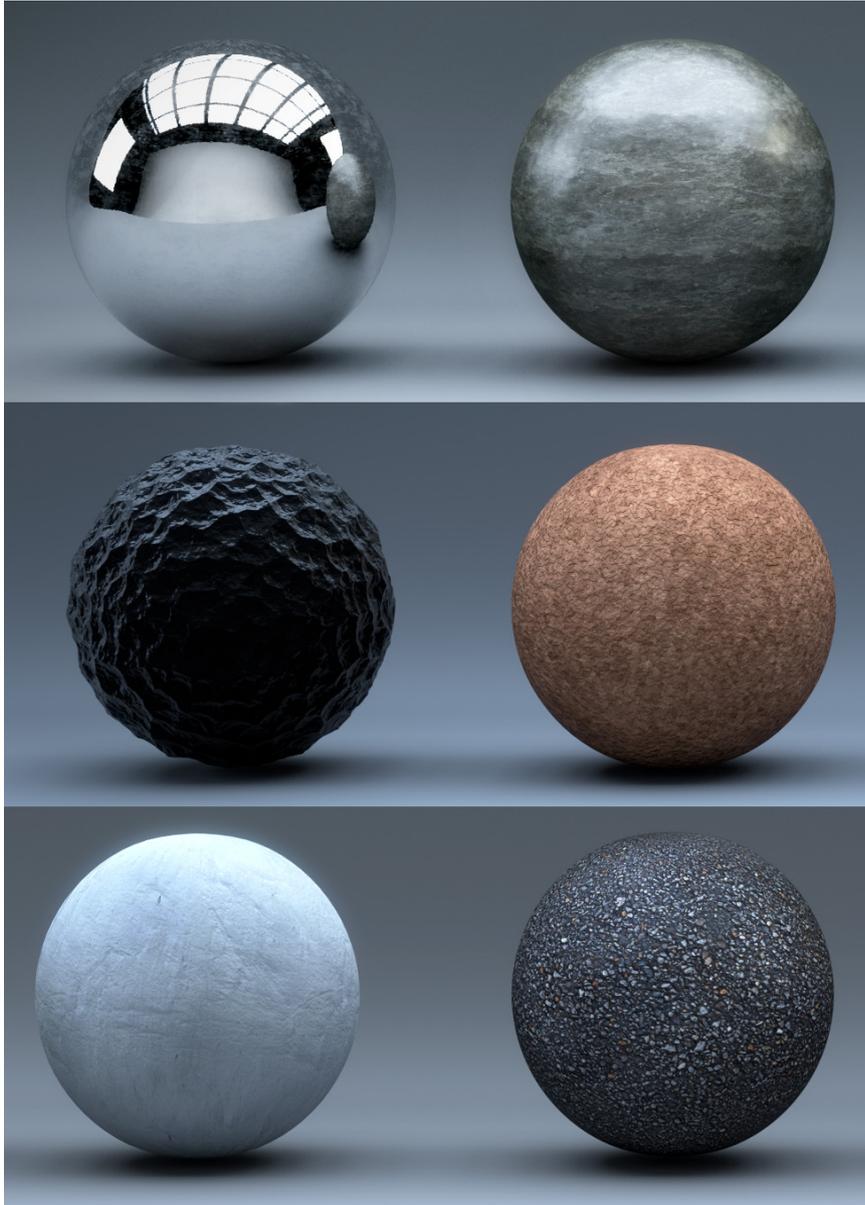
Figure 8. Examples of simulated materials and light fields for various different materials, rendered in Blender. Adapted from examples by Martinez (Martinez, 2015).

5.2.2 Isotrigon Discrimination in Clinical Diagnostics

If the texture discrimination performance functions described in Chapters 2-4 are representative of normal (healthy) subjects, deviations from the norm

may have clinical significance; such deviations could be measured in terms of maximally informative directions, or reduced learning for a texture subset (such as the ternary betas). There is a clinical rationale for this, as follows.

Several studies have attempted to identify which brain regions are involved in texture discrimination; there is some overlap between these regions and those affected by various neurological diseases. Beason-Held et al. used fMRI to measure responses to random textures and BoxEven binary isotrigon textures (Beason-Held et al., 1998a, Victor, 1985). The viewing of random textures increased activity in the striate cortex, with slight involvement of the cuneus and middle occipital, lingual and fusiform gyri. The viewing of BoxEven binary isotrigon textures resulted in activation of the same areas, but to a greater extent. In addition, activation was observed in the middle temporal region. Based on these findings, the authors suggest the presence of receptive field mechanisms in the ventral visual pathway that are sensitive to higher-order spatial correlations (Beason-Held et al., 1998a). Indeed, a subsequent fMRI study by the same authors identified a linear relationship between activation in the striate cortex and the density of higher-order spatial correlations in the textures presented (Beason-Held et al., 2000).

In another study by Beason-Held et al. (Beason-Held et al., 1998b) PET was used to evaluate regional cerebral blood flow (rCBF) in subjects viewing two random BoxEven binary isotrigon textures (as used in (Beason-Held et al., 1998a) and (Beason-Held et al., 2000)). The viewing of these textures resulted in increased rCBF along the occipito-temporal pathway, versus the viewing of random noise patterns. Significant activation was identified in

"…striate, extrastriate, lingual, and fusiform cortices as well as the hippocampus and brain stem" (Beason-Held et al., 1998b). Notably, increases in rCBF migrated from the occipito-temporal to the medial temporal areas (hippocampus) and frontal lobes following increased exposure to the isotrigon stimuli. This suggests the recruitment of higher-order brain areas during learning (Beason-Held et al., 1998b).

There is evidence that texture discrimination is negatively affected in multiple sclerosis (MS) patients, whilst other visual functions are spared (Regan and Simpson, 1995, Regan and Hong, 1994). Regan et al. found that MS produced a selective failure in the ability to identify texture-defined stimuli, whilst these patients retained the ability to identify comparable stimuli defined by luminance and motion (Regan and Simpson, 1995). The authors suggest that the decline in texture segregation ability may result from demyelination of long-range connections between orientation-tuned neurons in the striate cortex.

A pattern of higher-order vision impairment has also been observed in Alzheimer's disease (AD) patients (Rizzo et al., 2000). Rizzo et al. found that AD individuals performed significantly worse on tests of "…static spatial contrast sensitivity, visual attention, shape-from-motion, colour, visuospatial construction and visual memory" (Rizzo et al., 2000). Differences between AD and healthy subjects were not observed based on static visual acuity, stereoscopic acuity, dynamic visual acuity or motion direction discrimination. The results of the visual tests were found to be correlated with measures of cognitive functioning (Rizzo et al., 2000). The brain regions affected during AD vary considerably, but typically include higher-order association visual

areas such as the IT cortex and the hippocampus (Hof et al., 1997, von Gunten et al., 2006). These areas may be involved in learning higher-order texture discrimination (Beason-Held et al., 2000).

5.2.3 Illusions of 3D Shape

An interesting property of some ternary textures is the presence of cues which can confer the impression of 3D shapes. Maddess et al. previously reported this phenomenon for textures which had been produced by a variant of the deterministic cellular automata method (Maddess et al., 2004). Such illusions may be generated by false lighting cues, whereby the presence of grey is interpreted by the brain as shading, or false depth cues, whereby texture is interpreted as surface distortion (Maddess et al., 2004, Gibson, 1950). Some examples of these are shown in Figure 9.

In genuine 3D scenes, stereopsis and motion parallax provide important visual cues (Todd, 2004, Anzai and DeAngelis, 2010). However, in the absence of 3D information, humans are able to use other cues to form 3D percepts (Cavanagh and Leclerc, 1989, Todd, 2004, Georgieva et al., 2008). Numerous studies have investigated the perception of 3D shape from texture (Blake et al., 1993, Todd et al., 2004, Knill, 1998). Blake et al. describe the development of a statistical model for estimating shape from texture with an ideal observer (Blake et al., 1993). Knill et al. investigated the relative importance of perspective scaling, foreshortening and density for discriminating surface orientation (Knill, 1998). Using an ideal observer model, they found texture foreshortening to be the primary cue for perceiving

surface slant (Knill, 1998). In another study by Todd et al. (Todd et al., 2004) observers judged the surface depth profiles of randomly shaped textured objects. Observers consistently underestimated the depth of surface relief, but their evaluations of the overall shapes of surfaces were strongly correlated with the ground truth (Todd et al., 2004).

Perception of shape from image shading is affected, not only by surface geometry, but also reflectance and overall illumination (Koenderink). Perhaps unsurprisingly, perception of shape from shading is less tightly coupled to the ground truth than for other cues (Todd et al., 2004). Shading is particularly important for the analysis of curved surfaces according to Todd (Todd, 2004). Illusory shape from shading may result from an analogous process to that proposed by Purves et al. with regard lightness illusions (Purves et al., 2001, Purves and Lotto, 2011). When confronted with a shaded image, in the absence of other cues, the brain makes a probabilistic judgement based on prior experience and learned associations (Purves et al., 2001, Purves and Lotto, 2011). What neurological mechanisms are involved in the processing of 3D shape from shading and texture?
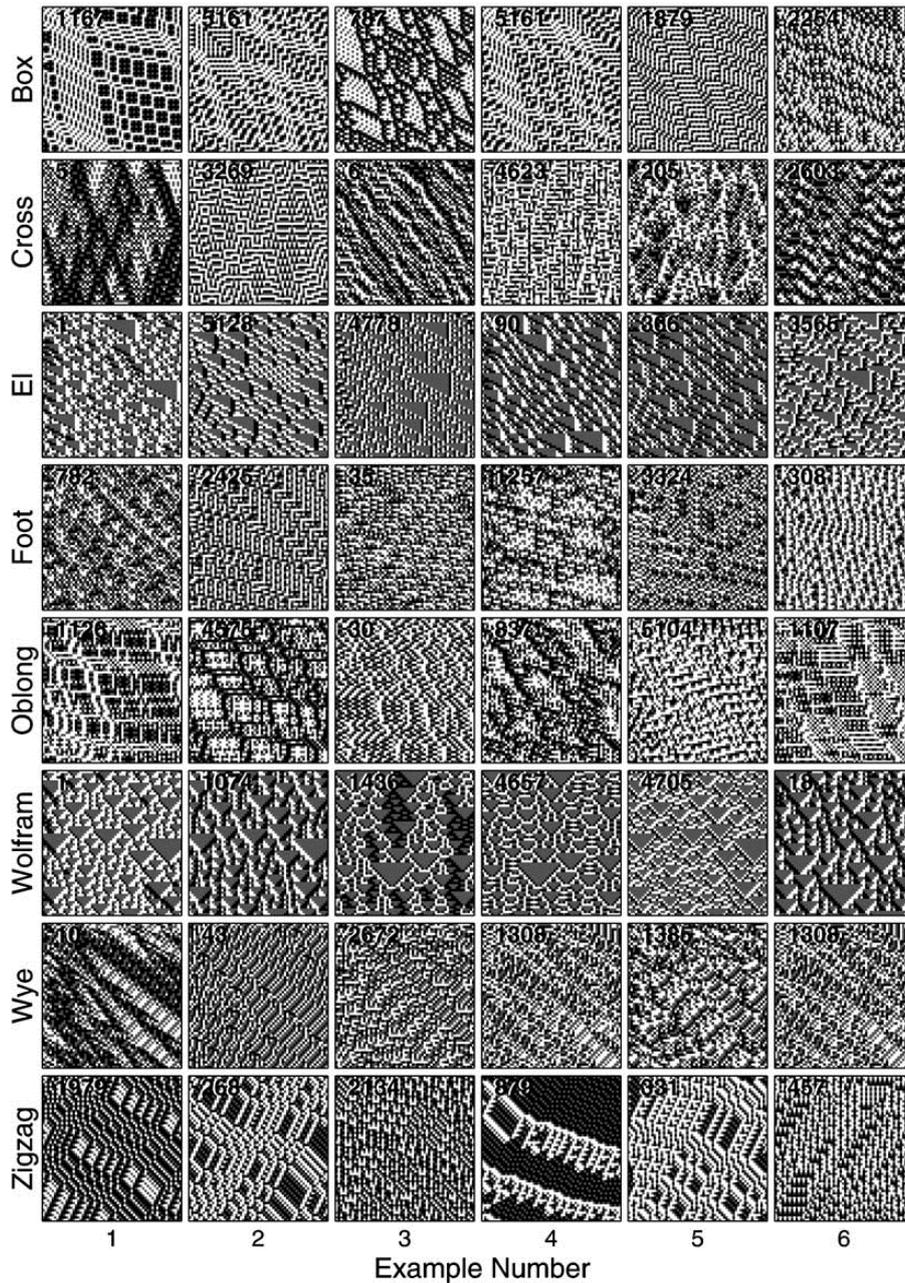
Figure 9. Examples of textures which show illusory depth effects of varying strengths. These textures have been created by a deterministic cellular automata method, but with more complex rules for combining the 3 input pixels, to form the 4th pixel in the recursion. The gliders and names are indicated on the y axis, and the example number on the x axis. Particularly strong depth effects have been

observed for the following examples: Box 1; El 2, 4, 5; Wolfram 1, 2, 6; Zigzag 1 (Maddess et al., 2004).

Early evidence came from lesional studies in macaques (Dean, 1976, Ungerleider and Mishkin). Bilateral lesions to the inferior temporal cortex impaired the macaques' ability to discriminate complex 2D patterns or shapes. Animals with lesions in the parietal cortex exhibited normal shape discrimination, but impaired ability to locate objects in space, thus suggesting distinct dorsal and ventral pathways (Todd, 2004). However, other fMRI studies in humans suggest that the perception of 3D shape involves *both* the dorsal and ventral pathways (Shikata et al., 2001, Murray et al., 2003, Taira et al., 2001). Judgments of 3D shape produces activation in ventral cortical regions, although interestingly these regions only partially overlap with those observed during viewing of 2D shapes (Kourtzi and Kanwisher, 2000).The analysis of 3D shape occurs at numerous locations within the dorsal pathway (Todd, 2004).

There has been relatively little research into the neurological mechanisms by which 3D shape is determined from texture and shading (Georgieva et al., 2008, Arcizet et al., 2009). One notable fMRI study by Georgieva et al. used shading or texture cues, whilst being careful to eliminate confounding cues such as edges and vertices (Georgieva et al., 2008). They found that the extraction of 3D shape from *texture* involves the bilateral caudal inferior temporal gyrus (caudal ITG), lateral occipital sulcus (LOS) and several sites along the intraparietal sulcus. It is notable that similar areas are involved in the processing of motion parallax and stereopsis cues (Todd, 2004).

328

However, the analysis of 3D shape from *shading* was restricted to the caudal ITG (Georgieva et al., 2008). Other fMRI studies on humans indicate a participation of both dorsal and ventral pathways (Taira et al., 2001).

Single-unit studies indicate that V4 neurons play an important role in shape from shading (Hanazawa and Komatsu, 2001, Pasupathy, 2006, Arcizet et al., 2009). A study by Pasupathy et al. found that contour features may be the basis of shape representation in the macaque V4 (Pasupathy, 2006). Single unit responses to V4 neurons were strongly tuned for such features; furthermore, responses to complex shapes were dictated by curvature at specific boundary locations (Pasupathy, 2006). Arcizet et al. found that V4 participated in the perception of complex shape from shading patterns in awake macaques (Arcizet et al., 2009). Interestingly, 78% of V4 neurons responded differently to 3D and 2D versions of the same stimuli (Arcizet et al., 2009). Caution should be used when extrapolating macaque studies to humans however, as there is some evidence of differences in 3D shape perception between the two models (Vanduffel et al., 2002).

An interesting question is whether the illusions of 3D shape described in complex textures contribute to discrimination performance and this remains to be investigated. The contribution of the grey token to 3D perception would also be interesting to explore; it may be that grey is essential to illusions of 3D shape as it provides the illusion of *shading*. Again, mTurk could be used to survey a large cohort of Workers for the emergence of perceived 3D shape from ternary textures. Once the most "3D-inducing" textures have been identified, it would also be interesting to explore the (presumably destructive) effects of decorrelation.

329

5.2.4 Probing Convolutional Neural Networks

The testing modality used in the HITs of Chapters 3 and 4 may have wider applications; for example, it might be used in the analysis of convolutional neural networks (CNN) (LeCun et al., 1998). CNNs are powerful universal function approximators which have found important applications in image classification (Krizhevsky et al., 2012), digit recognition (Ciresan et al., 2010), facial recognition (Lawrence et al., 1997), and video analysis (Simonyan and Zisserman, 2014). Their dependence on different image statistics is not understood, but will presumably vary depending on the specific classification task being undertaken. A particularly interesting classification task uses the ImageNet data set.

The ImageNet data set is a standard image library used to benchmark image classifiers (ImageNet, 2012). The data set consists of 150,000 natural images, hand labelled for the presence or absence of 1,000 object categories. Classification performance on the ImageNet data set has become something of an arms race, with the current classification record broken by a Microsoft Research CNN (4.94% error) (He et al., 2015), Google (4.80% error) (Ioffe and Szegedy, 2015), and most recently Baidu with 4.58% error rate (Wu et al., 2015), although they were subsequently found to have exceeded their weekly submissions. Human-level classification on this data set is estimated at 5.1% error rate (Russakovsky et al., 2014). Other highly effective ImageNet CNNs are available as open source software.

What image statistics these trained ImageNet CNNs utilise to make their classification decisions remains an open question. Zeiler et al. recently

investigated which gross image areas are important to the classifier. To do this, they plotted the probability of the class of interest as a function of the position of an occluding image mask (Zeiler and Fergus, 2014). The probability was then displayed as a colourmap and overlayed on the original image.

In another recent study using an ImageNet CNN, Long et al. explored how much spatial information is preserved by CNNs (Long et al., 2014). They found evidence that CNNs retain precise spatial information about object correspondences within images. Image features were found to localize at a much finer scale than the CNN receptive field sizes (Long et al., 2014). A study by Mahendran et al. is particularly pertinent (Mahendran and Vedaldi, 2014). They attempted to use the information in the deep layers of an ImageNet trained CNN by *inverting it*. Several layers in the CNN were found to retain photographically accurate image information, with different degrees of geometric invariance (Mahendran and Vedaldi, 2014).

We suggest that stochastic ternary textures could be used to directly explore which image statistics are important to trained ImageNet CNNs. After training an ImageNet CNN, the network would then be adapted to classify band positions within stochastic ternary textures. This process of adapting a pre-trained CNN is known as "transfer learning" and involves replacing the output layer of the network (Oquab et al., 2014). In this manner, the pre-trained CNN effectively operates as a *fixed feature extractor*. The modified CNN could then be subjected to the study protocol described in Chapters 3 and 4; in this case, the modified CNN would have near-human ImageNet classification performance, but be naïve to stochastic ternary textures.

If the trained ImageNet CNN is sensitive to similar directions within the texture space it would be a significant finding. If this is *not* the case, then the similarities and differences between the two would still be interesting to analyse, particularly considering the fact that their ImageNet classification performance is near equivalent. This approach might also indicate areas where performance can be improved. There is evidence that post-training inculcation of textural information may improve classification performance. For example, Dieleman recently reported using Haralick texture features (Haralick, 1979) to improve the performance of a CNN using a technique which he called "late fusion" (Dieleman, 2015).

5.3 Conclusion

The study of texture is a rich and challenging field. By studying textures and how they are perceived, we open up many leads into visual processing. Using artificially generated textures, with carefully controlled statistical properties, we can probe the sensitivities and limitations of the human visual system to higher-order correlations. Such studies may in turn inform us about the functioning of the cortex. Crowdsourcing platforms have great potential for implementing such experiments; the scale and efficiency savings they offer cannot be matched in traditional laboratory settings.

By identifying normative functioning in healthy individuals, tests of texture discrimination may be applied to clinical problems. Based on spatially localised patterns of neurodegeneration, and clinical symptoms, deficits in texture perception is a promising candidate for a diagnostic tool. We have

also considered how an extrapolation of 2D textures into the 3D space may inform theories of material perception. Lastly, the study of textural image statistics may have ramifications for image categorisation problems using CNNs. The testing modality outlined here may be applied to trained networks, which are functionally "black boxes", in order to elucidate their mode of action.

5.4 References

ANDERSON, B. L. 2011. Visual perception of materials and surfaces. *Current Biology,* 21**,** R978-R983.

ANZAI, A. & DEANGELIS, G. C. 2010. Neural computations underlying depth perception. *Curr Opin Neurobiol,* 20**,** 367-75.

ARCIZET, F., JOUFFRAIS, C. & GIRARD, P. 2009. Coding of shape from shading in area V4 of the macaque monkey. *BMC Neurosci,* 10**,** 140.

BARBOSA, M. S., BUBNA-LITIC, A. & MADDESS, T. 2013. Locally countable properties and the perceptual salience of textures. *Journal of the Optical Society of America A,* 30**,** 1687-1697.

BARLOW, H. 2001. Redundancy reduction revisited. *Network,* 12**,** 241-53.

BARLOW, H. B. 1961. Possible principles underlying the transformations of sensory messages.

BEASON-HELD, L. L., PURPURA, K. P., KRASUSKI, J. S., DESMOND, R. E., MANGOT, D. J., DALY, E. M., OPTICAN, L. M., RAPOPORT, S. I. & VANMETER, J. W. 2000. Striate cortex in humans demonstrates the relationship between activation and variations in visual form.

Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale, 130**,** 221-6.

BEASON-HELD, L. L., PURPURA, K. P., KRASUSKI, J. S., MAISOG, J. M., DALY, E. M., MANGOT, D. J., DESMOND, R. E., OPTICAN, L. M., SCHAPIRO, M. B. & VANMETER, J. W. 1998a. Cortical regions involved in visual texture perception: a fMRI study. *Brain research. Cognitive brain research,* 7**,** 111-8.

BEASON-HELD, L. L., PURPURA, K. P., VAN METER, J. W., AZARI, N. P., MANGOT, D. J., OPTICAN, L. M., MENTIS, M. J., ALEXANDER, G. E., GRADY, C. L., HORWITZ, B., RAPOPORT, S. I. & SCHAPIRO, M. B. 1998b. PET reveals occipitotemporal pathway activation during elementary form perception in humans. *Visual neuroscience,* 15**,** 503-10.

BLAKE, A., BULTHOFF, H. H. & SHEINBERG, D. 1993. Shape from texture: ideal observers and human psychophysics. *Vision Res,* 33**,** 1723-37.

BLAND, J. M. 2000. *An Introduction into Medical Statistics,* Oxford, Oxford University Press.

BLAND, J. M. & ALTMAN, D. G. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet,* 1**,** 307-10.

BLAND, J. M. & ALTMAN, D. G. 1999. Measuring agreement in method comparison studies. *Stat Methods Med Res,* 8**,** 135-60.

BUHRMESTER, M., KWANG, T. & GOSLING, S. D. 2011. Amazon's Mechanical Turk : A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science,* 6**,** 3-5.

BUZSAKI, G. & KANDEL, A. 1998. Somadendritic backpropagation of action potentials in cortical pyramidal cells of the awake rat. *Journal of neurophysiology,* 79**,** 1587-91.

CATTELL, R. B. 1966. The Scree Test For The Number Of Factors. *Multivariate Behavioral Research,* 1**,** 245-276.

CAVANAGH, P. & LECLERC, Y. G. 1989. Shape from shadows. *J Exp Psychol Hum Percept Perform,* 15**,** 3-27.

CHANDLER, M., DE LA CRUZ, F., DYDA, F., HICKMAN, A. B., MONCALIAN, G. & TON-HOANG, B. 2013. Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nat Rev Microbiol,* 11**,** 525-38.

CIRESAN, D. C., MEIER, U., GAMBARDELLA, L. M. & SCHMIDHUBER, J. 2010. Deep, big, simple neural nets for handwritten digit recognition. *Neural computation,* 22**,** 3207-3220.

COLE, F., SANIK, K., DECARLO, D., FINKELSTEIN, A., FUNKHOUSER, T., RUSINKIEWICZ, S. & SINGH, M. How Well do Line Drawings Depict Shape? ACM Transactions on Graphics, 2009.

COMSOL. 2014. *http://www.comsol.com/* [Online].

COY, D. G., BARBOSA, M., SEAMONS, J. W. G., & MADDESS, T. 2014. Learning Effects for Discrimination of Isotrigon Textures. (Honours Thesis; Journal Article In Preparation).

CRUMP, M. J., MCDONNELL, J. V. & GURECKIS, T. M. 2013. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS One,* 8**,** e57410.

DEAN, P. 1976. Effects of inferotemporal lesions on the behavior of monkeys. *Psychological bulletin,* 83**,** 41.

DEVELLIS, R. F. 2012. Factor Analysis. *Scale Development Theory and Applications.* 3rd ed. University of North Carolina, Chapel Hill, USA: SAGE Publications, Inc.

DIELEMAN, S. 2015. *Classifying Plankton with Deep Neural Networks* [Online]. Available: http://benanne.github.io/2015/03/17/plankton.html.

FLEMING, R. W. 2014. Visual perception of materials and their properties. *Vision research,* 94**,** 62-75.

FLEMING, R. W. & BÜLTHOFF, H. H. 2005. Low-level image cues in the perception of translucent materials. *ACM Transactions on Applied Perception (TAP),* 2**,** 346-382.

FLEMING, R. W., DROR, R. O. & ADELSON, E. H. 2003. Real-world illumination and the perception of surface reflectance properties. *Journal of Vision,* 3**,** 3.

FLEMING, R. W., JÄKEL, F. & MALONEY, L. T. 2011. Visual perception of thick transparent materials. *Psychological science,* 22**,** 812-820.

FLEMING, R. W., WIEBEL, C. & GEGENFURTNER, K. 2013. Perceptual qualities and material classes. *Journal of vision,* 13**,** 9.

FREEMAN, J., ZIEMBA, C. M., HEEGER, D. J., SIMONCELLI, E. P. & MOVSHON, J. A. 2013. A functional and perceptual signature of the second visual area in primates. *Nature neuroscience,* 16**,** 974-981.

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. *Cereb Cortex,* 18**,** 2416-38.

GIBSON, J. J. 1950. The perception of the visual world.

GIBSON, J. J. 2013. The ecological approach to visual perception.

GILBERT, E. N. 1980. Random Colorings of a Lattice of Squares in the Plane. *SIAM Journal on Algebraic and Discrete Methods,* 1, 152-159.

GORITZ, A. S., WOLFF, H. G. & GOLDSTEIN, D. G. 2008. Individual payments as a longer-term incentive in online panels. *Behavior Research Methods,* 40.

GOSLING, S. D., VAZIRE, S., SRIVASTAVA, S. & JOHN, O. P. 2004. Should we trust Web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist,* 59, 93–104.

HANAZAWA, A. & KOMATSU, H. 2001. Influence of the direction of elemental luminance gradients on the responses of V4 cells to textured surfaces. *J Neurosci,* 21, 4490-7.

HARALICK, R. M. 1979. Statistical and structural approaches to texture. *Proceedings of the IEEE,* 67, 786-804.

HAUSSER, M., SPRUSTON, N. & STUART, G. J. 2000. Diversity and dynamics of dendritic signaling. *Science,* 290, 739-44.

HAYTON, J. C., ALLEN, D. G. & SCARPELLO, V. 2004. Factor Retention Decisions in Exploratory Factor Analysis: a Tutorial on Parallel Analysis. *Organizational Research Methods,* 7.

HE, K., ZHANG, X., REN, S. & SUN, J. 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *arXiv preprint arXiv:1502.01852.*

HEER, J., BOSTOCK, M. & 2010. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. *ACM Human Factors in Computing Systems (CHI),* 203-212.

HO, Y.-X., LANDY, M. S. & MALONEY, L. T. 2008. Conjoint measurement of gloss and surface texture. *Psychological Science,* 19**,** 196-204.

HO, Y., LANDY, M. & MALONEY, L. 2006. How illuminant direction affects perceived visual roughness. *J Vision,* 6**,** 634-648.

HOF, P. R., VOGT, B. A., BOURAS, C. & MORRISON, J. H. 1997. Atypical form of Alzheimer's disease with prominent posterior cortical atrophy: a review of lesion distribution and circuit disconnection in cortical visual pathways. *Vision Res,* 37**,** 3609-25.

HORTON, J. J., RAND, D. G. & ZECKHAUSER, R. J. 2011. The online laboratory: conducting experiments in a real labor market. *Experimental Economics,* 14**,** 399-425.

IMAGENET. 2012. *Large Scale Visual Recognition Challenge 2012 (ILSVRC2012)* [Online]. Available: http://www.image-net.org/challenges/LSVRC/2012/ 2015].

IOFFE, S. & SZEGEDY, C. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.

IPEIROTIS, P. 2010a. Analyzing the Amazon Mechanical Turk Marketplace. *ACM XRDS (Crossroads),* 17.

IPEIROTIS, P. 2010b. Demographics of Mechanical Turk. *Working Paper CeDER-10-01.* NYU Center for Digital Economy Research.

JULESZ, B., GILBERT, E. N. & VICTOR, J. D. 1978. Visual discrimination of textures with identical third-order statistics. *Biological cybernetics,* 31**,** 137-40.

KINGDOM, F. A. 2011. Lightness, brightness and transparency: a quarter century of new ideas, captivating demonstrations and unrelenting controversy. *Vision Res,* 51**,** 652-73.

KITTUR, A., CHI, E. H. & SUH, B. Crowdsourcing User Studies with Mechanical Turk. Conference on Human Factors in Computing Systems, 2008. 453-456.

KNILL, D. C. 1998. Ideal observer perturbation analysis reveals human strategies for inferring surface orientation from texture. *Vision research,* 38**,** 2635-2656.

KOENDERINK, J. J. Shape and shading.

KOURTZI, Z. & KANWISHER, N. 2000. Cortical regions involved in perceiving object shape. *The Journal of Neuroscience,* 20**,** 3310-3318.

KRIZHEVSKY, A., SUTSKEVER, I. & HINTON, G. E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012. 1097-1105.

LANDY, M. S. & GRAHAM, N. 2004. Visual perception of texture. *The visual neurosciences,* 2**,** 1106-1118.

LAWRENCE, S., GILES, C. L., TSOI, A. C. & BACK, A. D. 1997. Face recognition: A convolutional neural-network approach. *Neural Networks, IEEE Transactions on,* 8**,** 98-113.

LECUN, Y., BOTTOU, L., BENGIO, Y. & HAFFNER, P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE,* 86**,** 2278-2324.

LONG, J. L., ZHANG, N. & DARRELL, T. Do Convnets Learn Correspondence? Advances in Neural Information Processing Systems, 2014. 1601-1609.

MADDESS, T. & KULIKOWSKI, J. J. 1999. Apparent fineness of stationary compound gratings. *Vision research,* 39**,** 3404-16.

MADDESS, T. & NAGAI, Y. 2001. Discriminating isotrigon textures [corrected]. *Vision research,* 41**,** 3837-60.

MADDESS, T., NAGAI, Y., JAMES, A. C. & ANKIEWCZ, A. 2004. Binary and ternary textures containing higher-order spatial correlations. *Vision research,* 44**,** 1093-113.

MADDESS, T., NAGAI, Y., VICTOR, J. D. & TAYLOR, R. R. 2007. Multilevel isotrigon textures. *Journal of the Optical Society of America. A, Optics, image science, and vision,* 24**,** 278-93.

MAHENDRAN, A. & VEDALDI, A. 2014. Understanding deep image representations by inverting them. *arXiv preprint arXiv:1412.0035*.

MARLOW, P. J., KIM, J. & ANDERSON, B. L. 2012. The perception and misperception of specular surface reflectance. *Current Biology,* 22**,** 1909-1913.

MARTINEZ, R. 2015. *Cycles Material Studies (Blender)* [Online]. Available: http://www.reynantemartinez.com/cycles-material-studies.html.

MASON, W. & SURI, S. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behav Res Methods,* 44**,** 1-23.

MASON, W. A. & WATTS, D. J. 2009. Financial incentives and the performance of crowds. *Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP '10)***,** 77–85.

MEL, B. W. 1993. Synaptic integration in an excitable dendritic tree. *Journal of neurophysiology,* 70**,** 1086-101.

MICHIELSEN, K. & DE RAEDT, H. 2001. Integral-geometry morphological image analysis. *Physics Reports,* 347**,** 461-538.

MOTOYOSHI, I., NISHIDA, S. Y., SHARAN, L. & ADELSON, E. H. 2007. Image statistics and the perception of surface qualities. *Nature,* 447**,** 206-209.

MURRAY, S. O., OLSHAUSEN, B. A. & WOODS, D. L. 2003. Processing shape, motion and three-dimensional shape-from-motion in the human cortex. *Cerebral Cortex,* 13**,** 508-516.

NAGAI, Y., TAYLOR, R. R., LOH, Y. W. & MADDESS, T. 2009. Discrimination of complex form by simple oscillator networks. *Network,* 20**,** 233-52.

NISHIDA, S. Y. & SHINYA, M. 1998. Use of image-based information in judgments of surface-reflectance properties. *JOSA A,* 15**,** 2951-2965.

NORRIS, M. & LECAVALIER, L. 2010. Evaluating the use of exploratory Factor analysis in developmental disability psychological research. *J Autism Dev Disord,* 40**,** 8-20.

OQUAB, M., BOTTOU, L., LAPTEV, I. & SIVIC, J. Learning and transferring mid-level image representations using convolutional neural networks. Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, 2014. IEEE, 1717-1724.

PADILLA, S., DRBOHLAV, O., GREEN, P. R., SPENCE, A. & CHANTLER, M. J. 2008. Perceived roughness of 1/< i> fβ</i> noise surfaces. *Vision research,* 48**,** 1791-1797.

PAOLACCI, G., CHANDLER, J. & IPEIROTIS, P. 2010. Running experiments on Amazon Mechanical Turk. *Judgment and decision making,* 5**,** 411-419.

PASUPATHY, A. 2006. Neural basis of shape representation in the primate brain. *Prog Brain Res,* 154**,** 293-313.

PIZLO, Z. 2001. Perception viewed as an inverse problem. *Vision Research,* 41**,** 3145-3161.

POGGIO, T., TORRE, V. & KOCH, C. 1985. Computational vision and regularization theory. *NATURE,* 317**,** 26.

PONT, S. C. & KOENDERINK, J. J. 2008. Shape, surface roughness, and human perception. World Scientific.

PONTIN, J. 2007. Artificial intelligence, with help from the humans. *New York Times*, 25 March.

PURVES, D., AUGUSTINE, G. J., FITZPATRICK, D., HALL, W. C., LAMANTIA, A.-S., MCNAMARA, J. O. & WHITE, L. E. 2001. Neuroscience. *Sunderland (MA): Sinauer Associates*.

PURVES, D. & LOTTO, R. B. 2011. *Why we see what we do redux: A wholly empirical theory of vision*, Sinauer Associates Sunderland, MA.

REGAN, D. & HONG, X. H. 1994. Recognition and detection of texture-defined letters. *Vision Res,* 34**,** 2403-7.

REGAN, D. & SIMPSON, T. 1995. Multiple sclerosis can cause visual processing deficits specific to texture-defined form. *Neurology,* 45**,** 809-15.

REYMENT, R. A., JORESKOG, K.G. 1996. Applied Factor Analysis in the Natural Sciences. Cambridge University Press.

RIZZO, M., ANDERSON, S. W., DAWSON, J. & NAWROT, M. 2000. Vision and cognition in Alzheimer's disease. *Neuropsychologia,* 38**,** 1157-69.

ROMEIRO, F. & ZICKLER, T. 2010. Inferring reflectance under real-world illumination. Technical Report No. TR-10-10). Cambridge, MA: Harvard School of Engineering and Applied Sciences.

ROSLI, Y., BEDFORD, S. M. & MADDESS, T. 2009. Low-spatial-frequency channels and the spatial frequency-doubling illusion. *Investigative ophthalmology & visual science,* 50**,** 1956-63.

ROSS, J., IRANI, I., SILBERMAN, M., ZALDIVAR, A. & TOMLINSON, B. 2010. Who are the CrowdWorkers? Shifting Demographics in Amazon Mechanical Turk. *CHI EA 2010***,** 2863-2872.

RUSSAKOVSKY, O., DENG, J., SU, H., KRAUSE, J., SATHEESH, S., MA, S., HUANG, Z., KARPATHY, A., KHOSLA, A. & BERNSTEIN, M. 2014. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision***,** 1-42.

SCHRODER-TURK, G. E., MICKEL, W., KAPFER, S. C., KLATT, M. A., SCHALLER, F. M., HOFFMANN, M. J., KLEPPMANN, N., ARMSTRONG, P., INAYAT, A., HUG, D., REICHELSDORFER, M., PEUKERT, W., SCHWIEGER, W. & MECKE, K. 2011. Minkowski tensor shape analysis of cellular, granular and porous structures. *Advanced materials,* 23**,** 2535-53.

SCHRODER-TURK, G. E., MICKEL, W., KAPFER, S. C. 2010. Minkowski Tensors of Anisotropic Spatial Structure. *arXiv,* 1009.

SEAMONS, J. W., BARBOSA, M. S., BUBNA-LITIC, A. & MADDESS, T. 2015a. A lower bound on the number of mechanisms for discriminating fourth and higher order spatial correlations. *Vision research,* 108**,** 41-48.

SEAMONS, J. W. G., BARBOSA, M. S., COY, D. & MADDESS, T. 2015b. Developing and Validating an Isotrigon Texture Discrimination Task using Crowdsourcing. Australian National University.

SEKULER, R., WILSON, H. R. & OWSLEY, C. 1984. Structural modeling of spatial vision. *Vision research,* 24**,** 689-700.

SHARAN, L., ROSENHOLTZ, R. & ADELSON, E. 2009. Material perception: What can you see in a brief glance? *Journal of Vision,* 9**,** 784-784.

SHARAN, L., ROSENHOLTZ, R. & ADELSON, E. H. 2008. Eye movements for shape and material perception. *Journal of Vision,* 8**,** 219-219.

SHIKATA, E., HAMZEI, F., GLAUCHE, V., KNAB, R., DETTMERS, C., WEILLER, C. & BÜCHEL, C. 2001. Surface orientation discrimination activates caudal and anterior intraparietal sulcus in humans: an event-related fMRI study. *Journal of Neurophysiology,* 85**,** 1309-1314.

SIMONYAN, K. & ZISSERMAN, A. Two-stream convolutional networks for action recognition in videos. Advances in Neural Information Processing Systems, 2014. 568-576.

SIMPSON, W. A. & MCFADDEN, S. M. 2005. Spatial frequency channels derived from individual differences. *Vision research,* 45**,** 2723-7.

STUART, G., SCHILLER, J. & SAKMANN, B. 1997. Action potential initiation and propagation in rat neocortical pyramidal neurons. *Journal of physiology,* 505 (Pt 3)**,** 617-32.

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**,** 959-66.

TAYLOR, R. R. 2013. *Neural Computation of Statistical Structure. Chapter 4: A Greater Purpose: Computation of Statistical Structure within Pyramidal Cells.* PhD Thesis, ANU.

TAYLOR, R. R., MADDESS, T. & NAGAI, Y. 2008. Spatial biases and computational constraints on the encoding of complex local image structure. *Journal of vision,* 8**,** 19 1-13.

THOMPSON, W., FLEMING, R., CREEM-REGEHR, S. & STEFANUCCI, J. K. 2011. *Visual perception from a computer graphics perspective*, CRC Press.

TKACIK, G., PRENTICE, J. S., VICTOR, J. D. & BALASUBRAMANIAN, V. 2010. Local statistics in natural scenes predict the saliency of synthetic textures. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 18149-54.

TODD, J. T. 2004. The visual perception of 3D shape. *Trends Cogn Sci,* 8**,** 115-21.

TODD, J. T., OOMES, A. H., KOENDERINK, J. J. & KAPPERS, A. M. 2004. The perception of doubly curved surfaces from anisotropic textures. *Psychol Sci,* 15**,** 40-6.

UNGERLEIDER, L. & MISHKIN, M. Two cortical visual systems, Ingle DJ, Goodale MA, Mansfield RJW, Analysis of visual behavior, 1982, 549-586. MIT Press, Cambridge, MA.

VANDUFFEL, W., FIZE, D., PEUSKENS, H., DENYS, K., SUNAERT, S., TODD, J. & ORBAN, G. 2002. Extracting 3D from motion: differences in human and monkey intraparietal cortex. *Science,* 298**,** 413-415.

VAZ, S., FALKMER, T., PASSMORE, A. E., PARSONS, R. & ANDREOU, P. 2013. The case for using the repeatability coefficient when calculating test-retest reliability. *PLoS One,* 8, e73990.

VICTOR, J. D. 1985. Complex visual textures as a tool for studying the VEP. *Vision research,* 25, 1811-27.

VICTOR, J. D. 1994. Images, Statistics, and Textures: Implications of Triple Correlation Uniqueness for Texture Statistics and the Julesz Conjecture. *Journal of the Optical Society of America A,* 11, 1680-1684.

VICTOR, J. D. & CONTE, M. M. 1989. Cortical interactions in texture processing: scale and dynamics. *Vis Neurosci,* 2, 297-313.

VICTOR, J. D. & CONTE, M. M. 1991. Spatial organization of nonlinear interactions in form perception. *Vision research,* 31, 1457-88.

VICTOR, J. D. & CONTE, M. M. 2012. Local image statistics: maximum-entropy constructions and perceptual salience. *Journal of the Optical Society of America. A, Optics, image science, and vision,* 29, 1313-45.

VICTOR, J. D., THENGONE, D. J. & CONTE, M. M. 2013. Perception of second- and third-order orientation signals and their interactions. *J Vis,* 13, 21.

VON GUNTEN, A., BOURAS, C., KOVARI, E., GIANNAKOPOULOS, P. & HOF, P. R. 2006. Neural substrates of cognitive and behavioral deficits in atypical Alzheimer's disease. *Brain Res Brain Res Rev,* 51, 176-211.

VON HELMHOLTZ, H. 1962. *Helmholtz's treatise on physiological optics,* Dover Publications.

WU, R., YAN, S., SHAN, Y., DANG, Q. & SUN, G. 2015. Deep image: Scaling up image recognition. *arXiv preprint arXiv:1501.02876*.

ZAIDI, Q. 2011. Visual inferences of material changes: color as clue and distraction. *Wiley Interdisciplinary Reviews: Cognitive Science,* 2, 686-700.

ZEILER, M. D. & FERGUS, R. 2014. Visualizing and understanding convolutional networks. *Computer Vision–ECCV 2014.* Springer.

ZWICK, W. R. & VELICER, W. F. 1986. Comparison of five rules for determining the number of components to retain. *Psychological Bulletin,* 99, 432-442.

Chapter 6: Appendix

6.1 Binary Isotrigon HIT Page Screen Capture

The following figure shows a screen capture of the binary isotrigon texture discrimination HIT used in Chapter 3. The HIT proper is one continuous webpage, but here it has been split to fit the page.

348

click here to hide/show the Instruction part.
Where is the oddball strip in an image?

Have a look at this slide show.
This HIT is a visual perception task that probes how the brain and particularly the visual system works. We have used this task in our labs among ourselves and with a few volunteers. Now we want to validate our findings with a large pool of subjects, so your help is very appreciated. Here are some samples of what you will see:

Left structured band on random background

Textures may be composed of a structured band over a random background, or the other way around (that is, a random band over a structured background). The band will appear in one of four places: top, bottom, left, or right. Your job is just to find which of these locations has the band.
The movie shows some examples. In the first few examples, the band is structured and the background is random; in the next examples, it is the other way around (a random band on a structured background). You can pause the slides by hovering the mouse over a slide, and navigate back and forth with the controllers. During the task, the position of the band will be randomized among these four locations (top, bottom, left, or right), and whether the band is structured or the background is structured will also be randomized. We expect that some trials will be as easy and some will be hard, just like in the movie. The movie indicates the location of the band with a red box; this is just to explain the task; it will not appear during the HIT.

First perform this quick calibration procedure.

Calibration procedure, click to hide/show this procedure;

Using a rule, pick the line bellow that is closest in length to an inch:

○ ————————————————264
       ○ ——————————135
              ○ ————————
120
○ ——————113
    ○ —————102
        ○ ————96
           ○ ————92
              ○ ———82

○ ———72

349
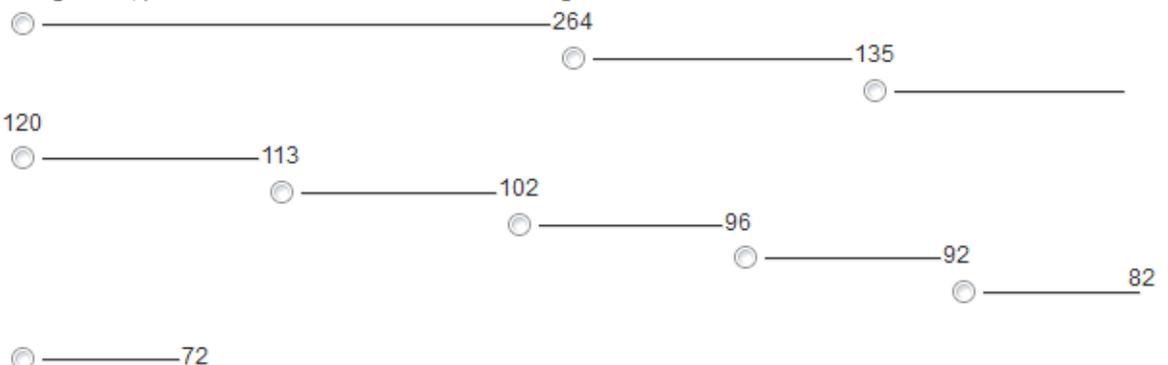
You must keep your eyes [____] inches from the computer screen.
Your IP seems to be:

124.171.108.22

Looks like your screen resolution is:

1366 x 768

And looks like your browser is :

Firefox 17 on Windows

This HIT should work on Chrome, Firefox and Safari. It may format badly in some versions of IE. Also, please make sure your browser zoom level is at 100%.

Lets start the task.
After you click the button (Ready?) at the center of the random image below, a texture will appear for about 2 seconds somewhere inside this image. After that the original random image will come back. The texture may be either a random or structured band, but, as a reminder, your job is just to find its location. Indicate your choice by clicking on the button corresponding to its location. If you are not sure, make your best guess. In some cases, it may be very hard to see where the band is, but we still prefer that you make your best guess. Please only click on the 'duh' button if you think your attention lapsed, as we are just as interested in your best guesses as in when you are pretty sure of where the band is. Submit only after you finish a batch of 20.

click to hide/show the Instruction part.

click to hide/show the Task part.

Click on the Ready? button at the center of the image bellow to start, then choose a location.

Top

ACCEPT the HIT to start

Left

Right

Bottom
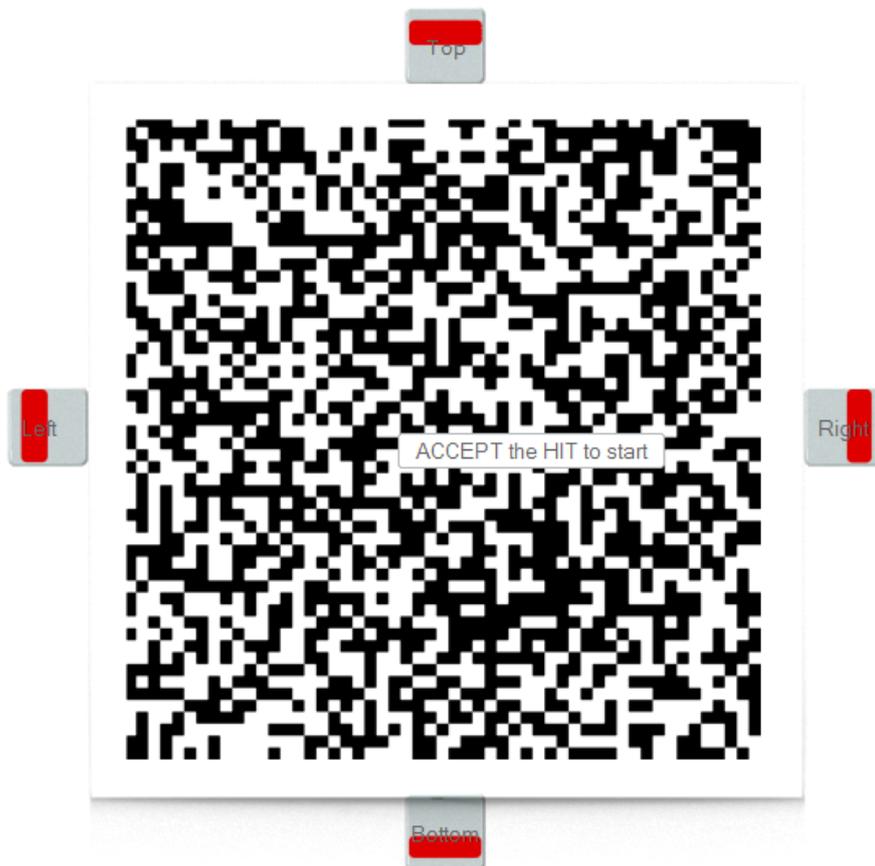
Missed

Don't abuse the **missed** button, the HIT may expire early...
You have [____] to go. Click here after you finish the batch [You must ACCEPT the HIT first.]

Want to work on this HIT?   Want to see other HITs?
Accept HIT   Skip HIT

350

Figure A.1: Screen capture of the binary isotrigon texture discrimination HIT that was used on Amazon Mechanical Turk (mTurk). The HIT page proper was one continuous webpage; the page has been split.

6.2 Lab Phase Platform Summary Table

Six different platforms (M1 to M6) were chosen on which to test the binary isotrigon texture discrimination HIT in the Lab experiments. Monitor luminance and other physical characteristics of the platforms were measured and recorded in the following table (Fig.A.2).

Machine	Platform	Rep0	Rep1	Rep2	Specification	Operating System	Browser	DPI	Diagonal (inch)	Resolution (pixels)	Distance (inch)	Power Settings	Monitor (cd/m^2)	Ambient (cd/m^2)
M1	Dell Latitude laptop (E6220)	21/08	29/08	2/09	Intel Core i7-2640M CPU @ 2.80 Ghz (8GB RAM)	Windows 7 (64 Bit)	Firefox ESR 17.0.7	120	12.5	1366x768	13	Mains power, "High Performance" setting, Max brightness	140	3.7
M2	Asus EeePC Netbook 1015PX	21/08	29/08	3/09	Intel Atom CPU N570 @ 1.66 Ghz (2GB RAM)	Windows 7 (32 Bit)	Google Chrome 29.0.1547.57m	113	10	1024x768	14	Mains power, "High Performance" setting, Max brightness	132	3.8
M3	Dell Latitude desktop D830 (E4600)	21/08	2/09	7/09	Intel Core 2 Duo @ 2.4 GHz (3.25GB RAM)	Windows XP Pro Version 2002 (Service Pack 3)	Firefox 12.0	82	27	1920x1200	19	Mains power, default brightness	468	153
M4	Ipad Mini 1 (WiFi 32GB)	22/08	29/08	3/09	Dual-core A5	IOS 6.0.1 (10A523)	Safari 6.0 Mobile	135	7.9	768x1024	12	Battery, Max brightness	292	4.3
M5	MacBook Pro 5,1	21/08	29/08	3/09	Intel Core 2 Duo @ 2.4 GHz (4GB RAM)	Mac OS X Snow Leopard 10.6.8	Firefox 15.0.1	113	15	1440x900	14	Mains power, "High Performance" setting, Max brightness	276	3.9
M6	Outside using Dell Latitude laptop (E6220)	22/08	29/08	3/09	Intel Core i7-2640M CPU @ 2.80 Ghz (8GB RAM)	Windows 7 (64 Bit)	Firefox ESR 17.0.7	120	12.5	1366x768	13	Battery, "High Performance" setting, Max brightness	140	1370 to 3000 (variable)

Figure A.2: Summary table for the six different platforms chosen for Lab testing using the binary isotrigon texture discrimination HIT. Rep0-2 indicates the repetition number and corresponding dates (day, month format, for the year 2013). The pixel resolution is given in

351

horizontal by vertical format. Thus, "768x1024" indicates a display in portrait format.

6.3 Bland-Altman Plots

Bland-Altman plots were produced for the Live and Lab data sets (Fig. A.3) (Bland, 2000, Bland and Altman, 1999, Bland and Altman, 1986, Vaz et al., 2013). The coefficient of repeatability for the DC versus the Live1 data was 14.3% and M1-M6 versus Live1 was 15.6%. These scores are quite low and they suggest that the live mTurk study has accurately reproduced the texture discrimination results that were produced by supervised laboratory trials. The coefficient of repeatability for the DC versus the Live2 data was 16.2% and test M1-M6 versus Live2 was 16.0%. These scores are slightly higher, but once again they are low, again indicating that the live mTurk study has accurately reproduced the supervised laboratory trials.

The CR between Live1 and Live2 was only 9.4% and this very low score indicates that the live study itself has a good repeatability. For completeness, the coefficient of repeatability between DC versus M1-M6 data was 17.2%.
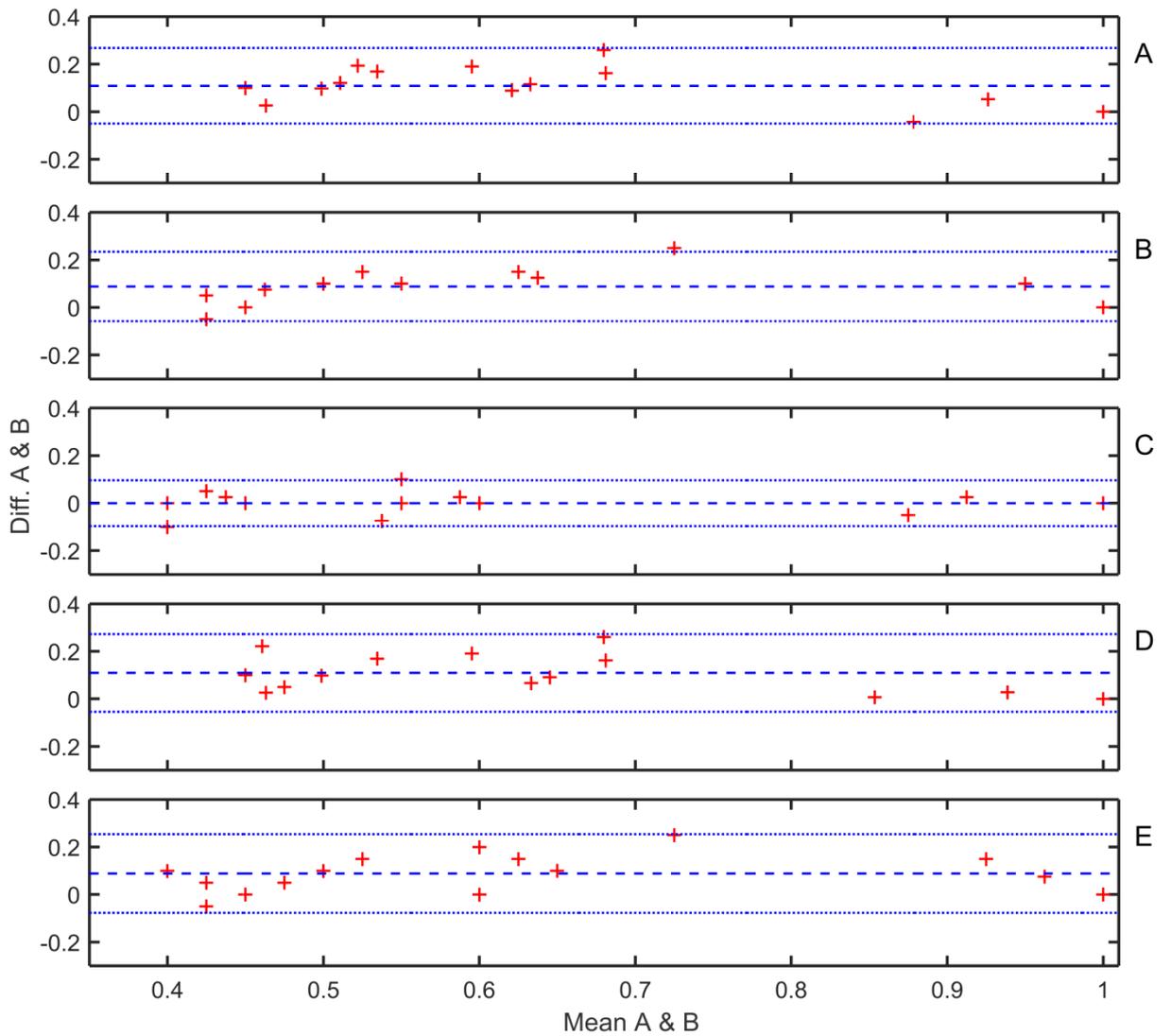
Figure A.3: M1-M6 vs Live1 is shown in A. DC vs Live1 is shown in B. Live1 vs Live2 is shown in C. M1-M6 vs Live2 is shown in D. DC vs Live2 is shown in E.

6.4 Factor Scores for the Machine Data Set (M1-M6)
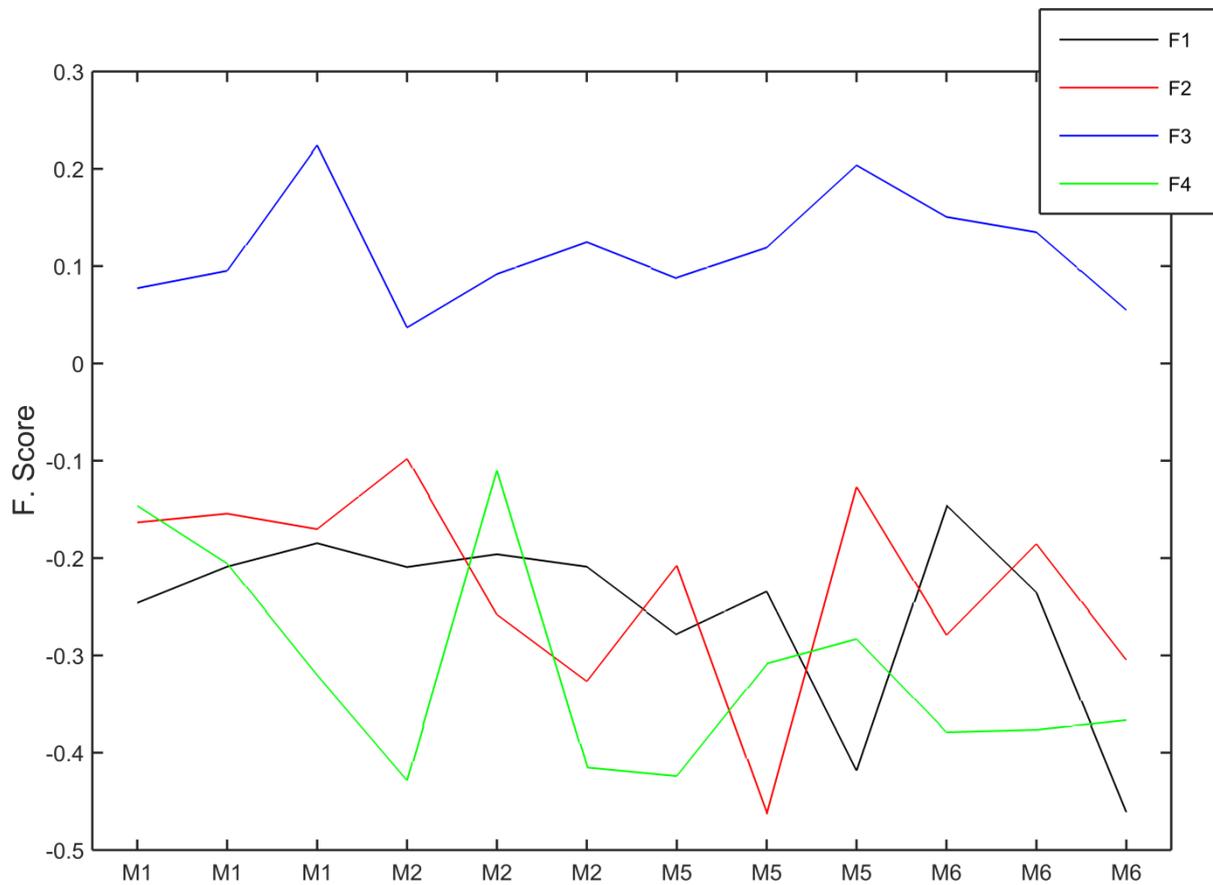
Figure A.4: Plot of factor scores against machine type for the M1-M6 data set. The effect of the top four factors is reasonably consistent across the 6 machines and 3 repeats on each evaluated. This suggests that variations between machines (such as monitor size, resolution, browser, etc) *do not significantly* alter the effect of the underlying factors, as we would expect. The relative strength of the four factors can also be inferred from this graph.

6.5 Effect of Physical Screen Size on Performance
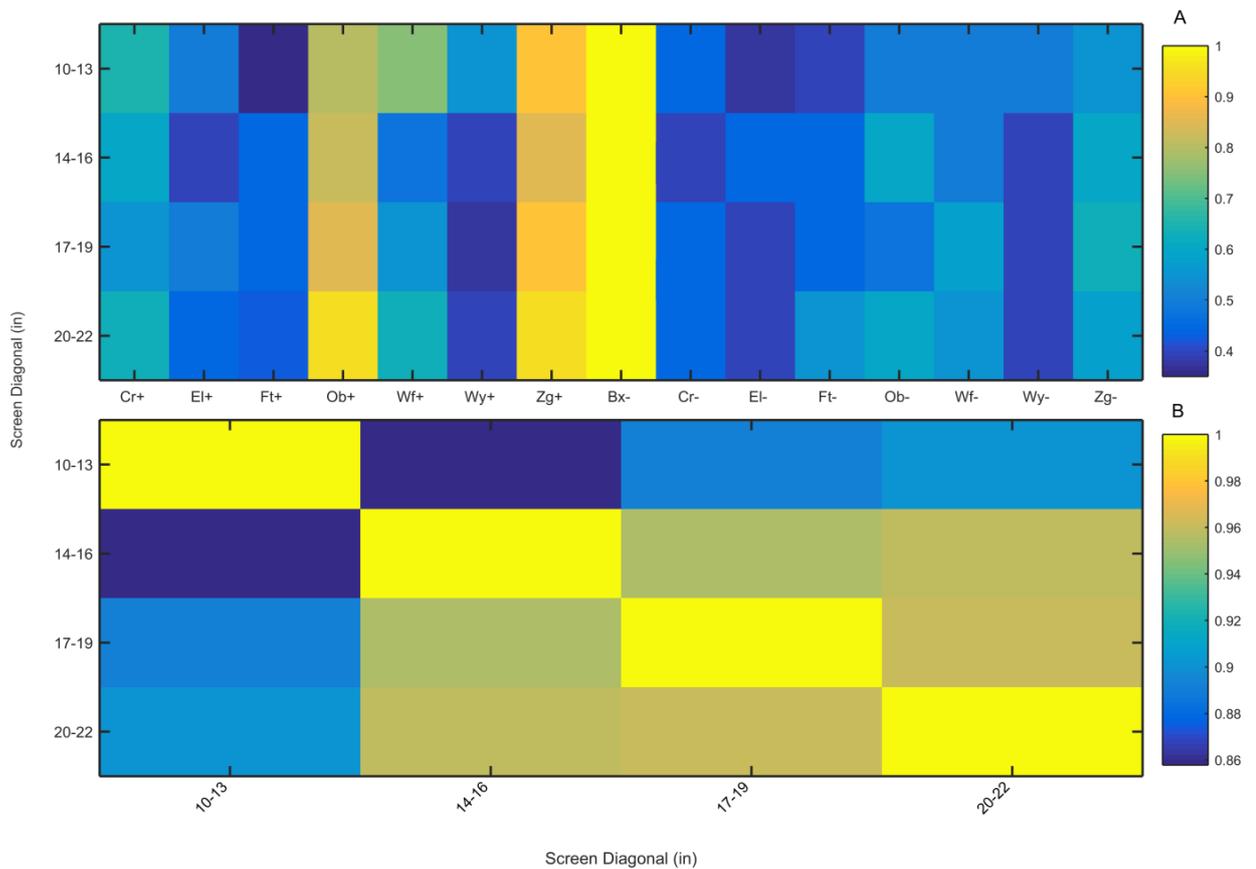
Figure A.5: Effect of physical screen size on performance. The Live performance data was stratified into four groups based on diagonal screen size: 10-13 in, 14-16 in, 17-19 in, and 20-22 in. This was calculated from DPI and screen resolution data, which was captured from the Live HITs. A: Colourmap showing median texture discrimination performance (probability correct) by physical screen size. B: Colourmap of correlations between median texture discrimination performance for each of four screen size groups.

The physical screen sizes (diagonal resolution in pixels) were calculated using the Pythagorean Theorem as follows:

$$d_p = \sqrt{w_p{}^2 + h_p{}^2}$$

$$PPI = d_p / d_i$$

…where: d_p is diagonal resolution in pixels; w_p is width resolution in pixels; h_p is height resolution in pixels; d_i is diagonal screen size in inches.

6.6 Comparison of Deterministic and Stochastic Ternary Textures using Histograms of Primitives

In the deterministic process for producing ternary textures, outlined by Maddess et al. (Maddess et al., 2007), each texture ensemble is generated by a cellular automata method combining a 3x3 pixel glider defining three input pixels (grey) and an output pixel (white) and 24 recursively applied arithmetic rules. This produces 3V3L ternary textures, with three input variables ("3V") and three levels ("3L"). In combination with 20 commonly used glider patterns, this process generates 216 texture classes (Maddess et al., 2007).

The recursive process itself is similar to that discussed for binary textures (Maddess and Nagai, 2001). The gliders and rules together specify how the values of each input pixel are combined to give the value for the output pixel. Each nascent texture is initialized with ternary noise. The glider is then moved across the matrix recursively, interacting with matrix elements according to the stated rules. Therefore, newly instantiated pixel values may

356

in turn feed back into the process (be selected as input pixels) as the glider moves across the nascent texture (Maddess et al., 2004, Maddess et al., 2007, Taylor et al., 2008). As with the binary isotrigon textures, the process is *deterministic* because the set of initial conditions (random boundary pixels) defines the textures obtained (Maddess et al., 2004, Maddess et al., 2007) (Figure A.6.1).
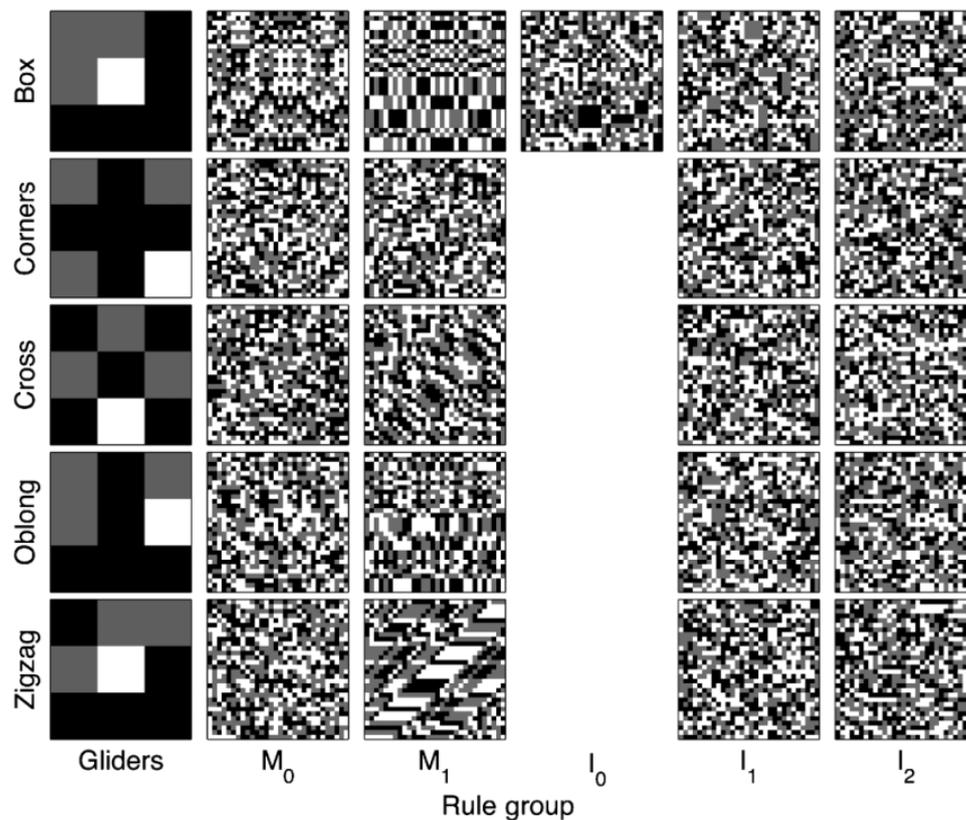

Figure A.6.1: Deterministic ternary texture examples from 21 ensembles (Taylor et al., 2008). Textures within each ensemble are generated via a set of arithmetic rules. The gliders are shown in the leftmost column. The grey squares indicate input pixels and the white square, the output pixel. The ternary texture is formed as the glider moves recursively across the surface of a ternary noise pattern. When the glider interacts with a randomly assigned pixel in the ternary noise

pattern, the value of that pixel may be changed. The result of the interaction depends on the value of the target pixel, the structure of the glider, and one of a set of rules. These rules are labelled on the X axis (M0, M1, I0, I1, and I2) and they insure the higher order properties of the ternary textures (Maddess et al., 2007).

We have discussed the stochastic and deterministic methods of ternary texture production (Chapter 4, Section 4.3.3 and Chapter 1, Section 1.5.2), but what is the relationship between these texture ensembles? Examination of the 12 types of stochastic ternary textures produced by constraining triplets of pixels (thetas) revealed that they were superficially similar to some V3L3 deterministic textures previously reported by Maddess et al. (Maddess et al., 2007).

To check that they were identical, their *histograms of primitives* (HOPs) were examined (Maddess, 2015) (Figure A.6.2). HOPs count the number of small, unique texture samples (primitives, typically 3x3 pixel samples) within a texture example and thereby uniquely identifies a texture type (Maddess et al., 2007). The stochastic ternary textures produced by our collaborator Prof. Jonathan Victor of the Weill Cornel Medical School (Victor and Conte, 2012) were found to be equal to the deterministic textures reported by Maddess in the 2007 study, which were made using the deterministic method, taking rotational and reflectional invariance into consideration (Maddess, 2015).

358
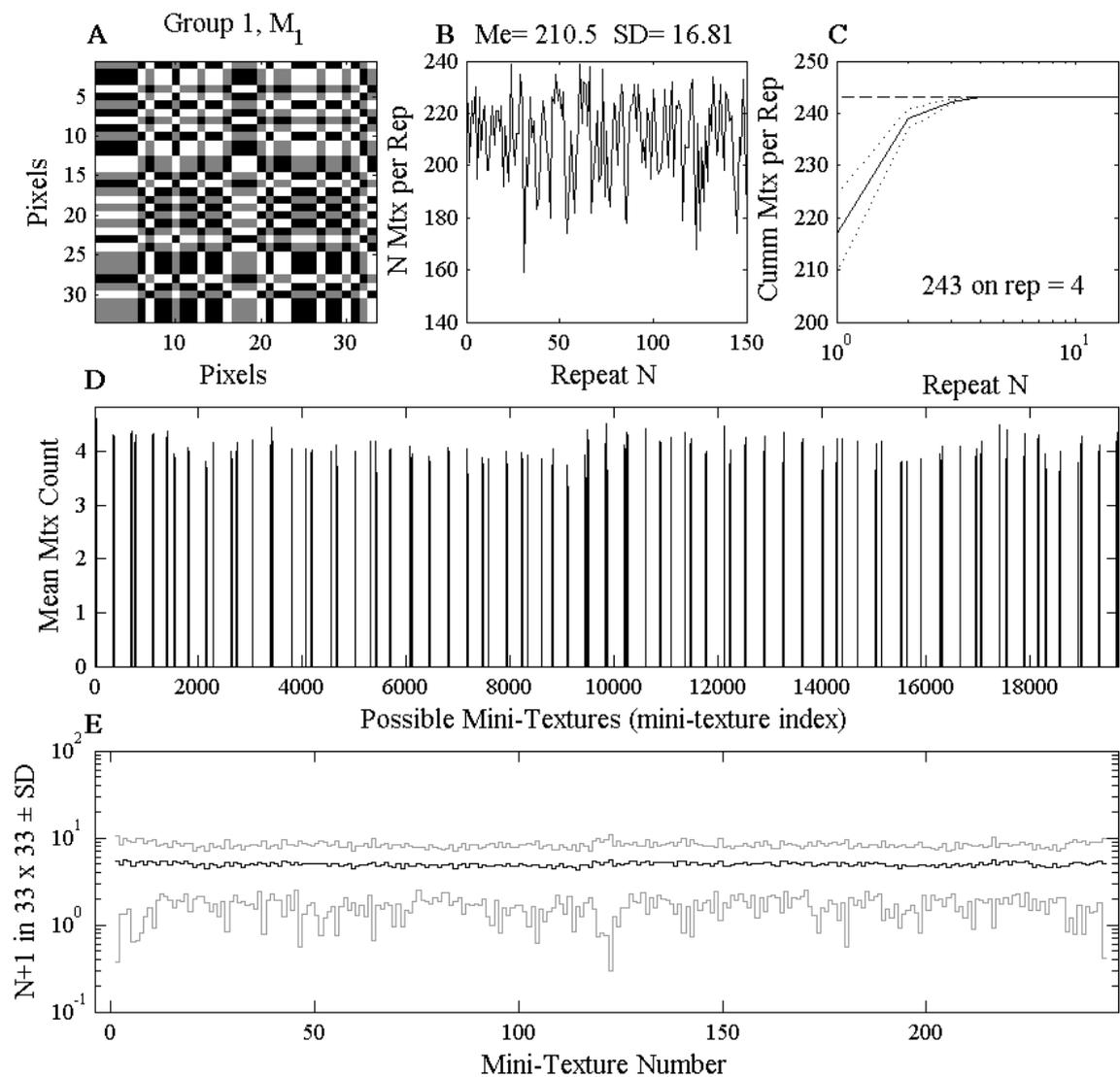
Figure A.6.2: Example of a Histogram of Primitives (HOPs). In this figure, an example of the ternary texture type being analysed is shown in A. The histogram of primitives for the texture is shown in D. The primitives are the 3x3 pixel "mini-textures" within the texture example shown in A. Each peak in the histogram identifies a single mini-texture type. Every ternary texture is defined by a specific pattern of mini-textures. Thereby, deterministic and stochastic V3L3 textures can be directly compared by comparing their HOPs. Some textures may appear to be visually similar, whilst their mini-texture content is distinct

and unique. Discounting rotations and reflections, the stochastic and deterministic textures were found to be equivalent (Maddess, 2015). Figure adapted from (Maddess et al., 2007).

6.7 Ternary HIT Page Screen Capture

The following figure shows a screen capture of the ternary texture discrimination HIT used in Chapter 4. The HIT proper is one continuous webpage, but has been split here to fit the page.

click here to hide/show the Instructions.

Where is the oddball strip in the image?

Please click on the following link to download the subject consent form. Read this document in full. If you do not agree to the terms outlined in the subject consent form, please do not accept this HIT. Download the PDF here

In the slide show below you will see some examples of the textures. The first four are very clear and easy, the last four are the actual textures assigned to this HIT.
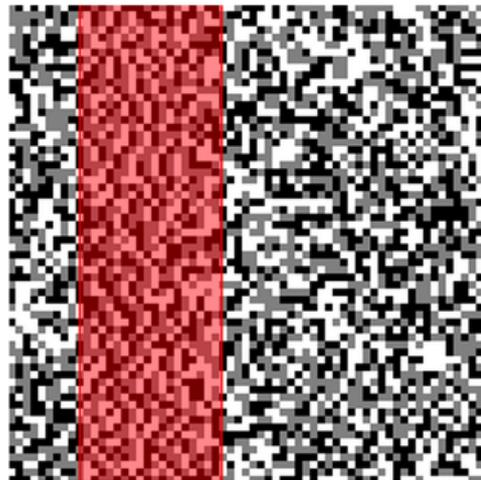
In each task you will first be presented with a square texture pattern in which the colorings of the pixels are random. Then, a strip of a different test texture will be presented in a band along one side of the square pattern as illustrated above. Your task is to detect which side the test texture is presented
The slide show above displays some examples. You can pause the slides by hovering the mouse over a slide, and navigate back and forth with the controllers. During the task, the position of the band will be randomized among these four locations (top, bottom, left, or right). We expect that some trials will be very easy and some will be hard, just like in the movie. The movie indicates the location of the band with a red box; this is just to explain the task; it will not appear during the HIT.
The examples in the image below show the specific types of textures you will be ask to spot. The textures come in 5 families called Gamma, Beta_hv, Beta_diag, Theta, and Alpha. Each family varies across the trico axis. In this HIT you will be asked to find Beta_diag_23 textures of the type highlighted, so please study them here to give you a feel for the task.

click to hide/show calibration procedure

Subject information and calibration procedure

Select your age range
○18-22 ○23-27 ○28-32 ○33-37 ○38-42 ○43-47 ○48 -52 ○53-58

Gender
○ male ○ female ○ other ○ I don't wish to answer this

Pixel density
Make sure your browser is at 100% magnification (press CTRL and 0 keys)
Using a rule, pick the line bellow that is closest in length to two inches (5.1cm) :

You must keep your eyes ☐ inches or ☐ centimeters from the computer screen.

Finally, look a the picture below while squinting your eyes.
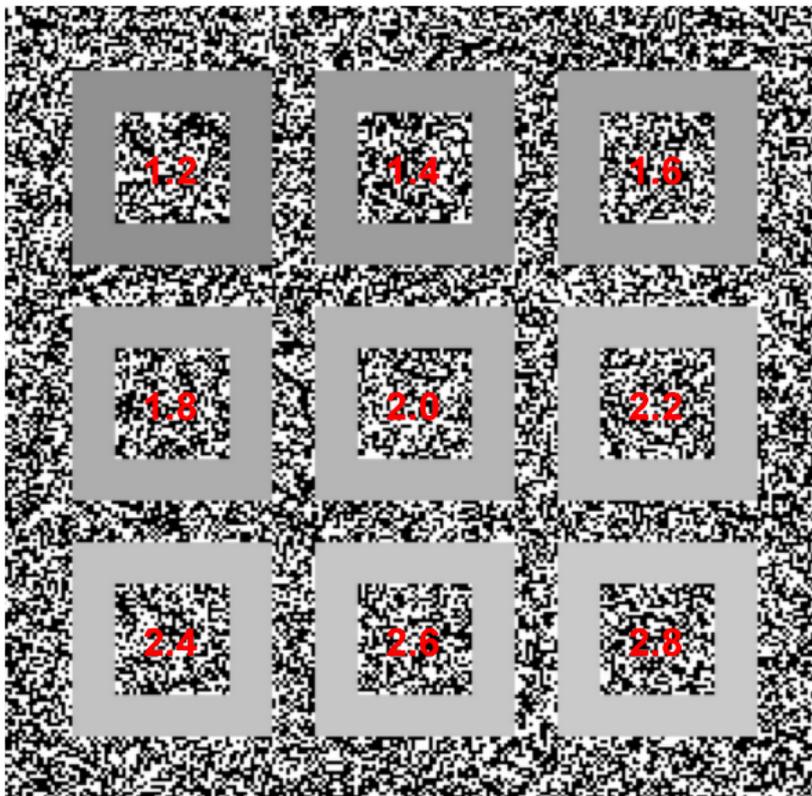
Now please pick the grey rectangle in the above image which you think is closest to the average brightness of the black and white surround, i.e. the mid-grey. It may help to defocus your eyes by squinting and looking through your eye lashes so as to blur the image while making this judgment.

○ 1.2 ○ 1.4 ○ 1.6
○ 1.8 ○ 2.0 ○ 2.2
○ 2.4 ○ 2.6 ○2.8

362

You can see periodic updates, outcomes and related background information on the project feedback page .

Lets start the task

After you click the button (Ready?) below, a texture band will appear for about 2 seconds somewhere inside this image. After that the original random image will come back. The texture may be either a random or structured band, but, as a reminder, your job is just to find its location. Indicate your choice by clicking on the button corresponding to its location. If you are not sure, make your best guess. In some cases, it may be very hard to see where the band is, but we still prefer that you make your best guess.

Click on the Ready? button at the center of the image bellow to start, then choose one of the location buttons.

Only use this [missed it!] button in case you have been distracted by something else. The HIT may be invalidated if there are too many missed responses. You have [____] to go. Click here after you finish a batch of 20 [Submit]

Figure A.7: Screen capture of the ternary texture discrimination HIT that was used on Amazon Mechanical Turk (mTurk) in Chapter 4. The HIT page proper was one continuous webpage; the HIT page shown here has been split.

6.8 Live1 Performance Data

All Live1 performance data was analysed *after* filtering the data by catch trial passes. i.e.: data from Workers that failed >= 4 catch trials was excluded. In the first instance, boxplots were produced for each ternary texture type (using the long condition only).

The three plots refer to the combined (top), long (middle) and short (bottom) presentation times. The dashed red lines divide the plots into the 6 rays with the following colour-biases: B (Black); BG (Black-grey); G (Grey); GW (Grey-white); W (White); WB (White-black). The step and ray values are indicated on the X axis, using the Victor nomenclature discussed in Chapter 4, Section 4.3.3.

Figure A.8.1: Performance data for **gamma_A1** ternary textures in **Live1**.
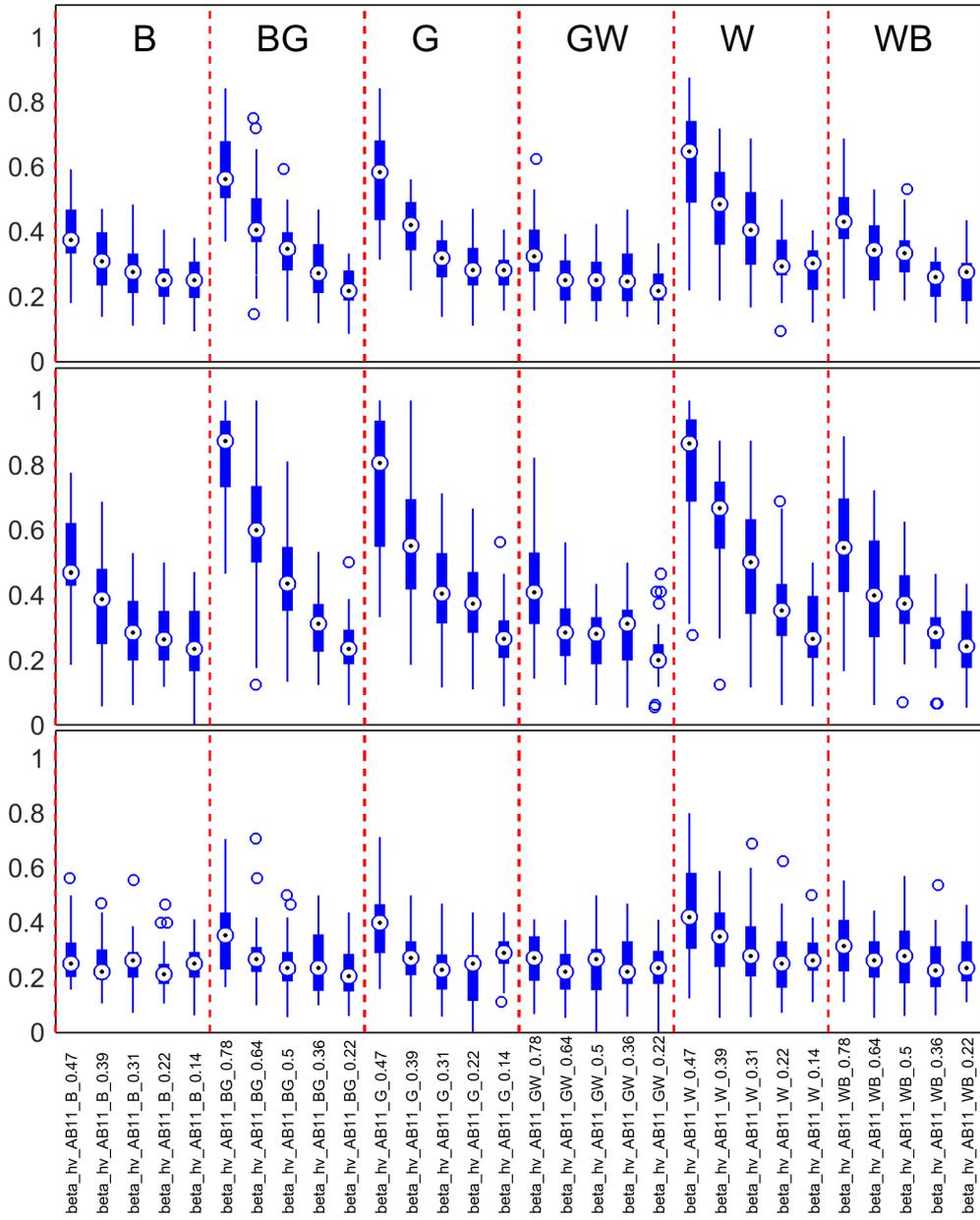
Figure A.8.2: Performance data for **beta_hv_AB11** ternary textures in **Live1**.
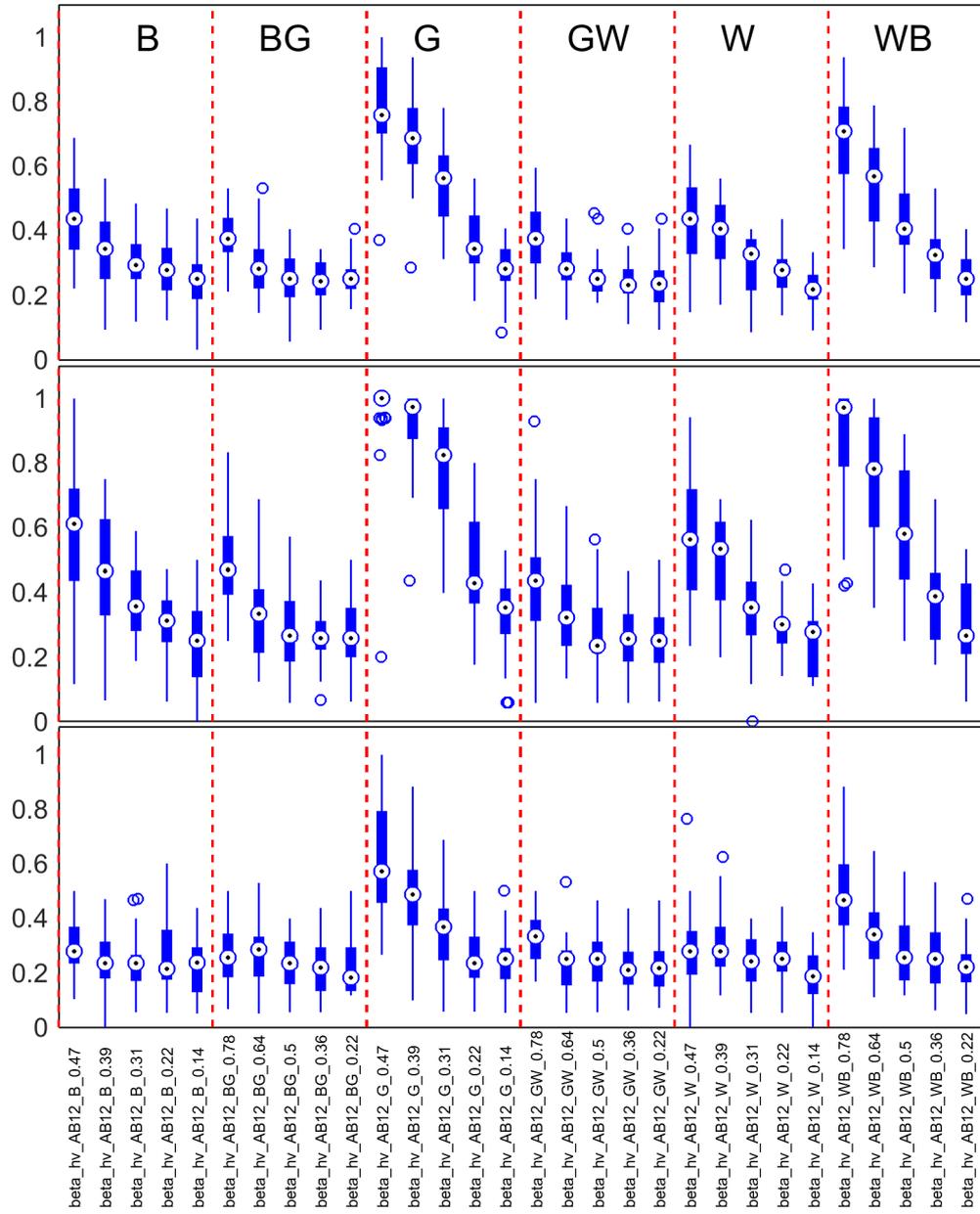
Figure A.8.3: Performance data for **beta_hv_AB12** ternary textures in **Live1**.
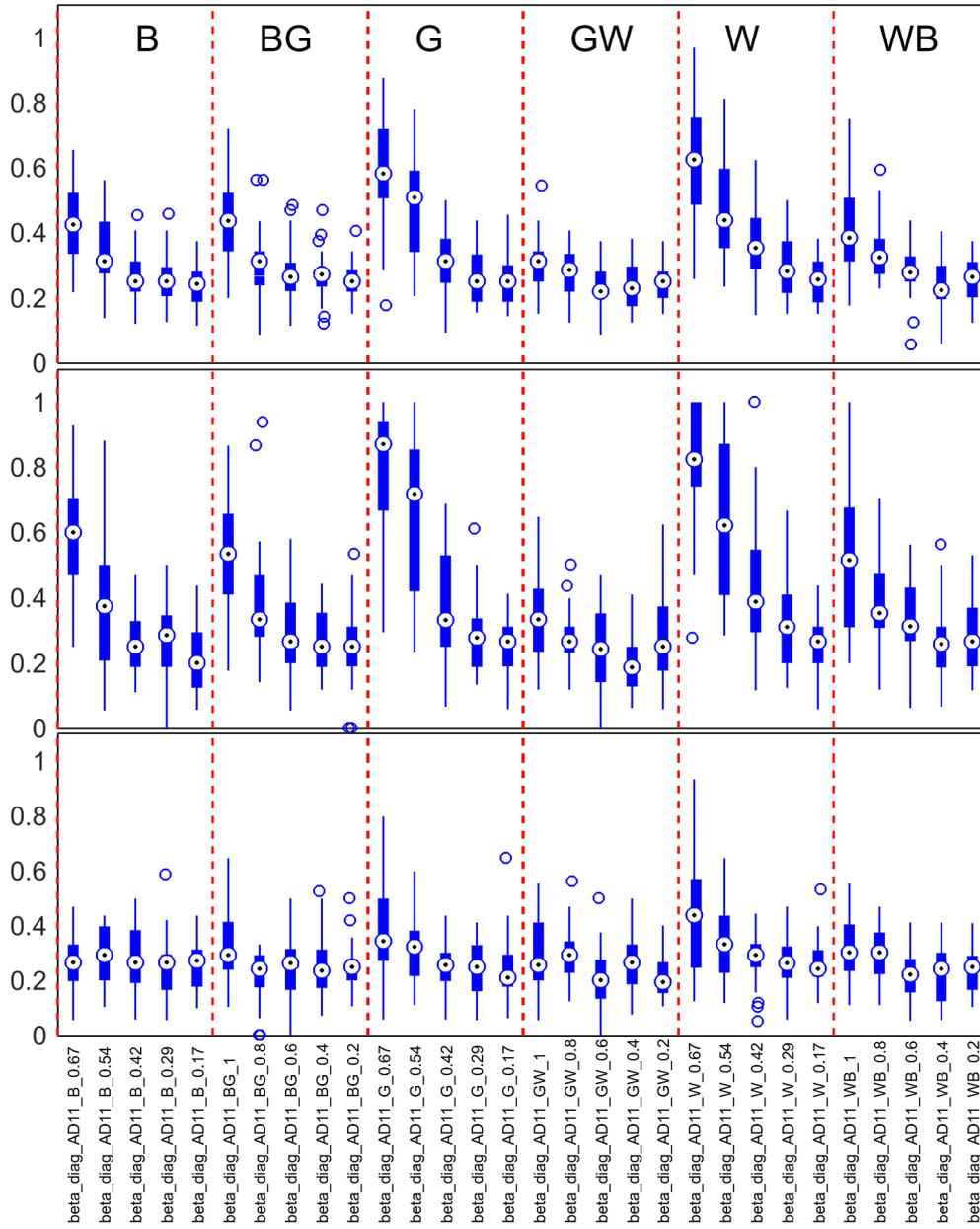
**Figure A.8.4**: Performance data for **beta_diag_AD11** ternary textures in **Live1**.
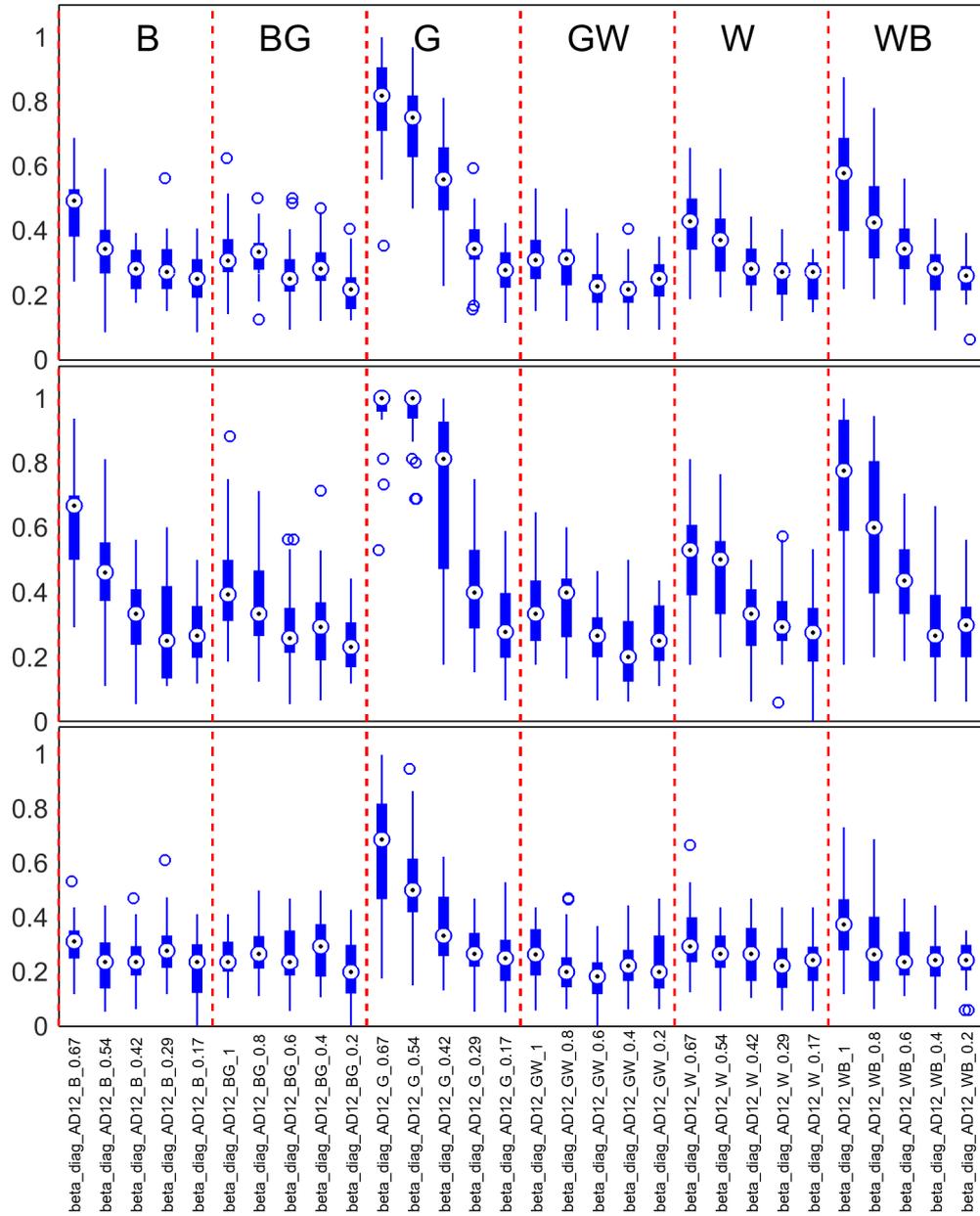
**Figure A.8.5**: Performance data for **beta_diag_AD12** ternary textures in **Live1**.
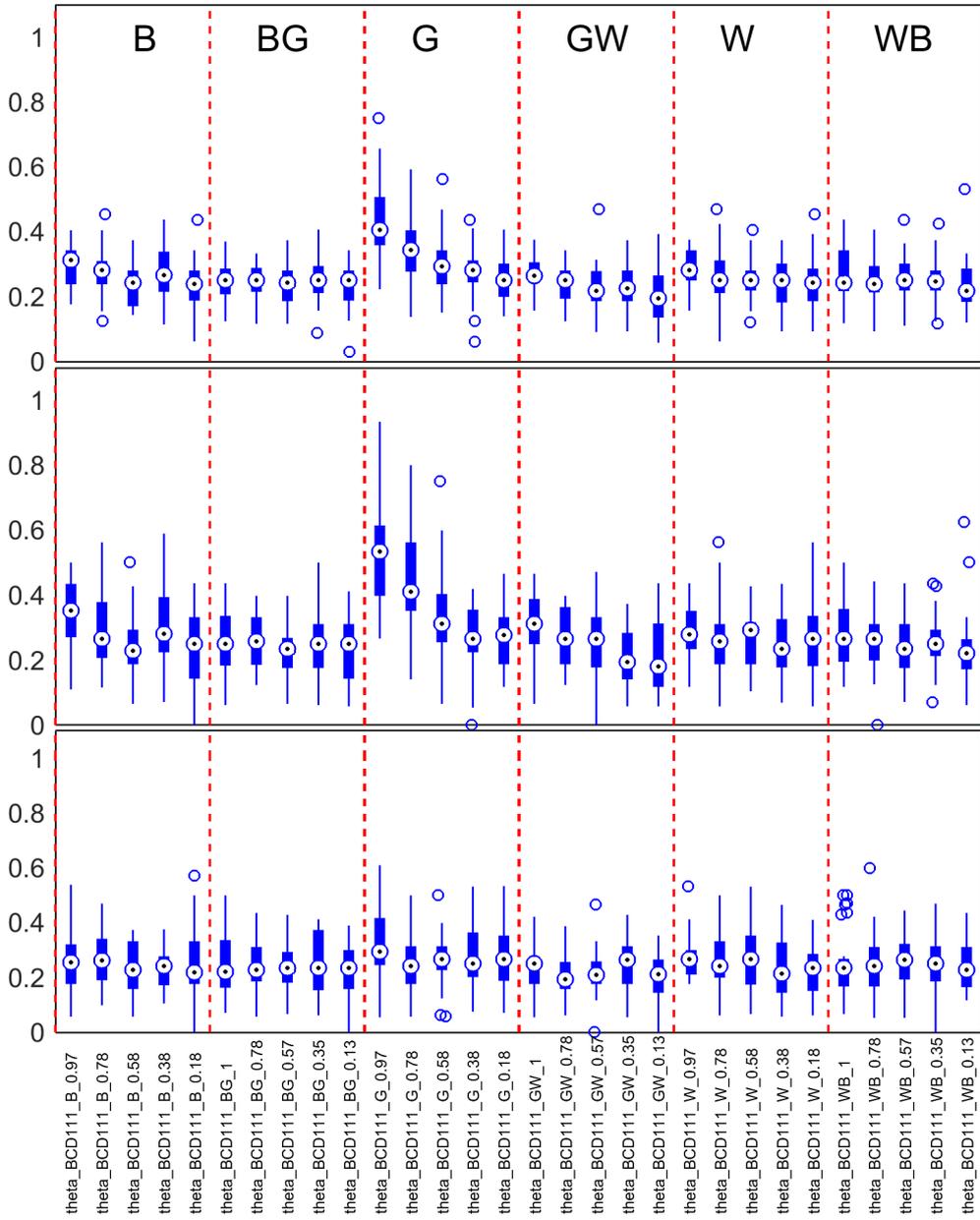
Figure A.8.6: Performance data for **theta_BCD111** ternary textures in **Live1**.
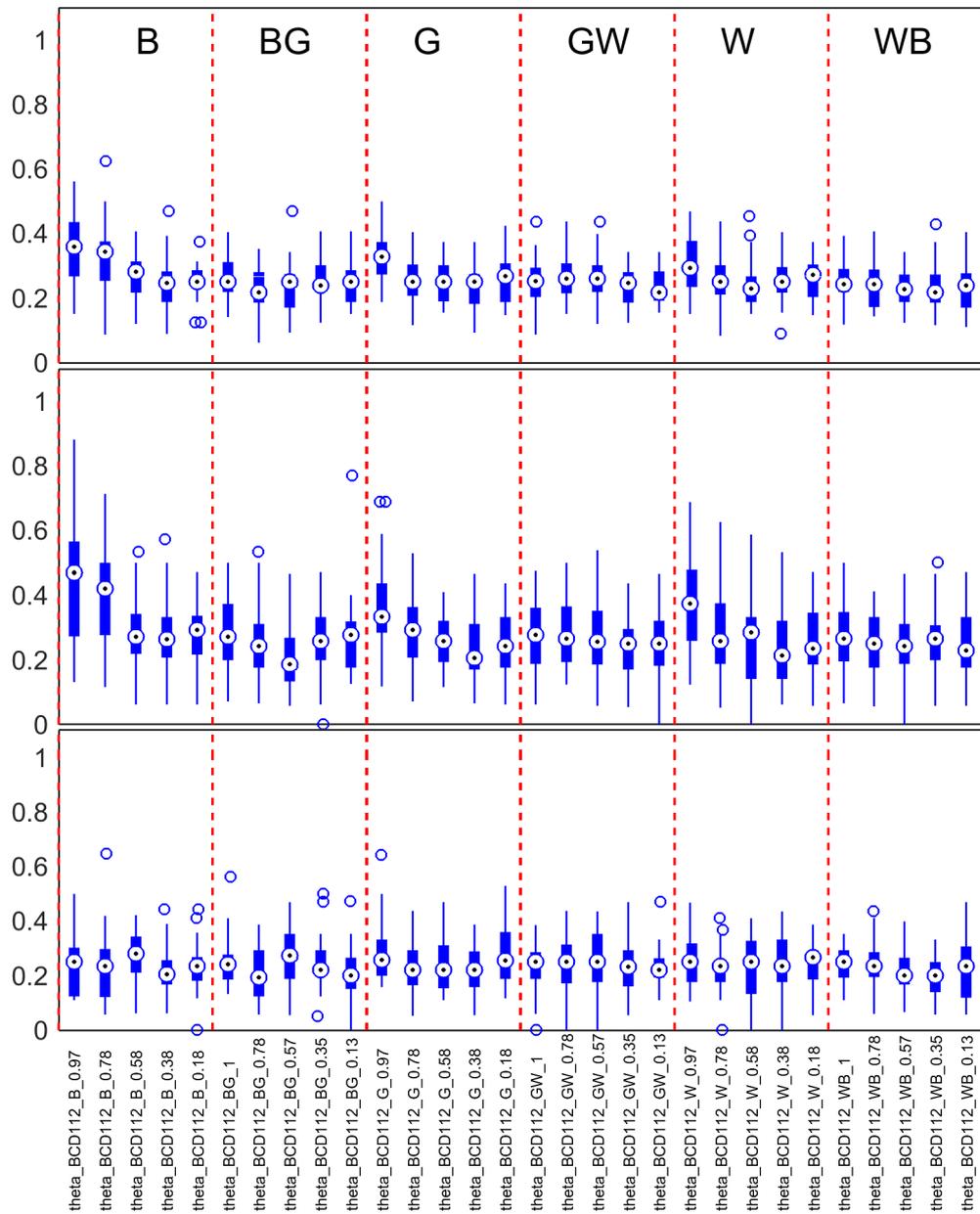
Figure A.8.7: Performance data for **theta_BCD112** ternary textures in **Live1**.
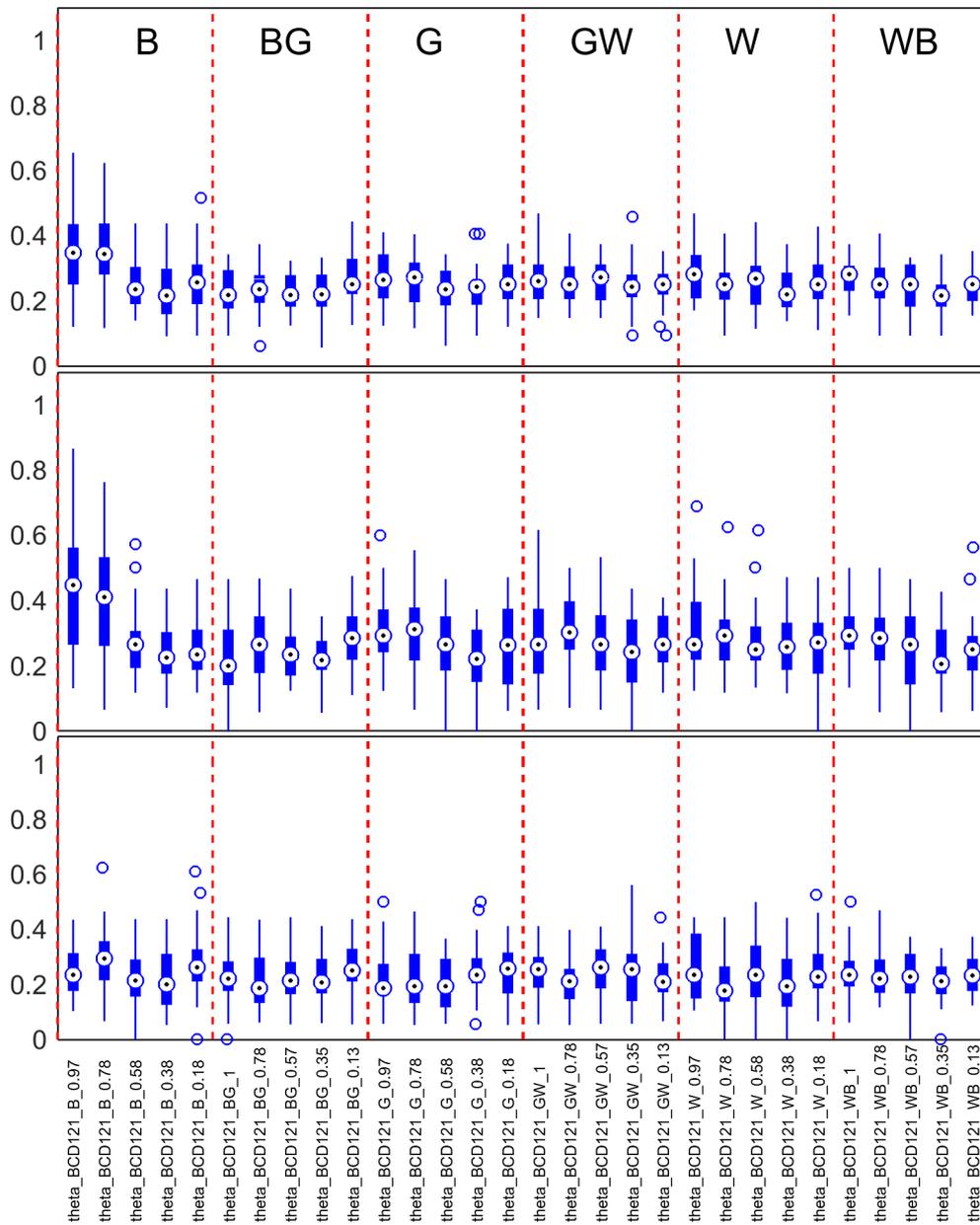
Figure A.8.8: Performance data for **theta_BCD121** ternary textures in **Live1**.
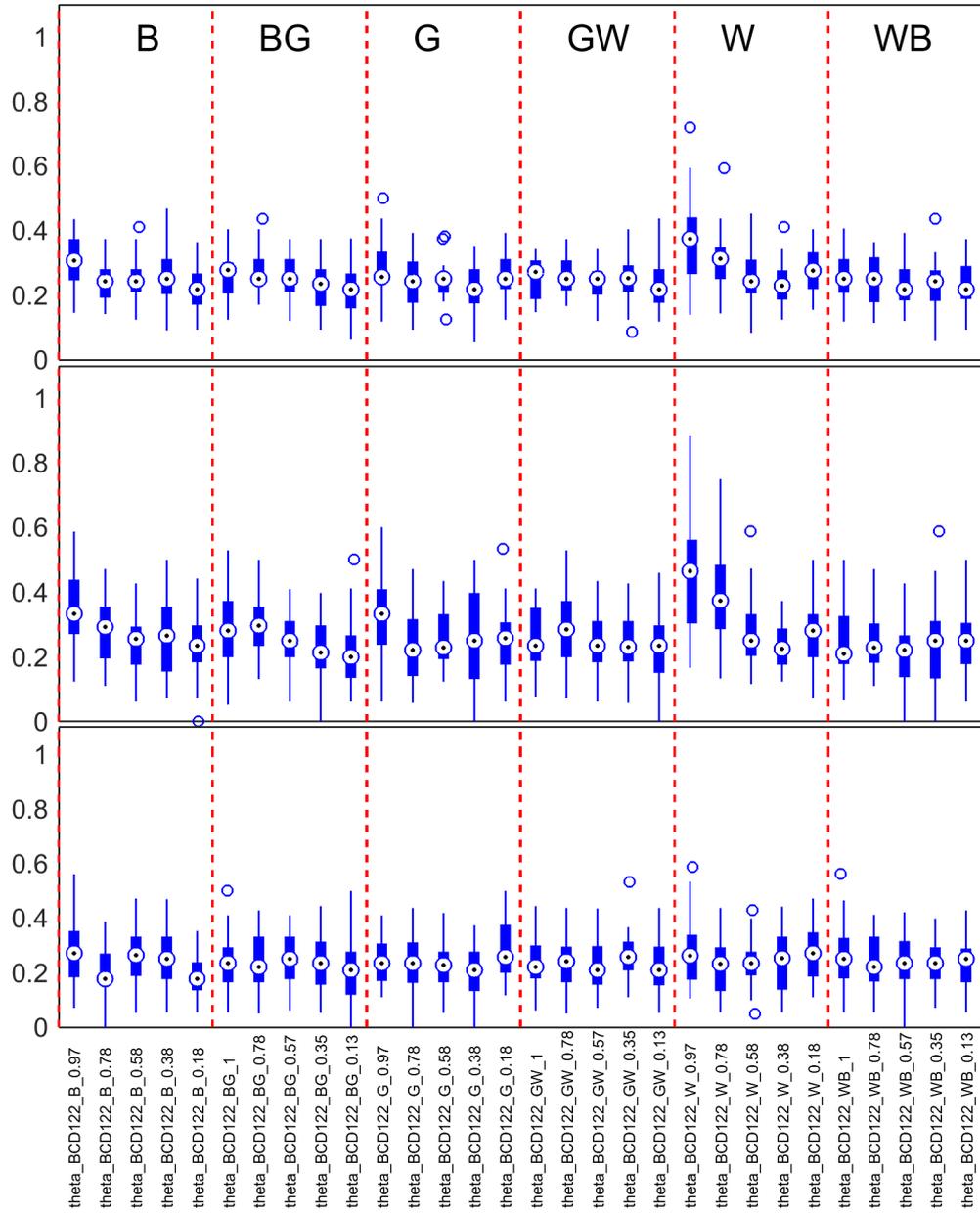
Figure A.8.9: Performance data for **theta_BCD122** ternary textures in **Live1**.
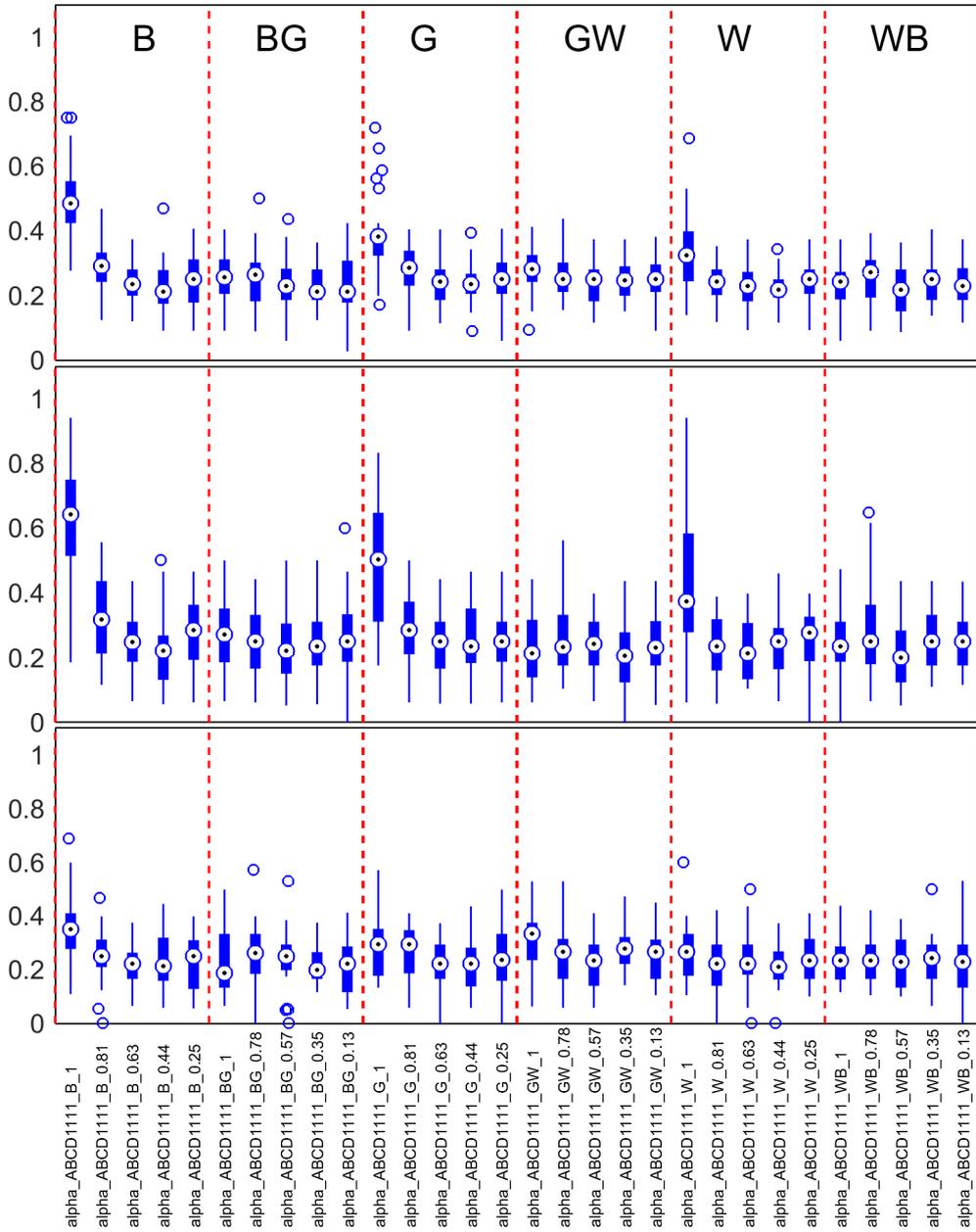
Figure A.8.10: Performance data for **alpha_ABCD1111** ternary textures in **Live1**.
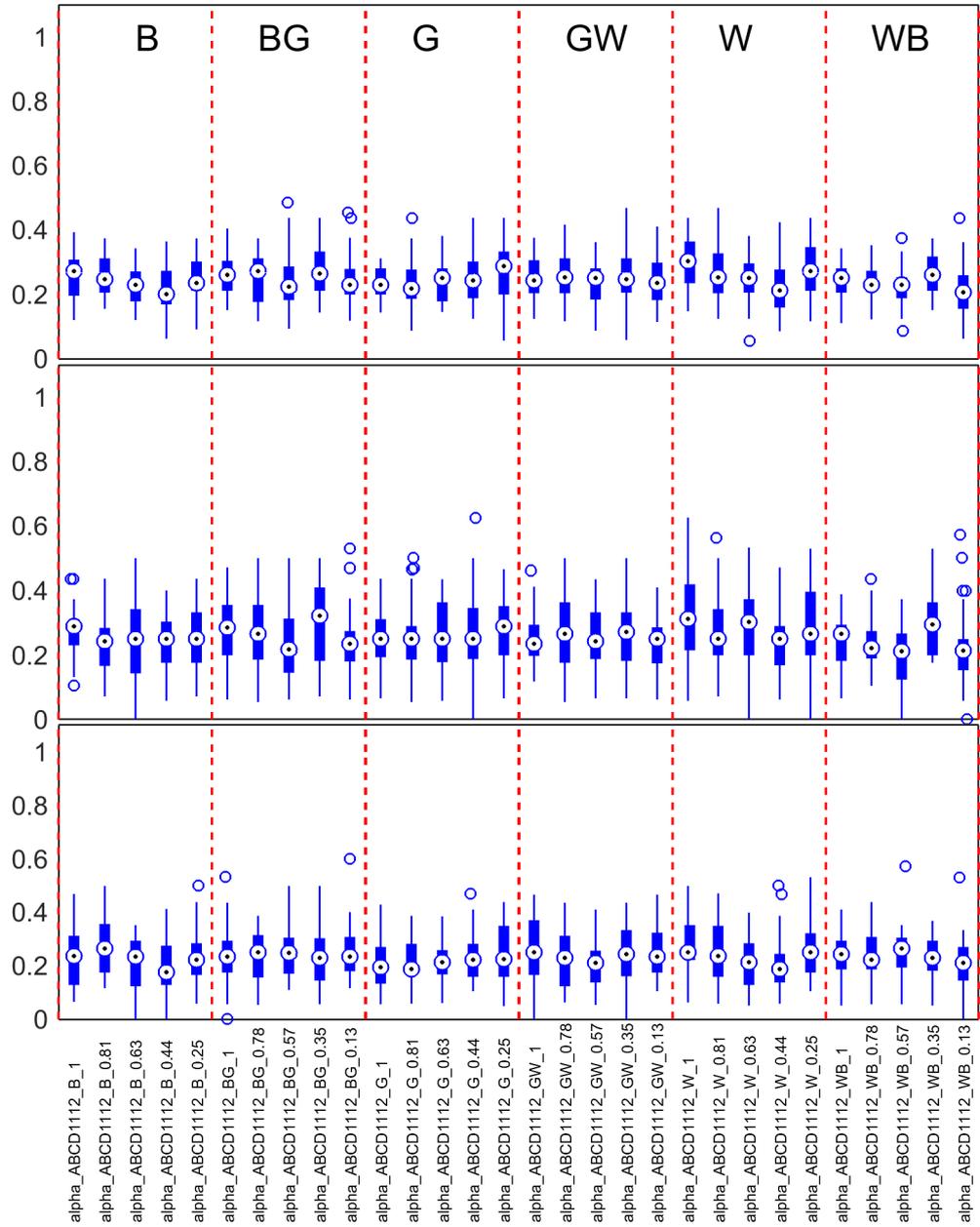
Figure A.8.11: Performance data for **alpha_ABCD1112** ternary textures in **Live1**.
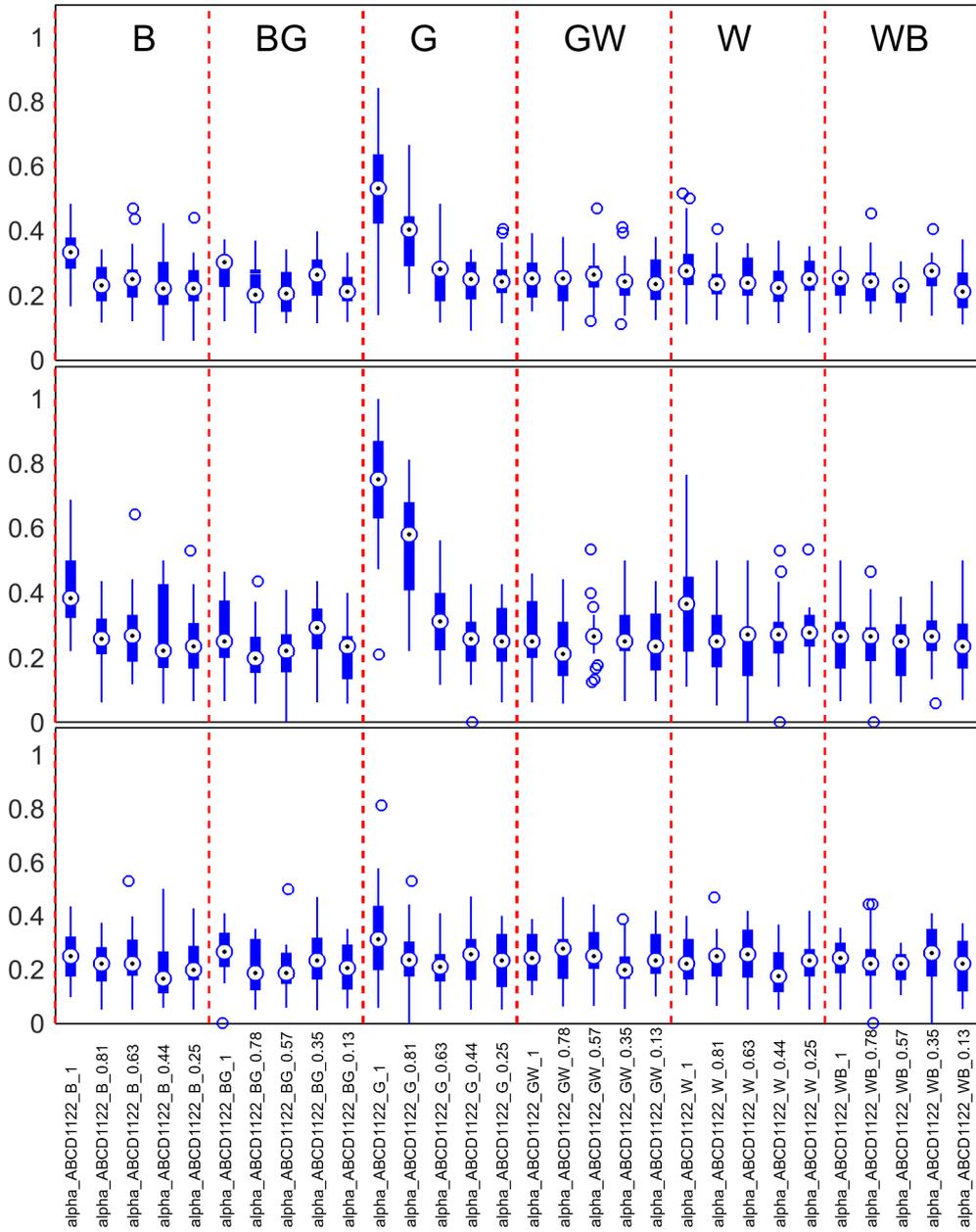
**Figure A.8.12**: Performance data for **alpha_ABCD1122** ternary textures in **Live1**.
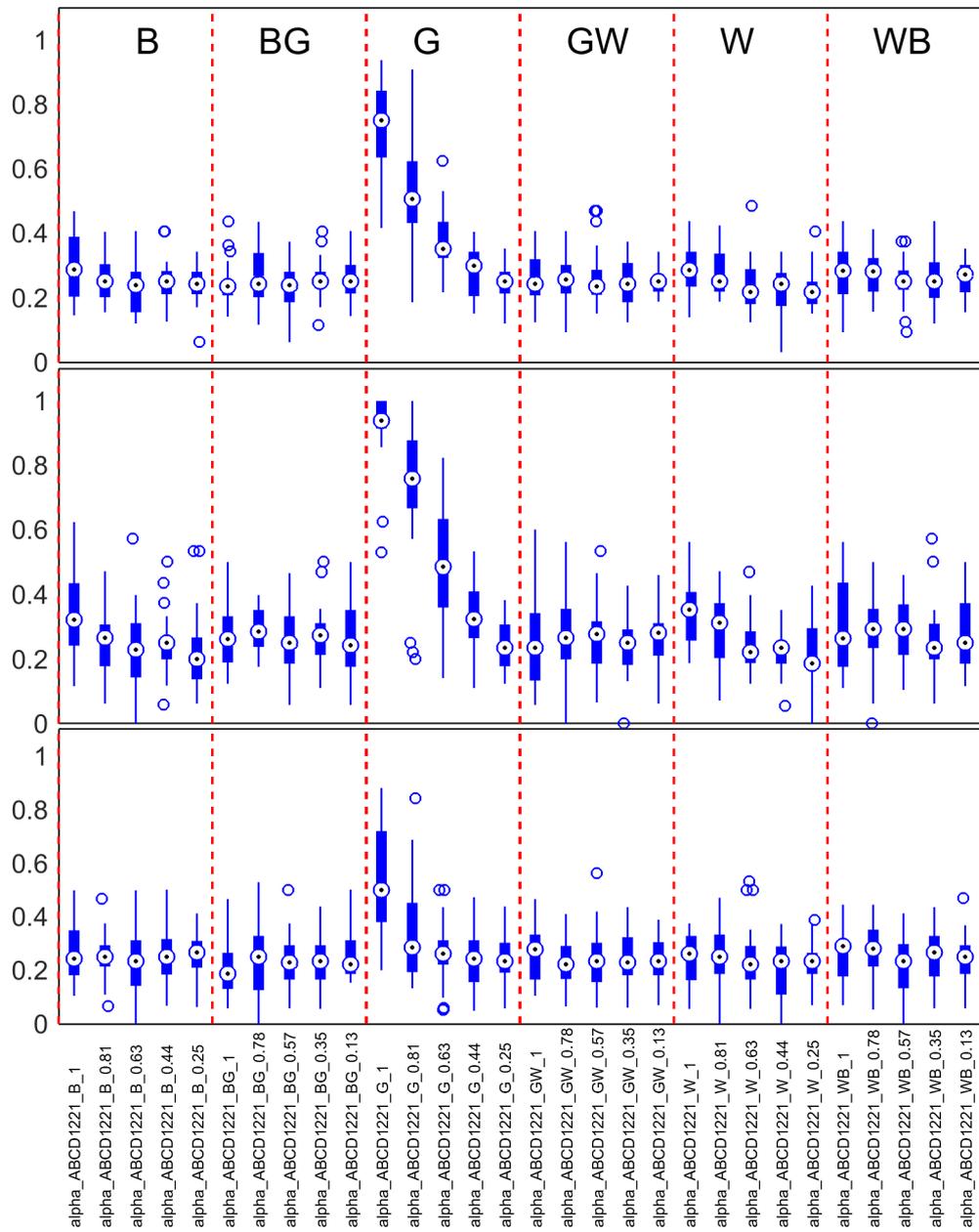
Figure A.8.13: Performance data for **alpha_ABCD1221** ternary textures in **Live1**.

6.9 Live2 Performance Data

All Live2 performance data was analysed *after* filtering the data by catch trial passes. i.e.: data from Workers that failed >= 4 catch trials was excluded. In the first instance, boxplots were produced for each ternary texture type (using the long condition only).

The three plots refer to the combined (top), long (middle) and short (bottom) presentation times. The dashed red lines divide the plots into the 6 rays with the following colour-biases: B (Black); BG (Black-grey); G (Grey); GW (Grey-white); W (White); WB (White-black). The step and ray values are indicated on the X axis, using the Victor nomenclature discussed in Chapter 4, Section 4.3.3.

Figure A.9.1: Performance data for **gamma_A1** ternary textures in **Live2**.
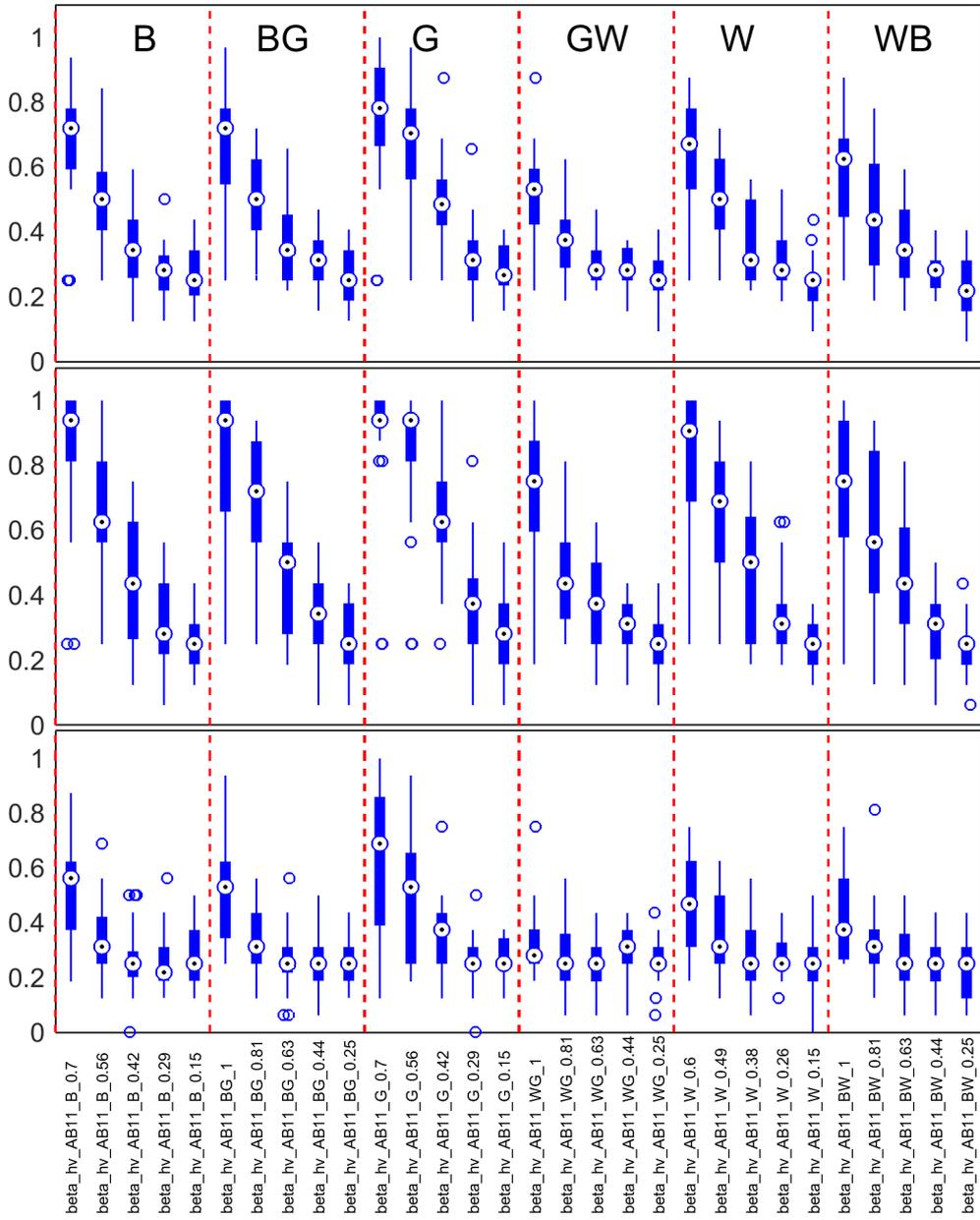
Figure A.9.2: Performance data for **beta_hv_AB11** ternary textures in **Live2**.
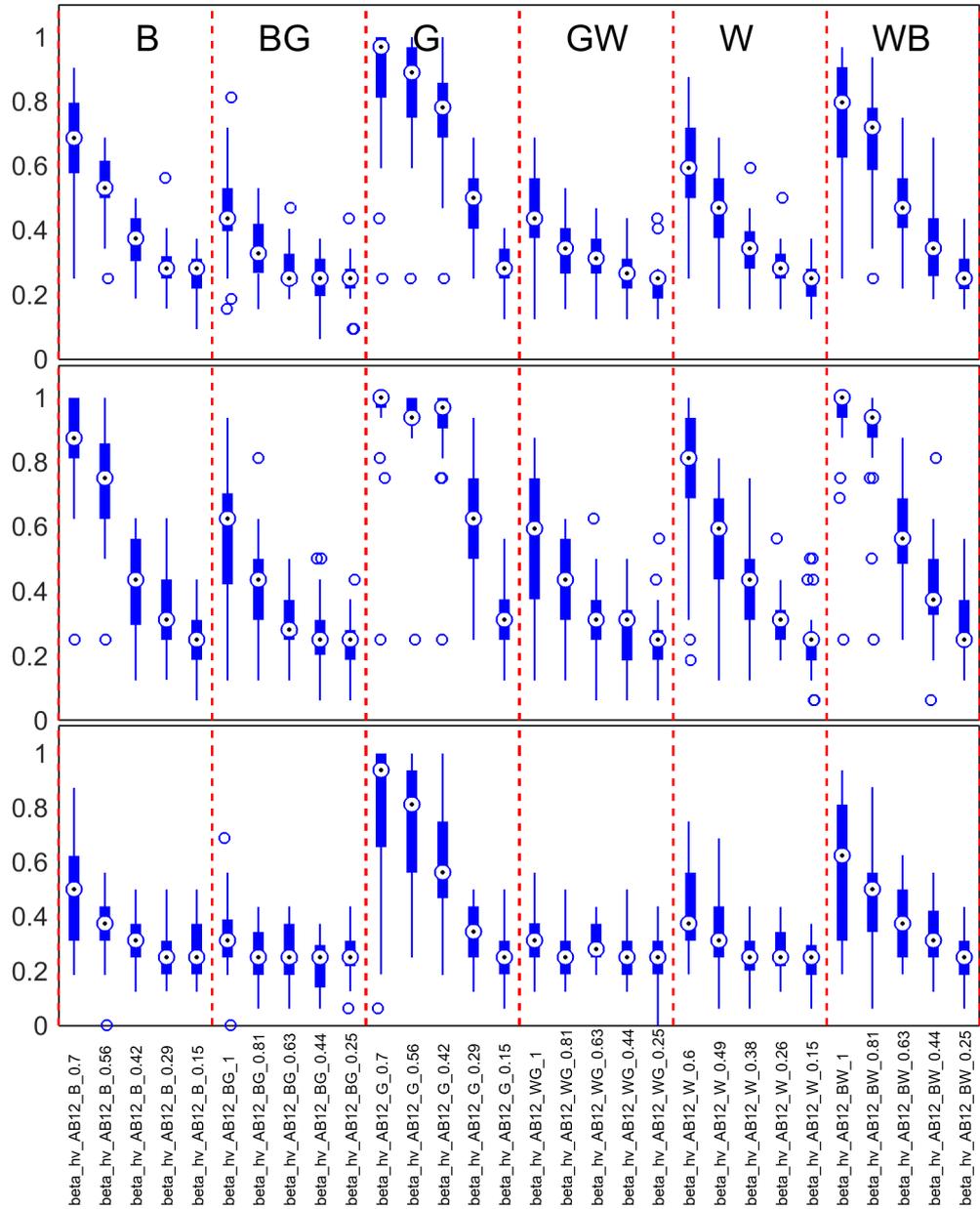
Figure A.9.3: Performance data for **beta_hv_AB12** ternary textures in **Live2**.
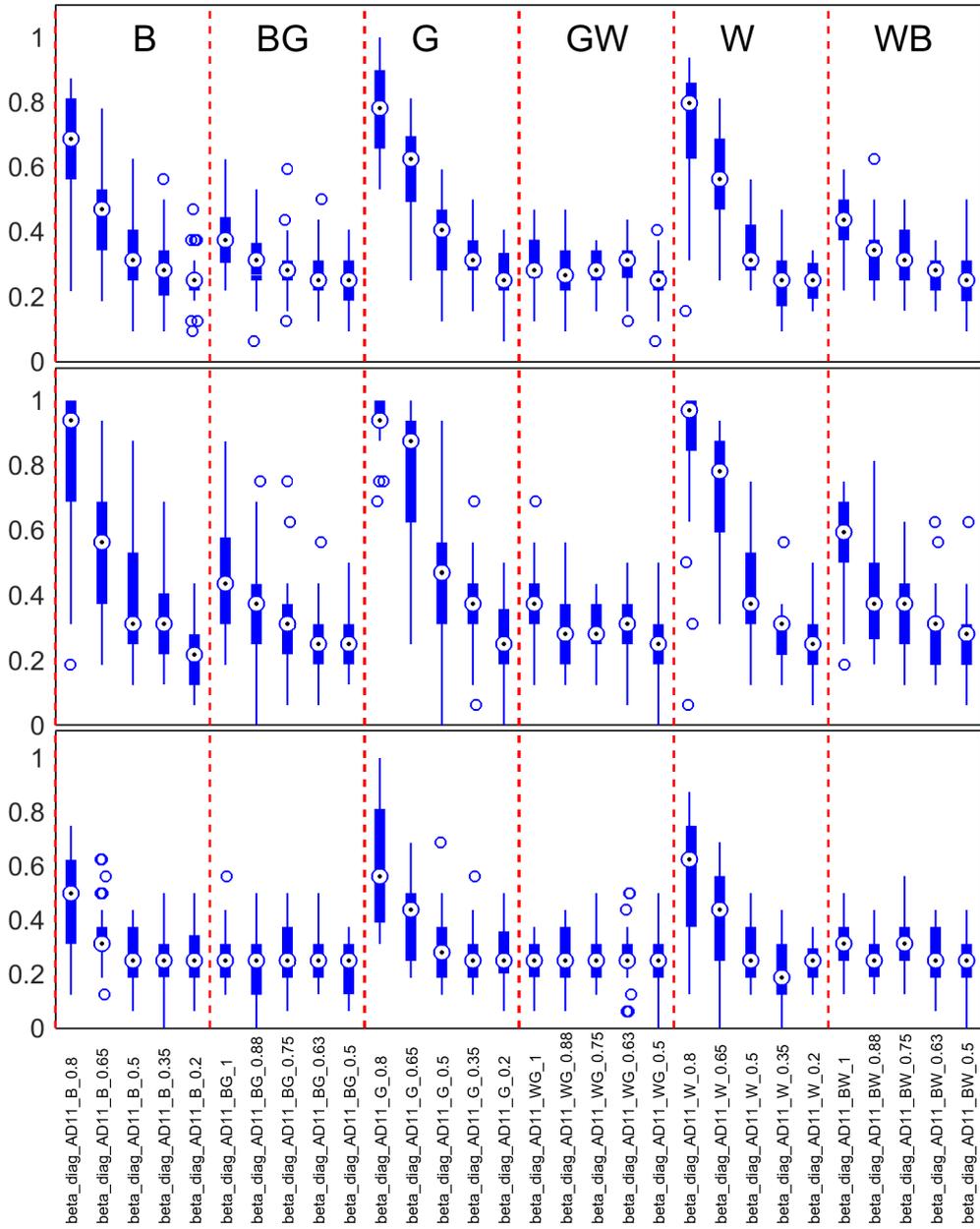
Figure A.9.4: Performance data for **beta_diag_AD11** ternary textures in **Live2**.
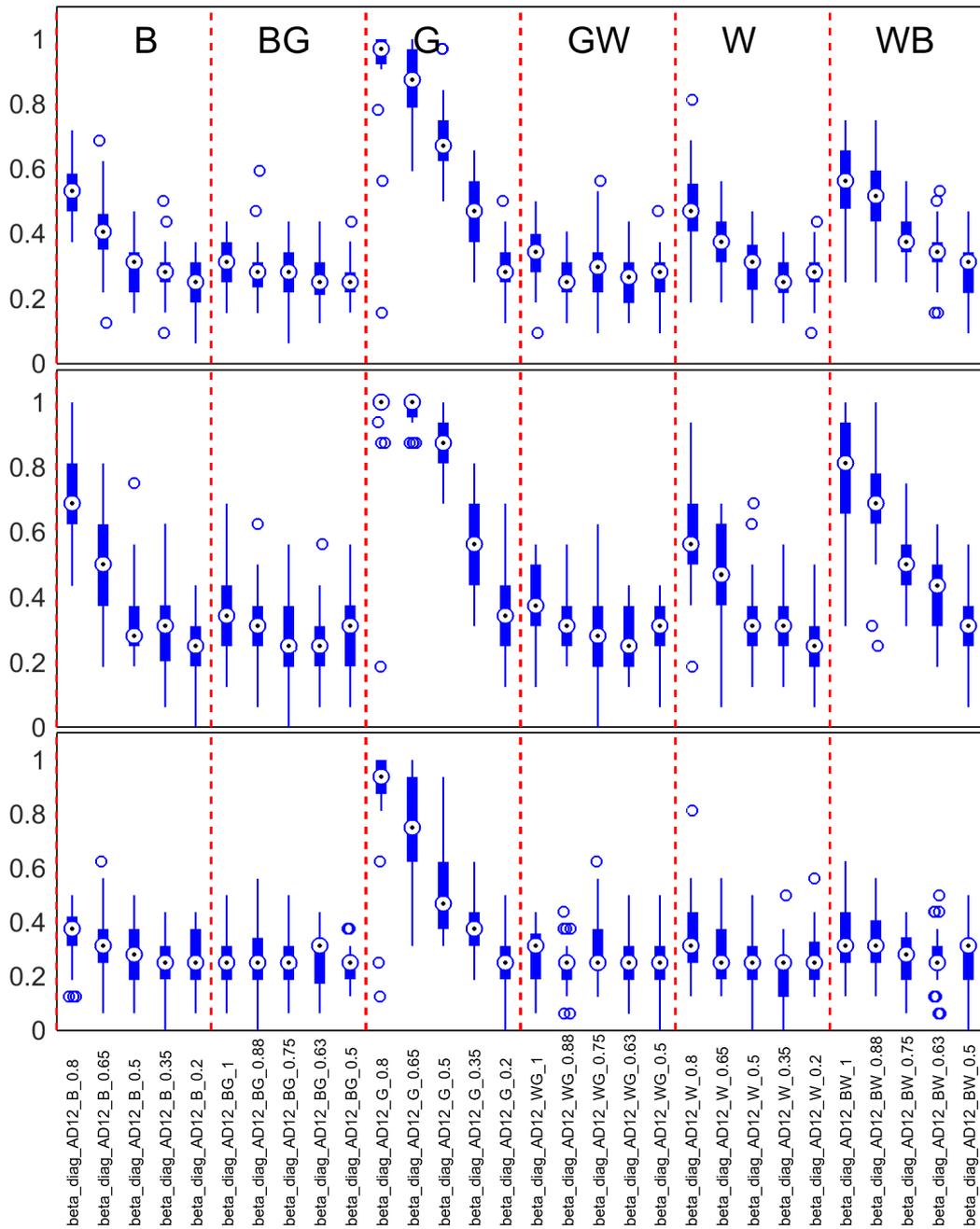
Figure A.9.5: Performance data for **beta_diag_AD12** ternary textures in **Live2**.
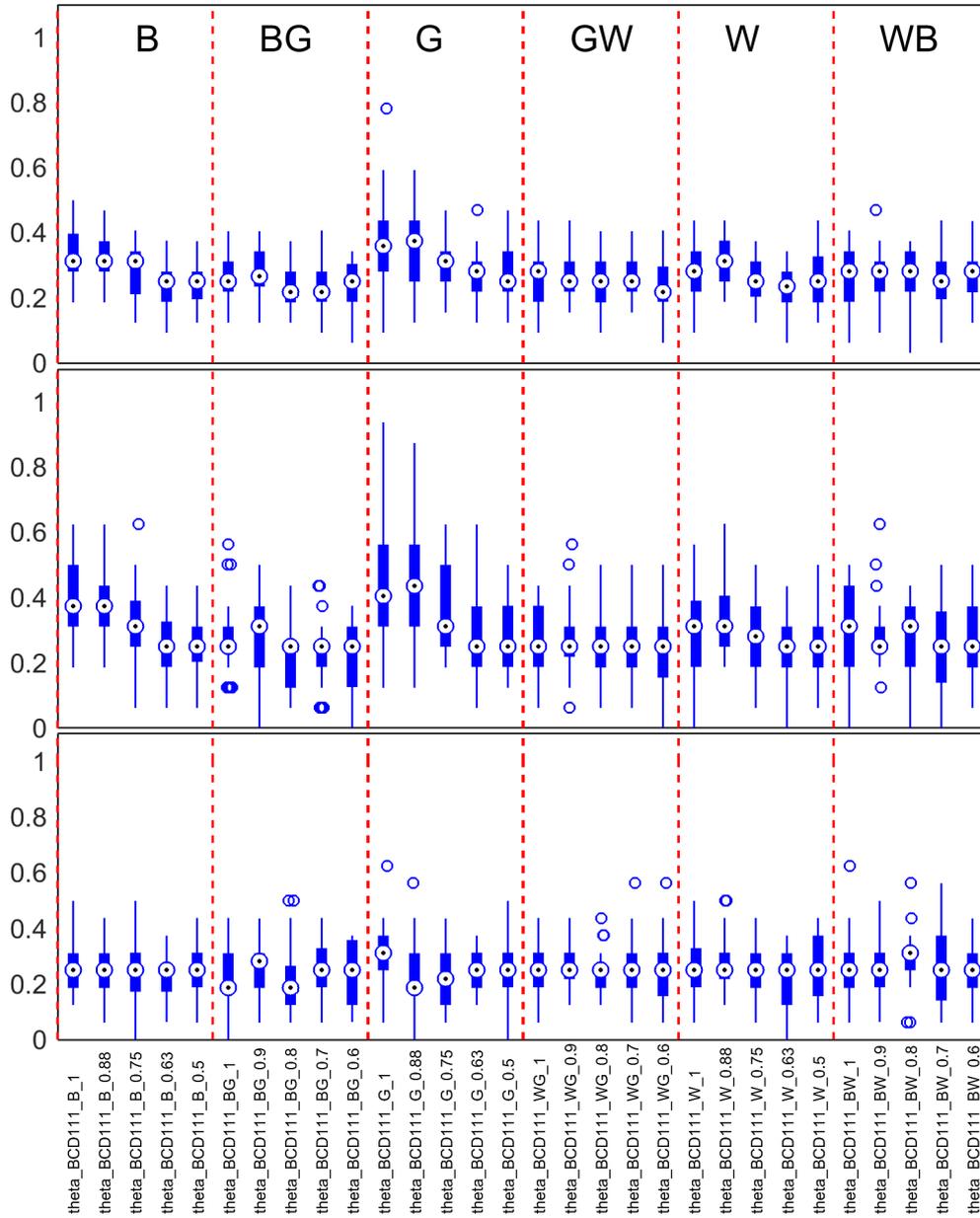
Figure A.9.6: Performance data for **theta_BCD111** ternary textures in **Live2**.
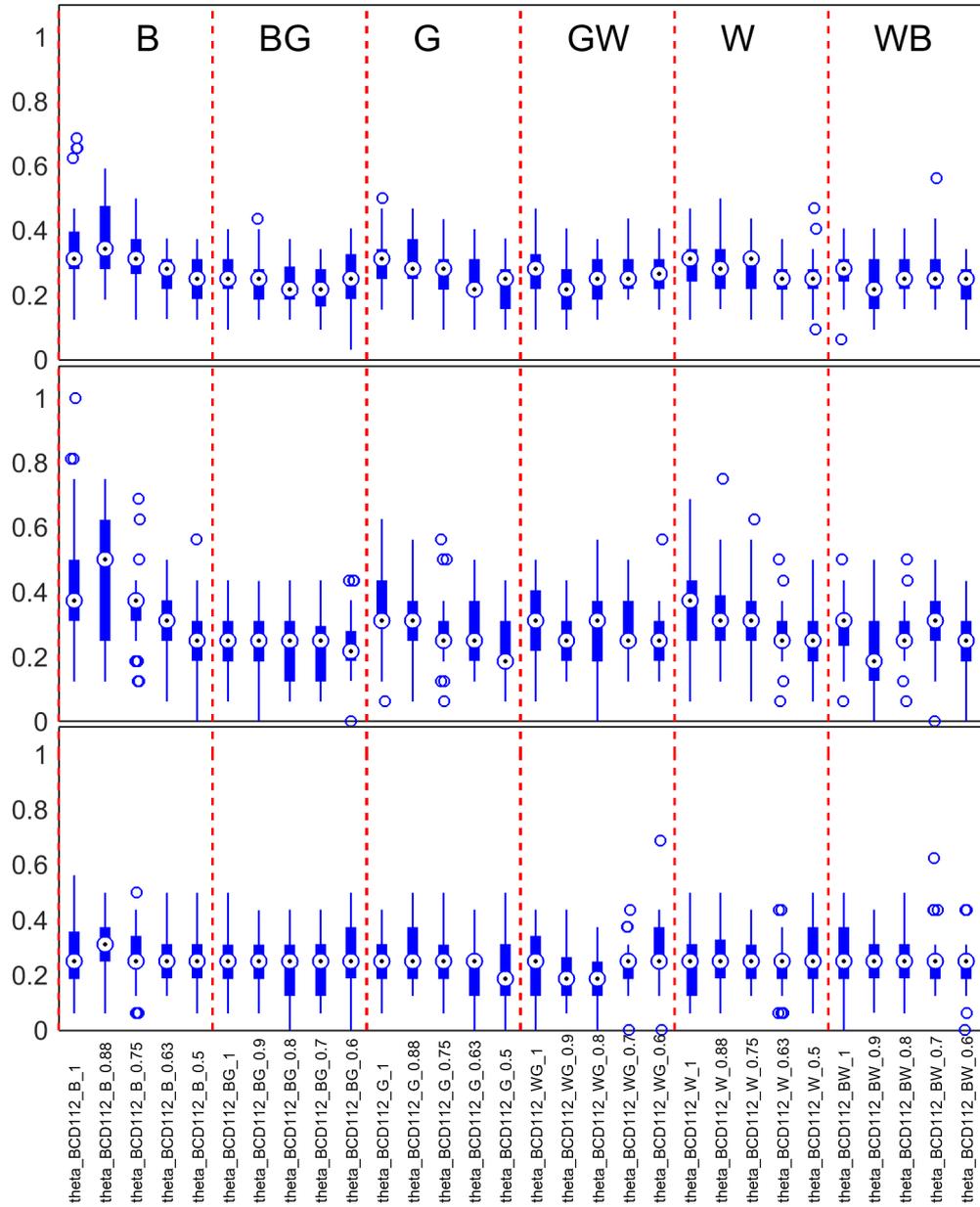
Figure A.9.7: Performance data for **theta_BCD112** ternary textures in **Live2**.
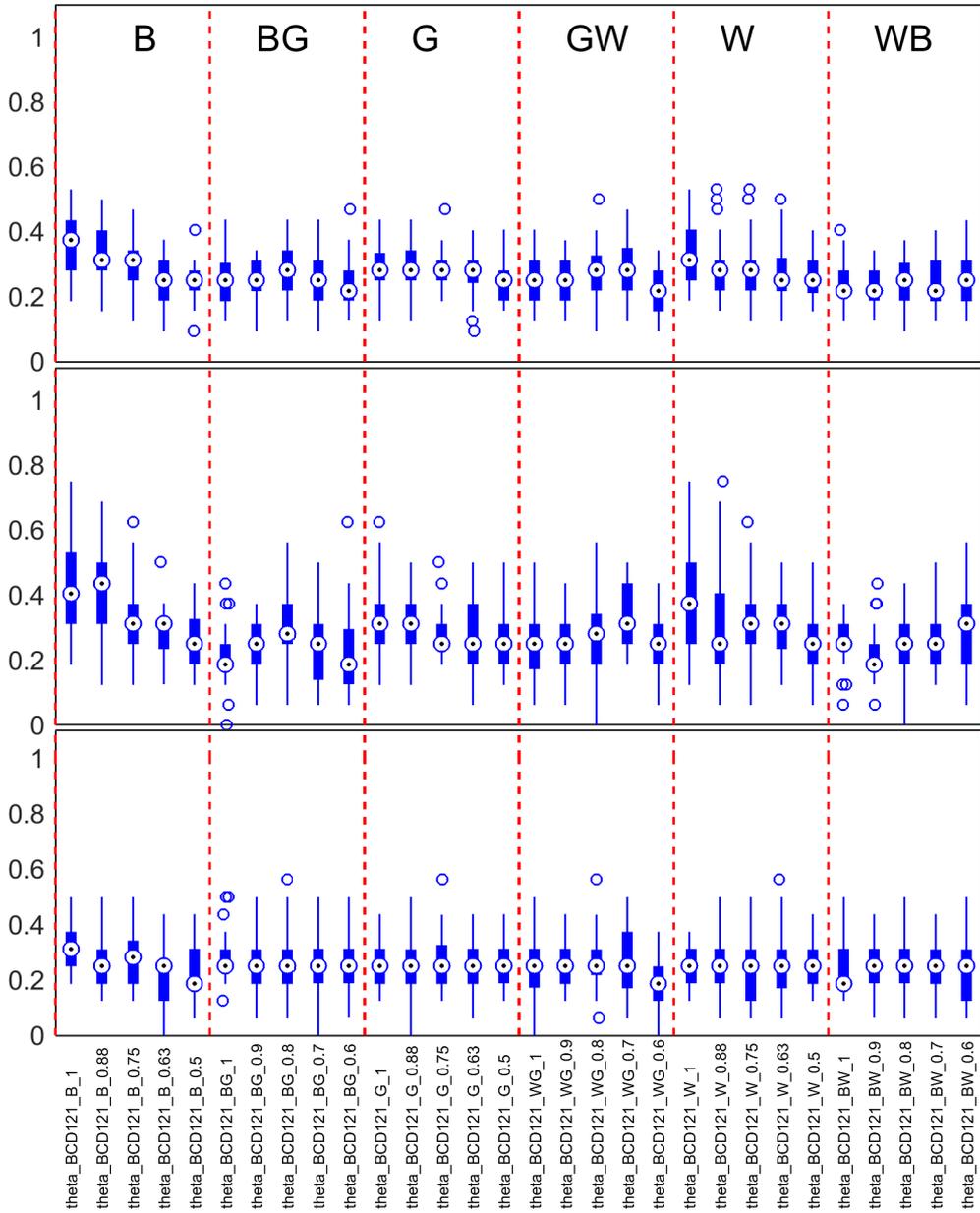
Figure A.9.8: Performance data for **theta_BCD121** ternary textures in **Live2**.
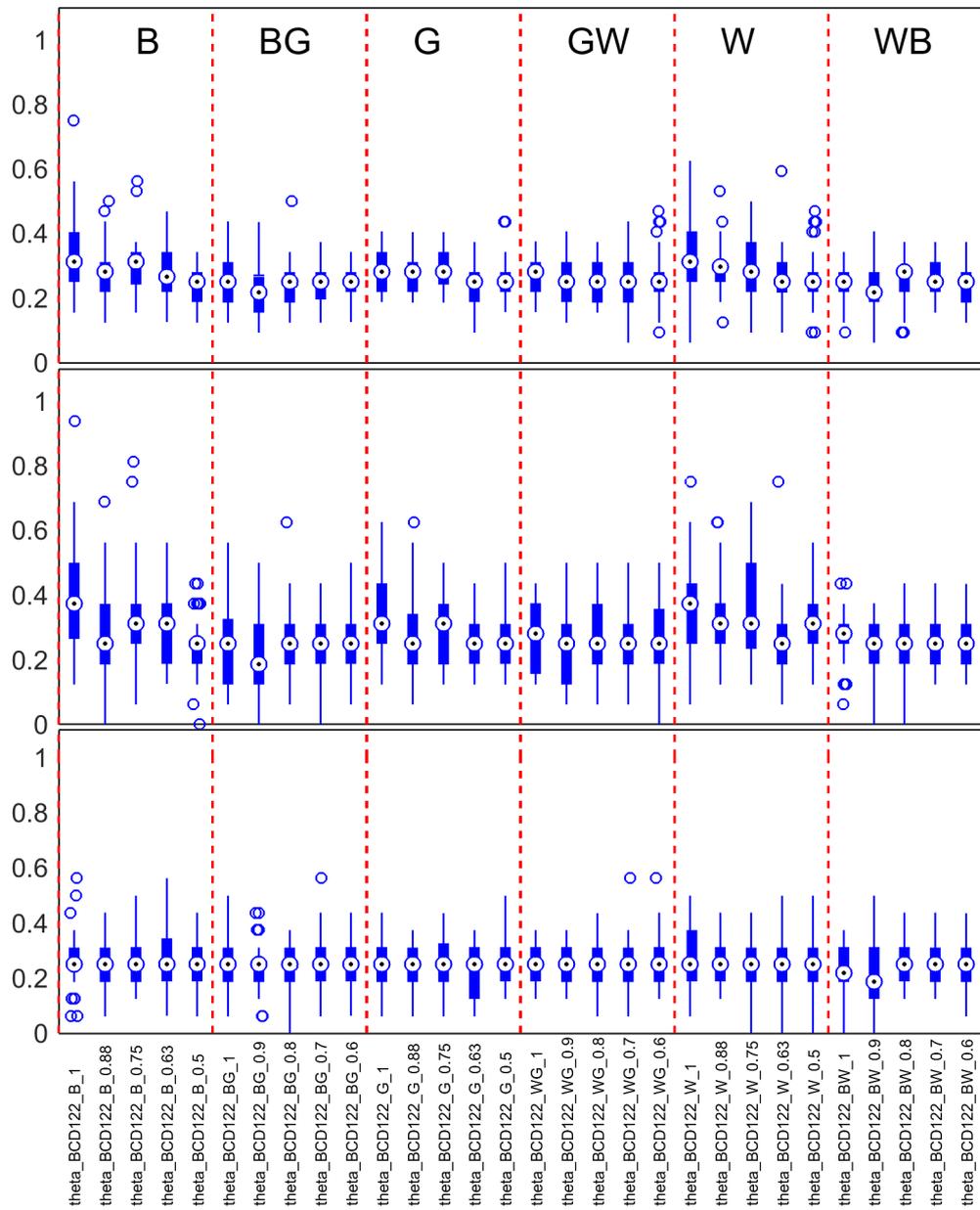
Figure A.9.9: Performance data for **theta_BCD122** ternary textures in **Live2**.
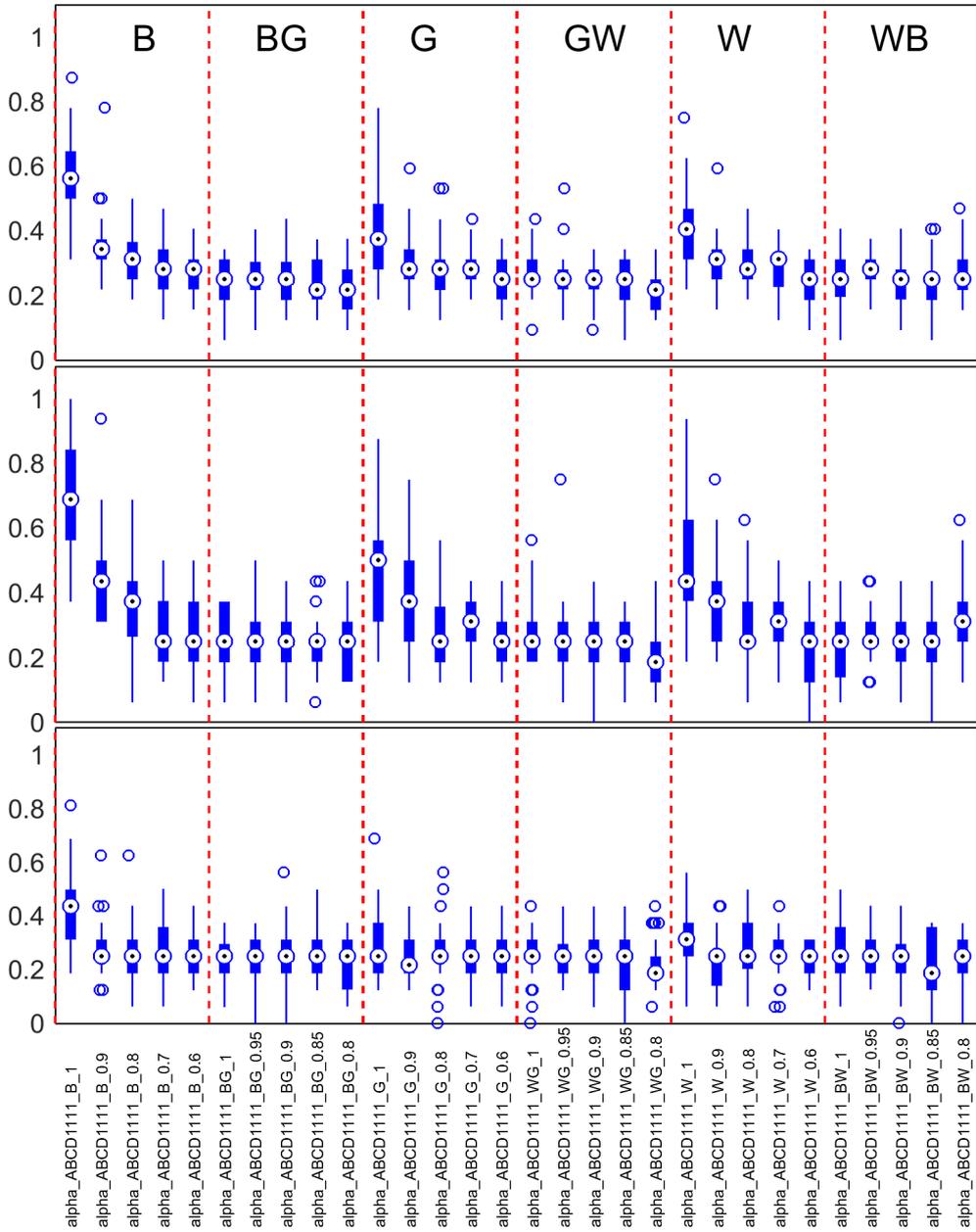
Figure A.9.10: Performance data for **alpha_ABCD1111** ternary textures in **Live2**.
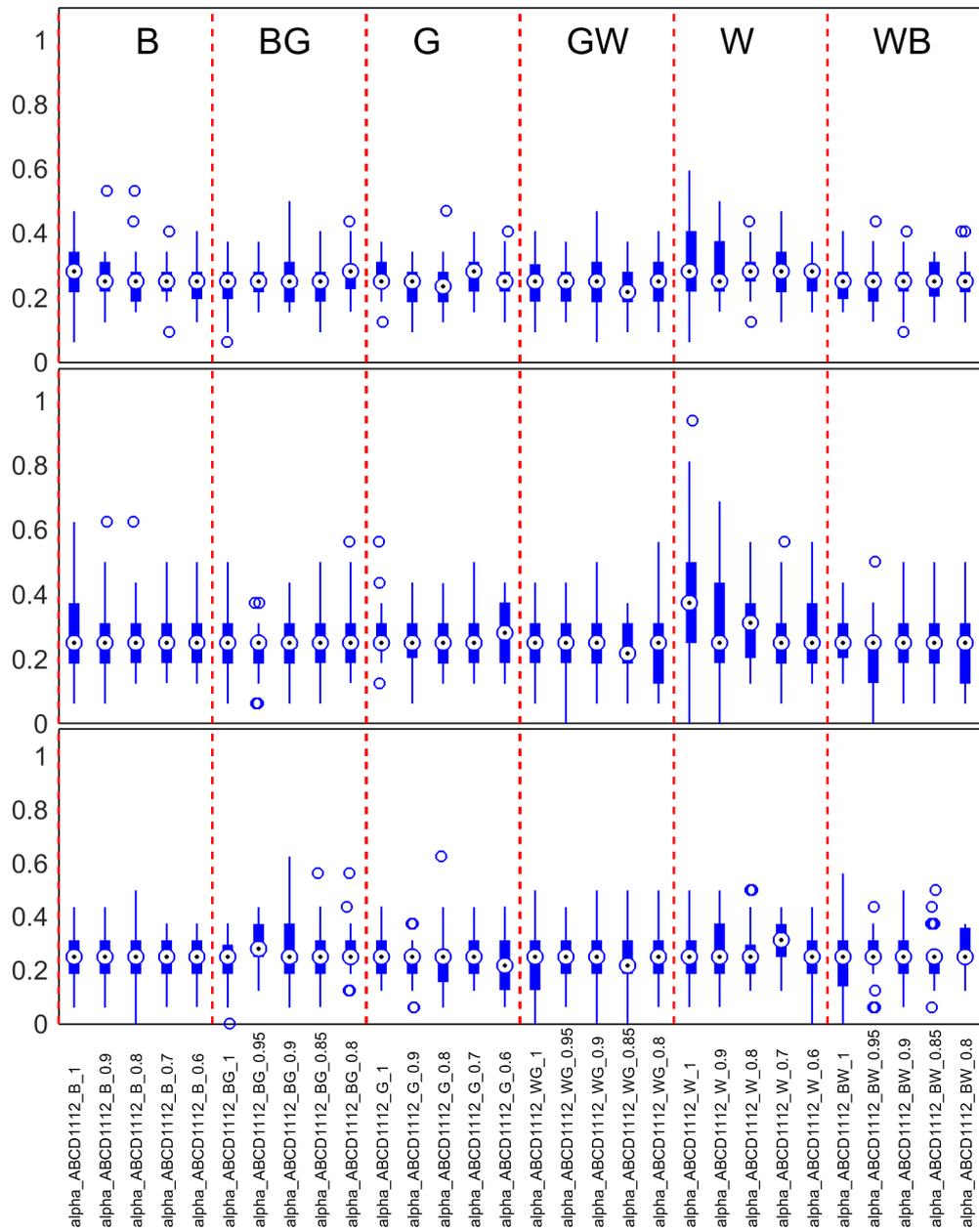
Figure A.9.11: Performance data for **alpha_ABCD1112** ternary textures in **Live2**.
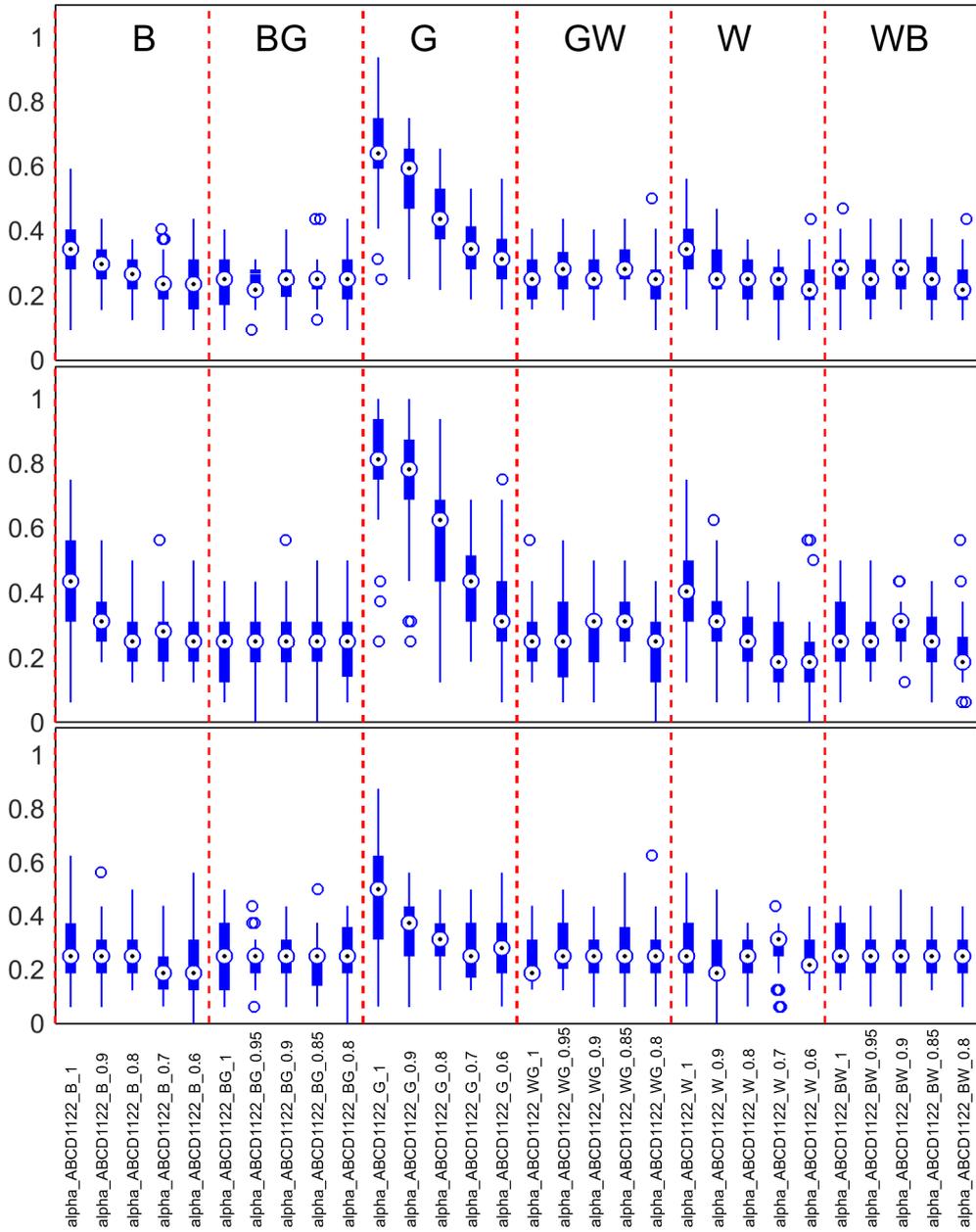
Figure A.9.12: Performance data for **alpha_ABCD1122** ternary textures in **Live2**.
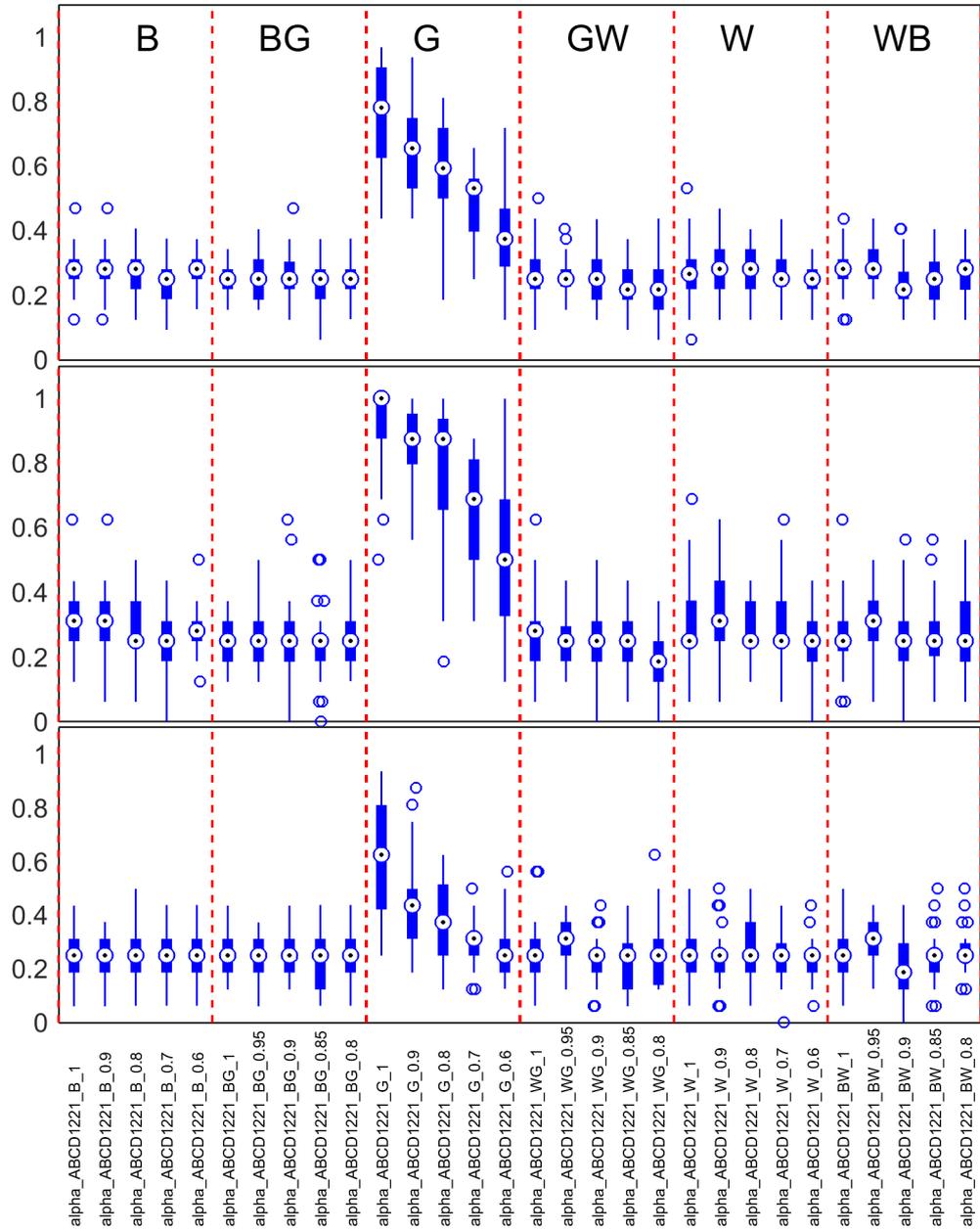
**Figure A.9.13:** Performance data for **alpha_ABCD1221** ternary textures in **Live2**.

6.10 References

BLAND, J. M. 2000. *An Introduction into Medical Statistics,* Oxford, Oxford University Press.

BLAND, J. M. & ALTMAN, D. G. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet,* 1**,** 307-10.

BLAND, J. M. & ALTMAN, D. G. 1999. Measuring agreement in method comparison studies. *Stat Methods Med Res,* 8**,** 135-60.

MADDESS, T. 2015. Histogram of Primitive Analysis of Deterministic and Stochastic Ternary Textures. *In Preparation*.

MADDESS, T. & NAGAI, Y. 2001. Discriminating isotrigon textures [corrected]. *Vision research,* 41**,** 3837-60.

MADDESS, T., NAGAI, Y., JAMES, A. C. & ANKIEWCZ, A. 2004. Binary and ternary textures containing higher-order spatial correlations. *Vision research,* 44**,** 1093-113.

MADDESS, T., NAGAI, Y., VICTOR, J. D. & TAYLOR, R. R. 2007. Multilevel isotrigon textures. *Journal of the Optical Society of America. A, Optics, image science, and vision,* 24**,** 278-93.

TAYLOR, R. R., MADDESS, T. & NAGAI, Y. 2008. Spatial biases and computational constraints on the encoding of complex local image structure. *Journal of vision,* 8**,** 19 1-13.

VAZ, S., FALKMER, T., PASSMORE, A. E., PARSONS, R. & ANDREOU, P. 2013. The case for using the repeatability coefficient when calculating test-retest reliability. *PLoS One,* 8**,** e73990.

VICTOR, J. D. & CONTE, M. M. 2012. Local image statistics: maximum-entropy constructions and perceptual salience. *Journal of the Optical Society of America. A, Optics, image science, and vision,* 29**,** 1313-45.

Thesis Bibliography

ABBOTT, L. & CHANCE, F. S. 2002. Rethinking the taxonomy of visual neurons. *Nature neuroscience,* 5**,** 391-392.

ABDULLAH, S. N., ALDAHLAWI, N., ROSLI, Y., BOON, M. Y. & MADDESS, T. 2012. Effect of contrast, stimulus density, and viewing distance on multifocal steady-state visual evoked potentials (MSVs). *Investigative ophthalmology & visual science,* 53**,** 5527-5535.

ADAMS, D. L. & HORTON, J. C. 2003. A precise retinotopic map of primate striate cortex generated from the representation of angioscotomas. *The Journal of neuroscience,* 23**,** 3771-3789.

ADELSON, E. H. On seeing stuff: the perception of materials by humans and machines. Photonics West 2001-Electronic Imaging, 2001. International Society for Optics and Photonics, 1-12.

ADELSON, E. H. & BERGEN, J. R. 1985. Spatiotemporal energy models for the perception of motion. *JOSA A,* 2**,** 284-299.

ADELSON, E. H. & BERGEN, J. R. 1991. The plenoptic function and the elements of early vision. *Computational models of visual processing,* 1.

ADESNIK, H., BRUNS, W., TANIGUCHI, H., HUANG, Z. J. & SCANZIANI, M. 2012. A neural circuit for spatial summation in visual cortex. *Nature,* 490**,** 226-31.

ALITTO, H. J. & USREY, W. M. 2003. Corticothalamic feedback and sensory processing. *Current opinion in neurobiology,* 13**,** 440-445.

ALITTO, H. J. & USREY, W. M. 2008. Origin and dynamics of extraclassical suppression in the lateral geniculate nucleus of the macaque monkey. *Neuron,* 57**,** 135-46.

ANDERSON, B. L. 2003. Perceptual organization and White's illusion. *PERCEPTION-LONDON-,* 32**,** 269-284.

ANDERSON, B. L. 2011. Visual perception of materials and surfaces. *Current Biology,* 21**,** R978-R983.

ANZAI, A. & DEANGELIS, G. C. 2010. Neural computations underlying depth perception. *Curr Opin Neurobiol,* 20**,** 367-75.

ANZAI, A., OHZAWA, I. & FREEMAN, R. D. 1999. Neural mechanisms for processing binocular information II. Complex cells. *Journal of Neurophysiology,* 82**,** 909-924.

ARCIZET, F., JOUFFRAIS, C. & GIRARD, P. 2009. Coding of shape from shading in area V4 of the macaque monkey. *BMC Neurosci,* 10**,** 140.

ATICK, J. J. & REDLICH, A. N. 1992. What does the retina know about natural scenes? *Neural computation,* 4**,** 196-210.

ATTNEAVE, F. 1954. Some informational aspects of visual perception. *Psychological review,* 61**,** 183.

AWS, A. W. S. 2014. *Amazon Mechanical Turk (mTurk)* [Online]. Available: https://www.mturk.com/mturk/welcome [Accessed 4 November 2014.

AZZOPARDI, P. & COWEY, A. 1996. The overrepresentation of the fovea and adjacent retina in the striate cortex and dorsal lateral geniculate nucleus of the macaque monkey. *Neuroscience,* 72**,** 627-639.

BALASUBRAMANIAN, V. & STERLING, P. 2009. Receptive fields and functional architecture in the retina. *The Journal of physiology,* 587**,** 2753-2767.

BARBOSA, M. S., BUBNA-LITIC, A. & MADDESS, T. 2013. Locally countable properties and the perceptual salience of textures. *Journal of the Optical Society of America A,* 30**,** 1687-1697.

BARLOW, H. 2001. Redundancy reduction revisited. *Network,* 12**,** 241-53.

BARLOW, H. B. 1961. Possible principles underlying the transformations of sensory messages.

BARLOW, H. B. 1963. The Information Capacity of Nervous Transmission. *Kybernetik,* 2**,** 1.

BEASON-HELD, L. L., PURPURA, K. P., KRASUSKI, J. S., DESMOND, R. E., MANGOT, D. J., DALY, E. M., OPTICAN, L. M., RAPOPORT, S. I. & VANMETER, J. W. 2000. Striate cortex in humans demonstrates the relationship between activation and variations in visual form. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale,* 130**,** 221-6.

BEASON-HELD, L. L., PURPURA, K. P., KRASUSKI, J. S., MAISOG, J. M., DALY, E. M., MANGOT, D. J., DESMOND, R. E., OPTICAN, L. M., SCHAPIRO, M. B. & VANMETER, J. W. 1998a. Cortical regions involved in visual texture perception: a fMRI study. *Brain research. Cognitive brain research,* 7**,** 111-8.

BEASON-HELD, L. L., PURPURA, K. P., VAN METER, J. W., AZARI, N. P., MANGOT, D. J., OPTICAN, L. M., MENTIS, M. J., ALEXANDER, G. E., GRADY, C. L., HORWITZ, B., RAPOPORT, S. I. & SCHAPIRO, M. B. 1998b. PET reveals occipitotemporal pathway activation during elementary form perception in humans. *Visual neuroscience,* 15**,** 503-10.

BELL, A. J. & SEJNOWSKI, T. J. 1997. The "independent components" of natural scenes are edge filters. *Vision research,* 37**,** 3327-3338.

BERGEN, J. R. & ADELSON, E. 1991. Theories of visual texture perception. *Spatial vision,* 10**,** 114-134.

BERGEN, J. R. & LANDY, M. S. 1991. Computational modeling of visual texture segregation. *Computational models of visual processing,* 1**,** 253-271.

BLAKE, A., BULTHOFF, H. H. & SHEINBERG, D. 1993. Shape from texture: ideal observers and human psychophysics. *Vision Res,* 33**,** 1723-37.

BLAKESLEE, B. & MCCOURT, M. E. 2004. A unified theory of brightness contrast and assimilation incorporating oriented multiscale spatial filtering and contrast normalization. *Vision research,* 44**,** 2483-2503.

BLAKESLEE, B. & MCCOURT, M. E. 2008. Nearly instantaneous brightness induction. *Journal of Vision,* 8**,** 15.

BLAKESLEE, B. & MCCOURT, M. E. 2012. When is spatial filtering enough? Investigation of brightness and lightness perception in stimuli containing a visible illumination component. *Vision research,* 60**,** 40-50.

BLAND, J. M. 2000. *An Introduction into Medical Statistics,* Oxford, Oxford University Press.

BLAND, J. M. & ALTMAN, D. G. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet,* 1**,** 307-10.

BLAND, J. M. & ALTMAN, D. G. 1999. Measuring agreement in method comparison studies. *Stat Methods Med Res,* 8**,** 135-60.

BOLZ, J. & GILBERT, C. D. 1986. Generation of end-inhibition in the visual cortex via interlaminar connections. *Nature,* 320**,** 362-5.

BOULTON, A. A., BAKER, G. B. & VANDERWOLF, C. H. 1990. *Neurophysiological techniques: applications to neural systems*, Humana Press.

BOVIK, A. C., CLARK, M. & GEISLER, W. S. 1990. Multichannel texture analysis using localized spatial filters. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* 12**,** 55-73.

BRACEWELL, R. 1965. The Fourier Transform and IIS Applications. *New York*.

BRIGGS, F. & USREY, W. M. 2008. Emerging views of corticothalamic function. *Current opinion in neurobiology,* 18**,** 403-407.

BRONKHORST, A. W. 2000. The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica united with Acustica,* 86**,** 117-128.

BUHRMESTER, M., KWANG, T. & GOSLING, S. D. 2011. Amazon's Mechanical Turk : A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science,* 6**,** 3-5.

BUZSAKI, G. & KANDEL, A. 1998. Somadendritic backpropagation of action potentials in cortical pyramidal cells of the awake rat. *Journal of neurophysiology,* 79**,** 1587-91.

CAELLI, T. & JULESZ, B. 1978. On perceptual analyzers underlying visual texture discrimination: Part I. *Biological Cybernetics,* 28**,** 167-175.

CAELLI, T., JULESZ, B. & GILBERT, E. 1978. On perceptual analyzers underlying visual texture discrimination: Part II. *Biological Cybernetics,* 29**,** 201-214.

CASAGRANDE, V. A. & XU, X. 2004. 31 Parallel Visual Pathways: A Comparative Perspective.

CATTELL, R. B. 1966. The Scree Test For The Number Of Factors. *Multivariate Behavioral Research,* 1**,** 245-276.

CAVANAGH, P. & LECLERC, Y. G. 1989. Shape from shadows. *J Exp Psychol Hum Percept Perform,* 15**,** 3-27.

CHANCE, F. S., NELSON, S. B. & ABBOTT, L. 1999. Complex cells as cortically amplified simple cells. *Nature neuroscience,* 2**,** 277-282.

CHANDLER, M., DE LA CRUZ, F., DYDA, F., HICKMAN, A. B., MONCALIAN, G. & TON-HOANG, B. 2013. Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nat Rev Microbiol,* 11**,** 525-38.

CHEVREUL, M. E. 1855. *The principles of harmony and contrast of colours, and their applications to the arts*, Longman, Brown, Green, and Longmans.

CHUBB, C. & LANDY, M. S. 1991. Orthogonal distribution analysis: A new approach to the study of texture perception. *Computational models of visual processing,* 12**,** 394.

CHUBB, C., LANDY, M. S. & ECONOPOULY, J. 2004. A visual mechanism tuned to black. *Vision Res,* 44**,** 3223-32.

CHUBB, C., OLZAK, L. & DERRINGTON, A. 2001. Second-order processes in vision: introduction. *JOSA A,* 18**,** 2175-2178.

CHUBB, C. & SPERLING, G. 1991. Texture quilts: Basic tools for studying motion-from-texture. *Journal of Mathematical Psychology,* 35**,** 411-442.

CIRESAN, D. C., MEIER, U., GAMBARDELLA, L. M. & SCHMIDHUBER, J. 2010. Deep, big, simple neural nets for handwritten digit recognition. *Neural computation,* 22**,** 3207-3220.

CLAASEN, T. & MECKLENBRAUKER, W. 1980. The Wigner distribution—A tool for time-frequency signal analysis. Part I: Continuous-time signals. *Philips J. Res,* 35**,** 217-250.

COHEN, M. A. & GROSSBERG, S. 1984. Neural dynamics of brightness perception: Features, boundaries, diffusion, and resonance. *Perception & psychophysics,* 36**,** 428-456.

COLE, F., SANIK, K., DECARLO, D., FINKELSTEIN, A., FUNKHOUSER, T., RUSINKIEWICZ, S. & SINGH, M. How Well do Line Drawings Depict Shape? ACM Transactions on Graphics, 2009.

COMON, P. 1994. Independent component analysis, a new concept? *Signal processing,* 36**,** 287-314.

COMSOL. 2014. *http://www.comsol.com/* [Online].

COREN, S. 2003. *Sensation and perception*, Wiley Online Library.

CORNSWEET, T. 1970. *Visual Perception,* New York, Academic Press.

COY, D. G., BARBOSA, M., SEAMONS, J. W. G., & MADDESS, T. 2014. Learning Effects for Discrimination of Isotrigon Textures. (Honours Thesis; Journal Article In Preparation).

CRUMP, M. J., MCDONNELL, J. V. & GURECKIS, T. M. 2013. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS One,* 8**,** e57410.

DAKIN, S. C. & BEX, P. J. 2003. Natural image statistics mediate brightness 'filling in'. *Proc Biol Sci,* 270**,** 2341-8.

DAUGMAN, J. G. 1985. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *JOSA A,* 2**,** 1160-1169.

DAVID, S. V. & GALLANT, J. L. 2005. Predicting neuronal responses during natural vision. *Network: Computation in Neural Systems,* 16**,** 239-260.

DE VALOIS, R. L., ALBRECHT, D. G. & THORELL, L. G. 1978. Cortical cells: bar and edge detectors, or spatial frequency filters? *Frontiers in visual science.* Springer.

DE VALOIS, R. L., ALBRECHT, D. G. & THORELL, L. G. 1982. Spatial frequency selectivity of cells in macaque visual cortex. *Vision research,* 22**,** 545-59.

DEAN, P. 1976. Effects of inferotemporal lesions on the behavior of monkeys. *Psychological bulletin,* 83**,** 41.

DEVALOIS, R. L. & BOTH PROFESSORS, B. K. K. D. 1988. *Spatial vision*, Oxford University Press.

DEVELLIS, R. F. 2012. Factor Analysis. *Scale Development Theory and Applications.* 3rd ed. University of North Carolina, Chapel Hill, USA: SAGE Publications, Inc.

DIELEMAN, S. 2015. *Classifying Plankton with Deep Neural Networks* [Online]. Available: http://benanne.github.io/2015/03/17/plankton.html.

DOBBINS, A., ZUCKER, S. W. & CYNADER, M. S. 1987. Endstopped neurons in the visual cortex as a substrate for calculating curvature. *Nature,* 329**,** 438-41.

DOW, S., KULKARNI, A., KLEMMER, S. & HARTMANN, B. Shepherding the Crowd Yields Better Work. ACM 2012 conference on Computer Supported Cooperative Work 2012 New York. ACM, 1013-1022.

DREHER, B. 1972. Hypercomplex cells in the cat's striate cortex. *Investigative ophthalmology,* 11**,** 355-6.

EL-SHAMAYLEH, Y. & MOVSHON, J. A. 2011. Neuronal responses to texture-defined form in macaque visual area V2. *The Journal of Neuroscience,* 31**,** 8543-8555.

EMERSON, R. C., BERGEN, J. R. & ADELSON, E. H. 1992. Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision research,* 32**,** 203-218.

EVANS, R. M. & BARTLEY, S. H. 1948. Introduction to color. *American Journal of Physics,* 16**,** 491-491.

FABRIGAR, L. R., WEGENER, D. T., MACCALLUM, R. C. & STRAHAN, E. J. 1999. Evaluating the use of exploratory Factor analysis in psychological research. *Psychological methods,* 4**,** 272.

FIELD, D. 1994. What is the goal of sensory coding? *Neural computation,* 6**,** 559-601.

FIELD, D. J. 1987. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America. A, Optics and image science,* 4**,** 2379-94.

FLEMING, R. W. 2014. Visual perception of materials and their properties. *Vision research,* 94**,** 62-75.

FLEMING, R. W. & BÜLTHOFF, H. H. 2005. Low-level image cues in the perception of translucent materials. *ACM Transactions on Applied Perception (TAP),* 2**,** 346-382.

FLEMING, R. W., DROR, R. O. & ADELSON, E. H. 2003. Real-world illumination and the perception of surface reflectance properties. *Journal of Vision,* 3**,** 3.

FLEMING, R. W., JÄKEL, F. & MALONEY, L. T. 2011. Visual perception of thick transparent materials. *Psychological science,* 22**,** 812-820.

FLEMING, R. W., WIEBEL, C. & GEGENFURTNER, K. 2013. Perceptual qualities and material classes. *Journal of vision,* 13**,** 9.

FOLDIAK, P. Y., M. P. 1995. Sparse Coding in the Primate Cortex. *In:* ARBIB, M. A. (ed.) *The Handbook of Brain Theory and Neural Networks.* Cambridge, MA: MIT Press.

FRANZ, M. O. & SCHÖLKOPF, B. 2005. Implicit Volterra and Wiener series for higher-order image analysis. *In:* SAUL, L. K., WEISS, Y. & BOTTOU, L. (eds.) *Advances in neural information processing systems.* Cambridge MA: MIT Press.

FREEMAN, J., ZIEMBA, C. M., HEEGER, D. J., SIMONCELLI, E. P. & MOVSHON, J. A. 2013. A functional and perceptual signature of the second visual area in primates. *Nature neuroscience,* 16**,** 974-981.

FREEMAN, R. D., OHZAWA, I. & WALKER, G. 2001. Beyond the classical receptive field in the visual cortex. *Progress in brain research,* 134**,** 157-170.

GABOR, D. 1946. Theory of communication. Part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering,* 93**,** 429-441.

GEISLER, W. S. 2008. Visual perception and the statistical properties of natural scenes. *Annu. Rev. Psychol.,* 59**,** 167-192.

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. *Cereb Cortex,* 18**,** 2416-38.

GIBSON, J. J. 1950. The perception of the visual world.

GIBSON, J. J. 2013. The ecological approach to visual perception.

GILBERT, C. D., DAS, A., ITO, M., KAPADIA, M. & WESTHEIMER, G. 1996. Spatial integration and cortical dynamics. *Proceedings of the National Academy of Sciences of the United States of America,* 93**,** 615-22.

GILBERT, E. N. 1980. Random Colorings of a Lattice of Squares in the Plane. *SIAM Journal on Algebraic and Discrete Methods,* 1**,** 152-159.

GOODALE, M. A. & MILNER, A. D. 1992. Separate visual pathways for perception and action. *Trends in neurosciences,* 15**,** 20-25.

GORITZ, A. S., WOLFF, H. G. & GOLDSTEIN, D. G. 2008. Individual payments as a longer-term incentive in online panels. *Behavior Research Methods,* 40.

GOSLING, S. D., VAZIRE, S., SRIVASTAVA, S. & JOHN, O. P. 2004. Should we trust Web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist,* 59**,** 93–104.

GRAHAM, N., BECK, J. & SUTTER, A. 1992. Nonlinear processes in spatial-frequency channel models of perceived texture segregation: Effects of sign and amount of contrast. *Vision research,* 32**,** 719-743.

GRAHAM, N. & SUTTER, A. 1996. Effect of spatial scale and background luminance on the intensive and spatial nonlinearities in texture segregation. *Vision Research,* 36**,** 1371-1390.

GRAHAM, N. & SUTTER, A. 1998. Spatial summation in simple (Fourier) and complex (non-Fourier) texture channels. *Vision research,* 38**,** 231-257.

GRAHAM, N. & SUTTER, A. 2000. Normalization: Contrast-gain control in simple (Fourier) and complex (non-Fourier) pathways of pattern vision. *Vision Research,* 40**,** 2737-2761.

GRIFFITH, D. A. 2003. *Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization*, Springer Science & Business Media.

HAIDER, B., KRAUSE, M. R., DUQUE, A., YU, Y., TOURYAN, J., MAZER, J. A. & MCCORMICK, D. A. 2010. Synaptic and network mechanisms of sparse and reliable visual cortical activity during nonclassical receptive field stimulation. *Neuron,* 65**,** 107-21.

HANAZAWA, A. & KOMATSU, H. 2001. Influence of the direction of elemental luminance gradients on the responses of V4 cells to textured surfaces. *J Neurosci,* 21**,** 4490-7.

HARALICK, R. M. 1979. Statistical and structural approaches to texture. *Proceedings of the IEEE,* 67**,** 786-804.

HARRIS, K. D. & THIELE, A. 2011. Cortical state and attention. *Nature reviews neuroscience,* 12**,** 509-523.

HARTLINE, H. K. 1938. The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *American Journal of Physiology,* 121**,** 400-415.

HAUSSER, M., SPRUSTON, N. & STUART, G. J. 2000. Diversity and dynamics of dendritic signaling. *Science,* 290**,** 739-44.

HAYTON, J. C., ALLEN, D. G. & SCARPELLO, V. 2004. Factor Retention Decisions in Exploratory Factor Analysis: a Tutorial on Parallel Analysis. *Organizational Research Methods,* 7.

HE, K., ZHANG, X., REN, S. & SUN, J. 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *arXiv preprint arXiv:1502.01852*.

HEEGER, D. J. 1992a. Half-squaring in responses of cat striate cells. *Visual neuroscience,* 9**,** 427-443.

HEEGER, D. J. 1992b. Normalization of cell responses in cat striate cortex. *Visual neuroscience,* 9**,** 181-197.

HEEGER, D. J. & BERGEN, J. R. Pyramid-based texture analysis/synthesis. Proceedings of the 22nd annual conference on Computer graphics and interactive techniques, 1995. ACM, 229-238.

HEER, J., BOSTOCK, M. & 2010. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. *ACM Human Factors in Computing Systems (CHI)***,** 203-212.

HEGDÉ, J. & VAN ESSEN, D. C. 2000. Selectivity for complex shapes in primate visual area V2. *J Neurosci,* 20**,** 61-66.

HEILBRONNER, R. & BARRETT, S. 2013. *Image analysis in earth sciences: microstructures and textures of earth materials*, Springer Science & Business Media.

HEINEMANN, E. G. 1955. Simultaneous brightness induction as a function of inducing-and test-field luminances. *Journal of experimental psychology,* 50**,** 89.

HEISENBERG, M. & WOLF, R. 1984. *Vision in Drosophila. Genetics of microbehaviour*, Springer Verlag.

HO, Y.-X., LANDY, M. S. & MALONEY, L. T. 2008. Conjoint measurement of gloss and surface texture. *Psychological Science,* 19**,** 196-204.

HO, Y., LANDY, M. & MALONEY, L. 2006. How illuminant direction affects perceived visual roughness. *J Vision,* 6**,** 634-648.

HOF, P. R., VOGT, B. A., BOURAS, C. & MORRISON, J. H. 1997. Atypical form of Alzheimer's disease with prominent posterior cortical atrophy: a review of lesion distribution and circuit disconnection in cortical visual pathways. *Vision Res,* 37**,** 3609-25.

HORTON, J. J., RAND, D. G. & ZECKHAUSER, R. J. 2011. The online laboratory: conducting experiments in a real labor market. *Experimental Economics,* 14**,** 399-425.

HOTELLING, H. 1933. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology,* 24**,** 417.

HUBEL, D. H. 1988. *Eye, brain, and vision*, Scientific American Library New York.

HUBEL, D. H. & WIESEL, T. N. 1959. Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology,* 148**,** 574-591.

HUBEL, D. H. & WIESEL, T. N. 1961. Integrative action in the cat's lateral geniculate body. *The Journal of physiology,* 155**,** 385-98.

HUBEL, D. H. & WIESEL, T. N. 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology,* 160**,** 106-54.

HUBEL, D. H. & WIESEL, T. N. 1965. Receptive Fields and Functional Architecture in Two Nonstriate Visual Areas (18 and 19) of the Cat. *Journal of neurophysiology,* 28**,** 229-89.

HUBEL, D. H. & WIESEL, T. N. 1968. Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology,* 195**,** 215-43.

HURRI, J. 1997. Independent component analysis of image data. *Master's thesis, Dept. of Computer Science and Engineering, Helsinki University of Technology, Espoo, Finland*.

HURRI, J., HYVÄRINEN, A., KARHUNEN, J. & OJA, E. Image feature extraction using independent component analysis. In Proc. NORSIG'96, 1996. Citeseer.

HYVÄRINEN, A. & OJA, E. 2000. Independent component analysis: algorithms and applications. *Neural networks,* 13**,** 411-430.

IBBOTSON, M. R., PRICE, N. & CROWDER, N. A. 2005. On the division of cortical cells into simple and complex types: a comparative viewpoint. *Journal of neurophysiology,* 93**,** 3699-3702.

IMAGENET. 2012. *Large Scale Visual Recognition Challenge 2012 (ILSVRC2012)* [Online]. Available: http://www.image-net.org/challenges/LSVRC/2012/ 2015].

IOFFE, S. & SZEGEDY, C. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.

IPEIROTIS, P. 2010a. Analyzing the Amazon Mechanical Turk Marketplace. *ACM XRDS (Crossroads),* 17.

IPEIROTIS, P. 2010b. Demographics of Mechanical Turk. *Working Paper CeDER-10-01.* NYU Center for Digital Economy Research.

IPEIROTIS, P. G. 2010c. Analyzing the amazon mechanical turk marketplace. *XRDS: Crossroads, The ACM Magazine for Students,* 17**,** 16-21.

JACOBSON, L. & WECHSLER, H. 1984. Invariant analogical image representation and pattern recognition. *Pattern recognition letters,* 2**,** 289-299.

JAIN, A., ROSS, A. & PRABHAKAR, S. Fingerprint matching using minutiae and texture features. Image Processing, 2001. Proceedings. 2001 International Conference on, 2001. IEEE, 282-285.

JEONG, D. H., ZIEMKIEWICZ, C., RIBARSKY, W., CHANG, R. & CENTER, C. V. 2009. Understanding principal component analysis using a visual analytics tool. *Charlotte Visualization Center, UNC Charlotte.*

JONES, J. P. & PALMER, L. A. 1987. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of neurophysiology,* 58**,** 1233-1258.

JULESZ, B. 1962. Visual pattern discrimination. *Information Theory, IRE Transactions on,* 8**,** 84-92.

JULESZ, B. 1975. Experiments in the visual perception of texture. *Scientific American,* 232**,** 34-43.

JULESZ, B., GILBERT, E. N., SHEPP, L. A. & FRISCH, H. L. 1973. Inability of humans to discriminate between visual textures that agree in second-order statistics-revisited. *Perception,* 2**,** 391-405.

JULESZ, B., GILBERT, E. N. & VICTOR, J. D. 1978. Visual discrimination of textures with identical third-order statistics. *Biological cybernetics,* 31**,** 137-40.

KASTNER, S., DE WEERD, P. & UNGERLEIDER, L. G. 2000. Texture segregation in the human visual cortex: a functional MRI study. *Journal of Neurophysiology,* 83**,** 2453-2457.

KERSTEN, D. 1987. Predictability and redundancy of natural images. *JOSA A,* 4**,** 2395-2400.

KINGDOM, F. A. 2011. Lightness, brightness and transparency: a quarter century of new ideas, captivating demonstrations and unrelenting controversy. *Vision Res,* 51**,** 652-73.

KIRK, R. E. 2007. *Statistics: An Introduction*, Cengage Learning.

KITTUR, A., CHI, E. H. & SUH, B. Crowdsourcing User Studies with Mechanical Turk. Conference on Human Factors in Computing Systems, 2008. 453-456.

KNILL, D. C. 1998. Ideal observer perturbation analysis reveals human strategies for inferring surface orientation from texture. *Vision research,* 38**,** 2635-2656.

KOENDERINK, J. J. Shape and shading.

KOURTZI, Z. & KANWISHER, N. 2000. Cortical regions involved in perceiving object shape. *The Journal of Neuroscience,* 20**,** 3310-3318.

KRIZHEVSKY, A., SUTSKEVER, I. & HINTON, G. E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012. 1097-1105.

KUFFLER, S. W. 1953. Discharge patterns and functional organization of mammalian retina. *Journal of neurophysiology,* 16**,** 37-68.

LANDY, M. S. & GRAHAM, N. 2004. Visual perception of texture. *The visual neurosciences,* 2**,** 1106-1118.

LANDY, M. S. & KOJIMA, H. 2001. Ideal cue combination for localizing texture-defined edges. *JOSA A,* 18**,** 2307-2320.

LANGLOIS, D., CHARTIER, S. & GOSSELIN, D. 2010. An introduction to independent component analysis: InfoMax and FastICA algorithms. *Tutorials in Quantitative Methods for Psychology,* 6**,** 31-38.

LAUGHLIN, S. 1981. A simple coding procedure enhances a neuron's information capacity. *Zeitschrift fur Naturforschung. Section C: Biosciences,* 36**,** 910-2.

LAWRENCE, S., GILES, C. L., TSOI, A. C. & BACK, A. D. 1997. Face recognition: A convolutional neural-network approach. *Neural Networks, IEEE Transactions on,* 8**,** 98-113.

LECUN, Y., BOTTOU, L., BENGIO, Y. & HAFFNER, P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE,* 86**,** 2278-2324.

LENNIE, P. 1998. Single units and visual cortical organization. *PERCEPTION-LONDON-,* 27**,** 889-936.

LEVITT, J. B. & LUND, J. S. 1997a. Contrast dependence of contextual effects in primate visual cortex. *Nature,* 387**,** 73-76.

LEVITT, J. B. & LUND, J. S. 1997b. Contrast dependence of contextual effects in primate visual cortex. *Nature,* 387**,** 73-6.

LONG, J. L., ZHANG, N. & DARRELL, T. Do Convnets Learn Correspondence? Advances in Neural Information Processing Systems, 2014. 1601-1609.

LU, C. S., CHUNG, P. C. & CHEN, C. F. 1997. Unsupervised texture segmentation via wavelet transform. *Pattern Recognition,* 30**,** 729-742.

MA, L., WANG, Y. & TAN, T. Iris recognition using circular symmetric filters. Pattern Recognition, 2002. Proceedings. 16th International Conference on, 2002. IEEE, 414-417.

MACMILLAN, N. A. & CREELMAN, C. D. 1991. *Detection theory: A User's Guide,* Cambridge, Cambridge University Press.

MADDESS, T. 2015. Histogram of Primitive Analysis of Deterministic and Stochastic Ternary Textures. *In Preparation*.

MADDESS, T., DAVEY, M. & YANG, E. 1999. Discrimination of complex textures by bees. *Journal of Comparative Physiology A,* 184**,** 107-117.

MADDESS, T. & KULIKOWSKI, J. J. 1999. Apparent fineness of stationary compound gratings. *Vision research,* 39**,** 3404-16.

MADDESS, T., MCCOURT, M., BLAKESLEE, B. & CUNNINGHAM, R. 1988. Factors governing the adaptation of cells in area-17 of the cat visual cortex. *Biological cybernetics,* 59**,** 229-236.

MADDESS, T. & NAGAI, Y. 2001. Discriminating isotrigon textures [corrected]. *Vision research,* 41**,** 3837-60.

MADDESS, T., NAGAI, Y., JAMES, A. C. & ANKIEWCZ, A. 2004. Binary and ternary textures containing higher-order spatial correlations. *Vision research,* 44**,** 1093-113.

MADDESS, T., NAGAI, Y., VICTOR, J. D. & TAYLOR, R. R. 2007. Multilevel isotrigon textures. *Journal of the Optical Society of America. A, Optics, image science, and vision,* 24**,** 278-93.

MAFFEI, L. & FIORENTINI, A. 1973. The visual cortex as a spatial frequency analyser. *Vision research,* 13**,** 1255-1267.

MAHENDRAN, A. & VEDALDI, A. 2014. Understanding deep image representations by inverting them. *arXiv preprint arXiv:1412.0035*.

MAHON, L. & DE VALOIS, R. 2001. Cartesian and non-Cartesian responses in LGN, V1, and V2 cells. *Visual neuroscience,* 18**,** 973-981.

MANTE, V. & CARANDINI, M. 2005. Mapping of stimulus energy in primary visual cortex. *Journal of neurophysiology,* 94**,** 788-798.

MARČELJA, S. 1980. Mathematical description of the responses of simple cortical cells*. *JOSA,* 70**,** 1297-1300.

MARESCHAL, I. & SHAPLEY, R. M. 2004. Effects of contrast and size on orientation discrimination. *Vision research,* 44**,** 57-67.

MARLOW, P. J., KIM, J. & ANDERSON, B. L. 2012. The perception and misperception of specular surface reflectance. *Current Biology,* 22**,** 1909-1913.

MARTIN, K. A. & WHITTERIDGE, D. 1984. Form, function and intracortical projections of spiny neurones in the striate visual cortex of the cat. *The Journal of physiology,* 353**,** 463-504.

MARTINEZ, L. M. & ALONSO, J.-M. 2001. Construction of complex receptive fields in cat primary visual cortex. *Neuron,* 32**,** 515-525.

MARTINEZ, R. 2015. *Cycles Material Studies (Blender)* [Online]. Available: http://www.reynantemartinez.com/cycles-material-studies.html.

MASON, W. & SURI, S. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behav Res Methods,* 44**,** 1-23.

MASON, W. A. & WATTS, D. J. 2009. Financial incentives and the performance of crowds. *Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP '10)***,** 77–85.

MATERKA, A. & STRZELECKI, M. 1998. Texture analysis methods–a review. *Technical university of lodz, institute of electronics, COST B11 report, Brussels*, 9-11.

MAUNSELL, J. H. & NEWSOME, W. T. 1987. Visual processing in monkey extrastriate cortex. *Annual review of neuroscience,* 10**,** 363-401.

MCCREADIE, R. M. C., MACDONALD, C. & OUNIS, I. 2010. Crowdsourcing a news query classification data set. *Proceedings of the ACM SIGIR 2010 Workshop on Crowdsourcing for Search Evaluation (CSE 2010)***,** 31-38.

MECHLER, F. & RINGACH, D. L. 2002. On the classification of simple and complex cells. *Vision research,* 42**,** 1017-1033.

MEGLEN, R. R. 1992. Examining large databases: a chemometric approach using principal component analysis. *Marine Chemistry,* 39**,** 217-237.

MEL, B. W. 1993. Synaptic integration in an excitable dendritic tree. *Journal of neurophysiology,* 70**,** 1086-101.

MICHIELSEN, K. & DE RAEDT, H. 2001. Integral-geometry morphological image analysis. *Physics Reports,* 347**,** 461-538.

MISHKIN, M., UNGERLEIDER, L. G. & MACKO, K. A. 1983. Object vision and spatial vision: two cortical pathways. *Trends in neurosciences,* 6**,** 414-417.

MITROFF, S. R., BIGGS, A. T., ADAMO, S. H., DOWD, E. W., WINKLE, J. & CLARK, K. 2015. What can 1 billion trials tell us about visual search? *Journal of experimental psychology: human perception and performance,* 41**,** 1.

MOTOYOSHI, I., NISHIDA, S. Y., SHARAN, L. & ADELSON, E. H. 2007. Image statistics and the perception of surface qualities. *Nature,* 447**,** 206-209.

MOVSHON, J. A., THOMPSON, I. D. & TOLHURST, D. J. 1978a. Receptive field organization of complex cells in the cat's striate cortex. *The Journal of physiology,* 283**,** 79-99.

MOVSHON, J. A., THOMPSON, I. D. & TOLHURST, D. J. 1978b. Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *The Journal of physiology,* 283**,** 53-77.

MURPHY, P. C. & SILLITO, A. M. 1987. Corticofugal feedback influences the generation of length tuning in the visual pathway. *Nature,* 329**,** 727-9.

MURRAY, S. O., OLSHAUSEN, B. A. & WOODS, D. L. 2003. Processing shape, motion and three-dimensional shape-from-motion in the human cortex. *Cerebral Cortex,* 13**,** 508-516.

NAGAI, Y., TAYLOR, R. R., LOH, Y. W. & MADDESS, T. 2009. Discrimination of complex form by simple oscillator networks. *Network,* 20**,** 233-52.

NISHIDA, S. Y. & SHINYA, M. 1998. Use of image-based information in judgments of surface-reflectance properties. *JOSA A,* 15**,** 2951-2965.

NORRIS, M. & LECAVALIER, L. 2010. Evaluating the use of exploratory Factor analysis in developmental disability psychological research. *J Autism Dev Disord,* 40**,** 8-20.

O'CARROLL, D. 1993. Feature-detecting neurons in dragonflies.

OLSHAUSEN, B. A. & FIELD, D. J. 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature,* 381**,** 607-9.

OLSHAUSEN, B. A. & FIELD, D. J. 1997. Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vision research,* 37**,** 3311-25.

OLSHAUSEN, B. A. & FIELD, D. J. 2004. Sparse coding of sensory inputs. *Current opinion in neurobiology,* 14**,** 481-7.

OPPENHEIM, A. V. & LIM, J. S. 1981. The importance of phase in signals. *Proceedings of the IEEE,* 69**,** 529-541.

OPPENHEIMER, D. M., MEYVIS, T. & DAVIDENKO, N. 2009. Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology,* 45**,** 867-872.

OQUAB, M., BOTTOU, L., LAPTEV, I. & SIVIC, J. Learning and transferring mid-level image representations using convolutional neural networks. Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, 2014. IEEE, 1717-1724.

ORBAN, G. A. 2008. Higher order visual processing in macaque extrastriate cortex. *Physiological Reviews,* 88**,** 59-89.

OZEKI, H., FINN, I. M., SCHAFFER, E. S., MILLER, K. D. & FERSTER, D. 2009. Inhibitory stabilization of the cortical network underlies visual surround suppression. *Neuron,* 62**,** 578-92.

PADILLA, S., DRBOHLAV, O., GREEN, P. R., SPENCE, A. & CHANTLER, M. J. 2008. Perceived roughness of 1/< i> fβ</i> noise surfaces. *Vision research,* 48**,** 1791-1797.

PALMER, L. A., JONES, J. P. & MULLIKIN, W. H. 1985. *Functional organization of simple cell receptive fields*, Wiley, New York.

PAOLACCI, G., CHANDLER, J. & IPEIROTIS, P. 2010. Running experiments on Amazon Mechanical Turk. *Judgment and decision making,* 5**,** 411-419.

PASUPATHY, A. 2006. Neural basis of shape representation in the primate brain. *Prog Brain Res,* 154**,** 293-313.

PEARSON, K. 1901. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science,* 2**,** 559-572.

PETERZELL, D. H. & TELLER, D. Y. 1996. Individual differences in contrast sensitivity functions: the lowest spatial frequency channels. *Vision Res,* 36**,** 3077-85.

PETERZELL, D. H., WERNER, J. S. & KAPLAN, P. S. 1993. Individual differences in contrast sensitivity functions: the first four months of life in humans. *Vision Res,* 33**,** 381-96.

PIZLO, Z. 2001. Perception viewed as an inverse problem. *Vision Research,* 41**,** 3145-3161.

POGGIO, T., TORRE, V. & KOCH, C. 1985. Computational vision and regularization theory. *NATURE,* 317**,** 26.

POLAT, U. & SAGI, D. 1993. Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vision research,* 33**,** 993-999.

PONT, S. C. & KOENDERINK, J. J. 2008. Shape, surface roughness, and human perception. World Scientific.

PONTIN, J. 2007. Artificial intelligence, with help from the humans. *New York Times*, 25 March.

PORTILLA, J. & SIMONCELLI, E. P. 2000. A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision,* 40**,** 49-70.

PRESS, W. H. 2007. *Numerical recipes 3rd edition: The art of scientific computing*, Cambridge university press.

PURPURA, K. P., VICTOR, J. D. & KATZ, E. 1994a. Striate cortex extracts higher-order spatial correlations from visual textures. *Proceedings of the National Academy of Sciences,* 91**,** 8482-8486.

PURPURA, K. P., VICTOR, J. D. & KATZ, E. 1994b. Striate cortex extracts higher-order spatial correlations from visual textures. *Proceedings of the National Academy of Sciences of the United States of America,* 91**,** 8482-6.

PURVES, D., AUGUSTINE, G. J., FITZPATRICK, D., HALL, W. C., LAMANTIA, A.-S., MCNAMARA, J. O. & WHITE, L. E. 2001. Neuroscience. *Sunderland (MA): Sinauer Associates*.

PURVES, D. & LOTTO, R. B. 2011. *Why we see what we do redux: A wholly empirical theory of vision*, Sinauer Associates Sunderland, MA.

PURVES, D., SHIMPI, A. & LOTTO, R. B. 1999. An empirical explanation of the cornsweet effect. *J Neurosci,* 19**,** 8542-51.

REDDY, K. R., BADAMI, S. & BALASUBRAMANIAN, V. 1994. *Oscillations and Waves*, Universities Press.

REGAN, D. & HONG, X. H. 1994. Recognition and detection of texture-defined letters. *Vision Res,* 34**,** 2403-7.

REGAN, D. & SIMPSON, T. 1995. Multiple sclerosis can cause visual processing deficits specific to texture-defined form. *Neurology,* 45**,** 809-15.

REICHARDT, W. & ROSENBLITH, W. Autocorrelation, a principle for evaluation of sensory information by the central nervous system. Symposium on Principles of Sensory Communication 1959, 1961. MIT Press, 303-317.

REYMENT, R. & JORESKOG, K. 1996. *Applied Factor Analysis in the Natural Sciences,* Cambridge, UK, Cambridge University Press.

REYMENT, R. A., JORESKOG, K.G. 1996. Applied Factor Analysis in the Natural Sciences. Cambridge University Press.

RIVEST, J. & CABANAGH, P. 1996. Localizing contours defined by more than one attribute. *Vision research,* 36**,** 53-66.

RIZZO, M., ANDERSON, S. W., DAWSON, J. & NAWROT, M. 2000. Vision and cognition in Alzheimer's disease. *Neuropsychologia,* 38**,** 1157-69.

ROMEIRO, F. & ZICKLER, T. 2010. Inferring reflectance under real-world illumination. Technical Report No. TR-10-10). Cambridge, MA: Harvard School of Engineering and Applied Sciences.

ROSENBLATT, M. & SLEPIAN, D. 1962. N th Order Markov Chains with Every N Variables Independent. *Journal of the Society for Industrial & Applied Mathematics,* 10**,** 537-549.

ROSLI, Y., BEDFORD, S. M. & MADDESS, T. 2009. Low-spatial-frequency channels and the spatial frequency-doubling illusion. *Investigative ophthalmology & visual science,* 50**,** 1956-63.

ROSS, J., IRANI, I., SILBERMAN, M., ZALDIVAR, A. & TOMLINSON, B. 2010. Who are the CrowdWorkers? Shifting Demographics in Amazon Mechanical Turk. *CHI EA 2010***,** 2863-2872.

RUBIN, N. 2001. The role of junctions in surface completion and contour matching. *PERCEPTION-LONDON-,* 30**,** 339-366.

RUSSAKOVSKY, O., DENG, J., SU, H., KRAUSE, J., SATHEESH, S., MA, S., HUANG, Z., KARPATHY, A., KHOSLA, A. & BERNSTEIN, M. 2014. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 1-42.

SANDLER, R. & LINDENBAUM, M. Gabor filter analysis for texture segmentation. Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on, 2006. IEEE, 178-178.

SCHRODER-TURK, G. E., MICKEL, W., KAPFER, S. C., KLATT, M. A., SCHALLER, F. M., HOFFMANN, M. J., KLEPPMANN, N., ARMSTRONG, P., INAYAT, A., HUG, D., REICHELSDORFER, M., PEUKERT, W., SCHWIEGER, W. & MECKE, K. 2011. Minkowski tensor shape analysis of cellular, granular and porous structures. *Advanced materials,* 23**,** 2535-53.

SCHRODER-TURK, G. E., MICKEL, W., KAPFER, S. C. 2010. Minkowski Tensors of Anisotropic Spatial Structure. *arXiv,* 1009.

SCHWARTZ, E. L. 1980. Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to perceptual coding. *Vision research,* 20**,** 645-669.

SEAMONS, J. W., BARBOSA, M. S., BUBNA-LITIC, A. & MADDESS, T. 2015a. A lower bound on the number of mechanisms for discriminating fourth and higher order spatial correlations. *Vision research,* 108**,** 41-48.

SEAMONS, J. W. G., BARBOSA, M. S., COY, D. & MADDESS, T. 2015a. Developing and Validating an Isotrigon Texture Discrimination Task using Crowdsourcing. Australian National University (In Preparation).

SEKULAR, R. & BLAKE, R. 1994. Perception (3rd edn). New York: McGraw-Hill.

SEKULER, R., WILSON, H. R. & OWSLEY, C. 1984. Structural modeling of spatial vision. *Vision research,* 24**,** 689-700.

SERENO, M. I., DALE, A., REPPAS, J., KWONG, K., BELLIVEAU, J., BRADY, T., ROSEN, B. & TOOTELL, R. 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science,* 268**,** 889-893.

SHANNON, C. E. 2001. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review,* 5**,** 3-55.

SHANNON, C. E. & WEAVER, W. 1948. *The Mathematical Theory of Communication*, University of Illinois Press.

SHAPLEY, R. & ENROTH-CUGELL, C. 1984. Visual adaptation and retinal gain controls. *Progress in retinal research,* 3**,** 263-346.

SHARAN, L., ROSENHOLTZ, R. & ADELSON, E. 2009. Material perception: What can you see in a brief glance? *Journal of Vision,* 9**,** 784-784.

SHARAN, L., ROSENHOLTZ, R. & ADELSON, E. H. 2008. Eye movements for shape and material perception. *Journal of Vision,* 8**,** 219-219.

SHEPHERD, G. M. 1998. *The synaptic organization of the brain*, Oxford University Press New York.

SHIKATA, E., HAMZEI, F., GLAUCHE, V., KNAB, R., DETTMERS, C., WEILLER, C. & BÜCHEL, C. 2001. Surface orientation discrimination activates caudal and anterior intraparietal sulcus in humans: an event-related fMRI study. *Journal of Neurophysiology,* 85**,** 1309-1314.

SIMONCELLI, E. P. & OLSHAUSEN, B. A. 2001. Natural image statistics and neural representation. *Annual review of neuroscience,* 24**,** 1193-216.

SIMONYAN, K. & ZISSERMAN, A. Two-stream convolutional networks for action recognition in videos. Advances in Neural Information Processing Systems, 2014. 568-576.

SIMPSON, W. A. & MCFADDEN, S. M. 2005. Spatial frequency channels derived from individual differences. *Vision research,* 45**,** 2723-7.

SKITKA, L. J. & SARGIS, E. G. 2006. The internet as psychological laboratory. *Annu Rev Psychol,* 57**,** 529-55.

SNOWDEN, R., SNOWDEN, R. J., THOMPSON, P. & TROSCIANKO, T. 2012. *Basic vision: an introduction to visual perception*, Oxford University Press.

SOLOMON, S. G., LEE, B. B. & SUN, H. 2006. Suppressive surrounds and contrast gain in magnocellular-pathway retinal ganglion cells of macaque. *The Journal of neuroscience : the official journal of the Society for Neuroscience,* 26**,** 8715-26.

SRINIVASAN, G. & SHOBHA, G. Statistical texture analysis. Proceedings of world academy of science, engineering and technology, 2008. 1264-1269.

SRINIVASAN, M., ZHANG, S. & WITNEY, K. 1994. Visual discrimination of pattern orientation by honeybees: Performance and implications forcortical'processing. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 343**,** 199-210.

SRINIVASAN, M. V., LAUGHLIN, S. B. & DUBS, A. 1982a. Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London B,* 216**,** 427-59.

SRINIVASAN, M. V., LAUGHLIN, S. B. & DUBS, A. 1982b. Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London. Series B, Containing papers of a Biological character. Royal Society,* 216**,** 427-59.

STONE, J. V. 2004. *Independent component analysis*, Wiley Online Library.

STONE, M. & BARTRAM, L. 2008. Alpha, Contrast and the Perception of Visual Metadata. *16th Color and Imaging Conference.* Society for Imaging Science and Technology.

STRZELECKI, M. & MATERKA, A. Markov random fields as models of textured biomedical images. Proc. 20th National Conf. Circuit Theory and Electronic Networks KTOiUE'97, 1997. 493-498.

STUART, G., SCHILLER, J. & SAKMANN, B. 1997. Action potential initiation and propagation in rat neocortical pyramidal neurons. *Journal of physiology,* 505 (Pt 3)**,** 617-32.

SUHR, D. D. 2005. Principal component analysis vs. exploratory Factor analysis. *SUGI 30 Proceedings***,** 203-230.

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**,** 959-66.

TAO, L., SHELLEY, M., MCLAUGHLIN, D. & SHAPLEY, R. 2004. An egalitarian network model for the emergence of simple and complex cells in visual cortex. *Proceedings of the National Academy of Sciences,* 101**,** 366-371.

TAYLOR, R. R. 2008. *Neural Computation of Statistical Structure.* PhD, The Australian National University.

TAYLOR, R. R. 2013. *Neural Computation of Statistical Structure. Chapter 4: A Greater Purpose: Computation of Statistical Structure within Pyramidal Cells.* PhD Thesis, ANU.

TAYLOR, R. R., MADDESS, T. & NAGAI, Y. 2008. Spatial biases and computational constraints on the encoding of complex local image structure. *Journal of vision,* 8**,** 19 1-13.

TEICHERT, T., WACHTLER, T., MICHLER, F., GAIL, A. & ECKHORN, R. 2007. Scale-invariance of receptive field properties in primary visual cortex. *BMC neuroscience,* 8**,** 38.

THOMPSON, W., FLEMING, R., CREEM-REGEHR, S. & STEFANUCCI, J. K. 2011. *Visual perception from a computer graphics perspective*, CRC Press.

TKACIK, G., PRENTICE, J. S., VICTOR, J. D. & BALASUBRAMANIAN, V. 2010. Local statistics in natural scenes predict the saliency of synthetic textures. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 18149-54.

TODD, J. T. 2004. The visual perception of 3D shape. *Trends Cogn Sci,* 8**,** 115-21.

TODD, J. T., OOMES, A. H., KOENDERINK, J. J. & KAPPERS, A. M. 2004. The perception of doubly curved surfaces from anisotropic textures. *Psychol Sci,* 15**,** 40-6.

TOOTELL, R. B., SWITKES, E., SILVERMAN, M. S. & HAMILTON, S. L. 1988. Functional anatomy of macaque striate cortex. II. Retinotopic organization. *The Journal of Neuroscience,* 8**,** 1531-1568.

TOYAMA, K., KIMURA, M. & TANAKA, K. 1981a. Cross-Correlation Analysis of Interneuronal Connectivity in cat visual cortex. *Journal of neurophysiology,* 46**,** 191-201.

TOYAMA, K., KIMURA, M. & TANAKA, K. 1981b. Organization of cat visual cortex as investigated by cross-correlation technique. *Journal of neurophysiology,* 46**,** 202-14.

TURANO, K. & PANTLE, A. 1989. On the mechanism that encodes the movement of contrast variations: velocity discrimination. *Vision research,* 29**,** 207-221.

UNGERLEIDER, L. & MISHKIN, M. Two cortical visual systems, Ingle DJ, Goodale MA, Mansfield RJW, Analysis of visual behavior, 1982, 549-586. MIT Press, Cambridge, MA.

VAN ESSEN, D. C., ANDERSON, C. H. & FELLEMAN, D. J. 1992. Information processing in the primate visual system: an integrated systems perspective. *Science,* 255**,** 419-423.

VAN HATEREN, J. 1992a. Real and optimal neural images in early vision. *Nature*.

VAN HATEREN, J. 1992b. Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation. *Journal of Comparative Physiology A,* 171**,** 157-170.

VAN HATEREN, J. 1992c. A theory of maximizing sensory information. *Biological cybernetics,* 68**,** 23-29.

VAN HATEREN, J. H. & VAN DER SCHAAF, A. 1998. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society of London. Series B: Biological Sciences,* 265**,** 359-366.

VAN KLEEF, J. P., CLOHERTY, S. L. & IBBOTSON, M. R. 2010. Complex cell receptive fields: evidence for a hierarchical mechanism. *The Journal of physiology,* 588**,** 3457-3470.

VAN SANTEN, J. P. & SPERLING, G. 1985. Elaborated reichardt detectors. *JOSA A, 2***,** 300-320.

VANDUFFEL, W., FIZE, D., PEUSKENS, H., DENYS, K., SUNAERT, S., TODD, J. & ORBAN, G. 2002. Extracting 3D from motion: differences in human and monkey intraparietal cortex. *Science,* 298**,** 413-415.

VAZ, S., FALKMER, T., PASSMORE, A. E., PARSONS, R. & ANDREOU, P. 2013. The case for using the repeatability coefficient when calculating test-retest reliability. *PLoS One,* 8**,** e73990.

VERMA, A. K., SRIVIDYA, A. & KARANKI, D. R. 2010. Basic Reliability Mathematics. *Reliability and Safety Engineering***,** 15-70.

VICTOR, J. D. 1985. Complex visual textures as a tool for studying the VEP. *Vision research,* 25**,** 1811-27.

VICTOR, J. D. 1988. Models for preattentive texture discrimination: Fourier analysis and local feature processing in a unified framework. *Spatial vision*.

VICTOR, J. D. 1994. Images, Statistics, and Textures: Implications of Triple Correlation Uniqueness for Texture Statistics and the Julesz Conjecture. *Journal of the Optical Society of America A,* 11**,** 1680-1684.

VICTOR, J. D. 1995. Isodipole Textures: A Window on Cortical Mechanisms of Form Processing. *In:* PAPATHOMAS, T. V. C., C.; GOREA, A; KOWLER, E (ed.) *Early Vision and Beyond.* The MIT Press.

VICTOR, J. D. & BRODIE, S. E. 1978. Discriminable textures with identical Buffon needle statistics. *Biological Cybernetics,* 31**,** 231-234.

VICTOR, J. D., CHUBB, C. & CONTE, M. M. 2005a. Interaction of luminance and higher-order statistics in texture discrimination. *Vision Res,* 45**,** 311-28.

VICTOR, J. D., CHUBB, C. & CONTE, M. M. 2005b. Interaction of luminance and higher-order statistics in texture discrimination. *Vision research,* 45**,** 311-328.

VICTOR, J. D. & CONTE, M. M. 1989. Cortical interactions in texture processing: scale and dynamics. *Vis Neurosci,* 2**,** 297-313.

VICTOR, J. D. & CONTE, M. M. 1991. Spatial organization of nonlinear interactions in form perception. *Vision research,* 31**,** 1457-88.

VICTOR, J. D. & CONTE, M. M. 1996. The role of high-order phase correlations in texture processing. *Vision Res,* 36**,** 1615-31.

VICTOR, J. D. & CONTE, M. M. 2005. Local processes and spatial pooling in texture and symmetry detection. *Vision research,* 45**,** 1063-73.

VICTOR, J. D. & CONTE, M. M. 2012a. Local image statistics: maximum-entropy constructions and perceptual salience. *Journal of the Optical Society of America. A, Optics, image science, and vision,* 29**,** 1313-45.

VICTOR, J. D. & CONTE, M. M. 2012b. Local image statistics: maximum-entropy constructions and perceptual salience. *J Opt Soc Am A Opt Image Sci Vis,* 29**,** 1313-45.

VICTOR, J. D. & KNIGHT, B. W. 2003. Simultaneously band and space limited functions in two dimensions, and receptive fields of visual neurons. *Perspectives and Problems in Nolinear Science.* Springer.

VICTOR, J. D., THENGONE, D. J. & CONTE, M. M. 2013. Perception of second- and third-order orientation signals and their interactions. *J Vis,* 13**,** 21.

VINJE, W. E. & GALLANT, J. L. 2000. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science,* 287**,** 1273-6.

VON GUNTEN, A., BOURAS, C., KOVARI, E., GIANNAKOPOULOS, P. & HOF, P. R. 2006. Neural substrates of cognitive and behavioral deficits in atypical Alzheimer's disease. *Brain Res Brain Res Rev,* 51**,** 176-211.

VON HELMHOLTZ, H. 1962. *Helmholtz's treatise on physiological optics*, Dover Publications.

WACHTLER, T., LEE, T.-W. & SEJNOWSKI, T. J. 2001. Chromatic structure of natural scenes. *JOSA A,* 18**,** 65-77.

WATSON, A. B. 1983. *Detection and recognition of simple spatial forms*, Springer.

WERKHOVEN, P., SPERLING, G. & CHUBB, C. 1993. The dimensionality of texture-defined motion: a single channel theory. *Vision research,* 33**,** 463-485.

WERKHOVEN, P., SPERLING, G. & CHUBB, C. 1994. Perception of apparent motion between dissimilar gratings: spatiotemporal properties. *Vision research,* 34**,** 2741-2759.

WHITE, M. 1979. A new effect of pattern on perceived lightness. *Perception,* 8**,** 413-416.

WIGNER, E. 1932. On the quantum correction for thermodynamic equilibrium. *Physical Review,* 40**,** 749.

WILLIAMS, S. M., MCCOY, A. N. & PURVES, D. 1998. The influence of depicted illumination on brightness. *Proceedings of the National Academy of Sciences,* 95**,** 13296-13300.

WOLFSON, S. S. & LANDY, M. S. 1995. Discrimination of orientation-defined texture edges. *Vision research,* 35**,** 2863-2877.

WU, R., YAN, S., SHAN, Y., DANG, Q. & SUN, G. 2015. Deep image: Scaling up image recognition. *arXiv preprint arXiv:1501.02876*.

YANG, E.-C. & MADDESS, T. 1997. Orientation-sensitive neurons in the brain of the honey bee (Apis mellifera). *Journal of Insect Physiology,* 43**,** 329-336.

YELLOTT, J. I. 1983. Spectral consequences of photorecptor sampling in the rhesus retina. *Science,* 221**,** 382-385.

YIN, Q., KIM, J.-N. & SHEN, L. 2009. Rotation-invariant texture classification using circular Gabor wavelets. *Optical Engineering,* 48**,** 017001-017001-13.

ZAIDI, Q. 2011. Visual inferences of material changes: color as clue and distraction. *Wiley Interdisciplinary Reviews: Cognitive Science,* 2**,** 686-700.

ZEILER, M. D. & FERGUS, R. 2014. Visualizing and understanding convolutional networks. *Computer Vision–ECCV 2014.* Springer.

ZETZSCHE, C., KRIEGER, G. & WEGMANN, B. 1999. The atoms of vision: Cartesian or polar? *JOSA A,* 16**,** 1554-1565.

ZETZSCHE, C. & NUDING, U. 2005. Nonlinear and higher-order approaches to the encoding of natural scenes. *Network,* 16**,** 191-221.

ZETZSCHE, C. & ROHRBEIN, F. 2001. Nonlinear and extra-classical receptive field properties and the statistics of natural scenes. *Network,* 12**,** 331-50.

ZWICK, W. R. & VELICER, W. F. 1986. Comparison of five rules for determining the number of components to retain. *Psychological Bulletin,* 99**,** 432-442.