COMPARING MODELS OF SYMMETRY PERCEPTION

Matthew James Dry

B.A. (Drama Studies), B.A. (Hons. Psychology)

Submitted for the award of

Doctor of Philosophy

in the School of Psychology

University of Adelaide

29-11-06

This thesis is dedicated to my supervisor Douglas Vickers who passed away in October 2004.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being made available in all forms of media, now or hereafter.

Matthew Dry 29-11-06

Abstract

It has been suggested that the perception of visual regularities plays a fundamental role in structuring the visual world. One form of regularity that has attracted a great deal of research in the visual perception literature is mirror symmetry. In human vision symmetry appears to play an important role in the low-level processing of visual scenes, object recognition, and shape representation. Although a large number of empirical studies have investigated symmetry perception it is as yet unclear how the human visual system manages to perform this process as quickly and efficiently as it does. One way that researchers have attempted to understand the processes underlying symmetry perception is through the use of formal models, or computer simulations. Models can be thought of as theories or explanations for a given psychological phenomena and are potentially useful because they force researchers to be highly specific about the types of processes underlying the phenomena of interest.

In the field of computational geometry one of the most powerful tools for describing spatial relations is Voronoi tessellation. Psychophysical and physiological evidence is reviewed suggesting that the visual system may be employing a process similar to Voronoi tessellation to extract the spatial relations necessary for the organisation of visual stimuli. Against this background a Voronoi tessellation based model of symmetry perception is outlined. The Voronoi model was developed in response to three key findings from the human symmetry perception literature: that symmetry perception is a graded response; that the perception of symmetry is more salient at, or close to, the axis of symmetry; and that the processes underlying symmetry perception can be generalised to other similar structure and regularity detection tasks.

A major weakness of previous papers outlining models of symmetry perception has been the lack of comparison between model output and empirical data sets. Furthermore, little effort has been made to compare the performance of different models and, where this has occurred, the comparisons have been qualitative rather than quantitative. This study addresses these weaknesses by employing Bayesian model selection to compare the performance of a number of models across a range of previously published and new empirical data sets.

The results of the model comparisons indicate that the Voronoi model is able to match and, for the most part, beat the performance of the rival models of symmetry perception. However, there are a number of problems with the Voronoi model that suggest that it provides an insufficient explanation for the processes underlying symmetry perception. The comparisons also provide useful insight into how the Voronoi model and the alternative models might be modified to address their specific weaknesses and suggest the need for a number of new empirical studies.

Acknowledgements

First I would like to thank my (many) supervisors: Douglas Vickers, Michael Lee, Nicholas Burns and Simon Dennis. This thesis is very much inspired by the work of two of my supervisors. First, Doug's relational approach to visual perception forms the backbone to this thesis, and although he never got to see this work I think he would have thought it was pretty cool. Second, Michael can largely take the blame for my interest in modelling psychological phenomena, and also for using Bayesian methods to evaluate them. Finally, extra special thanks go to Nick for taking on the role of principal supervisor at a fairly advanced stage in this thesis. Nick has provided vital support, advice and feedback. He has also managed to read an entire thesis in the time it takes a lot of supervisors to read a chapter.

Thanks also to Dan Navarro for advice relating to the analyses, and Anna Ma-Wyatt for reading a drafts of Chapters 2 and 3. Special mention needs to be made of my office-mate Jason McPherson. Four years is a very long time to spend in a small room with another person, but not only has it been fun, I think we have managed to learn a lot about each others research. Half the trick is to play the MC5 loudly on Friday afternoons.

I am indebted to Hugh Wilson for showing me how to implement spatial filter models in a MATLAB environment. Messing about with his code helped me to understand this class of models in a way that reading a text book or journal article never could.

This thesis would not have been possible without the support of the University of Adelaide School of Psychology. The School has provided me with financial support, gainful employment and more than ample resources. A special thankyou must go to my fellow PhD students for pretending to be as interested in dots as I am.

Similarly, I would never have gotten this thesis finished without the support of my friends and family. You have been very patient with me. Many, many thanks go to my long suffering partner Sonja.

Finally, I would like to thank God, without whom I might sometimes find myself drinking alone.

Table of Contents

List of Tables

List of Figures

1. INTRODUCTION

1.1. Why study symmetry?

The perception of visual regularities plays a fundamental role in structuring the visual world (Barlow, 2001; Koffka, 1935; Marr, 1982; Wertheimer, 1938). One form of regularity that has attracted a great deal of research in the visual perception literature is symmetry. An object is said to be symmetric if it demonstrates self-similarity across one or more of the Euclidian transformations (i.e., reflection, translation or rotation) and examples of the three different classes of symmetry are demonstrated in Figure 1.1. Figure 1.1a is an example of mirror symmetry and is generated by reflecting a pattern across a vertical axis located in the centre of the stimulus. Figure 1.1b is an example of repetition symmetry and is generated by subjecting a pattern to a horizontal translation. Finally, Figure 1.1c is an example of centric symmetry, and is generated by rotating a pattern through 180 degrees.

In human vision symmetry appears to play an important role in the low-level processing of visual scenes. For example, symmetry was recognised as one of the principal organisational factors in the perceptual grouping of discrete objects (Koehler, 1929; Koffka, 1935; Wertheimer, 1938). Symmetry appears to facilitate figure ground segmentation (Driver, Baylis, & Rafal, 1992) and it has been suggested that when a symmetric object is tilted, the degree of skew can be used as an indication of three-dimensional structure (Wagemans, 1993). Additionally, it has been suggested that symmetry plays an important role in object recognition and shape representation (Blum, 1973; Burbeck & Pizer, 1995; Kovacs, Feher, & Julesz, 1998; Marr, 1982).

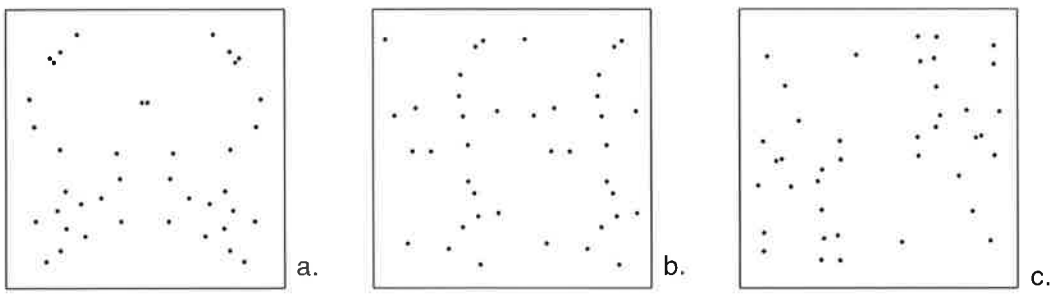

Figure 1.1 An example of mirror symmetry (a), repetition symmetry (b), and centric symmetry (c).

The ability to detect symmetry appears to be of importance not only for human vision, but also for visual systems at many different levels of sophistication. For example, because many animals tend to exhibit symmetric structure, the ability to detect symmetry within an otherwise randomly arranged scene may signal the presence of a predator or prey (Tyler & Hardage, 1996). Furthermore, because genetic defects and poor health can be signalled by asymmetry, symmetry detection appears to be an important cue in mate selection for many species (Moller, 1993; Watson & Thornhill, 1994). Similarly, insects show a preference for symmetric flowers, which tend to produce greater quantities of nectar (Moller, 1995).

1.2. Comparing models of symmetry perception

Although a large number of empirical studies have investigated symmetry perception it is as yet unclear how the human visual system manages to perform this arguably vital process as quickly and efficiently as it does. One way that researchers have attempted to understand the processes underlying symmetry perception is through the use of formal models, or computer simulations. Models can be thought of as theories or explanations for a given psychological phenomenon, and are potentially useful because they force researchers to be highly specific about the types of processes underlying the phenomena of interest. This thesis is primarily concerned with the comparison of models of symmetry perception.

It should be noted at the outset that, although there are three main classes of symmetry (see Figure 1.1), this thesis is solely concerned with mirror symmetry. Unless it is indicated, otherwise any references to symmetry should be interpreted as a reference to mirror symmetry. Furthermore, although there are many different types of stimuli that have been employed to measure the limits of symmetry detection in human observers, including line segments (e.g., Labonte, Shapira, Cohen, & Faubert, 1995; Locher & Wagemans, 1993), polygons (e.g., Palmer, 1985), and faces (e.g., Scognamillo, Rhodes, Morrone, & Burr, 2003; Tjan & Liu, 2005), this thesis focuses upon a single class of stimuli: dot patterns such as can be seen in Figure 1.1a. This specific focus on mirror symmetry in dot patterns is mainly driven by practical necessity. In the first place, a great deal of research has been conducted on symmetry perception and the scope of this body of research is far greater than can be dealt with effectively in a single thesis. Second, there are a number of models of symmetry perception that have been developed specifically to detect mirror symmetry in dot patterns. Given that this thesis is primarily concerned with model comparison it seemed logical to focus on research that provides numerous examples of models to implement and compare.

Broadly speaking, there have been two main approaches to modelling the processes underlying symmetry perception in the human visual system. By far the most common approach has been to assume that symmetry perception is the result of early spatial mechanisms filtering an image at different scales and orientations (Dakin & Hess, 1997; Dakin & Watt, 1996; Gurnsey, Herbert, & Kenemy, 1998; Osorio, 1996; Rainville & Kingdom, 1999, 2000; Rainville & Kingdom, 2002; Scognamillo et al., 2003; Tjan & Liu, 2005). An alternative approach to modelling symmetry perception has been to assume that the visual system is explicitly aware of relational information, or the relative proximity and orientation relationships that exist between individual objects, or elements, in the visual environment (Stevens, 1978; Wagemans, Van Gool, Swinnen, & Van Horebeek, 1993).

In a recent review of models of symmetry perception, Rainville and Kingdom (2000) contrasted these two approaches. The spatial filter-based models, which have largely developed from physiological studies of the receptive fields of cells in the retina (Kuffler, 1953) and visual cortex (Hubel & Wiesel, 1962, 1968) and from psychophysical investigations of spatial frequency channels (Campbell & Robson, 1968), were described as being "... psychophysically, physiologically, and theoretically motivated ..." (Rainville & Kingdom, 2000, p.2638). In contrast to this, models that did not specifically employ spatial filtering were construed as problematic because they relied upon "complex" grouping rules. It is therefore a major challenge to any model of symmetry perception that does not follow the classic spatial filtering approach to demonstrate a similar level of psychophysical and physiological plausibility.

This thesis attempts to meet this challenge by introducing an approach to the perception of visual regularities, such as mirror symmetry, based on the relational information revealed by a graphical structure known as Voronoi tessellation (Ahuja, 1982; Aurenhammer, 1991). In Chapter 2 evidence from a range of psychophysical and physiological studies is reviewed and it is suggested that a spreading activation or wavefront process (Kovacs et al., 1998; Lee, Mumford, Romero, & Lamme, 1998) could be used to encode relational structure in early vision. Furthermore, in Chapter 3 a Voronoi based model of symmetry perception is outlined and it is demonstrated that the model is capable of accounting for a number of fundamental characteristics of symmetry perception.

Taken together, the evidence suggests that a Voronoi based model is a theoretically plausible explanation for the processes underlying symmetry perception. However, it is argued that it is not sufficient to rely on theory or one-off demonstrations to test the utility of a model as an explanation for a cognitive or perceptual process. In order to test the utility of the Voronoi model it is necessary to directly compare the performance of the model to the performance of human observers across a range of empirical symmetry detection tasks. This form of comparison provides an indication of how closely the model responses correspond with the responses of the system it is proposed to be simulating. Furthermore, it is necessary to compare the performance of the Voronoi model to the performance of other models of symmetry perception. This provides an indication of the utility of the Voronoi model relative to the utility other potential explanations.

In Chapter 4, I discuss a number of issues related to model comparison and introduce a Bayesian approach to evaluating models of symmetry perception. Within psychology, Bayesian approaches have previously been applied to the comparison of models of a number of different phenomena including memory retention (Lee, 2004), information integration (Myung & Pitt, 1997), individual differences in cognition (Lee & Webb, 2005), response time distributions (Rouder, Lu, Speckman, Sun, & Jiang, 2005), decision making axioms (Myung, Karabatsos, & Iverson, 2005), stimulus representation (Navarro & Lee, 2004), and working memory failure (Carlin, Kass, Lerch, & Hugenard, 1992). It is argued that this approach meets the demands of Jacob and Grainger's (1994) criteria for model evaluation and provides a plausible and logical means of interpreting the relative utility of the models under comparison.

In Chapters 5 to 7, Bayesian model selection is employed to compare the performances of the Voronoi model to those of previously published models of symmetry perception across a range of empirical symmetry detection data sets. A number of the empirical data sets are drawn from seminal papers in the field of symmetry perception research. For example, Barlow and Reeves (1979) demonstrated that the perception of symmetry is not an all-or-none process and that the visual system is able to tolerate high levels of positional distortion or interference and still detect symmetry. Additionally, Jenkins (1982) demonstrated that symmetry perception is not uniform across the visual field and that there is a perceptual advantage for symmetric information located directly adjacent to the axis of symmetry (or line of symmetry reflection). These papers are highly informative in regards to the limits of

symmetry detection in human observers and represent a number of basic requirements that can be considered mandatory for models of symmetry detection.

1.3. Aims

To summarise, the aims of this thesis are as follows:

1. To develop a theoretically plausible model of symmetry perception based on the relational structure revealed by Voronoi tessellation.

2. To introduce Bayesian model selection as a methodology for comparing models of visual processes.

3. To compare the performances of the Voronoi-based model of symmetry perception to the performances of previously published models across a range of empirical symmetry detection tasks.

2. PERCEPTUAL ORGANISATION AND RELATIONAL STRUCTURE

This chapter introduces an approach to the perception of regularity and structure based on a graphical structure known as Voronoi tessellation (Ahuja, 1982; Aurenhammer, 1991). Voronoi tessellation and its dual graph the Delaunay triangulation are described and defined. Research suggesting that the perceptual system has access to relational structure based upon Voronoi tessellation and Delaunay triangulation is reviewed, and the potential strengths of a Voronoi tessellation based approach to perceptual organisation are discussed.

2.1. The perception of regularity and structure

The human visual system is highly adept at detecting regularity. For example, the perception of symmetric organisation in Figure 2.1 is immediate and appears to require little or no conscious effort on the part of the viewer. However, the apparent ease of detecting structure such as this belies the computational complexity of achieving this task.

In recent years a common approach to solving these types of structure detection problems has been to assume that perceptual grouping is the result of early spatial mechanisms filtering an image at different scales and orientations. For example, there are numerous models that apply a spatial filtering approach to the detection of structure in symmetrical stimuli (Dakin & Hess, 1997; Dakin & Watt, 1996; Gurnsey et al., 1998; Osorio, 1996; Rainville & Kingdom, 1999, 2000; Rainville & Kingdom, 2002; Scognamillo et al., 2003; Tjan & Liu, 2005). This

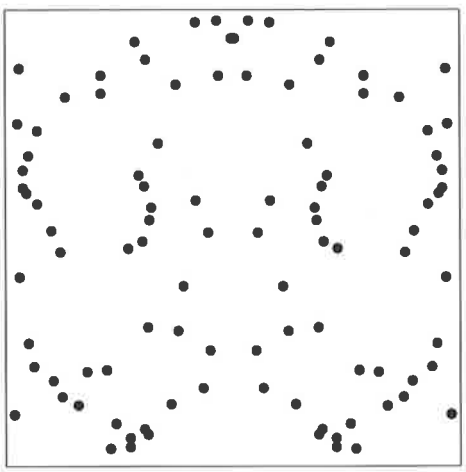

Figure 2.1 Example of a symmetrical dot pattern.

approach has largely developed from physiological studies of the receptive fields of cells in the retina (Kuffler, 1953) and visual cortex (Hubel & Wiesel, 1962, 1968), and psychophysical investigations of spatial frequency channels (Campbell & Robson, 1968). At the heart of this approach is the use of spatial filters that mimic the responses to image luminance in the receptive fields of individual neurons or retinal cells.

The following example illustrates a simple application of the spatial filtering approach to structure detection. The dot stimulus in 2.2a is usually seen as organised in columns rather than rows. Convolving the stimulus with an isotropic Laplacian-of-Gaussian spatial filter (i.e., the circularly symmetric, 'mexican hat' receptive field of a retinal ganglion cell) results in the fusion of the individual dots into vertical blobs, such as can be seen in 2.2b. The presence of vertical structure in the stimulus could be signalled by the activation of orientationally tuned neurons in V1. Importantly, according to this approach the perception of structure or regularity is not based upon the individual dots in the original image, but the blob-like features in the convolved image.

An alternative approach to detecting the symmetry in Figure 2.1 (or the vertical organisation in Figure 2.2a) is to assume that the visual system is explicitly aware of the relational structure within the stimulus. Relational structure is essentially an index of the relative proximity and orientation relationships that exist between individual stimulus elements. One of the fundamental tasks of early visual perception is the spatial organization of an image and it has been argued that the extraction of spatial relations such as relative proximity of image elements, orientation, collinearity, and inside/outside relations are a necessary step in solving essential perceptual organization 'problems' such as the spatial correspondence problem, figure/ground separation and the perception of texture boundaries (Marr, 1982; Ullman, 1984). The importance of relational structure was recognised in the principles of perceptual organisation proposed by Wertheimer (1938) and also plays an important role in many recent computational models of visual perception (Dry, Vickers, Lee, & Hughes, submitted; Jenkins, 1983a; Logan, 1996; Pizlo, Salach-Golyska, & Rosenfeld, 1997; Smits & Vos, 1986; Stevens, 1978; Stevens & Brookes, 1987; Wagemans et al., 1993).

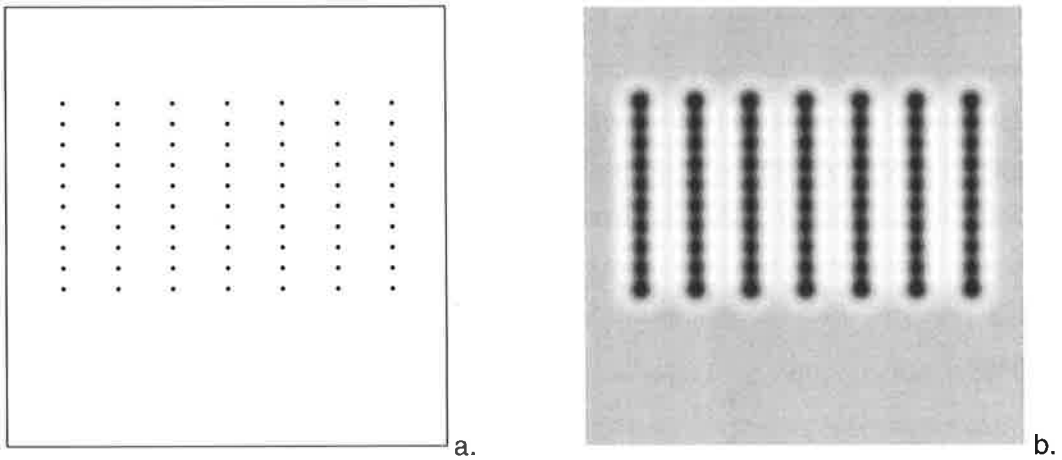

Figure 2.2 A stimulus containing dots arranged in columns (a) and the result of convolving the stimulus with an isotropic Laplacian-of-Gaussian filter (b).

In a recent review of models of symmetry perception, Rainville and Kingdom (2000) described spatial filter-based models as being "... psychophysically, physiologically, and theoretically motivated ..." (Rainville & Kingdom, 2000, p.2638). In contrast to this, models that did not specifically employ spatial filtering were construed as problematic because they relied upon "complex" grouping rules. It is therefore a major challenge to any model of symmetry detection that does not follow the classic spatial filtering approach to demonstrate a similar level of psychophysical and physiological plausibility.

In this chapter I will present evidence suggesting that the visual system may be accessing information about the relational structure of an image via a process similar to Voronoi tessellation. First, I will first describe and define Voronoi tessellation and a variant of Voronoi tessellation, the medial axis. I will then review psychophysical and neurophysiological evidence suggesting that a medial axis and Voronoi based representation may be calculated by the perceptual system in early visual processing. Third, I will describe and define Delaunay triangulation, the dual graph of Voronoi tessellation. Fourth, I describe how a spreading activation process could encode the relational information present in the Delaunay triangulation, and provide a review of theoretical and empirical evidence suggesting that the visual system has access to this information.

2.2. Voronoi tessellation

The detection of relational structure in dot patterns is not a problem that is unique to early visual processing. Aurenhammer (1991) provide a review of a wide variety of disciplines that are similarly concerned with the geometric properties of spatial point distributions including geography, crystallography, molecular chemistry, geophysics, ecology, anthropology, botany, and zoology. Furthermore, relational measures have been applied to computational and graph-theoretic problems such as information storage and retrieval, path planning, and site placement.

One measure of relational structure that has received particular attention in the computational geometry literature is Voronoi tessellation. Figure 2.3 is an example Voronoi tessellation for a random point set. As can be seen, the Voronoi tessellation divides the plane into a series of cells, with each cell containing a single point. Each cell corresponds to the area within the plane that is closer to the point located in the cell, than to any other point in the set. For example, in Figure 2.3 the area within the cell containing point p is closer to p than it is to q (or any of the other points in the set). In addition to this relative proximity information the Voronoi tessellation also contains 2-dimensional geometric properties such as cell perimeter, internal area, orientation of principle axis, elongation and ratio of side lengths (Ahuja & Tuceryan, 1989). It should also be noted that Voronoi tesselation is not restricted to dot points, but can be calculated for line lengths (Figure 2.4), regular or irregular shapes, and 'natural' objects or scenes (Aurenhammer, 1991; Tek & Kimia, 2003).

A Voronoi tessellation can be calculated for any set S of three or more co planar points, given that the points are not collinear or co-circular (Ahuja, 1982). Any point p of S is located inside a cell defined as the area within the plane that lies closer to p than to any other point. Aurenhammer's (1991) definition of Voronoi tessellation states that "for two distinct sites $p, q \in S$ the dominance of p over q is defined as the subset of the plane being at least as close to p as to q. Formally,

$$\mathrm{dom}(p,q) = \left\{x \in R^2 \mid \delta(x,p) \leq \delta(x,q)\right\},$$

for δ denoting the Euclidean distance function" (Aurenhammer, 1991). As can be seen in Figure 2.3, the edge shared by points p and q bisects the plane equidistant to p and q and indicates the portion of the plane in which all points will be closer to p than q and vice versa.

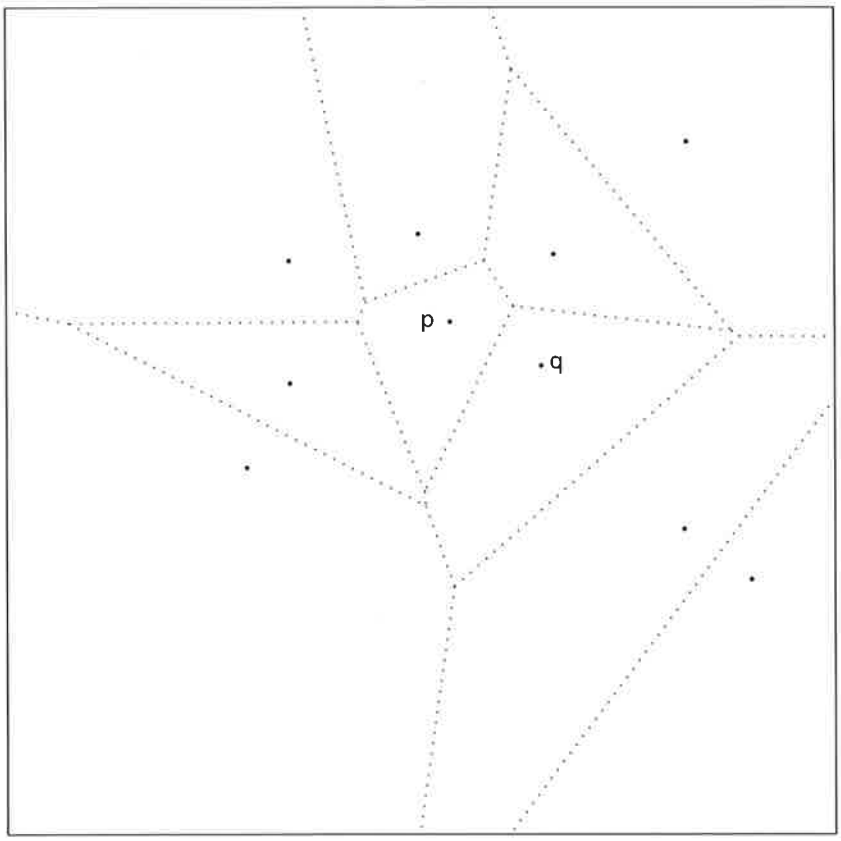

Figure 2.3 Example Voronoi tessellation for a set of randomly located sites.

The cell in which p is located ($p \in S$) is the portion of the plane in which p asserts dominance over all other points in the set:

$$\mathrm{reg}(p) = \bigcap_{q \in S-\{p\}} \mathrm{dom}(p.q).$$

The resulting tessellation can be thought of as the result of a uniform growth process (Ahuja, 1982), or 'grassfire' (Blum, 1973), beginning simultaneously at each point in the set. The circular wavefronts expand outwards from each point and are extinguished as they make contact with each other. The first point of contact between wave fronts will necessarily be the midpoint between two sites, with the edges expanding outwards until they make contact with additional wavefronts.

Voronoi tessellation can also be used to find the medial axis or skeleton of closed contours such as can be seen in Figure 2.5. This technique was pioneered by Blum (1973) as a means of describing biological objects and other irregular shapes. In this case the 'grassfire' spreads outwards from the contours rather than from discrete sites (such as the dots in Figure 2.3). It has been demonstrated that the medial axis is mathematically equivalent to Voronoi tessellation (Fabbri, Estrozi, & Costa, 2002) and so the two terms can be thought of as interchangeable. The importance of the interchageability of the medial axis and Voronoi tessellation will be explored more fully in the following section.

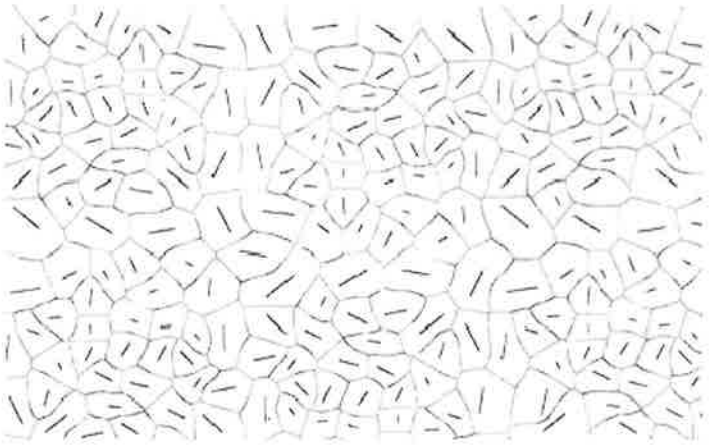

Figure 2.4 Example Voronoi tessellation for a set of randomly oriented and located line segments. Reproduced from Tek and Kimia (2003, p. 69).

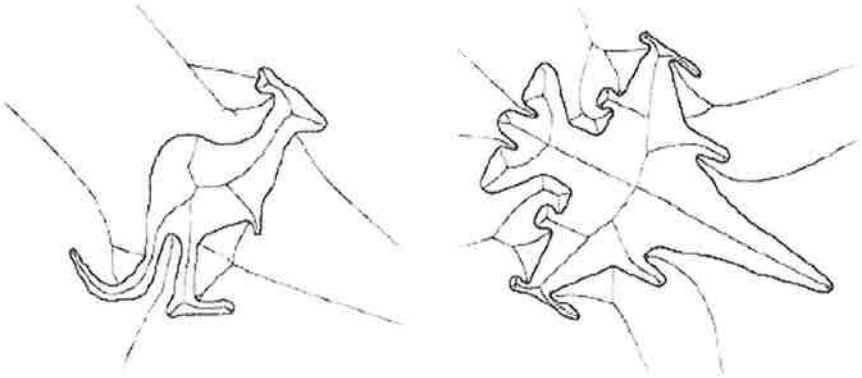

Figure 2.5 Example of the medial axis for two closed contours (a kangaroo and an aeroplane). Reproduced from Tek and Kimia (2003, p. 38).

2.3. Perceptual support for Voronoi tessellation and the medial axis

The medial axis plays a key role in a number of theoretical and computational models of shape and object recognition (Biederman, 1987; Blum, 1973; Leyton, 1992; Marr, 1982; Zhu & Yuille, 1996). There is also a growing body of experimental evidence suggesting that the visual system may be performing a process not unlike Voronoi tessellation or medial axis calculation at an early stage of processing. The following section provides a brief overview of this experimental research.

Burbeck and Pizer (1995) and Burbeck, Pizer, Morse and Ariely (1996) tested the ability of observers to bisect solid shapes with vertical edges that had been subjected to sinusoidal modulation. An example stimulus is shown in Figure 2.6 (left). The observers' task was to estimate whether a target dot that was located in horizontal alignment with a sinusoidal peak was positioned to the left or the right of the stimulus centre. Both the width of the stimuli and the degree of edge modulation were manipulated across a series of trials. The results indicated that the perceived central modulation decreased as a function of stimulus width and degree of edge modulation. It has been demonstrated that the empirical data are consistent with Burbeck et al's. (1995; 1996) core model, and Siddiqi, Kimia, Tannenbaum and Zucker's (2001) shock based model, both of which are based upon medial axis representations of shape.

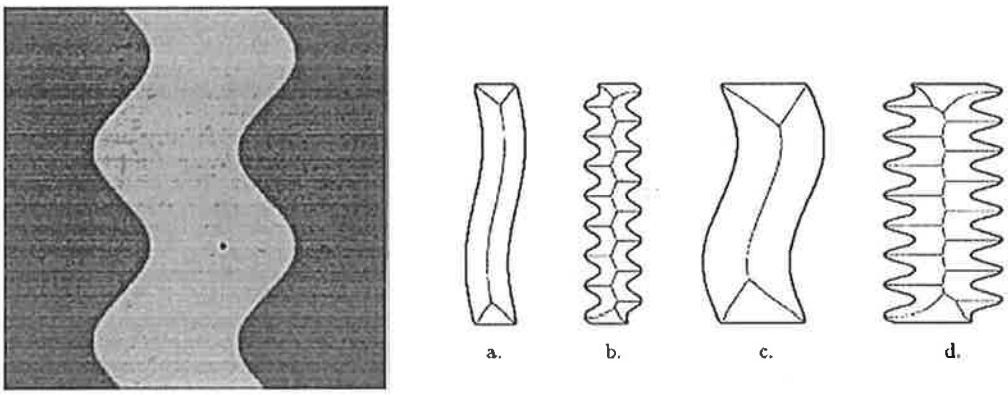

Figure 2.6 An example stimulus from the Burbeck and Pizer (1995) study (left), and examples of experimental stimuli and their respective medial axes (right). Reproduced from Siddiqi et al. (2001)

Figure 2.7 The height to width ratio of an ellipse determines the length of its medial axis. Reproduced from Siddiqi et al. (2001).

Siddiqi et al. (2001) also plausibly accounted for the results of an orientation discrimination experiment from Li and Westheimer (1997) in terms of a medial axis representation in early vision. Li and Westheimer measured the threshold for discriminating the orientation rotation of a series of stimuli ranging from a straight line, through vertical ellipses with varying height-to-width aspect ratios, to a circle. They found that the orientation discrimination for the ellipses with aspect ratios of less than 1:5 was almost identical to that of a straight line, after which performance degraded. Because the ellipses do not contain any explicit vertical edges Li and Westheimer speculated that the observers may have been using the central axis of the stimulus as a cue. Siddiqi et al. (2001) noted that the length of the vertical portion of a medial axis representation of the ellipses would decrease as the width of the ellipses increased (Figure 2.7) and suggest that this is consistent with Westheimer's (1981) finding that orientation discrimination for lines increases with line length.

In a series of experiments Kovacs and colleagues (Kovacs et al., 1998; Kovacs & Julesz, 1993, 1994) measured contrast sensitivity for a target Gabor patch embedded in a noise background of randomly oriented patches and surrounded by a contour defined by oriented patches (see Figure 2.8). The relative proximity of the target and contour was manipulated across trials and a number of different contour shapes (circle, ellipse, triangle, and cardioid) were employed across separate experiments. The contrast sensitivity for the target condition was compared to a baseline condition in which no contour was present and the difference between the two conditions was plotted, resulting in a contrast sensitivity map. The results

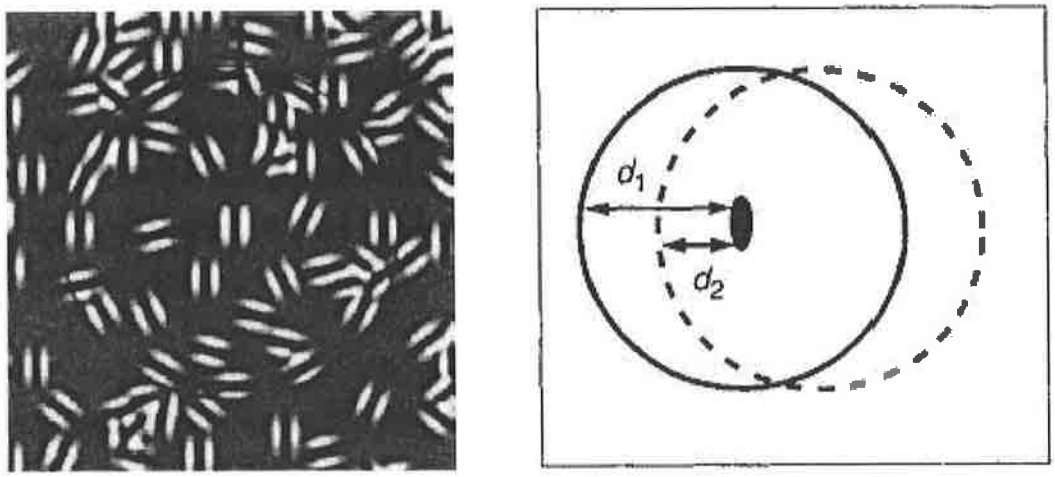

Figure 2.8 Example of the stimuli (left), and a schematic representation demonstrating how the relative proximity of the target and contour was manipulated (right) in the experiments by Kovacs et al (1998; 1993; 1994). Reproduced from Kovacs and Julesz (1994).

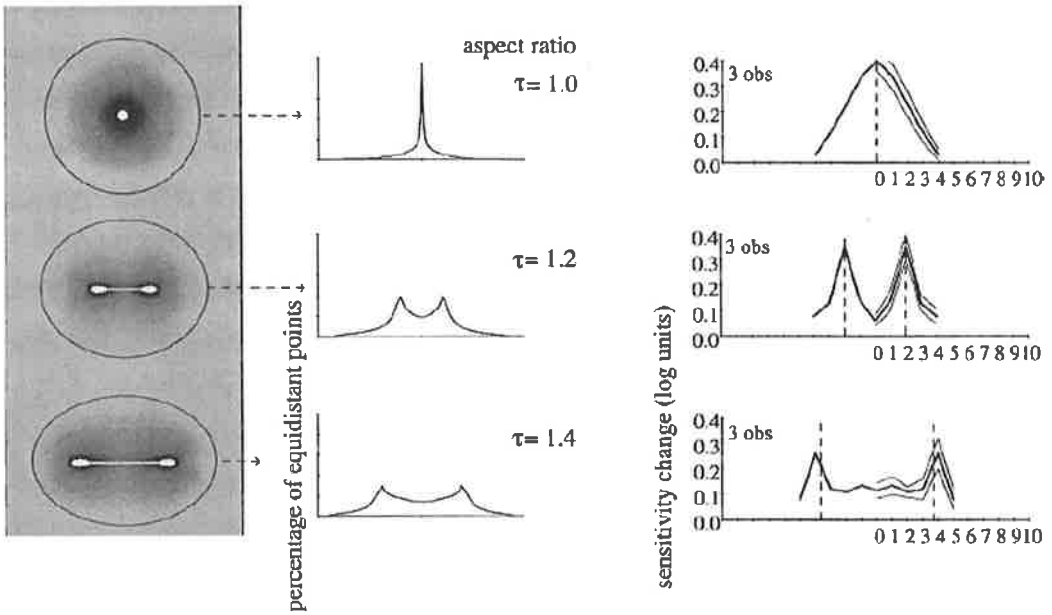

Figure 2.9 Results from Kovacs and Julesz (1998). The figures in left and middle columns show the predictions made by Kovacs and Julesz's model. The figures in the right column are the corresponding empirical sensitivity maps. Reproduced from Kovacs and Julesz (1998).

indicated that contrast sensitivity was non-uniform and dependant on the proximity of the target to the contour. Sensitivity peaks were found in two separate regions: close to the contour boundary, and within the contour. The peak at the boundary could be explained as a within-filter interaction; however, the second region of peak sensitivity was too far from the perimeter to be interpreted in the same way.

Kovacs and Julesz compared the empirical sensitivity maps for the within contour peaks to the peaks predicted by a medial axis representation (Figure 2.9). They found that the empirical data were well-predicted by the medial axis representation, for both the contours containing a single straight medial axis (such as the ellipses) and the more complex contours containing a branching medial axis (such as the triangle and cardioid). It was concluded that the data support the possibility that the visual system generates an explicit medial axis representation at an early stage in visual processing (Kovacs et al., 1998; Kovacs & Julesz, 1994).

In addition to the psychophysical studies described above, there is also neurophysiological evidence that supports the role of the medial axis in early visual processing. In a series of experiments Lee, Mumford, Romero and Lamme (1998) demonstrated that V1 cell responses at certain latencies correspond to the location of the medial axis. They recorded the responses of individual V1 neurons in awake behaving Rhesus monkeys that were presented with stimuli comprised of oriented texture strips or squares presented on an opposite orientation background (Figure 2.10, left). The position of the strip or square was shifted relative to the classical receptive field of each cell across a series of trials. The results indicated that the spatial response profiles were non-uniform and displayed peaks at the texture boundaries and the medial axis (Figure 2.10, right). Lee et al. commented on the similarity between their findings and those of Kovacs and Julesz (1994) and Burbeck and Pizer (1995) and suggested that their study adds further support to the plausibility of the medial axis being used to analyse visual objects and scenes.

Oka, van Tonder and Ejima (2001) investigated the relationship between medial axes and figure salience. They measured the differences in visual evoked potential (VEP) of human observers for stimuli with varying levels of maximal strengths in medial axis response as predicted by a model of image segmentation based on the medial axis (van Tonder & Ejima, 2000b). The stimuli were comprised of oriented Gabor patches, and five different shapes were

investigated (Figure 2.11). The results indicated significant trends in the N1 and P2 components of the VEP across the five stimulus classes, with shorter latencies being found for stimuli with greater peak medial axis strengths. Oka, van Tonder and Ejima interpreted their results in light of previous research (Lee et al., 1998; Kovacs, Feher and Julesz, 1998) and the predictions made by the image segmentation model and concluded that medial axis computation plays an important role in early visual processing.

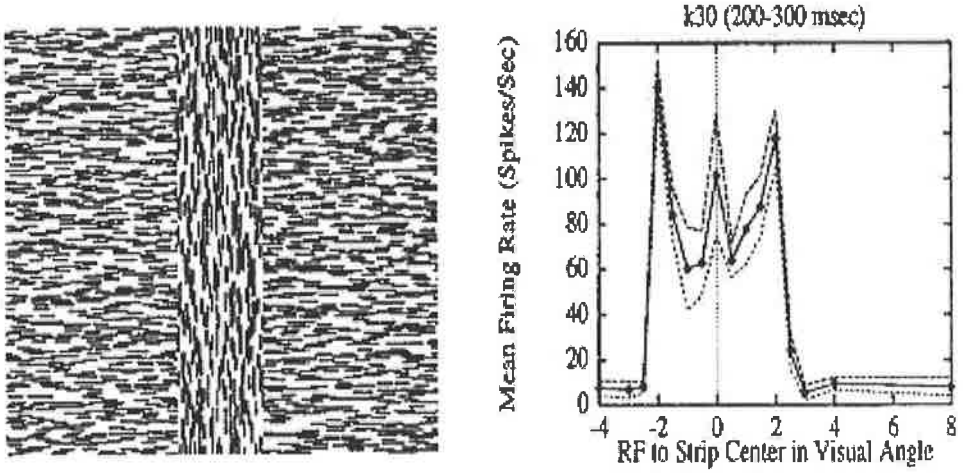

Figure 2.10 Example stimulus (left) and the corresponding spatial response profile. Reproduced from Lee, Mumford, Romero and Lamme (1998).

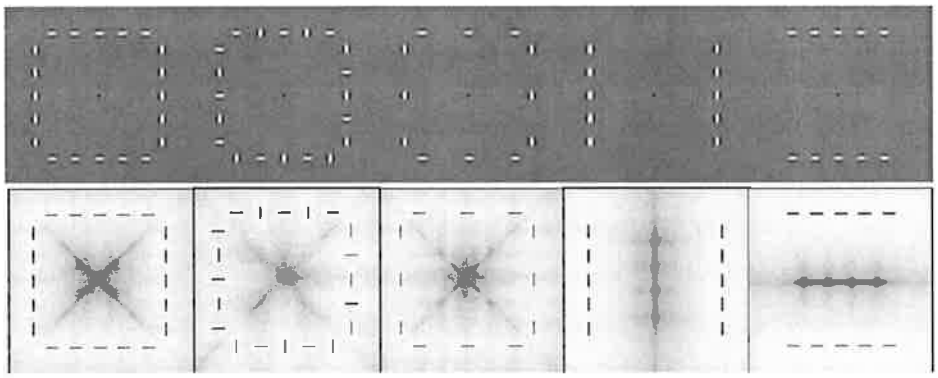

Figure 2.11 The five stimuli from the Oka, van Tonder and Ejima (2001) study (top row), and the corresponding medial axis representation predicted by a model of image segmentation based on the medial axis (bottom row). Reproduced from Oka, van Tonder and Ejima (2001).

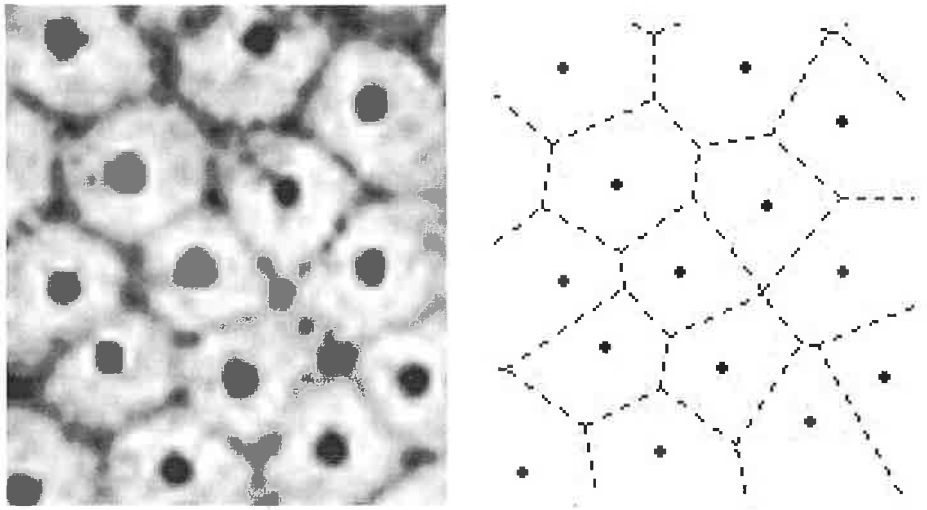

Figure 2.12 Detail from Figure 3 in van Tonder and Ejima (2000b) showing the tessellation predicted by a model of image segmentation based on the medial axis for a random point set (left), and the corresponding Voronoi tessellation given by replotting the points in MATLAB 7.0.4 (right).

Taken together, these studies provide cumulative evidence implicating the role of a medial axis-like representation in early vision for stimuli such as solid figures, contours, or texture defined regions. There do not appear to be any psychophysical or neurophysiological studies that have investigated whether a similar representation is derived for spatial point distributions such as Figure 2.3. However, a number of neurophysiologically plausible models of early vision based upon medial axis representations have been developed in recent years (Burbeck & Pizer, 1995; Burbeck et al., 1996; Hong & Pavel, 1996; Kimia, 2003a, 2003b; Kurbat, 1996; Tek & Kimia, 2003; van Tonder & Ejima, 2000b). Although these models have generally been used to find the medial axis or internal skeleton of solid objects or closed contours, they have also been applied to stimuli comprised of individual objects or features, in which case the resulting representation is equivalent to the Voronoi tessellation of the set of objects. For example, in van Tonder and Ejima (2000a), an image segmentation model was used to derive a tessellation of texton stimuli (Caelli & Julesz, 1978). The result of applying this model to a spatial point texture can be seen in Figure 2.12 (left panel). The points were replotted and the Voronoi tessellation was calculated (right panel). As can be seen, there is a

close correspondence between the tessellation predicted by the model and the standard Voronoi tessellation calculated by MATLAB 7.0.4.

Further support for existence of a Voronoi tesselation-like representation comes from research suggesting that the human visual system is aware of the relative proximity and orientation relationships that can be obtained from Delaunay triangulation, the dual graph of Voronoi tessellation. In the following sections Delaunay triangulation and a family of related relational structures will be defined and research implicating the role of these structures in low level perceptual organization will be reviewed.

2.4. Delaunay triangulation

Delaunay triangulation is the dual graph of Voronoi tessellation. Dual graphs can be thought of as paired graphs for the same point set, with each graph containing the information necessary to generate its dual. Figure 2.13 shows the Delaunay triangulation for the random point set in Figure 2.3 superimposed on the Voronoi tessellation, with the Delaunay edges shown as solid lines and the Voronoi edges as dashed lines. All of the points in the set that have Voronoi cells sharing a common edge are defined as Voronoi neighbours. Joining these neighbouring points creates the Delaunay triangulation. As can be seen, each of the edges in the triangulation runs orthogonal to their corresponding Voronoi edge but does not necessarily intersect it.

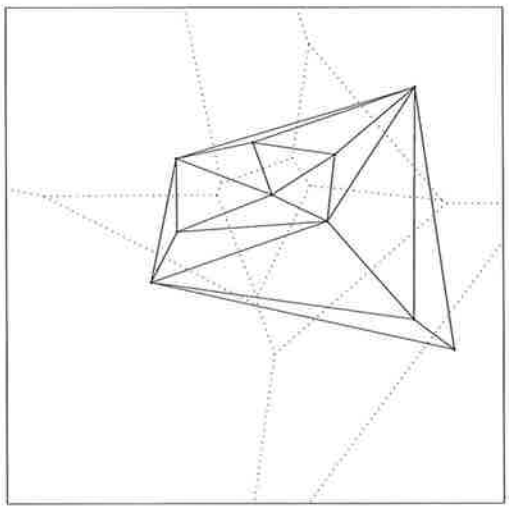

Figure 2.13 Delaunay triangulation of the points in Figure 2.5. The Delaunay edges are shown as solid lines and the Voronoi edges as dashed lines.

An important property of the Delaunay triangulation is that it is a super-graph of a number of other relational structures that have been implicated in research on structure detection (Ahuja, 1982; Ahuja & Tuceryan, 1989; Jarvis, 1972; Toussaint, 1980; Zahn, 1971). The *Gabriel graph* (Gabriel & Sokal, 1969) is the set of the Delaunay edges that directly intersect with their dual Voronoi edge (Howe, 1978, cited in Aurenhammer, 1991). Two points are joined in a Gabriel graph if no other point is closer to their mid-point than both of the points are. This means that a circle that is centred on the two points' mid-point, with a diameter equal to their inter-point distance, will contain no other points. The *relative neighbourhood graph* (Toussaint, 1980) connects points if no other point is closer to both of them than their inter-point distance. By definition this means that the intersection of the two circles, centred on each point, and with a radius equal to the inter-point distance, is free of any other points. A spanning tree is a structure that connects all of the points in a set with n-1 edges and contains no circuits. The *minimum spanning tree* is the structure that minimises the total length of the edges connecting points (Zahn, 1971). Finally, two points are joined as *nearest neighbours* if one of the points lies closer to the other than to any other point within the set. Note that this means that the nearest neighbour of one point is not necessarily the nearest neighbour of the other. If two points lie closer to each other than any other point they are defined as *reflexive nearest neighbours*. An example of the sub-graphs of the Delaunay triangulation is given in Figure 2.14. As can be seen, the graphs are hierarchically nested: the nearest neighbours are a subset of the minimum spanning tree, which is a subset of the relative neighbourhood graph, and so on, up to the Delaunay triangulation.

Aside from the Delaunay triangulation and its sub-graphs, a number of other methods have been employed to describe the relational structure in point sets, such as CODE (Compton & Logan, 1993; Logan, 1996; van Oeffelen & Vos, 1982, 1983), k-nearest neighbours and average nearest neighbours. Of particular interest are the k-nearest neighbour and average nearest neighbour approaches, both of which share a high degree of overlap with the point pairs predicted by the Delaunay triangulation. For the k-nearest neighbour (kNN) approach, the relative proximity relationship between each site in a set of k points is defined by the order of their kNN rank (i.e., nearest neighbour, second nearest neighbour, third nearest neighbour, through to k^{th} nearest neighbour). For the average nearest neighbour (*av.*NN) approach, all points that lie within a distance determined by the mean nearest neighbour distance of all points are joined in a pair. Furthermore, the average nearest neighbour distance is sometimes multiplied by a constant, for example Stevens (1978) used *av.*NN multiplied by 1.3 or 1.5.

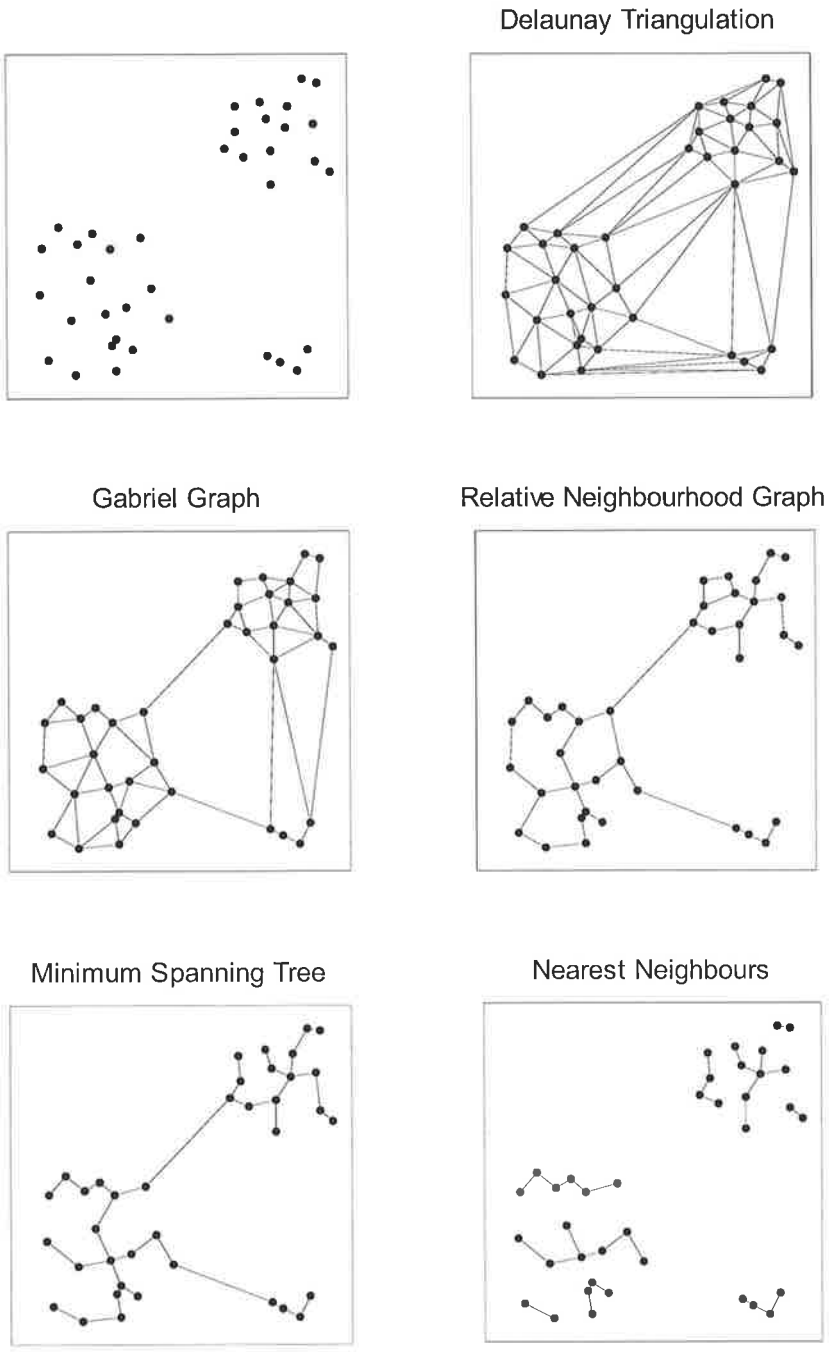

Figure 2.14 Set of randomly distributed points and the respective Delaunay triangulation, Gabriel graph, relative neighbourhood graph, minimum spanning tree, and nearest neighbour pairs.

Figure 2.15 shows a comparison of the *av.*NN multiplied by 1.5 and *k*NN relations for the set of points in Figure 2.14. Comparison with Figure 2.14 demonstrates that the links defined by *av.*NN*1.5 or *k*NN relations of an order of 2NN or higher are not necessarily a subset of the Delaunay triangulation. However, there is a high degree of overlap between these sets. Figure 2.16 shows the proportion of point pairs predicted by first, second and third nearest neighbour and average nearest neighbour methods that are also part of the set predicted by the Delaunay triangulation for random point patterns with 10 to 100 points. As can be seen, the pairs predicted by both the average NN distance and for low values of *k*, account for close to or over 90% of the pairs predicted by the Delaunay triangulation. Given this high degree of overlap, it could be argued that models employing methods such as *av.*NN and *k*NN relations are describing the same process as models employing Delaunay triangulation or its sub-graphs: an awareness of the proximity of a point relative to its immediate neighbours, based upon the Voronoi tessellation of the set of points.

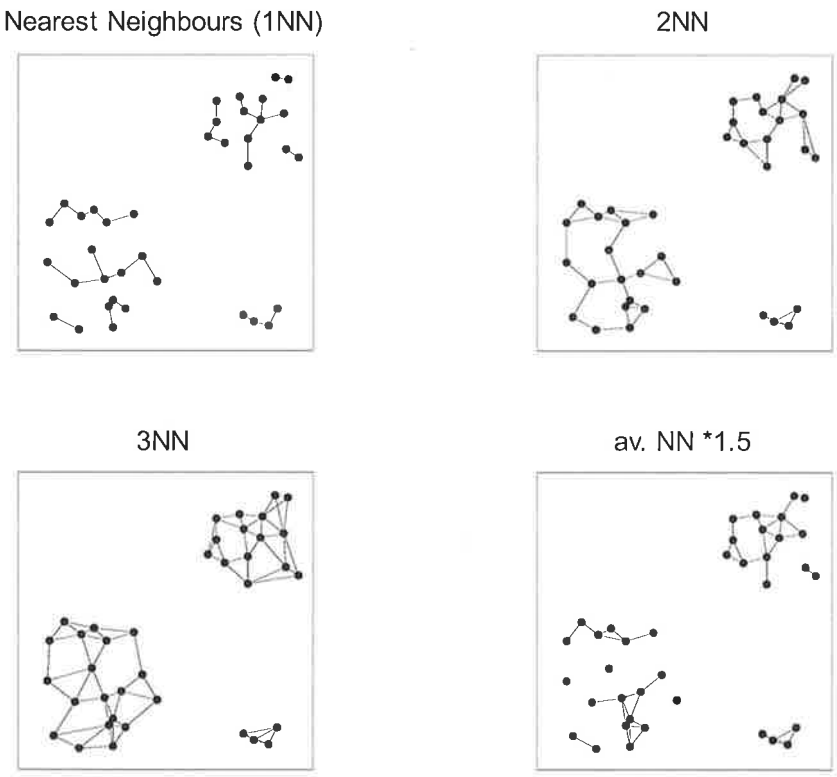

Figure 2.15 Comparison of point pair links predicted by Nearest Neighbours, 2NN, 3NN and average NN multiplied by 1.5.

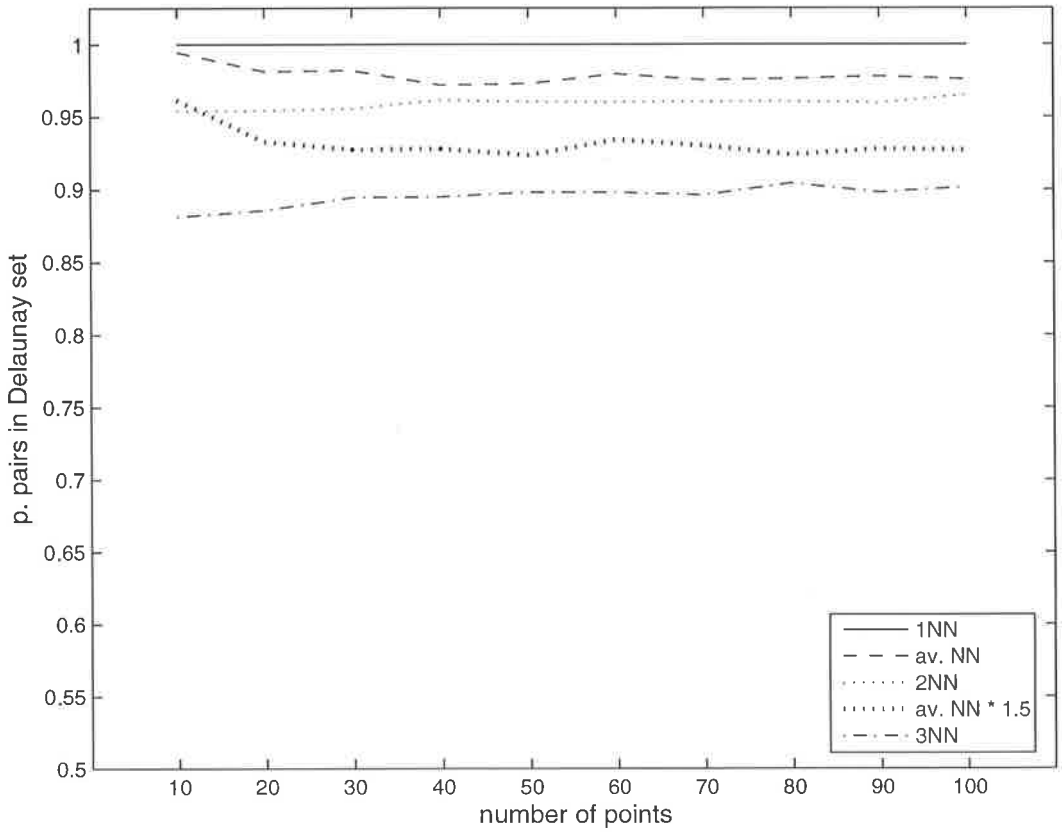

Figure 2.16 A comparison of the proportion of point-pairs predicted by low-order k *nearest neighbour (1-3NN) and average nearest neighbour methods (av. NN and av. NN *1.5) that are also part of the set of point-pairs predicted by Delaunay triangulation for random point patterns with 10 to 100 points. The proportional overlap is the average for 1000 randomly generated random point patterns. It should be noted that the scale on the y-axis is 1 to .5, not 1 to 0.*

2.5. Perceptual support for Delaunay triangulation.

In Section 2.3 research was reviewed indicating that the perceptual system may be performing a process similar to Voronoi tessellation at an early stage in visual processing. It was claimed that this is supported by additional evidence indicating the visual system is capable of exploiting the relational structure revealed by Delaunay triangulation, the dual graph of Voronoi tessellation. This section will review the theoretical and empirical support for this claim.

If the visual system is generating a Voronoi tessellation-like representation in early vision via some form of uniform growth or activation spread process (Kovacs et al., 1998; Lee et al., 1998), then this process could also provide a means of obtaining the relational structures present in Delaunay triangulation. For each point, the ordering of its Delaunay neighbours would be an index of the temporal ordering of wavefront extinction. The first point of contact between two wavefronts (and subsequent wavefront extinction) would give the relative location and proximity of a point's nearest neighbour. The next point of contact between two wave fronts would indicate the next nearest Delaunay neighbour, etc. This process would continue until all wave fronts were extinguished, by which point the Delaunay triangulation of the set would be fully defined.

The pairings predicted by Delaunay triangulation and its sub-graphs and related approaches such as the *k* nearest neighbour and average nearest neighbour methods, have all been used to account for the perception of structure in a wide variety of random point stimuli. For example, Zahn (1971) demonstrated that the minimum spanning tree could be used to detect the presence of separate clusters of dots and changes in dot density in random dot textures. Similar demonstrations have also been made by Jarvis (1972) using *k* nearest neighbours, Toussaint (1980) using the relative neighbourhood graph, minimum spanning trees and Delaunay triangulation, and by Ahuja (1982) and Ahuja and Tuceryan (1989) using Delaunay triangulation. In each of these papers it was demonstrated that algorithms based on the respective relational measures were able to detect the presence of what Toussaint described as "perceptually meaningful" structure.

A common weakness of these studies is that evidence in support of the success of the various algorithms was generally based upon a qualitative comparison of the reader's (and authors') perception of structure in the stimulus and the algorithms solution. However, there are a limited number of studies that have provided empirical support for these claims. Vickers, Mayo, Heitmann, Lee, and Hughes (2004) demonstrated that when participants are presented with random dot patterns and asked to create minimum spanning trees, the empirical solution lengths corresponded closely to the optimal solution. Furthermore, Pomerantz (1981) presented subjects with a range of five point random dot patterns and asked the participants to join the points together in the way they perceived them to be organised. They found that the participants' preferred organisation was the one with the minimum total link length, and that

the solutions tended not to contain circuits. In other words, even though the participants were not explicitly asked to create minimum spanning trees they behaved as if they had been.

Vickers, Preiss and Hughes (submitted) extended the Pomerantz study by presenting participants with higher density stimuli (10 to 100 dots). Whereas the stimuli in the Pomerantz study gave the appearance of a single cluster of points, the higher density stimuli appeared to contain several clusters of points (similar to Figure 2.14). The participants were asked to join up the dots that they perceived as belonging to separate clusters, as if they were joining up stars to form constellations. The results indicated that the empirical mean number of links detected, mean link length, and mean number of clusters were all well predicted by an organising process based on nearest neighbours (r = 0.998, 0.996, and 0.996, respectively). They suggest that the participants in this study did not create minimum spanning trees because they tended to avoid creating long links that joined up separate clusters of points.

Another source of evidence suggesting that the human visual system relies upon an awareness of relational structure to detect the organisation of random point textures comes from the literature on path-finding tasks such as the travelling salesperson problem, or TSP (Dry, Lee, Vickers, & Hughes, in press; Graham, Joshi, & Pizlo, 2000; MacGregor, Chronicle, & Ormerod, 2004; MacGregor & Ormerod, 1996; Vickers, Lee, Dry, & Hughes, 2003; Vickers, Lee, Dry, Hughes, & McMahon, 2006; Vickers et al., 2004). Solving a TSP involves finding the shortest pathway through an array of n nodes or 'cities', generally returning to the starting node. Vickers et al. (2003) proposed that the visual system may be solving the TSP task by exploiting spontaneous perceptual organization based on nearest neighbour links between nodes. The clusters formed by this nearest neighbour linking can be thought of as providing the initial solution for a TSP, and the TSP task as a hierarchical procedure of finding the most economical pathway connecting the clusters.

There is a growing body of evidence suggesting that such an approach might be able to provide a plausible account of human performance upon the TSP task. First, Vickers et al. (2003) demonstrated that TSP solution difficulty is related to number of potential intersections that a given problem contains. Because nearest neighbour links never intersect, reliance upon nearest neighbours would place a strong restriction on the number of plausible solution pathways within a problem. Furthermore, this would account for the general avoidance of crossings in TSP solutions (van Rooij, Stege, & Schactman, 2003).

Second, Vickers, Bovet, Lee & Hughes (2003) found that the solution length of both open and closed versions of the TSP was reliably correlated with the number of nearest neighbour links that a standard NN algorithm shared with the benchmark solution and Vickers, Mayo, Heitmann, Lee, and Hughes (2004) demonstrated that solution lengths for TSPs, minimum spanning trees and Steiner Trees (a clustering task closely related to the minimum spanning tree), were all highly correlated with a measure of path complexity indexing the 1^{st} to k^{th} order neighbour links within participant solutions.

Third, it has been demonstrated that the time taken to solve a TSP increases linearly as the number of problem nodes increases (Dry et al., in press; Graham et al., 2000). This finding suggests that human participants are only searching through a subset of potential solution pathways and imposes constraints upon the types of heuristics that participants might be employing to solve the TSP task. Because the number of clusters formed by nearest neighbour links increases as a linear function of n (Vickers, Bovet et al., 2003), we could expect the time taken by such a hierarchical nearest neighbour procedure to produce TSP solutions to similarly increase as a linear function of n.

Interestingly, an alternative explanation for human performance on TSPs, the convex hull model (MacGregor et al., 2004; MacGregor & Ormerod, 1996; MacGregor, Ormerod, & Chronicle, 2000), is also broadly consistent with an approach based upon Delaunay triangulation. This approach suggests the initial stage of solving a TSP involves finding the convex hull of the set of points. The convex hull is a boundary, outside of which no points are located. As can be seen in Figures 2.13 and 2.14, the Delaunay triangulation of a set of points always contains the convex hull. Therefore, if the visual system is aware of the relational structure based on Delaunay neighbours, it is also by proxy aware of the convex hull.

Each of the studies reviewed thus far have been primarily concerned with the perceptual organization of random or pseudo-random dot patterns. Delaunay triangulation and its related relational measures have also been applied to stimuli that contain some form of regularity. Of particular interest is the suggestion that the perception of structure in dot patterns such as Figure 2.1 results from the visual system explicitly representing proximity and orientation relations in terms of pair-wise connections between neighbouring stimulus elements. These connections have been variously referred to as virtual lines (Earle, 1991; Marr, 1982; Smits &

Vos, 1986; Stevens, 1978; Stevens & Brookes, 1987; Wagemans et al., 1993), chords (Moore & Parker, 1974) and dipoles (Caelli, 1981; Caelli & Julesz, 1978; Julesz, 1971). Marr (1982) used Glass patterns (Glass, 1969) to illustrate the necessity of virtual lines. Glass patterns are created by making a copy of a random dot pattern, subjecting the copy to some form of transformation and superimposing it on the original. An example of a rotational Glass pattern is given in Figure 2.17a. The individual points in Figure 2.17a have positional information, but do not carry any orientation information. Nonetheless, the stimulus gives a strong impression of oriented structure. It has been suggested that the perception of rotational structure must be due to the orientations of the 'virtual lines' joining the points in the original set to their transformed pair (see Figure 2.17b).

The perception of structure in stimuli such as Figure 2.17 appears to be dependant upon the relative proximity of the original and transformed point sets. For example, it has been demonstrated that the perception of structure in Glass patterns breaks down and eventually disappears as the extent of the transformation is increased (Caelli, 1981; Dakin, 1997; Glass, 1969; Glass & Perez, 1973; Jenkins, 1983b). Similarly, a number of studies have indicated that the perception of structure in symmetrical patterns appears to be restricted to a spatially limited area directly adjacent to the axis of symmetry (Barlow & Reeves, 1979; Jenkins, 1982; Rainville & Kingdom, 2002). These findings suggest that if the visual system is generating virtual lines then they are only being used to connect stimulus features that are located within a limited neighbourhood.

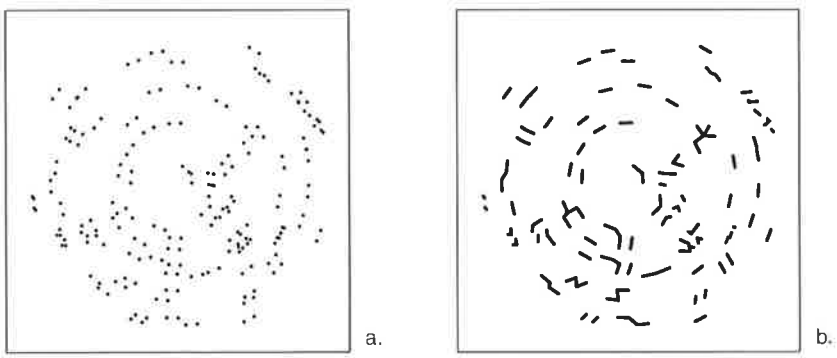

Figure 2.17 An example of a rotational Glass figure (a), and the structure revealed by virtual lines joining the transformed pairings (b).

A common approach to defining the extent of this limited neighbourhood has been in terms of some form of nearest neighbour measure. For example, Caelli (1981) suggested that the perception of structure in Glass patterns is reliant upon the distribution of nearest neighbour distances for the structured or paired dots, relative to the distribution of nearest neighbour distances for the unpaired or noise dots. Dry et al. (submitted) extended this research using a model that compared the orientations of point pairs formed between nearest neighbours, second nearest neighbours, through to k nearest neighbours. They found that performance on a Glass pattern discrimination task was best predicted when the model used just the first and second nearest neighbours. Wagemans, van Gool, Swinnen and Horebeck (1993) also employed a k nearest neighbour approach to defining a neighbourhood for a model of Glass and mirror symmetry perception; however, they did not provide any indication of the range of values of k that produced the best fits to the empirical data. Stevens (1978) and Stevens and Brookes (1987) described a model that compared the orientations of pairs within a neighbourhood defined by the average nearest neighbour distance multiplied by a constant between 1.3 and 1.5. They demonstrate that this model is able to account for the perception of structure in Glass patterns and Marroquin patterns (Marroquin, 1976, see Figure 2.18).

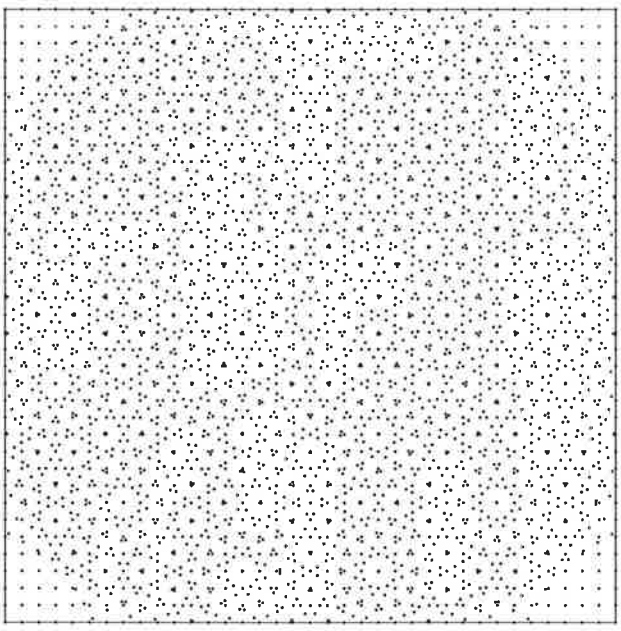

Figure 2.18 Example of a Marroquin pattern (Marroquin, 1976). These patterns are created by superimposing rotated copies of a regularly spaced grid of points. Circular structures appear at different locations, depending on the where the eye is fixated.

It should be noted that a number of alternative explanations have been put forward to account for each of the experimental dot pattern phenomena discussed thus far. For example, the CODE model has been applied to clustering and regularity detection problems (Compton & Logan, 1993; Logan, 1996; van Oeffelen & Vos, 1982, 1983) and it has been demonstrated that filter based models can also provide good fits to empirical Glass pattern data (Dakin, 1997, 1999; Wilson & Wilkinson, 1998, 2003; Wilson, Wilkinson, & Assad, 1997). Although the studies outlined in this section cannot be said to provide conclusive evidence that the perceptual organization of dot patterns is based upon an awareness of the point pairs formed by Delaunay triangulation, they suggest that such an approach is both theoretically and empirically plausible.

2.6. Advantages of a Voronoi based approach to perceptual organisation

It is worth briefly reviewing the evidence that has been presented so far. In the first place, research has been reviewed suggesting that the perceptual system is generating a medial axis representation at an early stage in visual processing. It has been suggested that this may be the result of a uniform growth or activation process based upon horizontal propagation in a layer of V1 (Kovacs et al., 1998). If this process is applied to discrete objects, the resulting representation closely resembles Voronoi tessellation. Furthermore, tracking the activation or growth process as it unfolds over time would provide a means of encoding the relative proximity and orientation relations present in Delaunay triangulation. Empirical evidence suggesting that observers are aware of and exploit the relational information in Delaunay triangulation (and related measures) can therefore be taken as evidence consistent with a Voronoi-like representation.

There are a number of advantages to adopting a Voronoi-based approach to perceptual organisation. In the first place, many perceptual phenomena demonstrate scale invariance: the phenomena remain constant regardless of changes in viewing distance (Pomerantz, 1981). For example, Kubovy, Holcombe and Wagemans (1998) demonstrated that the perceptual organisation of stimuli such as Figure 2.2 was not affected by viewing distance. Similar demonstrations have been made using Glass patterns (Maloney, Mitchison, & Barlow, 1987) and mirror symmetry (Rainville & Kingdom, 2002). It has been suggested that the scale invariance of the perceptual system is a reflection of the statistical structure of the environment, and that it may be a unifying psychological principle (Chater & Brown, 1999).

Importantly, unlike many other measures of relational structure Voronoi tessellation does not rely upon any fixed parameters (Ahuja, 1982; Aurenhammer, 1991), and is therefore also scale invariant. A model of structure detection based upon Voronoi tessellation should show an advantage over models based upon fixed distance tolerances (e.g., Pizlo et al., 1997) or spatial filter scales (Dakin, 1997; Dakin & Watt, 1996; Wilson & Wilkinson, 1998, 2003; Wilson et al., 1997) because it will be able to generalise to new stimuli regardless of their scale, without having to specify a new range of parameter values.

Second, the perceptual organisation of discrete objects into groups appears to be hierarchically structured (Palmer & Rock, 1994). For example, in Figure 2.19 the individual circles are organised into 9 X shapes. The X shapes are in turn organised into 3 small triangles, which together signal the presence of a single large triangle. Marr (1982) suggested that the perceptual organisation of more complex stimuli is built up in a similarly recursive fashion, with individual image primitives (such as blobs and bars) being actively grouped into larger-scale tokens reflecting the larger-scale structures in the image. The hierarchical nesting of the sub-graphs of the Delaunay triangulation (such as can be seen in Figure 2.14) may be able to provide insight into this hierarchical structuring. For example, in Figure 2.19, the presence of the X shaped structure is signalled by nearest neighbour connections, whereas the higher order structures (such as the triangles) are linked by higher order graphs (i.e., the relative neighbourhood graph). Under a Voronoi-based approach to perceptual organization, it could be assumed that the visual system has access to the full set of relational structures revealed by the Delauanay triangulation but that the resulting organisation is to a certain extent dependant upon whether attention is focused on local or global structures. One consequence of such an approach is that it offers a means of unifying the numerous theoretical and computer vision approaches to structure detection based on the various sub-graphs of Delaunay triangulation (Ahuja, 1982; Ahuja & Tuceryan, 1989; Jarvis, 1972; Toussaint, 1980; Zahn, 1971).

Finally, it is assumed that the same process is responsible for generating both the medial axis representation for solid objects and the Voronoi-like tessellation describing the relational structure of discrete objects. This has the advantage of potentially unifying research based upon simple stimuli such as random dot textures and more complex, real-world stimuli.

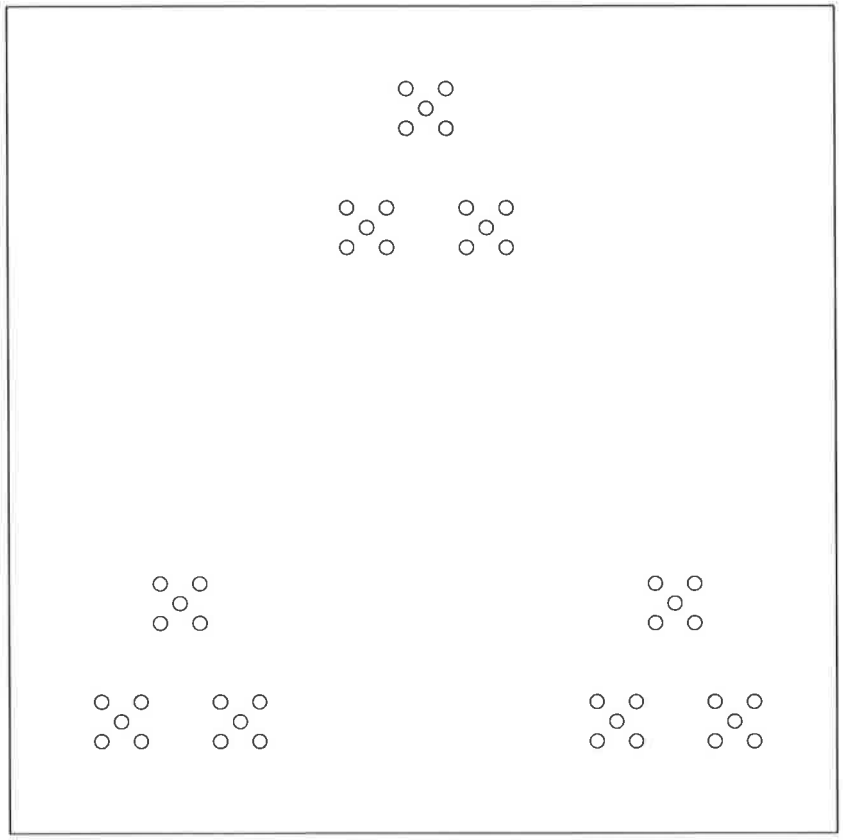

Figure 2.19 An example of the hierarchical structure of perceptual organisation.

2.7. Summary

The spatial organisation of scenes or images is one of the fundamental tasks of early vision. It has been argued that the extraction of spatial relations, such as the relative proximity and orientation of image features, is a necessary step in achieving this task (Marr, 1982; Ullman, 1984). In the field of computational geometry one of the most powerful tools for describing spatial relations is Voronoi tessellation. In this chapter it has been argued that the visual system may be employing a process similar to Voronoi tessellation to extract the spatial relations necessary for the organisation of visual stimuli.

Support for a Voronoi-like representation comes from two separate programs of perceptual research. First, psychophysical and physiological evidence was reviewed suggesting that the visual system is performing a process similar to medial axis calculation. The medial axis

describes the internal skeleton of solid objects, whereas Voronoi tessellation describes the relative proximity of discrete objects, but the two concepts are mathematically equivalent. It was demonstrated that when medial axis-based models are applied to discrete point sets, the resulting tessellation closely resembles a standard Voronoi tessellation.

The second source of evidence supporting a Voronoi-like representation is the body of research suggesting that the visual system has access to the relational structures revealed by Delaunay triangulation. Delaunay triangulation is the dual graph of Voronoi tessellation and having access to one graph provides enough information to generate the dual. It was suggested that the spreading activation or wavefront process that is believed to be generating the medial axis or Voronoi representation (Kovacs et al., 1998; Lee et al., 1998) could also encode the relational structures present in Delaunay triangulation.

Clearly, the evidence presented in this chapter does not constitute irrefutable evidence that a representation similar to Voronoi tessellation is being generated in early vision. Nonetheless, a review of research has been presented suggesting that such a process is theoretically, physiologically and empirically plausible. Furthermore, it has been argued that a Voronoi-based approach provides a powerful tool for understanding the processes underlying perceptual organisation and the perception of structure and regularities such as mirror symmetry.

In the next chapter I will review empirical symmetry perception research and outline a model of mirror symmetry perception based upon Voronoi tessellation. In subsequent chapters this model will be compared to a number of alternative models of symmetry perception.

3. A VORONOI MODEL OF SYMMETRY PERCEPTION

In Chapter 2 it was argued that the visual system may be generating a representation of relational structure in early vision that is similar to the Voronoi tessellation of an image. In the following chapter I will outline a model of symmetry perception that employs a Voronoi representation to detect symmetrical structure in dot patterns. First, I will outline three fundamental characteristics of human symmetry perception. This is followed by a general description of the Voronoi model and a discussion of a number of the model's potential strengths and weaknesses.

3.1. Key characteristics of symmetry perception.

The history of symmetry perception research stretches back more than 100 years to Mach's (Mach, 1886/1959) demonstration that mirror symmetry is more easily perceived than repetition or centric symmetry. Examples of mirror, repetition and centric symmetry are shown in Figure 3.1. Subsequent studies have uncovered a number of characteristics of human symmetry perception and, in an excellent review of research in this area, Wagemans (1997) lists the findings that he suggests have been the most informative in relation to the development of models of symmetry perception. He noted that although these empirical findings have been useful in terms of both informing and constraining the development of models of symmetry perception, the data have not yet converged upon a single universally accepted model. Rather than attempt to provide another comprehensive review of all of the

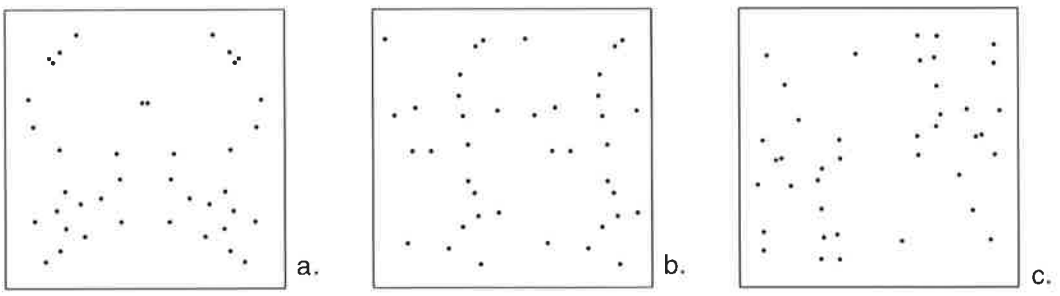

Figure 3.1 Example stimuli containing mirror symmetry (a), repetition symmetry (b), and centric symmetry (c).

empirical research in this area (see also: Wagemans, 1996), I will focus upon three key characteristics of symmetry perception that relate directly to and inform the Voronoi model of symmetry detection.

In the first place, it has been demonstrated that the perception of symmetry is not an all-or-nothing response but shows a gradual deterioration in the presence of added random noise or positional distortion (e.g., Barlow & Reeves, 1979; Jenkins, 1983a; Wagemans et al., 1993). Barlow and Reeves (1979) first demonstrated this effect in a series of experiments in which the discriminability of symmetrical dot patterns was manipulated. In one experiment Barlow and Reeves distorted symmetrical dot stimuli by randomly repositioning one half of each symmetrical dot pair within a square 'tolerance region'. In a second experiment, Barlow and Reeves manipulated the signal to noise ratio of the stimuli by replacing symmetrical dot pairs with randomly positioned points. In both cases observers were asked to discriminate between the symmetrical stimuli and the random dot stimuli, with the results indicating that observer discriminability declined gradually across the range of the experimental manipulations.

Somewhat surprisingly, the experiments revealed that the observers were able to tolerate quite high levels of distortion and interference. In the case of the distortion experiment, the observers were able to discriminate between the two classes of stimuli with a d' of close to 1 for tolerance regions of one third of the squared width of the stimulus. Similarly, for the signal to noise manipulation, the observers were still able to discriminate with a d' of close to 1 when 60% of the symmetrical pairs had been replaced with random noise. Barlow and Reeves' findings have been largely replicated in subsequent studies (e.g., Jenkins, 1983a; Wagemans et al., 1993).

Second, a number of studies have demonstrated that the relative proximity of stimulus elements has an effect upon the perceptual salience of symmetry. In particular, it has been suggested that the perception of symmetry is restricted to a spatially limited region located directly adjacent to the axis of symmetry. In an early study, Jenkins (1982) demonstrated that when symmetrical pairings located close to the central axis were replaced with a strip of random noise, performance on a discrimination task was impaired. As the width of the strip containing random noise was increased, performance decreased steadily and, beyond a given width, the observers were unable to discriminate between the stimuli comprised of the random

strip embedded in an otherwise symmetric stimulus and stimuli comprised entirely of random noise. The results of this experiment, and two similar experiments in Jenkins' (1982) study, suggested that the mechanism underlying the perception of symmetry was restricted to making short-range correspondences between stimulus elements. Additionally, it has been demonstrated that the spatial extent of the centrally located region of enhanced perceptual salience appears to be dependant upon the specific characteristics of the type of experimental stimuli being employed to measure the region's dimensions (Dakin & Herbert, 1998; Rainville & Kingdom, 2002). Dakin and Herbert (1998) demonstrated that the dimensions of the symmetry detection region varied as an inverse function of stimulus peak spatial frequency. The results of an experiment using band-limited textures indicated that stimuli with low peak spatial frequencies had a larger enhancement region than stimuli with high peak frequencies. Rainville and Kingdom (2002) questioned Dakin and Herberts' findings, indicating that spatial frequency had been confounded with stimulus density. Using a methodology similar to that used by Dakin and Herbert but also controlling for stimulus size, the number of points in each stimulus, and stimulus density (i.e., the number of points per degree of visual angle squared), Rainville and Kingdom demonstrated that the size of the enhancement region is determined by stimulus density but not spatial frequency, numerosity, or size.

The third key characteristic of human symmetry perception relates to the generality of the underlying process. There has been ongoing debate in the literature as to whether a specialised process drives symmetry detection, or whether it is the result of a more generalised regularity detection process. For example, early research indicated that vertical symmetry was more salient than symmetry presented at any other orientation (Barlow & Reeves, 1979). Furthermore, the salience of symmetry appeared to be dependant on the axis of symmetry being fixated centrally in the visual field (Barlow & Reeves, 1979; Julesz, 1971). These findings prompted some researchers to suggest that vertical symmetry is detected by a specialised mechanism based on the symmetrical organisation of the cerebral cortex (Herbert & Humphery, 1996; Julesz, 1971). Presumably, this suggests that additional separate mechanisms are needed to detect symmetry with orientations other than vertical.

An alternative approach assumes that a more general mechanism is responsible for the perception of symmetry (Baylis & Driver, 1994; Dakin & Hess, 1997; Dry, 2005; Labonte et al., 1995; Rainville & Kingdom, 2000; Royer, 1981; van der Helm & Leeuwenberg, 1996;

Wagemans et al., 1993; Wilkinson & Halligan, 2003). According to this approach the perception of symmetry is a result of some form of low-level grouping process that operates automatically on all visual stimuli. In other words, it is assumed that the mechanism underlying symmetry perception is not specific to a single type of stimuli but appears to be related to the perception of all other types of visual regularities. As Rainville and Kingdom (2002) indicate, in order to detect symmetry, the visual system can be considered to be performing a type of spatial correspondence calculation whereby it employs the inherent statistical information contained in the stimulus to determine which of the stimulus elements are grouped together or paired. The perception of a number of other visual phenomena, such as apparent motion, Glass patterns, and stereopsis, appear to be reliant upon solving a similar correspondence problem (Papathomas, Kovacs, Gorea, & Julesz, 1995). Furthermore, a number of researchers have demonstrated that the same, or very similar, computational models are able to detect the structure in both symmetrical stimuli and related stimuli such as Glass patterns (Dakin, 1997; Dakin & Watt, 1996; Wagemans et al., 1993). Taken together these studies suggest that symmetry perception is part of a general structure detection process rather than a specialised process.

3.2. A Voronoi model of symmetry perception

The three key characteristics of human symmetry perception outlined above place informative constraints upon the development of a computational model of symmetry detection. In the first place, the model must be able to tolerate reasonably high levels of distortion and interference. Second, the model must include some means of simulating the region of enhanced perceptual saliency immediately adjacent to the axis of symmetry. Third, the model must be based upon a general process that would be capable of detecting other types of structure and regularity, not just symmetrical structure. In light of these constraints, I propose a Voronoi model of symmetry detection. In this section I will describe the basic features and assumptions of this model.

Using Voronoi tessellation as the base representation from which symmetrical structure is detected meets the demands of the generality constraint. In Chapter 2 evidence was presented indicating that the visual system may be generating a representation of relational structure in early vision that is similar to the Voronoi tessellation of an image. Empirical evidence suggests that the relational structures that are revealed by the Voronoi tessellation of a set of discrete elements (i.e., the Delaunay triangulation and its sub-graphs such as nearest

neighbour relations) play an important role in the perception of structure and regularity in a number of different visual phenomena (Pomerantz, 1981; Tripathy, Mussap, & Barlow, 1999; Vickers, Lee et al., 2003; Vickers et al., 2004). Furthermore, these relational structures are integral to a number of computational models of structure detection for stimuli such as Glass patterns (Dry et al., submitted; Stevens, 1978; Stevens & Brookes, 1987; Wagemans et al., 1993). Finally, it has been suggested that the perception of symmetry in solid forms is based upon an awareness of the medial axis (Hong & Pavel, 1996; Kurbat, 1996). Given that the process of generating the Voronoi tessellation for a set of discrete objects would also generate a medial axis for solid forms, basing a model upon this type of representation presents the opportunity of providing a unifying explanation for the perception of symmetry in these two very different classes of visual stimuli.

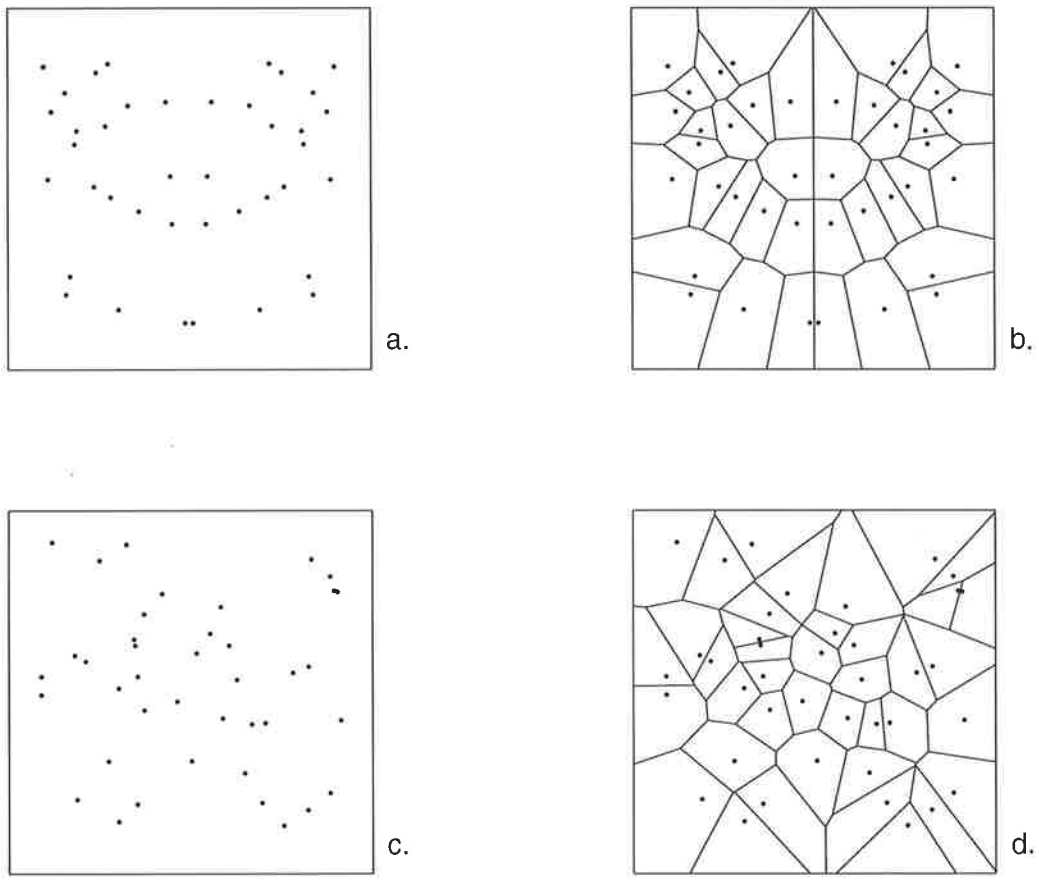

Figure 3.2 Examples of a symmetric dot pattern (a) and its Voronoi tessellation (b), and a randomly distributed pattern (c) and its Voronoi tessellation (d).

Figure 3.2 provides an example of a symmetrical dot pattern (Figure 3.2a) and its resultant Voronoi tessellation (Figure 3.2b). As can be seen, Voronoi tessellation divides the field into a series of cells, with each cell containing exactly one point. Furthermore, because the distribution of dots in Figure 3.2 is symmetrical the tessellation is also symmetrical. In contrast to this, in the case of a random dot pattern (Figure 3.2c) and its Voronoi tessellation (Figure 3.2d), it can be seen that neither the distribution of points nor the tessellation is symmetrical.

Jenkins (1983a) suggested that the perception of symmetry may arise as a consequence of the visual system detecting pairs of points that are uniformly orientated and have collinear midpoints. The Voronoi model can be thought of as a formal instantiation of Jenkins' model. The Voronoi model searches for symmetry by attempting to pair each point in a stimulus with another point located in roughly the same position on the opposite side of the axis of symmetry. For stimuli that contain little or no symmetry, the model will generally be unable to find matching points in areas where they are expected or, alternatively, for any given point the model may find more than one potential matching point. On the other hand, for perfectly symmetrical stimuli and stimuli with high levels of symmetry, the point-pair matching process will tend to find a single unambiguous match for each point.

The matching process operates by reflecting each Voronoi cell across the stimulus midline and counting the number of points that are located in that region. Figure 3.3 compares this process for a symmetrical pattern and a random pattern. The top row of Figure 3.3 shows the Voronoi tessellation for the two patterns. The bottom row shows the reflection of each of the cells across the stimulus midline. For the symmetrical pattern (3.3c) each reflected cell contains exactly one point. However, for the randomly distributed pattern (3.3d) the number of points in the region corresponding to a reflected cell will vary from region to region. As can be seen, for a random dot pattern, the region corresponding to a reflected cell may contain a single point but it may also contain no points, or more than one point.

Barlow and Reeves (1979) demonstrated that the visual system is able to tolerate a reasonably high degree of positional distortion or interference and still detect symmetry. The Voronoi model meets this constraint by assuming that the degree of tolerance afforded to each point in a stimulus is determined by the size of its Voronoi cell. For example, Figure 3.4a shows a symmetrical stimulus that has been subjected to a degree of positional distortion, and its

resulting tessellation. Figure 3.4b shows the same dot pattern but with a reflected tessellation. As can be seen, although the original pattern contains a substantial amount of distortion, the reflected tessellation shows that the majority of cells contain a single point, indicating an unambiguous point-pair match. As the degree of positional distortion in a stimulus is increased the number of unambiguous point-pair matches drops, until it reaches the level found in completely random stimuli (such as in Figure 3.3d). A similar process would occur in symmetrical stimuli to which interference is added in the form of randomly positioned points, with the number of unambiguous matches dropping as a function of increasing noise.

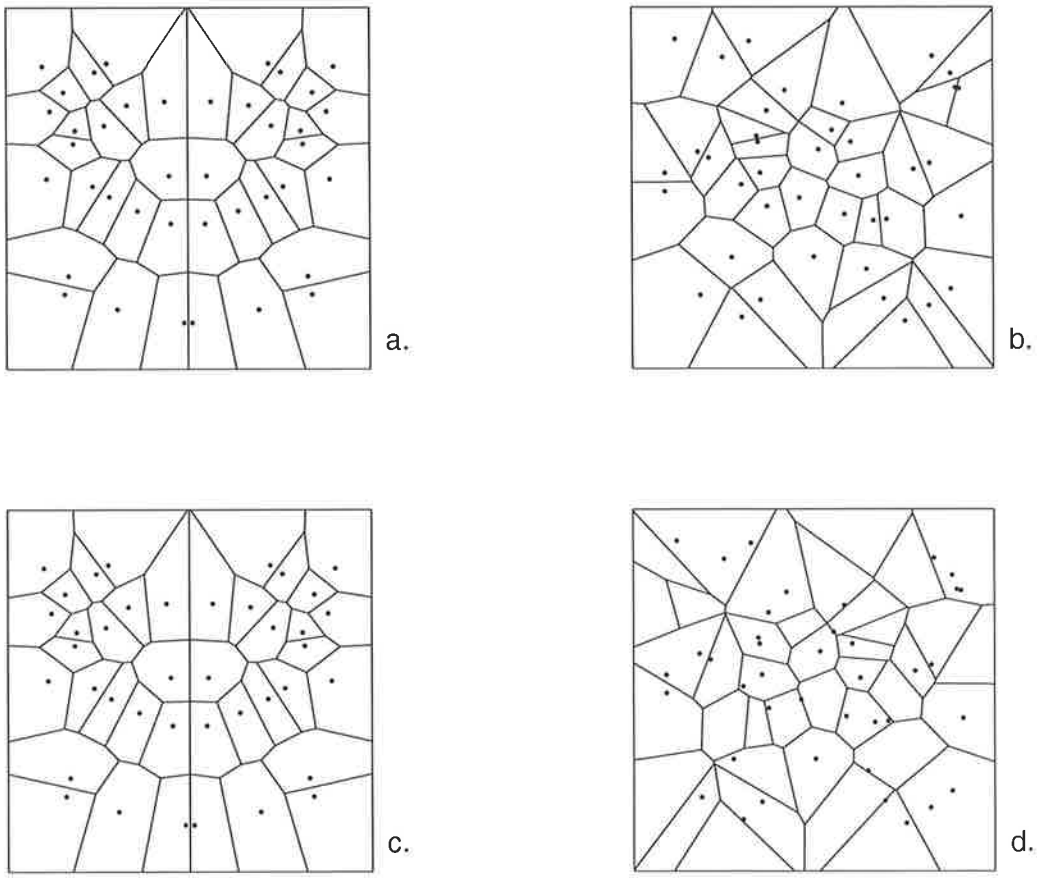

Figure 3.3 The Voronoi tessellation for a symmetric dot pattern (a) and its corresponding reflected tessellation (c), and a randomly distributed dot pattern (b), and its corresponding reflected tessellation (d). For the symmetrical pattern the original and reflected tessellations are identical. In contrast to this, the reflected tessellation of the random pattern does not correspond closely to the original, with very few cells containing a single point.

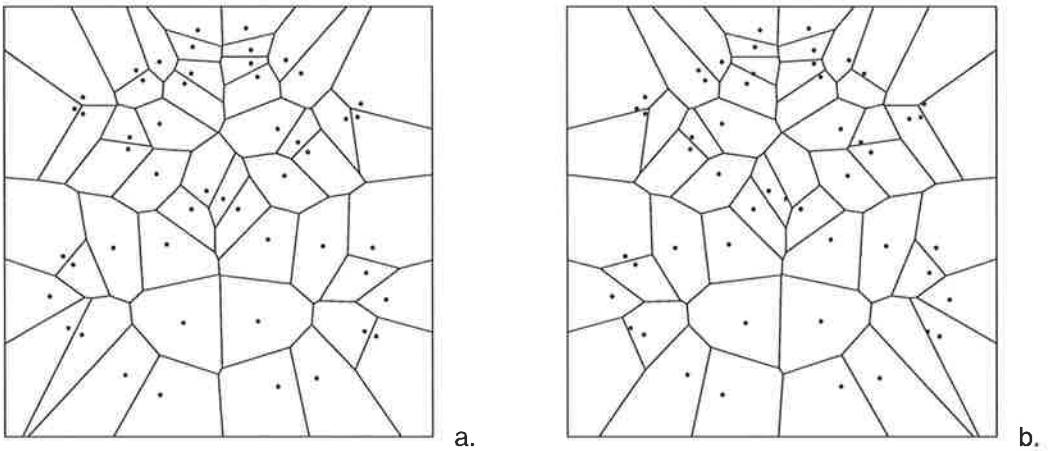

Figure 3.4 An example of a distorted symmetrical stimulus and its Voronoi tessellation (a), and the reflected tessellation (b). The majority of cells in the reflected tessellation still contain a single point.

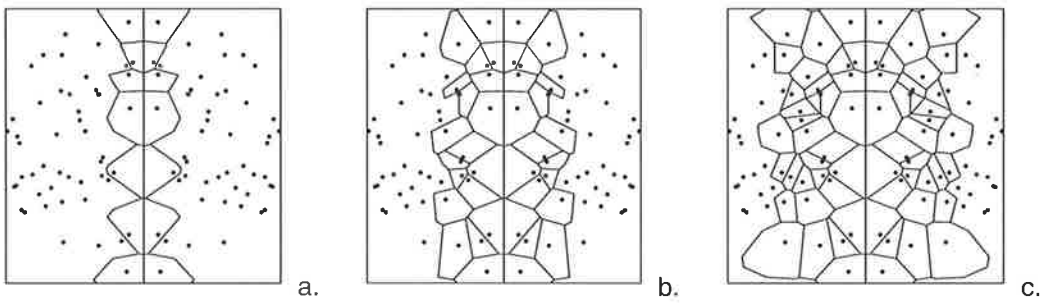

Figure 3.5 An example of a symmetrical stimulus and the region that is considered to contribute to the detection of symmetry, defined by a single column of cells located either side of the stimulus midpoint (a), two columns of cells (b), and three columns of cells (c). The number of columns included in this detection region is the only free parameter in the Voronoi model.

The Voronoi model also allows a plausible means of meeting the constraint of restricting the detection of symmetry to a region directly adjacent to the axis of symmetry (i.e., Dakin & Herbert, 1998; Jenkins, 1982; Rainville & Kingdom, 2002). Each stimulus can be divided into a series of columns of Voronoi cells which can be differentiated according to their relative proximity to the stimulus midline. For example, Figure 3.5a shows the model's representation

of a symmetrical stimulus under the assumption that only a single column of cells either side of the stimulus midpoint contributes to the detection of symmetry. Figure 3.5b shows the representation under the assumption that two columns of cells either side of the midpoint contribute to detection, and Figure 3.5c shows the representation for three columns either side of the midpoint. In each case, only the points that are located within a cell are considered to contribute to the point-pair matching process. It is assumed that all of the unbounded points fall outside of the visual system's zone of attention and are effectively ignored.

Because the extent of this region of detection is unclear, the number of columns included in the symmetry calculation process of the Voronoi model is a free parameter. The minimum size of the region of detection is a single column either side of the midpoint and, at the upper limit, all columns are considered to contribute. This is the only free parameter in the basic version of the Voronoi model. Although the number of columns included in the detection region is free to vary, the size of the region can still be seen to scale with density and is therefore consistent with the findings of Rainville and Kingdom (2002). In other words, for any given column-number assumption, the overall size of the region of detection (relative to the size of the overall stimulus) will vary according to stimulus density. For high density stimuli, the individual Voronoi cells are small, so the detection region will be narrow. Conversely, when stimulus density is low, individual cells are (on average) larger and the region will be broader.

Once the number of columns in the detection region has been defined, an overall level of stimulus symmetry can be calculated. Each cell within the detection region is reflected across the stimulus midpoint and the number of points falling within the reflected cell is counted. For each cell, the number of expected points (i.e., one) is divided by the number of observed points. As has been indicated, it is assumed that this process reflects the visual system performing a form of correspondence calculation in which it is attempting to match up point pairs that have uniform orientation and collinear midpoints. When the number of expected points is equal to the number of observed points, the match is unambiguous. It follows that when the expected and observed values differ there is some degree of ambiguity in the matching process, indicating a lower level of symmetry. The values calculated for each cell are summed giving an overall symmetry measure for the stimulus. For perfectly symmetrical stimuli this value will be high, but as the level of symmetry in a stimulus is perturbed (either

by the addition of random noise or distortion) this value will drop to the level found in purely random stimuli.

3.3. Potential strengths and weaknesses of the Voronoi model.

In the following section I will outline a number of the potential strengths and weaknesses of the Voronoi model. As has been mentioned, in theory the Voronoi model should be able to account for the key empirical findings that the perception of symmetry is a graded response (Barlow & Reeves, 1979) and appears to be restricted to a region directly adjacent to the axis of symmetry (Jenkins, 1982). Furthermore, because Voronoi tessellation scales with stimulus density, the Voronoi model should also be able to explain density-dependant changes in the extent of the detection region (Rainville & Kingdom, 2002).

Data from Rainville and Kingdom (2002) also suggest that stimulus density should have an effect upon the amount of positional distortion that an observer can tolerate and still reliably discriminate between random and symmetrical stimuli. Specifically, as stimulus density rises, the extent of the positional shift that observers can tolerate drops. Rainville and Kingdom's finding is theoretically consistent with the predictions made by the Voronoi model. For example, Figure 3.6 compares two symmetrical patterns with differing dot densities. As can be seen, the average size of the Voronoi cells in the high density pattern (3.6a) is much smaller in comparison to the average size of the cells in the low density pattern (3.6b).

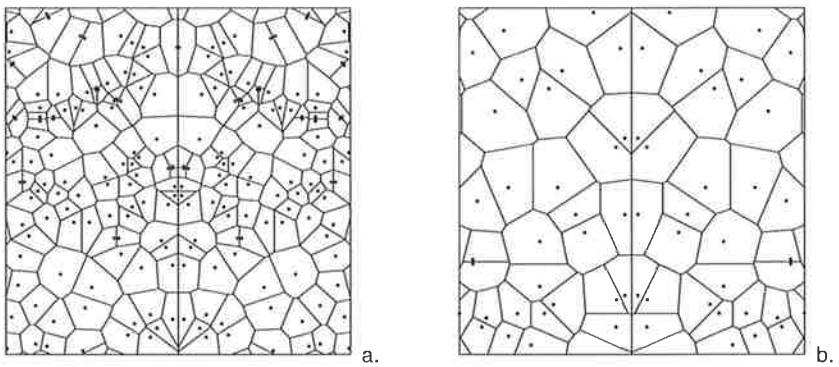

Figure 3.6 A comparison of the Voronoi cells for high density (a) and low density (b) symmetric dot patterns. On average the cells in the high density pattern are smaller than in the low density pattern.

Because the size of each point's Voronoi cell determines the degree of positional distortion that the Voronoi model can tolerate when matching point pairs, then, just like the human observers, the model would be able to tolerate less positional distortion in the high density pattern compared to the low density pattern.

One potential problem for the Voronoi model is that it appears to rely upon an *a priori* knowledge of the location of the axis of symmetry. In other words, it could be argued that the Voronoi model is guilty of a computational sleight-of-hand: in order to detect symmetry the model must reflect cells across the axis of symmetry, and (presumably) knowledge of the location of the axis is reliant upon the perception of symmetry. Wenderoth (1996a) levels a similar charge at Barlow and Reeves (1979) model but it could be applied to nearly all computational models of symmetry detection. This is actually not as large a problem as it might initially seem. First, in most experimental settings, the observers are made explicitly aware of the location of the axis by the presence of some form of fixation point presented immediately prior to the presentation of the stimulus. In this case, there is no reason to assume that the model would not have access to the same information that the human observers have access to.

Of course, this does not explain how the Voronoi model might generalise to situations in which the axis is not explicitly revealed (as, it could be argued, is the case for symmetry perception in real world environments). In this second situation, it is assumed that the model would need to actively seek the reflection line (essentially a proxy axis) that maximises the symmetry within the stimulus. It should be noted that a similar proposal has also been put forward by Dakin and Watt (1996) in relation to their feature alignment model. An active search process might seem computationally unfeasible but there are a number of factors that would serve to constrain this task. For example, the visual system appears to be biased towards detecting vertical symmetry (Barlow & Reeves, 1979; Julesz, 1971; Mach, 1886/1959, etc), but it has been demonstrated that the salience of symmetry presented at varying orientations is to a certain extent modulated by expectation (Wenderoth, 1994). This suggests that a model might plausibly initiate the symmetry detection process by searching for potential axes at the most likely or expected orientations and thereby constrain the number of comparisons that need to be made.

Moreover, it is possible that the visual system might use regularities such as the co-alignment of the short-range connections revealed by Delaunay triangulation to initially signal the presence of the axis, after which longer-range comparisons could be made. For example, in Figure 3.7a, symmetric structure is quickly and easily identified in the stimulus, despite the fact that the axis is not centred. Figure 3.7b shows Delaunay pairings in the stimulus that have uniform orientation and collinear midpoints. A slightly more sophisticated version of the Voronoi model could use this regularity to signal the presence of a potential axis, around which longer-range point-pair comparisons could be made.

Finally, the size of the detection region is also a major constraint upon the number of potential reflection axes that the Voronoi model would be required to search through. In this sense the detection region can be thought of as a localised zone of attention, centred on the current fixation point. It is assumed that the model only has access to the relational information of stimulus features located within this region and, therefore, all other potential axes are not considered in the model's search. This would also explain the results of studies that have

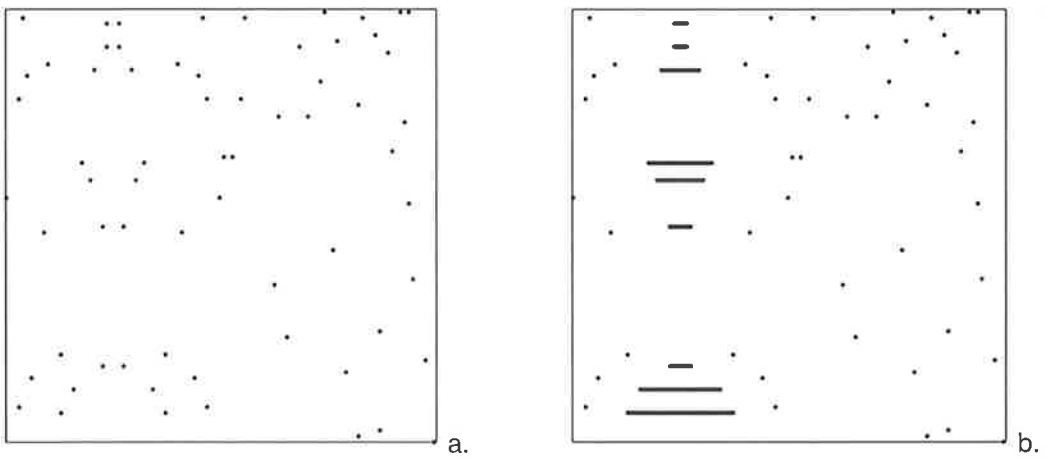

Figure 3.7 Symmetric structure is easily perceptible in the pattern in 3.7a even though the axis is not centred. 3.7b shows the Delaunay pairings in the stimulus that have uniform orientation collinear midpoints. It is suggested that the Voronoi model could be extended to include a search for stimulus features such as these, which would signal the presence of a potential axis of symmetry.

found that the perceptual salience of symmetry decreases as the axis of symmetry is moved further from fixation (Barlow & Reeves, 1979; Gurnsey et al., 1998). Beyond a certain point, the perception of symmetry disappears altogether, presumably because the axis falls entirely outside of the region of detection. Figure 3.8 demonstrates the response of the Voronoi model to symmetrical stimuli with offset axes. Figures 3.8a and 3.8b show the same symmetrical dot pattern with varying degrees of axis offset. In Figure 3.8a the axis is only slightly offset from the stimulus centre (indicated by the X), whereas the pattern in Figure 3.8b shows a considerably larger axis offset. Figure 3.8c and 3.8d show the corresponding detection regions for the two patterns (under the assumption that the detection region is comprised of two

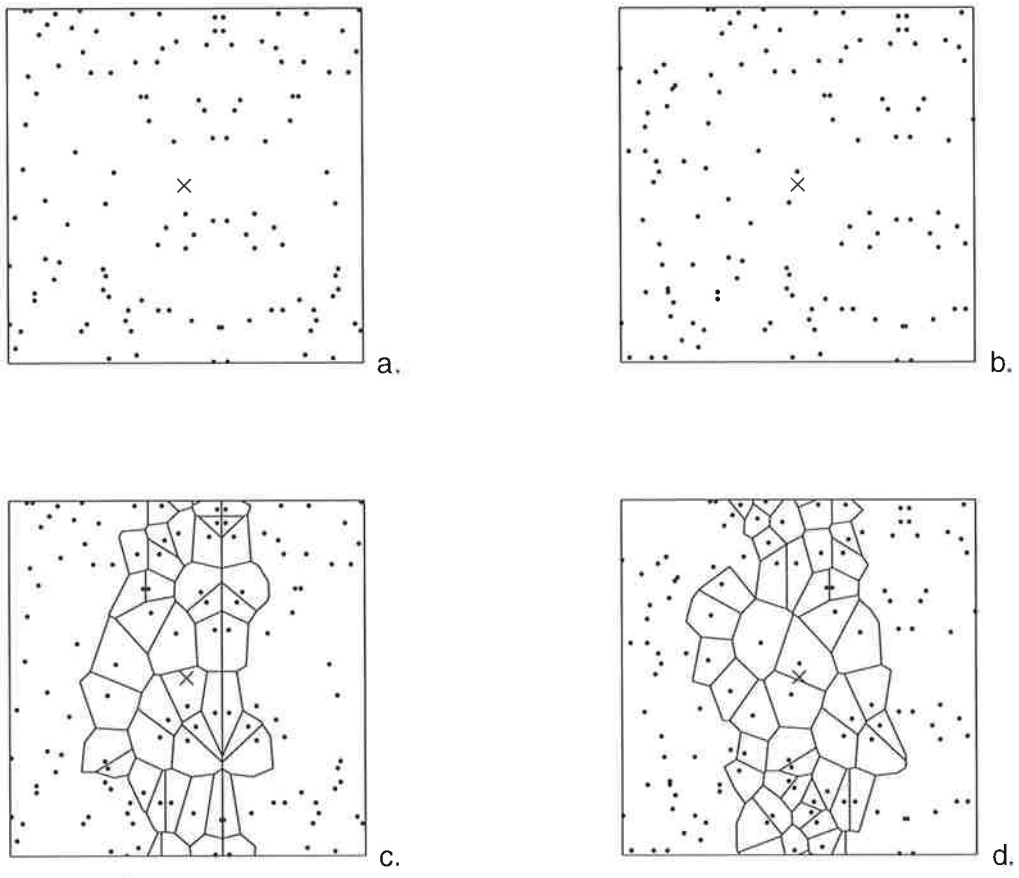

Figure 3.8 An example of the response of the Voronoi model to symmetrical stimuli in which the axis of symmetry is offset from the stimulus midline. The detection region is equal to two columns of cells either side of the midline.

columns of cells either side of the stimulus midline). For Figure 3.8c, the axis falls within the detection zone allowing the model to match up symmetrical point pairs. In contrast, for Figure 3.8d, the detection region contains no symmetrical pairings; therefore, the symmetry would go undetected. Interestingly, the distance of axis offset from fixation at which symmetry detection fails appears to scale with stimulus density (Barrett, Whitaker, McGraw, & Herbert, 1999; Sally & Gurnsey, 2001), a finding that is also consistent with the properties of the Voronoi model.

It is recognised that the Voronoi model's detection region is something of an over-simplification. A number of studies have demonstrated that the properties of the detection region are more complex than was initially suggested by early studies (such as Jenkins, 1982). For example, it has been shown that the extent of the detection region is related to the length of time for which symmetrical stimuli are presented (Tyler, Hardage, & Miller, 1995). However, a subsequent study has also shown that there is a region of enhanced sensitivity located at, or close to, the axis of symmetry that is not dependant on short presentation times (Dry, 2005). Using a methodology similar to Jenkins (1982), Dry (2005) demonstrated that when observers were allowed unlimited presentation times, symmetry was detectable regardless of its location in the stimulus. The results indicated that observer performance peaked when the symmetry information was located directly adjacent to the central axis, but beyond this region performance was relatively stable. The results also suggested the presence of an additional but much weaker region of perceptual salience located at the stimulus outline, a finding that has also been reported in a number of previous studies (Barlow & Reeves, 1979; Wenderoth, 1995).

Additionally, studies by Dakin and Herbert (1998) and Rainville and Kingdom (2002) have provided evidence suggesting an approximately 2:1 height to width ratio for the detection region. In light of this, a detection region comprised of columns of Voronoi cells spanning the entire height of the stimulus may be inappropriate. This, and other weaknesses related to the detection region, could be easily addressed in a more sophisticated version of the Voronoi model without violating the fundamental assumptions of the basic model. For example, it would be simple to implement a version of the Voronoi model in which the height to width ratio of the detection region was free to vary. An example of a Voronoi model detection region with a 2:1 height to width ratio is given in Figure 3.9. Similarly, the contributions that

individual cells within the detection region provide to the overall perception of symmetry could be weighted according to their relative proximity to the stimulus midline (or outline, under the assumption that the detection region covers the entire stimulus). As with the height to width ratio, this would necessarily be a free parameter because there is not yet enough evidence to put a definite value upon the strength of the weighting. However, because these free parameters add complexity to the Voronoi model (see Chapter 4 for a discussion of model complexity) it is believed that initial tests of the Voronoi model's utility should focus upon the simplest version of the model.

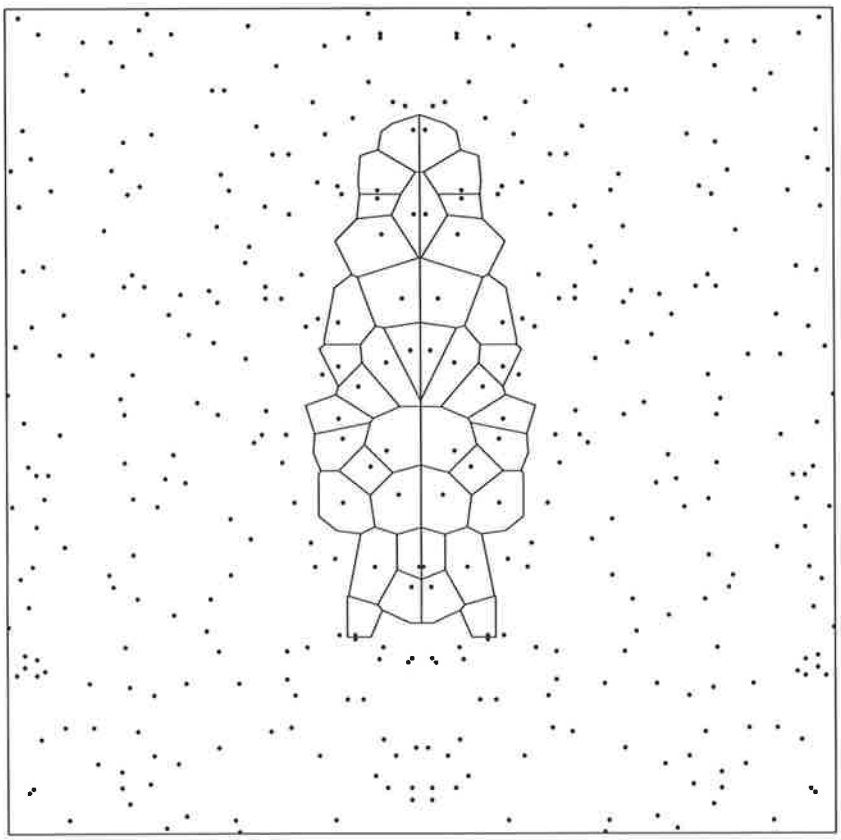

Figure 3.9 An example of the Voronoi detection region with a 2:1 height to width ratio (under the assumption that the width of the region is equal to two cells either side of the stimulus midline).

One final potential problem that faces the Voronoi model relates to the speed at which the visual system is able to generate the underlying Voronoi representation. The process of symmetry perception happens very quickly. Using static dot patterns, Barlow & Reeves (1979) demonstrated that symmetry perception is possible using presentations of 100 milliseconds and Tyler, Hardage & Miller (1995) showed that near-perfect performance could be achieved for presentations as short as 70 milliseconds. Additionally, Julesz (1971) found that detection was possible in dynamic dot textures at 40 to 50 milliseconds and Carmody, Nodine and Locher (1974) employed simple geometric shapes and found symmetry was detectable at 25 milliseconds.

In contrast to this, Lee, Mumford, Romero and Lamme's (1998) research (see Chapter 2) indicated that the peak response corresponding to the location of the medial axis appears at around 80-100 milliseconds. Because it is assumed that the same process is responsible for generating the medial axis and Voronoi representation, this is potentially problematic for the Voronoi model. Nonetheless, although Lee et al.'s study is useful in terms of tracking the changing developmental time-course of a medial axis representation, it does not necessarily follow that this time-course should share a one-to-one relationship with the development of a similar representation in attentive human observers. Furthermore there is evidence to suggest that there are differences between monkeys and humans in terms of the magnitude of the activation induced by symmetrical stimuli. In an fMRI study Sasaki, Vanduffel, Knutsen, Tyler and Tootell (2005) reported that a symmetry response could be found in analogous cortical areas for monkeys and humans, but that the intensity of the response was far weaker for monkeys. In light of this, it could be argued that while Lee et al's (1998) findings pose a challenge to the Voronoi model, they cannot directly rule it out as a potential model of symmetry detection.

3.4. Summary

In this chapter a model of symmetry detection based upon a Voronoi representation was introduced. The Voronoi model was developed in response to three important findings from the human symmetry perception literature: that symmetry perception is a graded response; that the perception of symmetry is more salient at or close to the axis of symmetry; and that the processes underlying symmetry perception can be generalised to other similar structure and regularity detection tasks.

A number of the strengths and weaknesses of the Voronoi model were discussed. It was suggested that the model had the potential to account for a number of different empirical symmetry perception findings. Furthermore, it was argued that some of the potential weaknesses of the model could be overcome by developing the model and that these refinements would not violate the underlying assumptions of the basic model.

The background literature and simple demonstrations provided in this chapter suggest that the Voronoi model is a theoretically plausible model of symmetry perception. However, it is not sufficient to rely on theory or one-off demonstrations to test the utility of a model as an explanation for a cognitive or perceptual process. In order to test the utility of the Voronoi model it is necessary to compare the performance of the model to the performance of human observers across a range of empirical symmetry detection tasks. This will provide an indication of how closely the model responses correspond with the responses of the system it is proposed to be simulating. Furthermore, it is necessary to compare the performance of the Voronoi model to the performance of other models of symmetry perception. This will provide an indication of the utility of the Voronoi model relative to the utility other potential explanations.

In the next chapter I will outline a quantitative methodology for comparing model performance. This will be followed by the first of the studies in this thesis, in which the performance of the Voronoi model is compared to the performance of a number of alternative models across a range of representative symmetry detection tasks.

4. MODEL SELECTION

The primary aim of his thesis is to develop a model of symmetry perception based upon relational information, and to compare it to other models of symmetry perception. In this chapter I outline what is meant by the term 'model' and discuss why models might be of interest to researchers seeking to understand psychological phenomena. A number of philosophical and methodological issues are discussed and a Bayesian approach to comparing models of symmetry perception is introduced.

4.1. What are models?

The term 'model' has been attached to a wide range of applications including flowcharts, simulations, formulae, statistical hypotheses, conceptualisations, and verbal descriptions. In the most general sense of the word a model is a theory and the terms 'model' and 'theory' are often used interchangeably in the literature.

Some researchers find it useful to distinguish between models in terms of degree of specificity or formality. They contrast informal models (such as flowcharts) with formal mathematical models in which specific assumptions are made concerning the interrelationships of the model components (Hintzman, 1991; Uttal, 1990). Unfortunately, this distinction tends to ascribe a degree of complexity to formal models that is not necessarily warranted: a *t*-test is an example of the use of a simple formal model; and the Australian Labor Party's 2001 Knowledge Nation model (Figure 4.1) is an example of a complex informal model. An alternative means of phrasing this distinction is to ask whether the model is sufficiently detailed enough to be, in principal, empirically testable. If the various components of a model are detailed enough that falsifiable predictions can be formulated then the model can be considered to be formal. Following this logic any informal model can (theoretically) be transformed into a formal model.

4.2. Types of formal models

As a blanket term the idea of formal models can be seen to encompass a disparate collection of methodologies. It is worthwhile, therefore, attempting to classify the different classes of formal model found in psychological research.

A primary goal of psychological research is to try to understand the perceptual and cognitive processes that give rise to some form of observable behaviour. One way to differentiate between classes of formal models is to place them on a continuum in terms of the degree to which they are able to provide insight into these internal processes (Figure 4.2). At one end of the continuum are models such as the group comparison hypotheses generally associated with the analysis of variance (ANOVA). These models are able to describe relationships in the data, but are unable to explain how these relationships might arise.

Further along the continuum are mathematical functions such as Weber's law and the models associated with Factor Analysis or Structural Equation Modelling. Here the models are able to provide a more detailed description of the inter-relations between variables, and this generally places stronger constraints upon the types of processes that might be underlying the behaviour in question. For example, a *t*-test may indicate that there is a substantive difference in the number of items retained in memory when measured at two distinct time periods but fitting mathematical functions to an item retention data series tells us more about the nature of forgetting over a continuous time period.

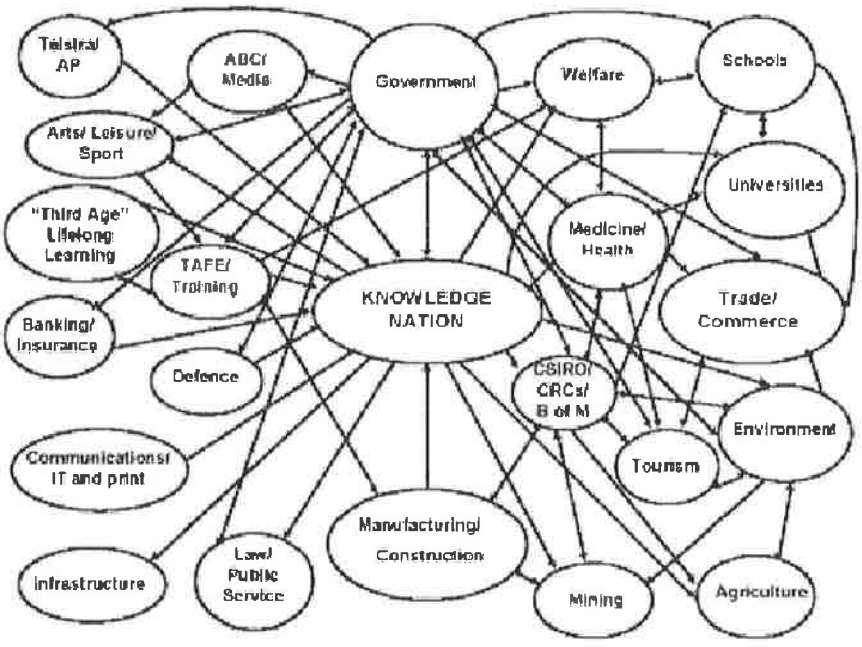

Figure 4.1 The infamous 'Knowledge Nation' model, an example of a complex informal model.

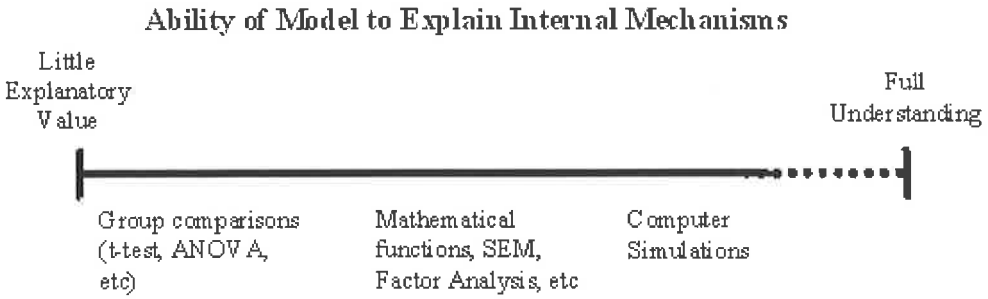

Figure 4.2 A representation of the explanatory value of different types of formal models along a continuum. The models become increasingly specific as they move down the continuum from left to right. Note that no model can provide a full understanding of any internal mechanism.

Further still along the continuum are models known variously as computer simulations, implementations, or process models. These models are an attempt to simulate the processes that give rise to a given behaviour. In other words, these models are a set of rules for transforming some form of input, or experimental stimulus, into an output, or behavioural response. These models differ from the previous two classes of model in that they are generative rather than descriptive. Rather than describing the relationships in an empirical data set, simulations generate data sets which are then (usually) compared to the empirical data set. The degree to which the simulated data correspond with the empirical data is commonly taken as a measure of the degree to which the model corresponds with the actual cognitive or perceptual processes that are supposedly being simulated. Examples of these types of models can be found in most areas of cognitive research including (but not restricted to) language acquisition, memory, decision-making, attention, cognitive abilities, problem solving, and visual perception. It is this class of models with which this thesis is primarily concerned.

Computer simulations are largely considered to be useful because they force researchers to explicitly outline the process or processes that they believe to be responsible for transforming an experimental stimulus into behavioural response (Hintzman, 1991; Lewandowsky, 1993; Uttal, 1990). They increase our understanding of psychological mechanisms by facilitating the transition from vague theories to specific formulations. For example, on the basis of a series of experiments Jenkins (1983a) theorised that symmetry perception arises from the detection

and fusion of stimulus features with orientational uniformity and midpoint colinearity. However, he did not explicitly specify the nature of these stimulus features or how they are 'fused', nor the processes by which orientational uniformity and midpoint colinearity might be detected. Dakin and Watt's (1996) model and the Voronoi model proposed later in this thesis are two very different examples of computer simulations that are partly based upon Jenkins' informal model of symmetry perception.

This high degree of specificity is valuable for a number of reasons. In the first place, it reduces any confusion surrounding what a researcher actually means when they propose a theory concerning a given psychological process. As Broadbent (1987) indicated, theories based on vague and imprecise terminology (such as 'fusion' or 'stimulus features') can be misunderstood or misinterpreted. Second, it is often unclear what sorts of tests or experimental manipulations would be necessary to either support or falsify imprecise theories, whereas the highly specific nature of computer simulations allows them to be replicated and quantitatively compared (Lewandowsky, 1993).

4.3. Model comparison

Uttal (1990) notes that there is a tendency for some researchers to wrongly interpret computer simulations as reductive explanations of the internal workings of the brain. He argues that there are a number of barriers (both practical and philosophical) that place limitations upon the ability of any single simulation to provide "... a precise, unique, and valid statement of the particular internal mechanisms (neural, logical, or cognitive) by means of which a system carries out its function" (p. 189). One particularly salient limitation that Uttal highlights is the black box problem, which he illustrates by quoting Moore's (1956) Theorem 2:

> *Given any machine S and any multiple experiments performed on S, there exist other machines experimentally distinguishable from S for which the original experiment would have had the same outcome.*

This theorem suggests that there are numerous unique, distinguishable models that are capable of generating an empirical data set equally well. Given that the brain is a closed system, it is impossible to determine which model is the 'true' model.

Uttal's objective was not to argue against the use of computer simulations in cognitive research, but to clarify what simulations can and cannot tell us about internal mechanisms. He

indicated that it is important to recognise that any given simulation is just one of a set of possible theories. If we are willing to accept this, then it follows that any simulation, in and of itself, can tell us little about the true nature of the internal processes that give rise to a behavioural response. In other words, the output of a simulation is only of value in relation to other data.

It is common for authors to present examples of model output and rely upon scrutiny by the reader to determine if it accords with their perception of structure in a stimulus (e.g., Ahuja & Tuceryan, 1989; Bonneh, Reisfeld, & Yeshurun, 1996; Hong & Pavel, 1996; Kurbat, 1996; Labonte et al., 1995; Osorio, 1996; Stevens, 1978; Stevens & Brookes, 1987; Zahn, 1971). Typically the model output is shown side by side with the original stimulus such as in Labonte et al. {, 1995 #245; see Figure 4.3}. Less commonly, authors will present some form of function indicating model performance across a range of experimental stimuli. For example, Gurnsey et al. (1998) demonstrated that the performance of their symmetry detection model degraded as the proportion of noise within a symmetrical stimulus is increased. However, Gurnsey et al. did not provide any empirical data to compare the model performance against and presumably the sigmoid shape of the model performance function is meant to be taken as evidence of the model's plausibility. In each of these cases the examples

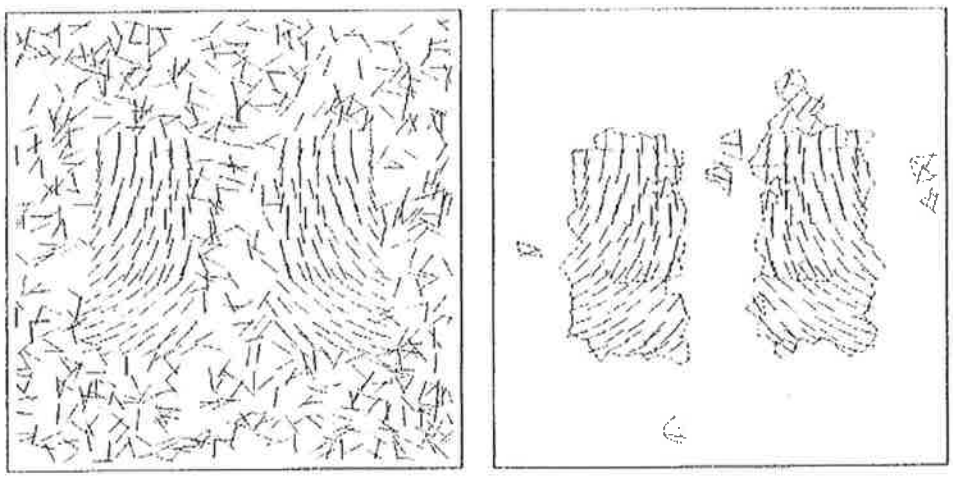

Figure 4.3 Example model output from Labonte et al. (1995). The original text reads "Original image ... and the corresponding results of the grouping process (with border elements)" (pp 282).

that the authors present are of limited value beyond serving to demonstrate that the model in question is capable of producing some form of output. This may be of interest in its own right, and is an excellent first step, but it does not offer any particular insight into psychological processes.

At the very least, simulation output should be compared with appropriate empirical data. This (most basic) test will indicate whether or not the model is capable of capturing the general qualitative shape of an observer's responses. However, it is important to never lose sight of the fact that any model is just one of a set of potential explanations for the phenomena in question. An even better test, therefore, is to compare the output of two or more rival models to an empirical data set. In this way the utility of each model can be seen relative to the utility of the rival explanation(s). Given that we cannot distinguish the true nature of any psychological process, the next best thing is to be able to distinguish between plausible and implausible explanations (Uttal, 1990).

4.4. Evaluation criteria

In order to compare models it is necessary to adopt some forms of evaluation criteria. Jacobs and Grainger (1994) suggested that models should be judged in terms of descriptive adequacy, complexity, generality, and explanatory adequacy. In the following section each of these issues will be dealt with in turn.

Descriptive adequacy is essentially a measure of a model's goodness-of-fit. It quantifies how closely a model is able to approximate an empirical data set. Typical measures of descriptive adequacy that can be found in the literature include the sum of squared errors (SSE), proportion of variance accounted for by the model (R^2), and the maximum likelihood, in which the likelihood of observed data having arisen due to a particular model is maximised in relation to the model's parameters (Myung & Pitt, 1997). Underlying the use of descriptive adequacy as a measure of model utility is the assumption that the closer a model can fit an empirical data set, the better the model is as an explanation of the process that generated the data.

Because it is easily quantified, descriptive adequacy is the most widely used and accepted model evaluation criterion. Nonetheless, in recent years there has been a growing criticism of

the use of goodness-of-fit as the sole measure of model utility (Lee, 2004; Myung & Pitt, 1997; Pitt & Myung, 2002; Roberts & Pashler, 2000). As Roberts and Pashler (2000) noted, when a model has free parameters, the best fit of a model to a data set is just one example of what the model predicts. The single best fitting parameterisation of a model does not give an indication of everything that the model does and does not allow. In contrast to this, *complexity* encompasses the entire range of responses that a model is capable of producing, taking into account the number of free parameters a model has, the values that these parameters are allowed to take, and the way they interact (also known as functional form).

Complexity is an important issue in model evaluation. It has been demonstrated that complex models are able to produce good fits to data sets even if they are not responsible for having generated the data (Myung & Pitt, 1997; Pitt & Myung, 2002). In other words, if goodness-of-fit is used as the sole criterion for model selection, researchers run the risk of selecting a complicated model over a simpler model even if the simpler model is actually a better explanation for the psychological mechanism of interest. Furthermore, the increased flexibility of complex models means that not only do they tend to fit data more closely than a simple model; they are also more likely to be able to make predictions that are implausible, such as response curves that move in the opposite direction to the empirical response curves. When considering the range of responses that highly complex models can make it is sometimes difficult to determine if there are any relationships that the model *wouldn't* be able to predict. If this is the case, then the ability of the model to closely fit an actual empirical data set is hardly interesting, and not at all informative.

Figure 4.4 provides an example of the descriptive adequacy (top row) and complexity (bottom row) of three fictitious models. Model A is a simple model that manages to capture the basic sigmoid shape of the empirical data but its overall descriptive adequacy is poor. Model B has better descriptive adequacy than Model A but the greater number of responses that it is able to produce is an indication of its increased complexity. Model C is able to fit the empirical data set perfectly, but it is also produces numerous responses that bear little resemblance to the empirical data. If we were to evaluate these three models purely on the basis of descriptive adequacy, then the perfect fit of Model C would make it the clear winner. However, the range of responses that Model C produces across all of its parameters suggests that it is overly complex because it produces responses that we would not expect the mechanism underlying

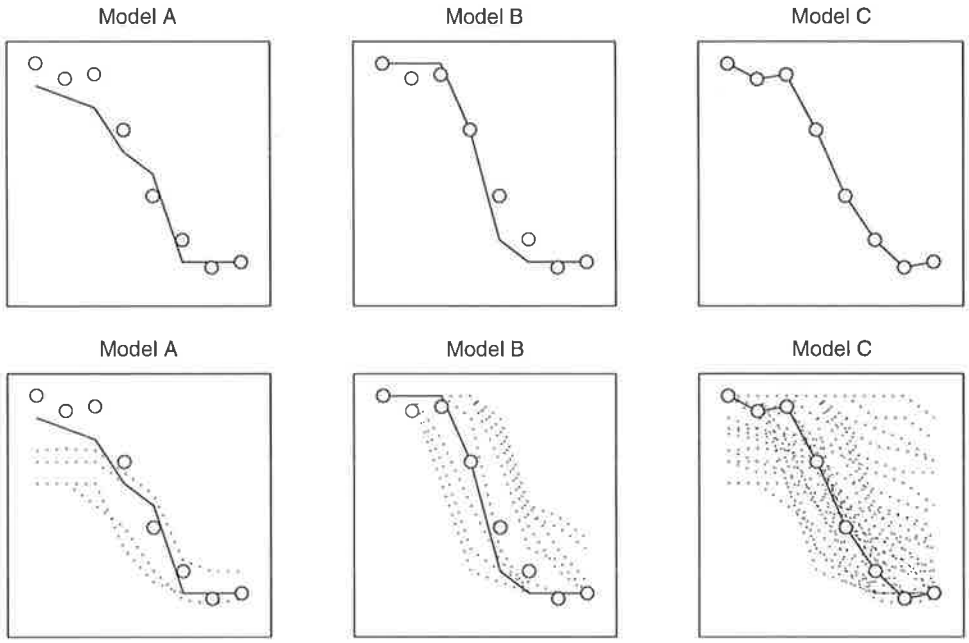

Figure 4.4 Descriptive adequacy (top row) and complexity (bottom row) of three fictional models to an empirical data set. The empirical data is represented by circles. The best fitting parameterisation for each model is represented by a solid line and all other parameterisations by a dashed line.

the empirical data to be capable of producing (such as near perfect performance). Although the fit to the empirical data is not as good, Models A and B are less complex than Model C because they produce a limited, and more plausible, range of responses.

Generality refers to the ability of models to predict new (as yet unseen) data sets (Pitt, Myung, & Zhang, 2002) and to perform well across different stimuli, tasks and response measures (Roberts & Pashler, 2000). The presence of random noise in psychological data means that it is particularly important for models of psychological processes to be able to generalise across data sets. Although variance in the form of individual differences, random participant error, or the fluctuation of attentional processes can be minimised or controlled for in the experimental design, it is difficult (or impossible) to completely eliminate. Comparing models across data sets provides a measure of the model's ability to match the variance in

performance that is due to the underlying psychological mechanism of interest rather than the nuisance or noise variance.

Furthermore, model performance should not be dependent upon a single type of stimulus or task. For example, a number of different experimental manipulations have been employed to determine the properties of symmetry perception. Additionally, there are a number of differences between the types of stimuli used in the experiments (such as stimulus density, element size, position of stimulus in visual field, etc). It is reasonable to assume that a good model of symmetry detection should be able to generalise across these different tasks and the different types of stimuli employed in them.

It is also useful to consider the ability of models to generalise across different types of response measures. This is particularly important when researchers have employed reduced indices such as threshold estimates or *d'*. Model fits based upon these measures may be misleading because the measures are derived by reducing a more detailed data set to a single datum. For example, a model might be able to match an empirical threshold, but the shapes of the model and empirical performance functions from which the thresholds are derived might be very different. Similarly, model and empirical *d'* might be the same, but the hits and false alarms used to generate these measures could be quite different when comparing model and empirical data. Because it is common to fit models to previously published data sets it is not always possible to obtain access to the 'raw' performance measures. However, by fitting models to data sets employing different types of response measure it is still possible to determine if a model has more than limited utility.

The last of Jacobs and Grainger's (1994) model evaluation criteria is *explanatory adequacy.* This refers to the plausibility of a model and asks whether its underlying assumptions are in agreement with conventional theory and knowledge. Of the four criteria, it is the hardest to apply but it is arguably the most important. A number of authors have suggested that the quantifiable criteria are useful tools in model selection but they should not replace explanatory adequacy as a means of determining the utility of a model (Browne & Cudeck, 1992; Myung & Pitt, 1997). The reason for this is simple enough: a model may be able to satisfy the demands of fit, complexity and generality and yet be based upon physically or biologically implausible assumptions (for example, see Barlow & Reeves (1979) model of

symmetry detection which relies upon the visual field being divided into a number of equal-sized square regions).

The obvious difficulty with using explanatory adequacy is that it is subjective and open to interpretation. Furthermore, new models are sometimes controversial and may be in direct conflict with established theory and knowledge because the established theories are wrong (i.e., it was once common knowledge that the world was flat). Nonetheless, explanatory adequacy is still a useful evaluation criterion for a number of reasons. In the first place, it can be used to rule out any highly implausible or logically impossible models. Second, it forces researchers to avoid the use of *ad hoc* assumptions when developing a model (Jacobs & Grainger, 1994). In particular, all model parameters should be meaningfully interpretable, not just some form of distributional assumption designed to increase goodness-of-fit.

4.5. Bayesian model selection

In the following section I outline a Bayesian approach to evaluating models of symmetry perception. This approach accounts for the competing pressures of descriptive adequacy, complexity, and generality and provides a plausible and logical means of interpreting the relative utility of the models under comparison. Within psychology, Bayesian approaches have previously been applied to the comparison of models of a number of different phenomena including memory retention (Lee, 2004), information integration (Myung & Pitt, 1997), individual differences in cognition (Lee & Webb, 2005), response time distributions (Rouder et al., 2005), decision making axioms (Myung et al., 2005), stimulus representation (Navarro & Lee, 2004), and working memory failure (Carlin et al., 1992). Although I will be applying this methodology to models of symmetry perception, there is no reason why it could not be used to compare any formal models of visual perception.

One of the problems with the literature on model selection techniques is that it generally assumes that the reader has a high level of mathematical knowledge. This is unfortunate, because the use of esoteric symbols and highly technical language tends to make the techniques appear complex and inaccessible to non-specialists. The aim of this section is to provide a non-specialist reader with a basic understanding of the Bayesian approach to model selection. First, I will use a simple example to outline how Bayesian model selection operates in a general sense. Following this, I will discuss the theory underlying Bayesian model selection in more depth and provide the necessary technical details.

4.6. An example application of Bayesian model selection

Introductions to Bayesian model selection (hereafter BMS) usually begin with a discussion of Bayes' theorem. Although this theorem is undoubtedly central to BMS (and will be outlined shortly), I believe it is useful to begin at a more practical level.

Let us imagine that we have collected an empirical data set that demonstrates an increase in discriminability across some form of experimental manipulation. As has been discussed in the previous sections, one way to gain insight into the psychological processes underlying this relationship is to compare the performance of two or more models that seek to simulate these processes. In the following example I will demonstrate how BMS can be used to make this comparison.

At a basic level, BMS operates by comparing likelihood functions. Specifically, we are interested in the likelihood that an observed empirical data set was generated by a model, taking into account all of the model's parameterisations. Figure 4.5 compares the fits of two simple one-parameter models (Model A and Model B) to our imaginary empirical data set and their associated likelihood functions. In each case, the model's sole parameter takes seven discrete values. The top left panel (4.5a) shows the fits of each parameterisation of Model A to the empirical data. Each of the seven parameterisations is represented by a separate line, and the empirical data are represented by the circles. As can be seen, parameter value four provides the best fit to the empirical data. Correspondingly, parameter value four contributes the most to Model A's likelihood function (4.5b), and this peak is known as the maximum likelihood (Myung, 2003). In contrast to this, Model A's parameter values one, two, six and seven provide fits that are very poor and as a result they contribute very little to the likelihood function. The bottom left panel (4.5c) shows the fits of each parameterisation of Model B to the empirical data. In this instance the fits of parameter values one and seven are similarly poor as those of Model A but the other parameterisations provide fits that are all clustered closely around the empirical data. As a result, the corresponding likelihood function (4.5d) has a much greater density than Model A's likelihood function.

In this example the likelihood functions are represented by a curve because the models only have one parameter. For two parameter models the likelihood function becomes a surface located above a two-dimensional parameter space. Similarly, but more difficult to visualise,

for *k*-parameter models the function becomes a surface located above a *k*-dimensional plane. By integrating across each of the values in the likelihood functions we obtain the marginal likelihood. The marginal likelihood is essentially equal to the area under the curve (or surface) representing the likelihood function (Myung & Pitt, 1997). Although Models A and B have equal maximum likelihoods (the peaks of the likelihood functions are equal) the marginal likelihood of Model B is greater than that of Model A. Once we have obtained the marginal likelihoods of each model we can compare them in the form of Bayes' factors. These express the likelihoods of the two models in the form of a ratio. If the ratio given by the Bayes factors for Models A and B were 10:1 then Model A would be ten times less likely to have generated the observed empirical data than Model B.

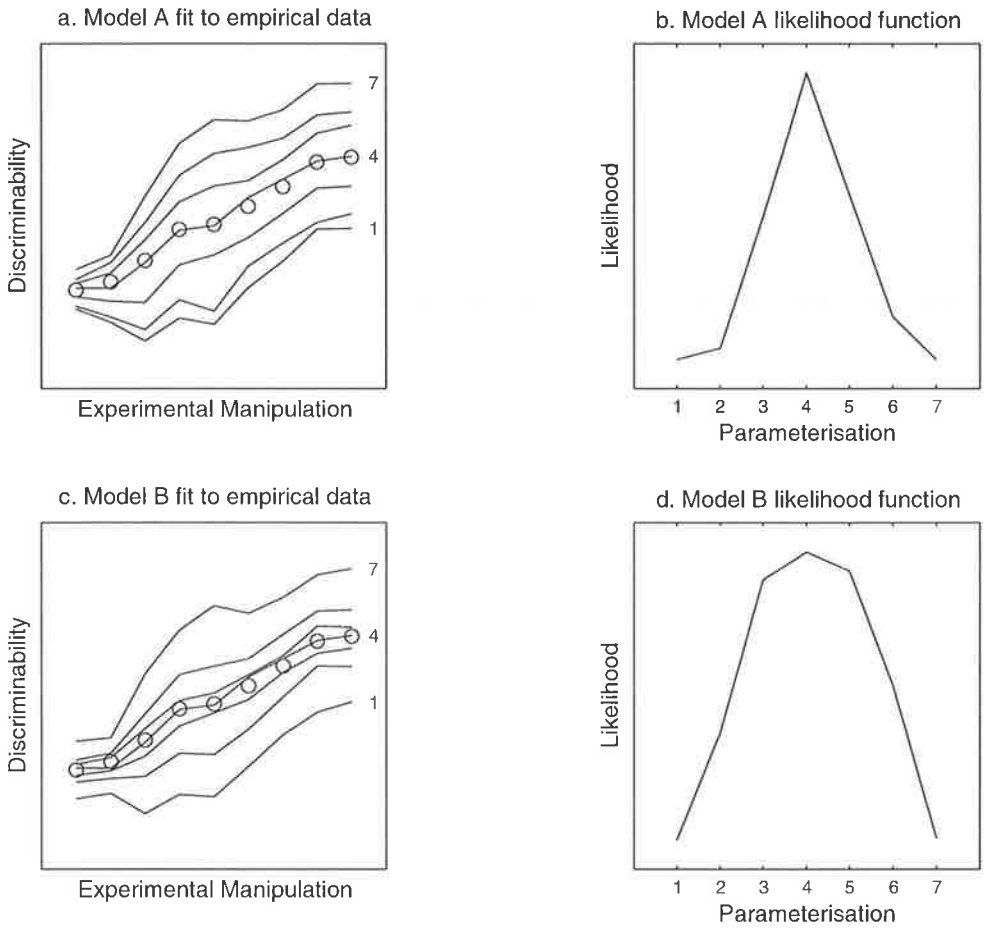

Figure 4.5 The fits of Models A and B to an empirical data set and their associated likelihood functions (see text for details).

This example highlights a number of the advantages of BMS. In the first place, it demonstrates the way that BMS deals with complexity. Some alternative model selection methods (such as the AIC (Akaike, 1983), BIC (Schwarz, 1978) and RMSEA (Steiger, 1990)) penalise models with a large number of parameters more than models with few parameters by adjusting the best fitting parameterisation according to some form of weighted parameter count. However, the number of model parameters is just one aspect of complexity. Our example demonstrates that models with the same number of parameters can have different functional forms. BMS accounts for this additional source of complexity by basing the overall likelihood of the model not only on the best fitting parameterisation but on all parameterisations. Model B is a better model than Model A because more of its parameterisations provide a close fit to the data. We could extend our example by also considering a model with greater parametric complexity. The best fits would probably be closer than those of Models A and B, but it would also be likely to produce numerous parameterisations that do not fit the data well (in a similar manner to Model C in Figure 4.4). The marginal likelihood of the new model would not necessarily be any greater than that of the simpler models because the numerous poorly fitting (and therefore unlikely) parameterisations contribute just as much to the marginal likelihood as the small number of closely fitting parameterisations.

A second advantage of BMS is that it provides a philosophically sound and interpretable means of evaluating the utility of a model. Thus, the ability of a single model to fit a data set tells us very little about the actual psychological processes that generated the data. Similarly, in BMS a model's marginal likelihood is meaningless without reference to another model's marginal likelihood. In BMS it is explicitly understood that the utility of a model is quantified relative to the model or models with which it is being compared. In our example Model B is a more likely explanation of the processes that generated the data than Model A but this tells us nothing about the relative likelihood of Model B in comparison to the universe of potential models that exist as explanations for the data. By expressing model utility in the form of a likelihood ratio (Bayes' factors) we can satisfy a fundamental philosophical requirement of model evaluation (i.e., judging the utility of a given model as an explanation of a psychological process relative to the utility of alternative explanations). Furthermore, because

Bayes' factors lie on a scale that can be understood in terms of betting they provide an easily interpreted and understood means of quantifying the relative likelihoods of models.

4.7. Fundamentals of Bayesian model selection

In the next section I will discuss the theoretical basis of the Bayesian approach and outline in more detail the methodology I will be employing. For a more technical discussion of Bayesian methods see Myung and Pitt (1997), or Kass and Raftery (1995). The following outline of BMS is for the most part drawn from these two sources and the highly accessible and concise overview in Lee (2004).

The Bayesian approach to model selection uses Bayes' theorem to determine the probability that a given model generated a data set. This probability is known as the posterior probability $p(\mathrm{M} \mid \mathrm{D})$ and is dependant upon the way that the prior probability of a model $p(\mathrm{M})$ is affected by the evidence provided by the data:

$$p(\mathrm{M} \mid \mathrm{D}) = \frac{p(\mathrm{D} \mid \mathrm{M})}{p(\mathrm{D})} p(\mathrm{M}).$$

The evidence provided by the data can be seen to be the probability that the data was generated by the model $p(\mathrm{D} \mid \mathrm{M})$ normalised by the probability that the data could have arisen under any set of circumstances $p(\mathrm{D})$.

When we compare two models (M_1, M_2) the relationship takes the form of a posterior odds ratio:

$$\frac{p(\mathrm{M}_1 \mid \mathrm{D})}{p(\mathrm{M}_2 \mid \mathrm{D})} = \frac{p(\mathrm{D} \mid \mathrm{M}_1)}{p(\mathrm{D} \mid \mathrm{M}_2)} \cdot \frac{p(\mathrm{M}_1)}{p(\mathrm{M}_2)}.$$

This is essentially the ratio of the evidence provided by the data (or Bayes' factor) multiplied by the prior odds ratio. It should be noted that the normalising term $p(\mathrm{D})$ is absent when Bayes' theorem takes this form. The reason for this is that the probability of the data having arisen under any set of circumstances is the same for all models. Similarly, the prior odds ratio is only included in this relationship because it has theoretical significance. The prior probability of a model may not exist or it may be impossible to calculate, so a common simplifying assumption in BMS is that the prior probabilities are uniform (Myung & Pitt, 1997). Although there are circumstances in which model priors can be defined (Kass &

Raftery, 1995), in the case of the models compared in this thesis there is no theoretically justifiable reason to suggest anything other than equal priors (i.e., $p(M_1)/p(M_2) = 1$).

Because the models are parameterised, $p(D \mid M)$ becomes a marginal probability: the probability that the data were generated by a model, across each of the model parameterisations $p(D \mid M,\theta)$. In order to compare two models we need to obtain a probabilistic fit between each model and the data across all possible parametric combinations. Returning to the example application of BMS in Figure 4.5, we can see that the fit between model and data is the difference between some measure of observed empirical discriminability, y_i, and the discriminability that the model predicts, $\hat{y}_i$, over the n data points. Under the assumption that each of the empirical data points comes from a Gaussian distribution with a known variance s^2, the probabilistic fit for single model parameterisation is given by:

$$p(D \mid M,\theta_i) = \prod_{i=1}^{n} \frac{1}{(2\pi s^2)^{\frac{1}{2}}} \exp\left(-\frac{(y_i-\hat{y}_i)^2}{2s^2}\right)$$
$$= \frac{1}{(2\pi s^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2s^2}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2\right).$$

The set of probabilistic fits between model and data across each parameterisation provides the likelihood function (such as in Figure 4.5b or d), and by integrating across these values we obtain the marginal likelihood. As we have seen, the ratio of the marginal likelihoods is the Bayes' factor.

There is no hard and fast rule for interpreting Bayes factors. As Kass and Raftery (1995) indicate, interpretation may depend upon context: the weight of evidence needed to send an accused criminal to the electric chair is (hopefully) going to be different to that needed to choose between eating a chicken sandwich or a cheese sandwich for lunch. Jeffreys (1961) guidelines are outlined in Table 4.1 and these will be used as the framework for evaluating differences between the models under comparison in this thesis.

Table 4.1 Jeffreys' (1961) suggested interpretation scale for Bayes' factors.

Bayes Factor	Evidence
< 3.2	Not worth more than a bare mention
3.2 to 10	Substantial
10 to 100	Strong
> 100	Decisive

4.8. Summary

In this chapter I have provided a general overview of the different classes of model employed in psychological research and indicated that this thesis is primarily concerned with the class of models known variously as computer simulations, implementations, or process models. These types of models are generally considered to be interesting because they make highly specific hypotheses about the nature of psychological mechanisms.

It has been argued that, in and of itself, a single model of perceptual or cognitive processes can tell us little about the actual psychological phenomena that it is said to be simulating. Because the true state of these processes is unknowable, the only meaningful interpretation of a model's utility is in relation to the utility of rival models. It was suggested that in order to gain insight into the processes underlying symmetry perception it is necessary to not only physically implement symmetry perception models, but also to compare the performance of rival models in relation to empirical data.

Towards this end Jacob and Grainger's (1994) criteria for evaluating models (descriptive adequacy, complexity, generality, and explanatory adequacy) were outlined and discussed in detail. Finally, a Bayesian approach to evaluating models of symmetry perception was introduced. It was argued that this approach accounts for the competing pressures of descriptive adequacy, complexity, and generality and provides a plausible and logical means of interpreting the relative utility of the models under comparison.

In the following chapter I will present the first of three studies in which Bayesian model selection is employed to compare the performance of the Voronoi model against those of a number of previously published models of symmetry perception across a range of empirical symmetry detection data sets.

5. STUDY 1

5.1. Introduction

A number of models of symmetry perception have been proposed in recent years and, in some cases, it has been demonstrated that a model is able to provide a reasonable account of the performance of human observers on an empirical symmetry detection task (see, for example, Barlow & Reeves, 1979; Rainville & Kingdom, 2002; Tjan & Liu, 2005). Generally speaking, if a model is able to replicate the performance of human observers, then it is considered to be a good potential explanation for the types of processes responsible for the perception of symmetry. Unfortunately, it is often quite difficult to assess the utility of these models relative to each other, because little attempt has been made to compare the performance of different models on the same task. Furthermore, because the comparison of model and empirical data tends to be restricted to a single task or experimental manipulation, it is difficult to determine how well a model is able to generalise across different situations.

A notable exception is Dakin and Watt's (1996) study. In this paper Dakin and Watt compare the performance of four theoretically motivated models of symmetry detection across a series of six previously published experimental data sets from Barlow and Reeves (1979) and Jenkins (1982). They employ a 2 x 2 design, implementing two forms of spatial filtering mechanisms (isotropic and oriented filters), and two forms of symmetry detection (cross-correlation and feature alignment). After comparison of the four models it is concluded that the alignment model with oriented filters provides the best account of human performance across the six symmetry detection experiments. More generally, Dakin and Watt argue that this finding has important implications for human form vision because it suggests that the detection of salient structure and objects in real world environments is realised via the output of oriented filters and a general-purpose feature alignment mechanism.

However, there are three fundamental problems in the Dakin and Watt (1996) study that cast some doubt upon the reliability of these conclusions. In the first place, Dakin and Watt base their conclusions about the relative fits of the models purely upon qualitative inspections of the data fits. At no point do they provide any statistical evidence to support their conclusions. Second, there appears to be a discrepancy between the method employed by Dakin and Watt to calculate *d'* (i.e., stimulus discriminability) and the method employed in Barlow and Reeves' (1979) original studies. This methodological difference causes Dakin and Watt to

overestimate the values of *d'* attained by their models and makes it difficult to make any meaningful comparison of the simulated and empirical response data. Third, there appears to be some discrepancy between the data point values quoted in several of the original experiments and those displayed in Dakin and Watt's figures. For example, the value of the datum point at 0.5 on the horizontal axis in Dakin and Watt's Figure 2 is given as 1.1. However, in Barlow and Reeves' (1979) Figure 4, the corresponding datum point has a value closer to 1.6.

In light of these problems, Dakin and Watt's conclusions concerning the relative adequacy of the four models are thrown into doubt. Nonetheless, the overall aim of Dakin and Watt (1996) in comparing the performance of rival models of symmetry perception is theoretically sound and has the potential to provide insight into the processes underlying perceptual organization. It is therefore proposed that Dakin and Watt's study should be replicated in a manner that avoids the abovementioned problems. Replication also presents the opportunity to compare the performance of models that were not included in the original study.

In this chapter I will present a replication of Dakin and Watt (1996). The performance of six models of symmetry detection; four from Dakin and Watt (1996), Barlow and Reeves'(1979) model, and the Voronoi model (outlined in Chapter 3) will be compared to empirical data originally published in Barlow and Reeves (1979) and Jenkins (1982). In the next section I will outline the methods and findings of the original studies. This will be followed by a description of the six models and the theory underlying the decision-making process that the models all employ. The simulation results for each of the six experiments will then be dealt with in turn, followed by a general discussion.

5.2. The empirical data sets.

The empirical data sets used in this study were originally published in Barlow and Reeves (1979) and Jenkins (1982). These experiments have informed numerous models of symmetry perception (including the Voronoi model) and, as a result, they have been mentioned elsewhere in this thesis (e.g., Chapter 3). Nonetheless, because the experiments form the backbone of this study it is worthwhile spending some time outlining their main findings. Specific details relating to experimental methodology and stimulus design are provided elsewhere in this chapter.

Experiments 1 to 3 in this study are taken from Barlow and Reeves (1979). These experiments are of particular interest because they demonstrate that the perception of symmetry is not an all-or-nothing response. Rather, it appears to degrade gracefully as symmetrical structure is perturbed via positional distortion or the addition of random noise. Figure 5.1 displays empirical data for Experiments 1 to 3, replotted from the original figures in Barlow and Reeves (1979). In Experiment 1 the observers were asked to discriminate between symmetrical dot patterns that had been subjected to varying degrees of positional distortion and randomly distributed dot patterns. The stimuli were distorted by randomly relocating one half of each symmetrical pairing within a square region centred on each point's original position. As can be seen in Figure 5.1a the observers were able to tolerate reasonably high levels of positional distortion, and were still discriminating with a *d'* of close to 1 for tolerance regions of one third of the width of the stimulus squared.

Experiments 2 and 3 both involved the addition of random noise to symmetrical dot patterns. In Experiment 2 the observers were required to discriminate between symmetrical dot patterns with varying proportions of symmetrical pairings replaced by random noise and patterns comprised of randomly distributed points. As could be expected, observer performance was positively related to the proportion of symmetrical pairings in the stimuli (Figure 5.1b). As with Experiment 1, the observers were able to tolerate a high degree of interference, and could discriminate between the two classes of stimuli with a *d'* of close to 1 when only 40% of the points were symmetrically aligned.

Experiment 2 indicated that observers were able to discriminate between random noise and symmetry with added noise. It should be noted that this does not necessarily imply that observers would be able to discriminate with a similar degree of accuracy between perfect symmetry and symmetry to which noise had been added. It is possible that the flexibility of the visual system (in terms of being able to perceive symmetry despite the presence of high levels of noise) might come at the cost of sensitivity. Experiment 3 tested the ability of observers to discriminate between perfect symmetry and symmetry to which varying amounts of noise had been added. As can be seen in Figure 5.1c the observers were sensitive to small amounts of interference, discriminating with a d' of greater than 1 when 20% of the points in the stimuli were randomly distributed. Taken together Experiments 1 to 3 indicate that the mechanism underlying symmetry perception is both robust *and* sensitive to interference.

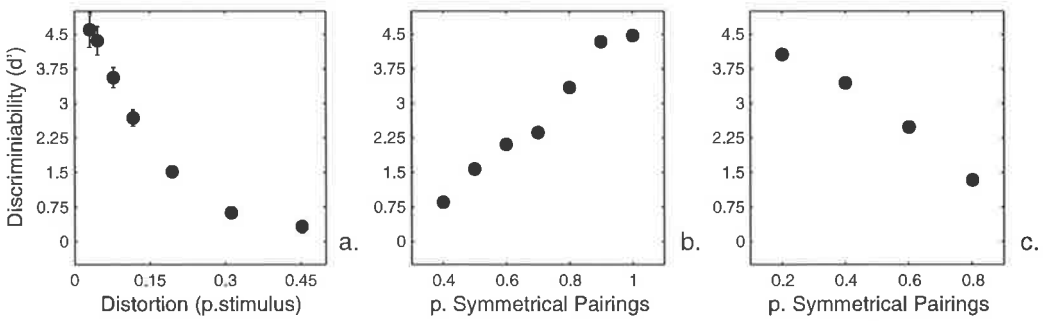

Figure 5.1 Empirical data for Experiment 1 (5.1a), Experiment 2 (5.1b) and Experiment 3 (5.1c). Data replotted from figures originally published in Barlow and Reeves (1979). For 5.1a the scale on the x-axis indicates degree of positional distortion in terms of the size of the square region within which points were randomly relocated. The size of this region is given as a proportion of the width of the stimulus. For 5.1b and c the scale on the x-axis indicates the proportion of points in each stimulus that were symmetrically positioned.

The data from Experiments 4 to 6 were originally published in Jenkins (1982). Earlier studies had indicated that the perception of symmetry appeared to be restricted to a spatially limited region located directly adjacent to the axis of symmetry (Barlow & Reeves, 1979; Julesz, 1971). Jenkins investigated the relative contribution that symmetrical point-pairs located at varying distances from the central axis made to the perception of symmetry. This was achieved by restricting the symmetrical point pairs to strips of varying widths located either at the stimulus centre or outer edge, with the remainder of the stimulus comprised of randomly positioned points. It should be noted that a centrally located fixation point preceded the presentation of all stimuli, and that the observers were instructed to not shift their gaze from this position.

Experiment 4 investigated the minimum amount of symmetry information needed for observers to discriminate between random noise stimuli and stimuli that contained a strip of symmetry located at the stimulus centre. As can be seen in Figure 5.2a, the observers were able to discriminate at a level greater than chance when the width of the strip containing symmetry was only around 2% of the overall width of the stimulus, with performance rapidly increasing to perfect levels as the width of the strip was increased.

Experiments 5 and 6 investigated the degree of redundancy in symmetrical stimuli. In Experiment 5 the observers were required to discriminate between perfectly symmetrical stimuli and stimuli comprised of a strip of symmetry located at the axis that was flanked on either side by random noise. In this case the observers' task was to detect the presence of randomly located points. Figure 5.2b shows that as the width of the symmetrical strip was increased the observers' ability to detect the random noise points decreased. In Experiment 6 the observers were required to discriminate between random noise stimuli and stimuli comprised of a strip of centrally located random noise flanked by symmetry. The observers' task was to detect the presence of symmetrical structure, and Figure 5.2c indicates that as the width of the noise strip increased observer performance decreased.

Taken together, the results of Jenkins' (1982) investigations indicate that the visual system is highly sensitive to the presence of symmetry if it is located immediately adjacent to the axis of symmetry (Experiment 4). Furthermore, the perception of symmetry (and of deviations from perfect symmetry) appears to be limited to a spatially restricted region immediately adjacent to the central axis (Experiments 5 and 6). Any information located outside of this region appears to be redundant, and does not contribute to the perception of symmetry.

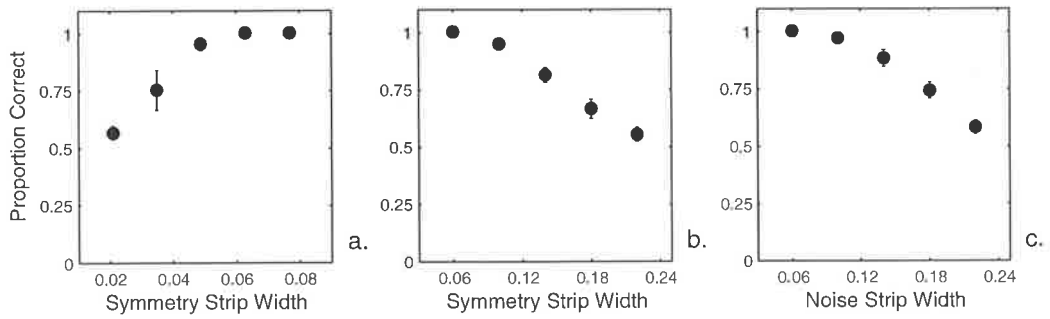

Figure 5.2 Empirical data for Experiment 4 (5.2a), Experiment 5 (5.2b) and Experiment 6 (5.2c). Strip width is given as a proportion of total stimulus width. Data replotted from figures originally published in Jenkins (1982). The scale on the x-axis indicates the width of the central strip containing symmetry (Experiments 4 and 5) or noise (Experiment 6), expressed as a proportion of the total stimulus width. Note the difference in x-axis scale for Experiment 4 compared to Experiments 5 and 6.

5.3. The models

In the following section I outline the six models under comparison in this study: the Voronoi model, Barlow and Reeves (1979) model, and the four filter-based models from Dakin and Watt (1996). Because the Voronoi model has already been outlined in detail in Chapter 3, the model's description has been restricted to a brief summary of its key characteristics. For the remaining models I provide an overview of both the theoretical and computational aspects.

5.3.1. The Voronoi model.

The Voronoi model is based on the assumption that the visual system employs relational information to detect regularities such as symmetric structure. The model employs the Voronoi tessellation of an image as the base representation from which symmetric structure is detected. Figure 5.3 provides a comparison of the Voronoi tessellation for a symmetric dot pattern (5.3a) and a randomly distributed dot pattern (5.3b).

Jenkins (1983a) suggested that the perception of symmetry may arise as a consequence of the visual system detecting pairs of points that are uniformly orientated and have collinear midpoints. The Voronoi model can be thought of as a formal instantiation of Jenkins' model. The Voronoi model searches for symmetry by attempting to pair each point in a stimulus with another point located in roughly the same position on the opposite side of the axis of symmetry. For stimuli that contain little or no symmetry, the model will generally be unable to find matching points in areas where they are expected, or alternatively, for any given point

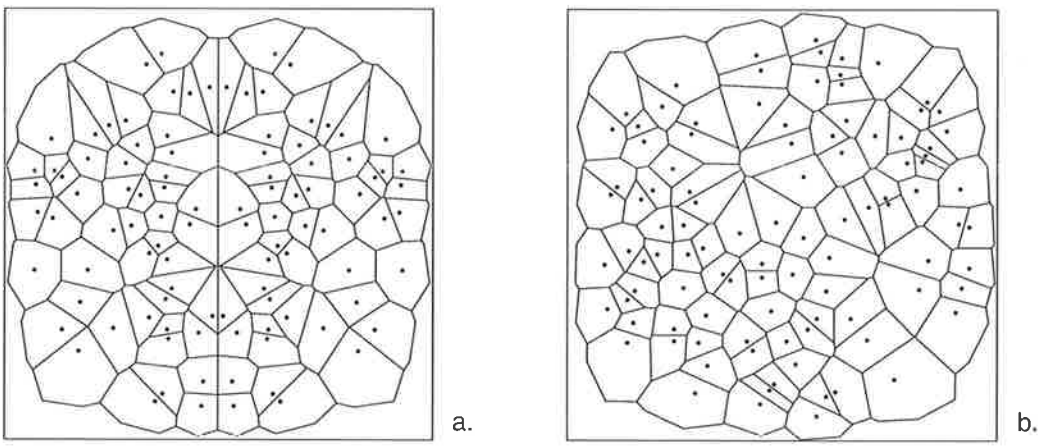

Figure 5.3 Comparison of the Voronoi tessellation for a symmetric dot pattern (5.3a) and a randomly distributed dot pattern (5.3b).

the model may find more than one potential matching point. On the other hand, for perfectly symmetrical stimuli and stimuli with high levels of symmetry, the point-pair matching process will tend to find a single unambiguous match for each point.

The matching process operates by reflecting each Voronoi cell across the stimulus midline and counting the number of points that are located in that region. Barlow and Reeves (1979) demonstrated that the visual system is able to tolerate a reasonably high degree of positional distortion or interference and still detect symmetry. The Voronoi model meets this constraint by assuming that the degree of tolerance afforded to each point in a stimulus is determined by the size of its Voronoi cell. The Voronoi model also allows a plausible means of meeting the constraint of restricting the detection of symmetry to a region directly adjacent to the axis of symmetry (Jenkins, 1982). Each stimulus is divided into a series of columns of Voronoi cells which can be differentiated according to their relative proximity to the stimulus midline. Figure 5.4 provides a comparison of the model representation for one, two and three columns of cells either side of the stimulus midline. In each case, only the points that are located within a cell are considered to contribute to the point-pair matching process. It is assumed that all of the unbounded points fall outside of the visual system's zone of attention and are effectively ignored.

Because the extent of this region of detection is unclear, the number of columns included in the symmetry calculation process is a free parameter, and this is the only free parameter in the Voronoi model. The minimum size of the region of detection is a single column either side of

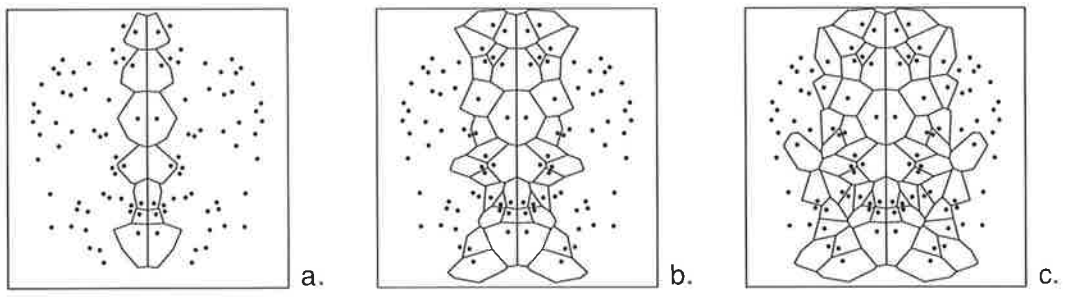

Figure 5.4 An example a symmetric stimulus and the region that is considered to contribute to the detection of symmetry, defined by a single column of cells located either side of the stimulus midline (5.4a), two columns of cells (5.4b), and three columns of cells (5.4c).

the midpoint, and at the upper limit all columns are considered to contribute. In the current study the width of the detection region ranged from 1 to 7 columns either side of the stimulus midline.

Once the number of columns in the detection region has been defined, an overall level of symmetry can be calculated for each stimulus. Each cell within the detection region is reflected across the stimulus midpoint and the number of points falling within the reflected cell is counted. The symmetry value for a stimulus is given by the expected number of points in the reflected cell e_j (where $e_j = 1$) divided by the observed number of points o_j, summed across all j cells in the detection region[1]:

$$\text{Symmetry Value} = \sum_j \frac{e_j}{o_j}$$

As has been indicated, it is assumed that this process reflects the visual system performing a form of correspondence calculation in which it is attempting to match up point pairs that have uniform orientation and collinear midpoints. When the number of expected points is equal to the number of observed points, the match is unambiguous. It follows that when the expected and observed values differ there is some degree of ambiguity in the matching process, indicating a lower level of symmetry. The values calculated for each cell are summed giving an overall symmetry measure for the stimulus. For perfectly symmetrical stimuli this value will be high, but as the level of symmetry in a stimulus is perturbed (either by the addition of random noise or distortion) this value will drop to the level found in purely random stimuli.

5.3.2. Barlow & Reeves' (1979) model.

The Barlow and Reeves model was first outlined in the same paper from which Experiments 1 to 3 are drawn (Barlow & Reeves, 1979). The results of these experiments indicated that symmetry detection was a graded response that declined as symmetry pairs were either perturbed, or replaced with randomly positioned points. Barlow and Reeves suggested that because observers are able to tolerate large inaccuracies in the placement of point-pairs it is unlikely that the visual system is making comparisons of individual spatial locations within a stimulus. They used a simple model to demonstrate that it is possible to detect symmetry in dot patterns by making comparisons between the dot densities of symmetrically positioned sub-regions of a stimulus, and they argued that this is the kind of activity that a neuron with a fixed receptive field might be able to perform.

[1] Note: when the number of observed points equals zero, the Voronoi model assigns a value of zero, not infinity.

Barlow and Reeves' (1979) model operates by dividing a stimulus into a number of equal sized square sub-regions. Each sub-region is paired with a second sub-region positioned equidistant from the axis of symmetry. If a stimulus is perfectly symmetric then the number of points in paired sub-regions will be equal. Alternatively, if a stimulus does not contain any symmetry, then the variance in number of points in the paired regions will be equal to the variance between the non-paired regions.

The number of points in each sub-region is counted and χ^2 is calculated using:

$$\chi^2 = \sum_j \frac{(N_j - N_j')}{2\bar{N}},$$

Where N_j and N_j' are the number of points in a pair of sub-regions, and $\bar{N}$ is the mean. Random arrays will have values of χ^2 equal to the degrees of freedom (*df* = number of paired sub-regions – 1). For stimuli that contain symmetrical dot pairings χ^2 will be less than *df*, dropping to zero for perfectly symmetrical stimuli.

The Barlow and Reeves model has a single variable parameter: the number of symmetrically positioned sub-regions. Because the sub-regions must be equal sized, have a symmetrically positioned pair and cannot overlap, the number of sub-regions in a stimulus can only take on values equal to the squares of positive numbers. This necessarily places restrictions upon the range of values over which the model can be tested. Barlow and Reeves (1979) reported model fits for 16 and 36 sub-regions. For the current experiments the model was implemented at 7 levels of sub-region numerosity: 4, 16, 36, 64, 100, 144, and 196 sub-regions. Figure 5.5 demonstrates the representation of a symmetrical stimulus derived by the Barlow and Reeves model for 16 and 36 sub-regions.

Barlow and Reeves demonstrated that their model was able to provide a reasonable qualitative fit to the data from Experiment 1 using both 16 and 36 sub-regions. Nonetheless, it is a fairly unlikely candidate for the process underlying human perception of symmetry, in that it is implausible that the visual system could be neatly dividing stimuli into equal-sized symmetrically positioned sub-regions. Barlow and Reeves acknowledged the general

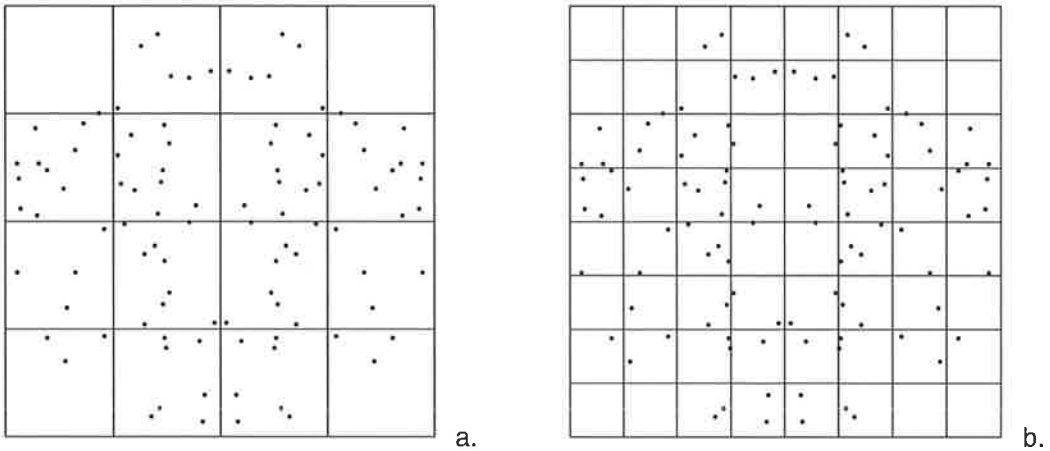

Figure 5.5 Comparison of the Barlow and Reeves model representation of a symmetric dot stimulus for 16 (5.5a) and 36 (5.5b) sub-regions.

implausability of the model in its present form, but argued that the goodness of the model fits to the empirical data suggest that the visual system might be detecting symmetry by comparing the expected and observed dot densities of regions equidistant from the central axis. In this sense the Barlow and Reeves model is somewhat similar to the Voronoi model, which is also reliant upon a comparison of sub-region dot densities.

The general implausibility of the Barlow and Reeves model is the primary reason that it has been included in the set of models being compared in this study. In essence it is a yardstick or 'strawman' model against which to measure the performance of the other models. Jacobs and Grainger (1994) suggested that models should be judged in terms of descriptive adequacy, complexity, generality, and explanatory adequacy. Bayesian model selection provides an indication of the relative likelihood of a model based upon the first three of these criteria, but it is unable to account for a model's explanatory adequacy. Explanatory adequacy refers to the plausibility of a model and asks whether its underlying assumptions are in agreement with conventional theory and knowledge. It has been argued that the Voronoi model has some claim to plausibility, in that there is a body of psychophysical and physiological evidence suggesting that the visual system might be generating a Voronoi-like representation in early vision (Chapter 3). Bayesian model selection might find that the Voronoi model is the most likely of the set of models under comparison in this study. However, if the arguably implausible Barlow and Reeves model performs at a level that approaches the performance of

the Voronoi model, then we may have grounds to seriously question the utility of the Voronoi representation based approach.

5.3.3. Dakin & Watt's (1996) filter models.

Dakin and Watt (1996) interpret Barlow and Reeves (1979) finding that the visual system is resistant to the positional distortion of symmetric stimuli as evidence of the involvement of low frequency spatial filtering mechanisms in the perception of symmetry. Locher and Wagemans (1993) had previously suggested that convolving a stimulus with isotropic, Laplacian-of-Gaussian filters would generate a representation largely resistant to positional distortion. They argue that such an approach is in line with the assumptions of the primal sketch based visual processing models of Marr (1982) and Watt (1988).

In contrast to this, Dakin and Watt (1996) suggested that the visual system may be deriving a stimulus representation from the output of anisotropic filters such as the orientationally tuned neurons in the primary visual cortex. They noted that when a symmetric stimulus with a vertical axis is filtered with a horizontally oriented filter, the resulting image is comprised of a series of parallel stripes with collinear midpoints (see Figure 5.7b and c). They argued that such a representation is analogous with Jenkins' (1983a) suggestion that the perception of symmetry is reliant upon the visual system detecting stimulus features with uniform orientation and collinear midpoints.

In order to compare these two approaches, Dakin and Watt employed both oriented and isotropic filters to generate a stimulus representation. The isotropic filter was defined as a Laplacian-of-Gaussian function:

$$f(x,y,s)=\frac{1}{s^2}\left(1-\frac{x^2+y^2}{s^2}\right)e^{\frac{-(x^2+y^2)}{2s^2}}.$$

The oriented filter model was a horizontally aligned Difference-of-Gaussian function:

$$f(x,y,s)=\left(e^{-y^2/2s^2}-\frac{1}{2.23}e^{-y^2/2(2.23s)^2}\right)e^{\frac{-x^2}{2(3s)^2}}.$$

For both the isotropic and oriented filters, the space constant of the spread function is represented by s, and is given in pixels. The isotropic filter model was implemented using

values of s ranging from 2 to 16 pixels in 8 steps of 2 pixels, and the oriented filter model was implemented using s values ranging from 1.41 to 11.31 pixels in 8 steps of 1.41 pixels. This is the same range of values used by Dakin and Watt (1996).

Following the filtering stage, stimuli were standardised so that the maximum and minimum grey-scale values were equal to 1 and 0.5 respectively. The stimuli were then doubly half-wave rectified by setting pixels with grey-scale values within one standard deviation of the mean value to 0 (black). The resulting thresholded output retains both the positive and negative portions of the filtered stimulus. Example representations derived from the isotropic and oriented filters are given in Figures 5.6 and 5.7 respectively.

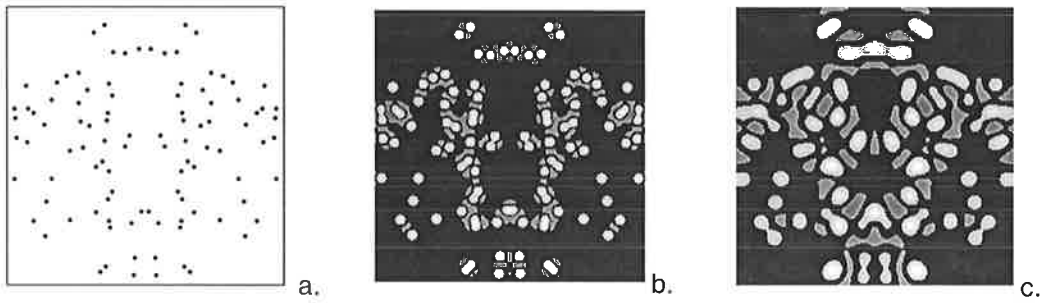

Figure 5.6 A symmetric dot pattern (5.6a), and a comparison of the output of the isotropic filter using a spread function of s = *4 pixels (5.6b) and* s = *8 pixels (5.6c).*

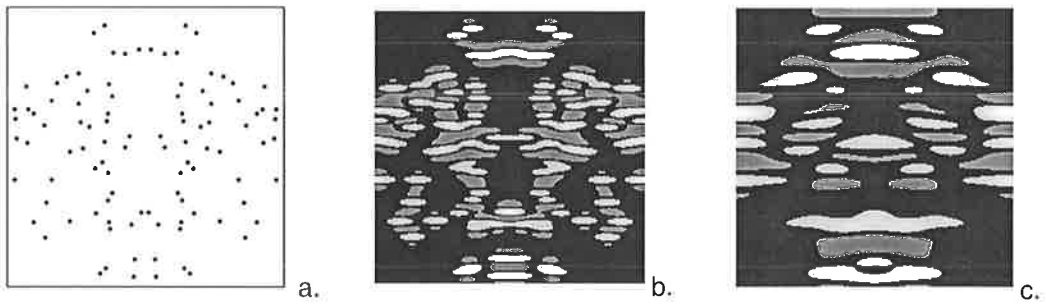

Figure 5.7 A symmetric dot pattern (5.7a), and a comparison of the output of the oriented filter using a spread function of s = *2.82 pixels (5.7b) and* s = *5.65 pixels (5.7c).*

Dakin and Watt (1996) implemented two separate approaches to calculating the degree of symmetry within a stimulus: cross-correlation and feature alignment. A number of researchers have proposed a cross-correlation based approach to symmetry detection (Barlow & Reeves, 1979; Julesz, 1971). In essence, this involves correlating one half of an image with a mirror reflection of the other half. For example, the Barlow and Reeves (1979) model can be thought of as correlating local dot densities. The cross-correlation approach tested by Dakin and Watt bears a closer resemblance to Julesz's (1971) suggestion that the visual system may be performing a point-by-point comparison of luminance values.

Dakin and Watt described the cross-correlation measure as:

$$\text{Cross-correlation} = 4\left(\frac{I(x_a,y_a)I(-x_a,y_a)}{(I(x_a,y_a)+I(-x_a,y_a))^2}\right),$$

where $I(x_a,y_a)$ indicates the luminance value of an individual pixel location within an image. The cross-correlation measure is calculated for each location within an image, and an average is obtained across all locations. This process provides symmetry values that are equal to 1 when a stimulus is perfectly correlated with its mirror image, and that approach zero when a stimulus is randomly distributed.

While cross-correlation makes use of all of the available information within a stimulus, feature alignment produces a measure of stimulus symmetry based solely upon the stimulus features or filtered 'blobs' that intersect with the central axis. In this sense, the models employing feature alignment calculation are more likely to be sensitive to effects based on the relative proximity of stimulus features (such as in Experiments 4 to 6) than the cross-correlation based models.

Following Watt (1991), zero-bounded stimulus features (i.e., the grey 'blobs' in the filtered outputs shown in Figures 5.6 and 5.7) were described in terms of their centroid (cx,cy), length (λ), and mass (μ). The alignment A(x) of stimulus features that intersect the axis was calculated using:

$$\mathrm{A}(x) = \frac{1}{M}\sum_{i=1}^{N_c} e\left[-\left(\frac{x^2-cx_i^2}{2\lambda_i^2}\right)\right]\mu_i,$$

which measures the distance from each centroid to the axis (x), weighted in inverse proportion to feature length, and in direct proportion to its mass. The summed error of the relevant features is normalized by their total mass (M) to ensure that A(x) lies between 1 and 0.

Based upon the two forms of filtering mechanism and two forms of symmetry calculation Dakin and Watt (1996) tested four models in a 2x2 design. The four models were: oriented filter with feature alignment (Oriented/Alignment), oriented filter and cross-correlation (Oriented/Correlation), isotropic filter and feature alignment (Isotropic/Alignment), and isotropic filter and cross-correlation (Isotropic/Correlation).

5.4. The decision process

The models' task was essentially the same as that of the human observers in the original experiments: they were required to discriminate between two classes of stimuli. For each experiment a model produced two sets of responses, one for the experimental stimuli (e.g., the symmetric point patterns with varying levels of randomly distributed noise points) and one for the standard stimuli (e.g. the randomly distributed noise patterns). All models employed the same decision mechanism to choose between the two sets of responses.

Barlow and Reeves (1979) reported that for Experiments 1 to 3 they employed a single-interval, two-alternative forced-choice design, with the stimuli presented in blocks containing a single level of the experimental stimulus manipulation. Observer performance was expressed as d', a measure of discriminability based on signal detection theory (Green & Swets, 1966). Here model discriminability was calculated following the method outlined in Dakin and Watt (1996). Each model has two sets of responses, one for the standard stimuli and one for the experimental stimuli. These responses have means μ^s and μ^e, and standard deviations σ^s and σ^e respectively. From these values a decision criterion value can be calculated for each separate experimental stimulus signal level parameterisation (i):

$$\text{Criterion}_i = \mu_i^e + \frac{\sigma_i^e}{\sigma_i^e + \sigma^s}\left(\mu^s - \mu_i^e\right).$$

The two sets of model responses can be compared to the criterion value to determine the probabilities of hits $p(H)$ and false alarms $p(F)$. From these two values d' can be calculated using signal detection theory (Green & Swets, 1966):

$$d' = \Phi^{-1} p(H) - \Phi^{-1} p(F),$$

where $\Phi^{-1}(y)$ is the inverse phi or Gaussian probability function. This enables the values associated with *p(H)* and *p(F)* to be converted into *z*-scores.

Theoretically, the maximum and minimum values that *d'* can take are +/-∞, respectively (Stanislow & Todorov, 1999). However, as Dakin and Watt indicated *d'* can be calculated to an arbitrary precision level by restricting the maximum and minimum values of *p(H)* and *p(F)*. Barlow and Reeves (1979) reported that *d'* was calculated using Elliots' (1964) table of values of *d'*. The maximum and minimum values of *d'* that can be attained using this table are +/- 4.64, as *p(H)* and *p(F)* are restricted to a range of 0.01 to 0.99. In contrast to this, Dakin and Watt report model *d'*s as high as 6.2 (presumably this is a result of allowing the *p*-values to range from 0.001 to 0.999). This has the unfortunate effect of making a number of the models appear to perform at a level far greater than that of the human observers. It is fair to suggest that if a meaningful comparison is to be made between previously published data sets and model performance, then *d'* should be calculated using the same level of precision in both cases. Therefore, in order to correspond with the precision levels of the original experiments the upper and lower bounds of *p(H)* and *p(F)* were capped at 0.01 and 0.99.

A slightly different decision making mechanism was used for Experiments 4 to 6, reflecting the difference in psychophysical procedures employed in Jenkins (1982). The three experiments from Jenkins (1982) used a two-alternative, temporal forced choice design, in which each trial was comprised of the presentation of two stimuli (one experimental stimulus and one standard stimulus) separated by a one-second inter-stimulus interval. The order of stimulus presentation within each trial was randomised, and the observers' task was to indicate whether the experimental stimulus was presented in the first or second interval. The stimuli were presented in blocks containing equal numbers of each level of the experimental stimulus manipulation. In other words, unlike the experiments from Barlow and Reeves, the experimental stimulus signal levels were free to vary *within* a block of trials. Observer performance was expressed as the proportion of correctly identified trials for each stimulus level.

It is assumed that the observers' decisions were based upon the relative strength of the *perceived* signal component (i.e. level of symmetry) in each of the two stimuli comprising a

single trial, and that the interval containing the stimulus which gave the stronger impression of symmetry would be classified as the target interval (Green & Swets, 1966). In order to simulate this procedure every possible combination of experimental and standard stimulus pairings was considered. The probability of a model correctly identifying an individual experimental stimulus (e_i) across all (j) possible pairings with standard stimuli (s) is given by:

$$\text{Probability of Correct Identification} = \frac{\sum_j \left(e_i > s_j\right) + \frac{1}{2}\left(e_i = s_j\right)}{N},$$

where N is equal to the total number of unique pairings. Averaging across these values provides the proportion of correct responses for each level of the experimental stimulus manipulation.

5.5. General methods

MATLAB 7.0.1. was used to generate stimuli and run the simulations. The stimuli were generated in a manner that replicated as closely as was practically possible the physical characteristics of the stimuli described in the original experiments. The only major point of difference between the current and the original experimental stimuli was that Jenkins (1982) employed dynamic dot textures comprised of 16446 point pairs that were plotted sequentially for 1.5 milliseconds each. According to Jenkins the resulting stimulus was perceived as a static texture with a dot density of 26 points/deg^2 of visual angle (650 points). All of the current experimental stimuli were static textures.

The symmetric stimuli were always comprised of vertical bilateral symmetry pairs with an axis centred in the middle of the stimulus. In order to negate the potential effects of density cues, equal numbers of points were located on each side of the central axis. For Experiments 1, 2 and 3 the stimuli were comprised of 100 2x2 pixel black dots presented within a white circular region with a 256 pixel diameter. For Experiments 4, 5 and 6 the stimuli were comprised of 650 1 pixel^2 black dots presented within a 256x256 pixel white background.

A total of 100 stimuli were generated for each stimulus parameterisation, half of which were the experimental stimuli (i.e., symmetric patterns with varying levels of noise) and half of which were the standard stimuli (i.e., random noise stimuli). The same stimuli were presented

to all of the models in order to ensure that any difference between the responses of the models was a result of differences between the models and not differences between the stimuli.

Empirical data points were obtained from the figures published in the original papers using Data Thief III (Tummers, van der Laan, & Huyser, 2005).

5.6. Experiment 1: Discriminating distorted symmetry from random noise.

The first of the six experiments explored the effect of spatial distortion upon symmetry detection. Barlow and Reeves (1979) measured observers' ability to discriminate between stimuli comprised of perturbed or distorted symmetrical pairings, and stimuli that were comprised entirely of randomly positioned dots.

5.6.1. Stimuli

The symmetrical stimuli were distorted by repositioning one half of each symmetrical pair a random distance within a given horizontal and vertical 'tolerance area' centred on the point's original position. The size of the tolerance area was manipulated across nine levels: 8, 12, 20, 30, 50, 80 and 116 pixels^2. This corresponds to widths of 0.03, 0.04, 0.07, 0.11, 0.19, 0.31 and 0.45 of the diameter of the stimulus. It should also be noted that the repositioning process was subject to two main restrictions: First, in order to maintain equal density across both halves of the stimulus points could not be repositioned across the axis and, second, points could not be moved outside of the 256 pixel diameter presentation circle. Typical examples of the stimuli employed in Experiment 1 are presented in Figure 5.8.

5.6.2. Results.

The best fits of the six models to the empirical data are displayed in Figure 5.9. All six models match the basic pattern of the empirical data, showing a gradual decrease in discriminability as the level of distortion is increased. A qualitative comparison of the data fits suggests that the Voronoi model, Barlow and Reeves model, and the two cross-correlation models provide a better account of the data than the two feature alignment models. However, as has been discussed in Chapter 4, in order to assess the relative utility of competing models it is necessary to account for the fits of each model across the entire range of their parameterisations. In this way models that are able to provide a close (or reasonably close) fit to the empirical data across all of their parameterisations are favoured over models that can provide a close fit at one parameter level but provide poor fits at other parameterisations.

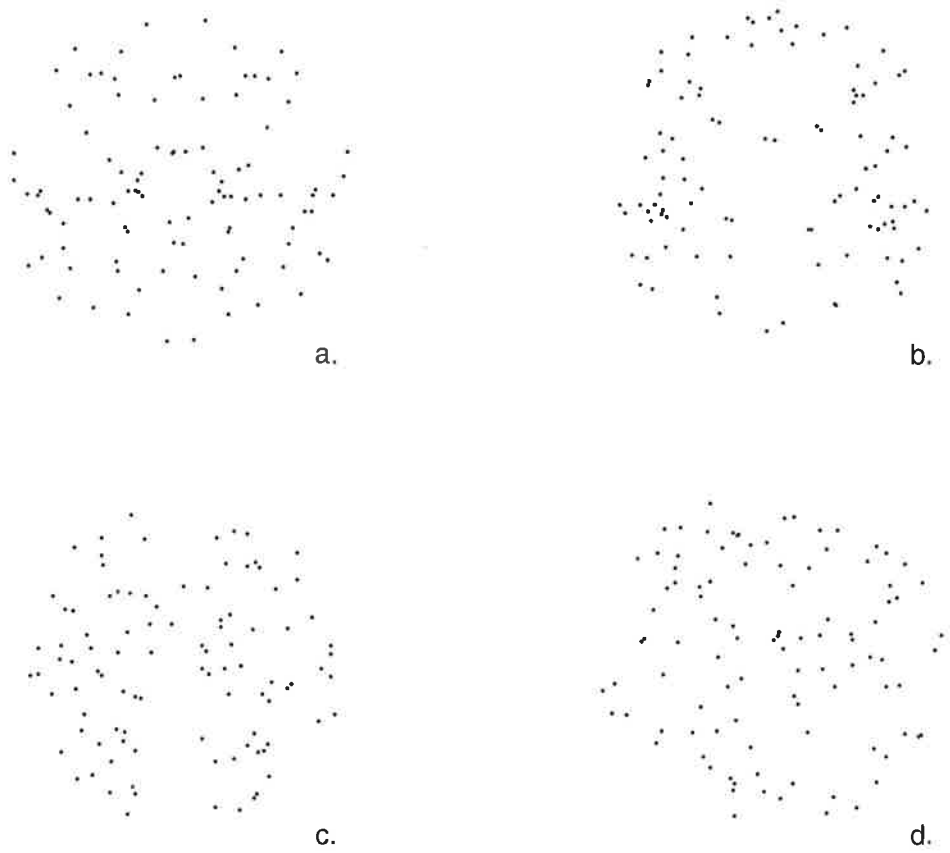

Figure 5.8 Examples of the stimuli presented in Experiment 1. The level of distortion is given as the width of the tolerance area in relation to the diameter of the stimulus: 0.03 distortion (a), 0.07 distortion (b), and 0.19 distortion (c). Figure 5.8d is an example of a random noise stimulus.

Figure 5.10 provides a comparison of the six model fits to the empirical data across all of their parameterisations. As can be seen the fits of the Barlow & Reeves and the two Correlation models are tightly clustered around the empirical data points across all their parameterisations. In contrast to this, the Voronoi model systematically underperforms for stimuli with high levels of distortion, and the two feature alignment models tend to under-perform across all parameterisations.

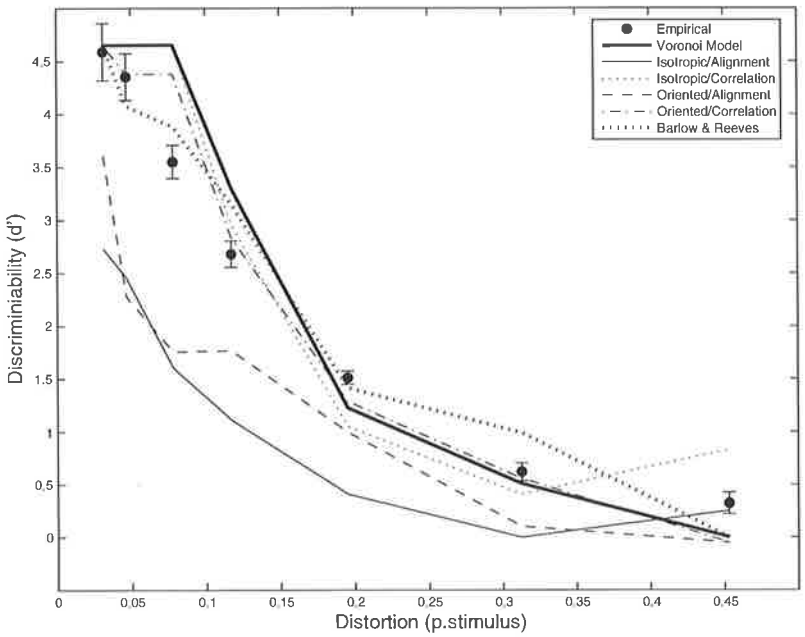

Figure 5.9 Best fits for the six models to the empirical Experiment 1 data.

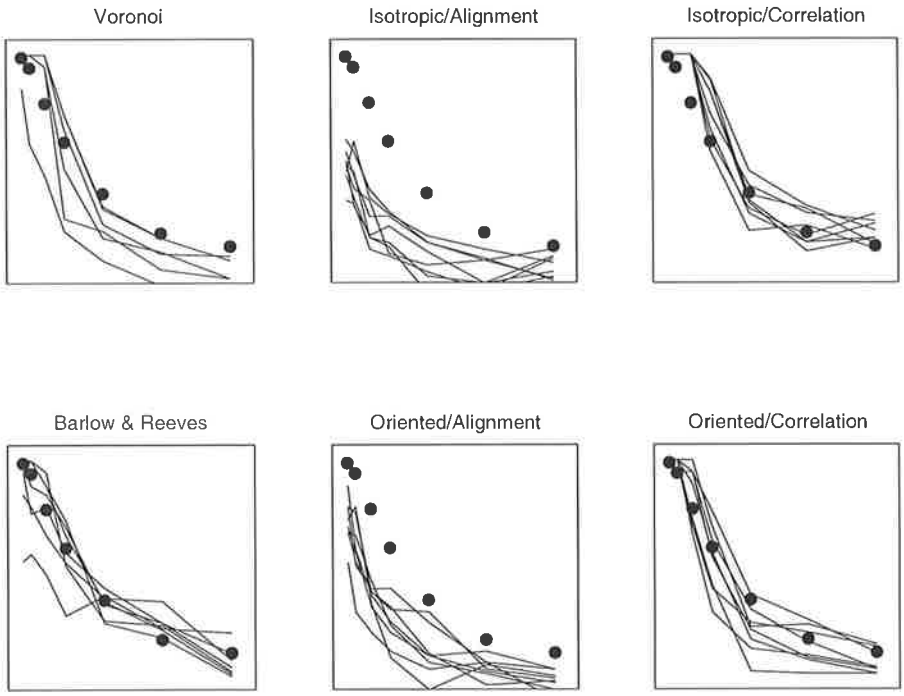

Figure 5.10 Fits of the six models across all parameterisations to the empirical Experiment 1 data.

Table 5.1 Log marginal likelihoods and Bayes Factors for the six models in Experiment 1.

Model	log marginal likelihood	Bayes Factor
Voronoi	3.03	54.00
Barlow & Reeves	7.02	1.00
Isotropic/Alignment	-39.75	2.06×10^{20}
Isotropic/Correlation	2.06	143.26
Oriented/Alignment	-20.82	1.24×10^{12}
Oriented/Correlation	5.46	4.75

Note- Bayes Factors are taken in relation to the most likely model, which in this case is the Barlow & Reeves model.

From the weighted sum squared error of these data fits it is possible to calculate the marginal log likelihood for each model. These can then be expressed as Bayes factors which gives an indication of the likelihood of each model in relation to the most likely model. Table 5.1 summarises the log marginal likelihoods and Bayes factors of the six models for the data from Experiment 1. The data suggest that the Barlow & Reeves model is the most likely of the six models tested. At first glance this result is somewhat surprising given the general implausibility of this model. However, it should be noted that Barlow and Reeves have previously demonstrated that their model is able to provide good fits to this empirical data set (Barlow & Reeves, 1979).

5.7. Experiment 2: Discriminating symmetry with added noise from random noise.

Experiment 2 explored the effect of manipulating the proportion of symmetrical pairings within a stimulus when discriminating between experimental stimuli comprised of a mixture of symmetrical and random dots, and standard stimuli that were comprised entirely of randomly positioned dots.

5.7.1. Stimuli

The proportion of symmetrical pairings ranged from 0.4 to 1.0 in seven 0.1 increments. Equal numbers of randomly positioned points were located on either side of the stimulus midline. Figure 5.11 provides typical examples of the stimuli used in Experiment 2. Figures 5.11a, b and c are examples of experimental stimuli, and Figure 5.11d is an example of the standard stimuli.

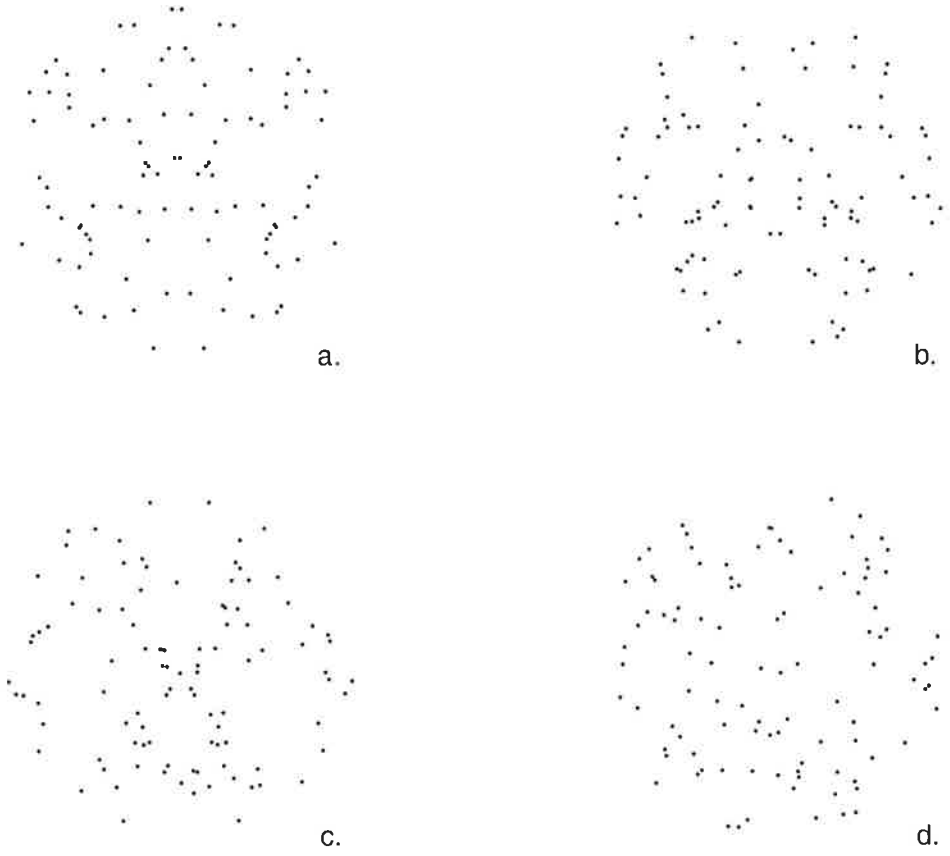

Figure 5.11 Examples of stimuli presented in Experiments 2 and 3: 100% symmetry (a), 80% symmetry (b), 40% symmetry (c), and 0% symmetry (d).

5.7.2. Results

The best fits of the six models to the empirical data are displayed in Figure 5.12. As with Experiment 1, the six models match the basic pattern of the empirical data, showing a gradual increase in discriminability as the proportion of symmetrical pairings is increased. However, Figure 5.13 demonstrates that (compared to the Experiment 1 data) the model fits across the full range of parameterisations tend to be less tightly clustered around the empirical data. In particular we can see that the two Correlation models tend to over-perform, and the two feature alignment models under-perform.

This data set provides an additional challenge for implementing Bayesian model selection in that the published data provide no indication of the variance associated with each data point.

In the absence of variance data, the log marginal likelihoods and their associated Bayes factors must be calculated across a range of data precision assumptions (Lee, 2001; Lee, 2004). A pilot experiment replicating Experiment 2 indicated that the standard error of *d'* ranged from 0.1 to 0.23, with a mean of 0.15. However, as Lee (2004) suggested, it is useful to compare the models across a wide range of precision assumptions. In this case log marginal likelihoods were calculated for standard errors ranging from 0.01 to 1.0.

Figure 5.14 presents Bayes factors for the fits of the six models to the empirical data from Experiment 2 across the full range of data precision assumptions. Bayes factors are given in relation to the most likely model, which in this case is the Voronoi model for precision assumptions of less than approximately 4.5, and the Oriented/Alignment model for precision assumptions greater than 4.5. There are two points of interest in Figure 5.14. First, it should be noted that the Isotropic/Correlation model falls outside of the range of values displayed in the figure. Second, it can be seen that the relative likelihoods of the models tend to converge as the width of the standard error assumption is relaxed. In other words, if we assume that the data are imprecise, then it is harder to distinguish between the performances of the models. This makes sense, because imprecise data ought not carry the information needed to be able to distinguish between competing models.

In the absence of empirical variance data it is difficult to make a judgement about the relative likelihoods of the six model fits to the Experiment 2 data. However, if we take the upper and lower bounds provided by the pilot experiment as estimates of the sample variance, then we can obtain each model's average log marginal likelihood within this range. Table 5.3 summarises the log marginal likelihoods and Bayes factors for the six models under the assumption that the data precision in Experiment 2 ranges from 0.1 to 0.23 standard errors. A cautious interpretation of the Bayes factors would suggest that the Oriented/Correlation, Barlow & Reeves, and the two Isotropic models can be ruled out. Of the two remaining models, the Oriented/Alignment model is only 2.78 times less likely than the Voronoi model. Given that Jeffreys' (1961) guidelines for interpreting Bayes' factors (see Table 4.1) indicate that differences of less than 3.2 are 'not worth more than a bare mention', we can conclude that no meaningful distinction can be made between the likelihoods of these two models.

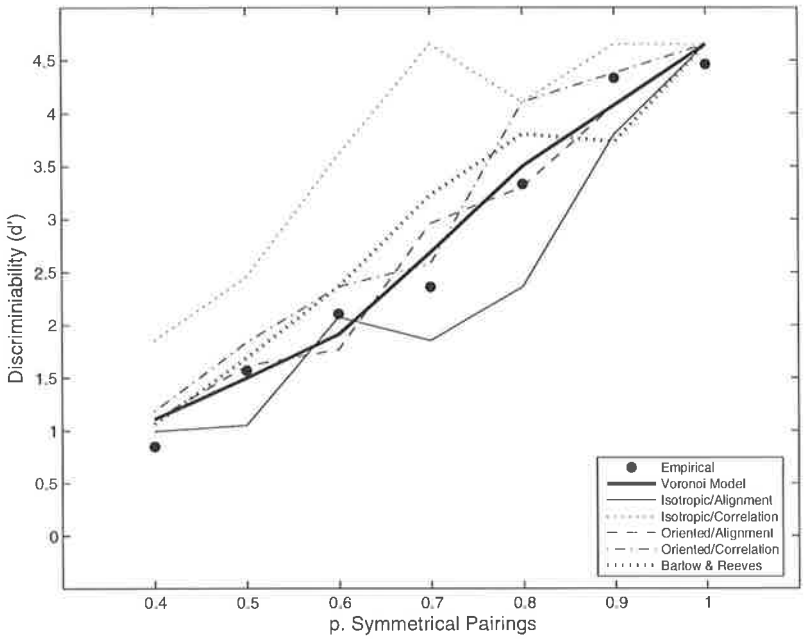

Figure 5.12 Best fits for the six models to the empirical Experiment 2 data.

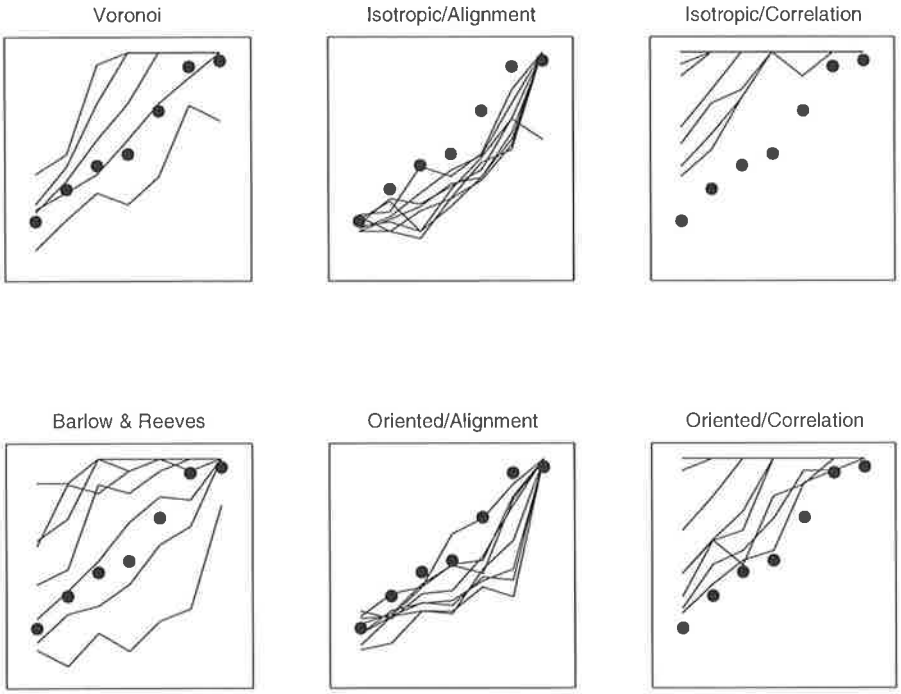

Figure 5.13 Fits of the six models across all parameterisations to the empirical Experiment 2 data.

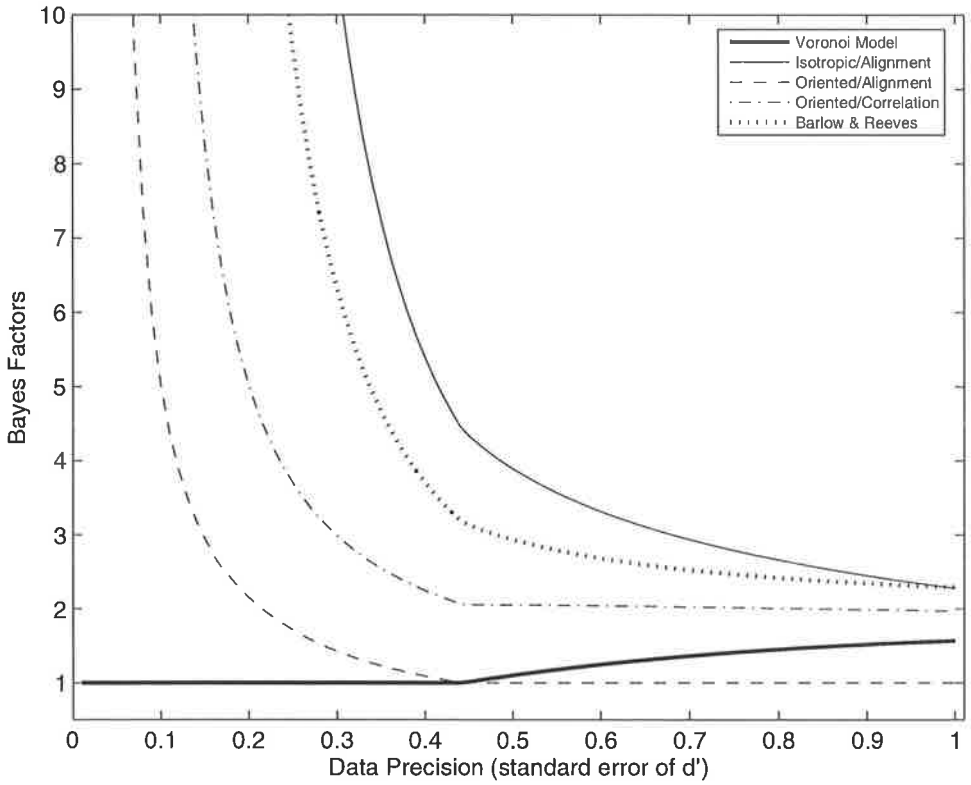

Figure 5.14 Bayes factors for the six models fits to the Experiment 2 data as a function of data precision. The Bayes factors are taken in relation to the most likely model, which in this case is the Voronoi model for precision assumptions of less than approximately 4.5, and the Oriented/Alignment model for precision assumptions greater than 4.5. Note that the Isotropic/Correlation model has minimum Bayes factor values that lie outside the range of this figure.

Table 5.2 Log marginal likelihoods and Bayes Factors for the six models in Experiment 2 assuming a data precision range of 0.1 to 0.23.

Model	log marginal likelihood	Bayes Factor
Voronoi	4.88	1.00
Barlow & Reeves	1.16	41.40
Isotropic/Alignment	0.10	119.77
Isotropic/Correlation	-26.65	4.99×10^{13}
Oriented/Alignment	3.86	2.78
Oriented/Correlation	2.83	7.79

Note- Bayes Factors are taken in relation to the most likely model, which in this case is the Voronoi model.

5.8. Experiment 3: Discriminating symmetry with added noise from perfect symmetry.

Experiment 3 explored the effect of manipulating the proportion of symmetrical pairings within a stimulus when discriminating between experimental stimuli comprised of a mixture of symmetrical and random dots, and standard stimuli that were perfectly symmetrical.

5.8.1. Stimuli

The proportion of symmetrical pairings in the experimental stimuli ranged from 0.2 to 0.8 in four 0.2 increments. As with the Experiment 2 stimuli, to avoid the effect of density cues an equal number of randomly positioned points were located on either side of the stimulus midline. Typical examples of stimuli used in Experiment 3 are also given in Figure 5.11. In this case, however, Figures 5.11b, 5.11c and 5.11d are examples of experimental stimuli, and 5.11a is an example of a standard stimulus.

5.8.2. Results

The best fits of the models to the Experiment 3 data are much less convincing when compared to the fits achieved for the previous two experimental data sets. Figure 5.15 indicates that only three of the models (the Voronoi, Barlow & Reeves and Isotropic/Alignment models) come close to matching observer performance. However, it should be noted that the quality of these fits is poor at best. Comparison of the model fits across the range of their parameterisations (Figure 5.16) shows that all of the models tend to over-perform on this task. In particular, the two Correlation models fail to show any decrease in discriminability as the level of symmetric pairings is increased.

Similar to the previous experiment, the published data provide no indication of the variance associated with each empirical data point. Following the same approach applied in Experiment 2, log marginal probabilities for each model were calculated across a range of data precision estimates (standard errors ranging from 0.01 to 1.0). Figure 5.17 shows the probabilities expressed as Bayes factors. In this case the most likely model is the Voronoi model, regardless of the assumptions made regarding level of precision present in the Experiment 3 data. Given the inadequacy of the fits in Figure 5.16 it is unsurprising that the

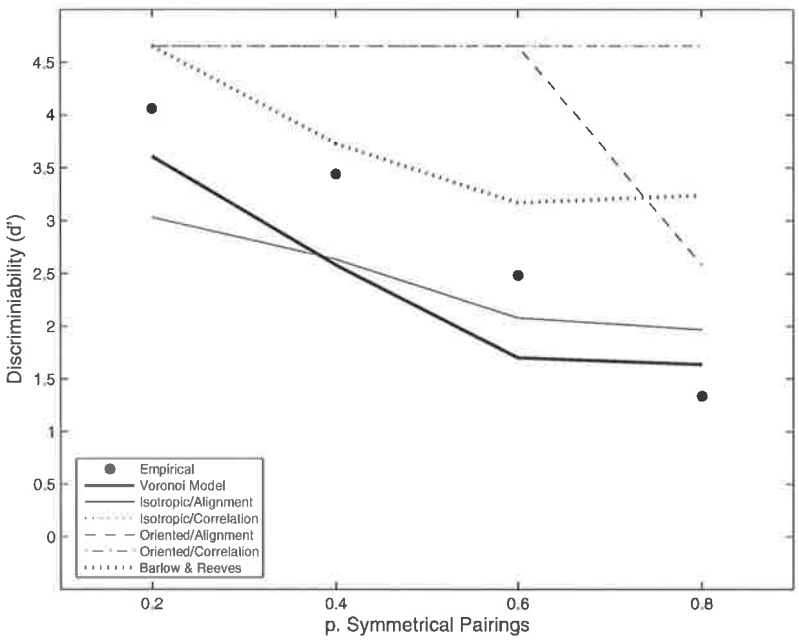

Figure 5.15 Best fits for the six models to the empirical Experiment 3 data.

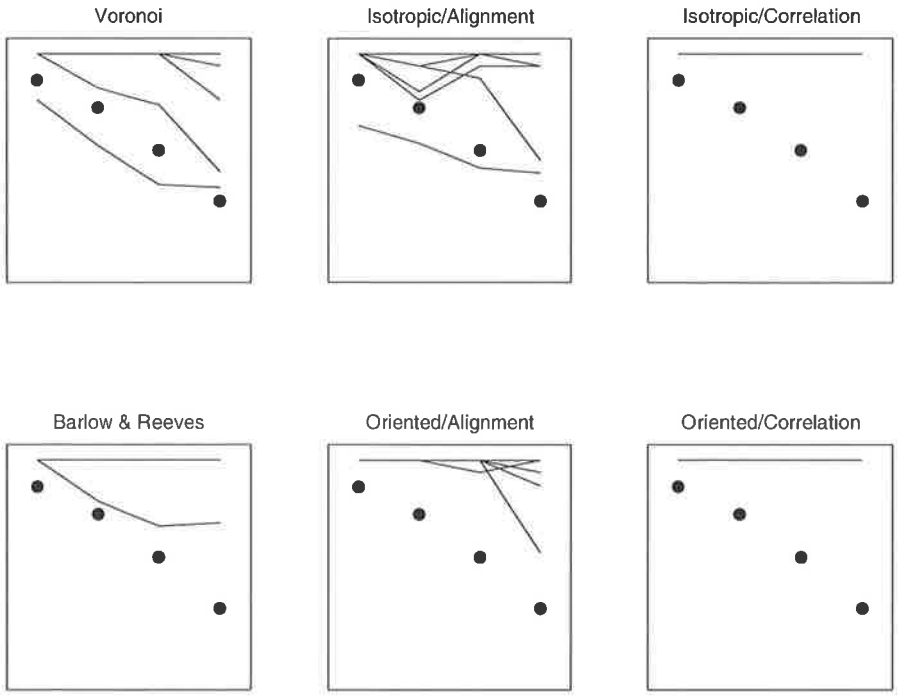

Figure 5.16 Fits of the six models across all parameterisations to the empirical Experiment 3 data.

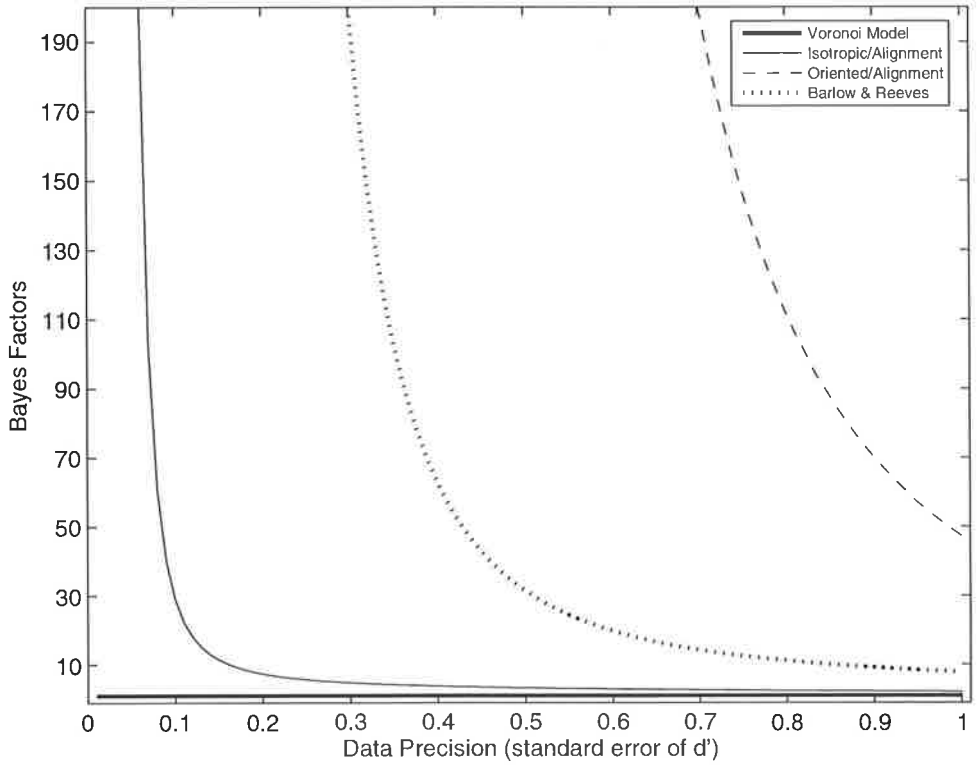

Figure 5.17 Bayes factors for the six models fits to the Experiment 3 data as a function of data precision. The Bayes factors are taken in relation to the most likely model, which in this case is the Voronoi model. Note that both the Isotropic/Correlation and Oriented/Correlation models have minimum Bayes factor values that lie outside the range of this figure.

two correlation models fall outside of the range of values presented in Figure 5.17. The Barlow & Reeves and Isotropic/Alignment models approach the likelihood of the Voronoi model if high variance levels are assumed. However, a pilot experiment replicating Experiment 3 suggests that the standard errors lie in the range of 0.09 to 0.13, with a mean value of 0.11.

As with Experiment 2 it is possible to take the upper and lower bounds provided by the pilot experiment as estimates of the sample variance and obtain each model's average log marginal likelihood within this range. The log marginal likelihoods and Bayes factors for the six

Table 5.3 Log marginal likelihoods and Bayes Factors for the six models in Experiment 3 assuming a data precision range of 0.09 to 0.13.

Model	log marginal likelihood	Bayes Factor
Voronoi	-1.73	1.00
Barlow & Reeves	-15.45	9.12×10^{5}
Isotropic/Alignment	-4.88	23.41
Isotropic/Correlation	-73.73	1.85×10^{31}
Oriented/Alignment	-32.03	1.45×10^{13}
Oriented/Correlation	-73.73	1.85×10^{31}

Note- Bayes Factors are taken in relation to the most likely model, which in this case is the Voronoi model.

models are summarised in Table 5.3. Although these results should be interpreted cautiously it is clear that all of the models are substantially less likely than the Voronoi model. Nonetheless, the poor fit of the Voronoi model across the range of its parameterisations (Figure 5.16.) is evidence of the general inadequacy of *all* of the models to simulate observer performance on this task.

5.9. Experiment 4: Central symmetry strip in noise vs. noise.

Experiment 4 was designed to measure the minimum amount of information needed to detect symmetry. The observers were required to discriminate between random noise stimuli and stimuli comprised of random noise with a strip of symmetrical pairings located at the axis.

5.9.1. Stimuli

Five sizes of symmetry strip width were employed. Expressed as a proportion of the width of the stimulus the symmetry strip widths were 0.021, 0.035, 0.049, 0.063 and 0.077. Examples of the stimuli used in Experiment 4 are given in Figure 5.18.

5.9.2. Results

Figure 5.19 indicates that the best fits of the models to the Experiment 4 empirical data appear (with the exception of the Barlow & Reeves model) to capture the basic pattern of results: as the width of the strip containing symmetric pairings increases the proportion of correct responses that the models make also increases. However, Figure 5.20 shows that across most parameterisations the models all tend under-perform. In particular, the Barlow and Reeves model and the two Oriented filter models are performing at close to chance level across most parameterisations

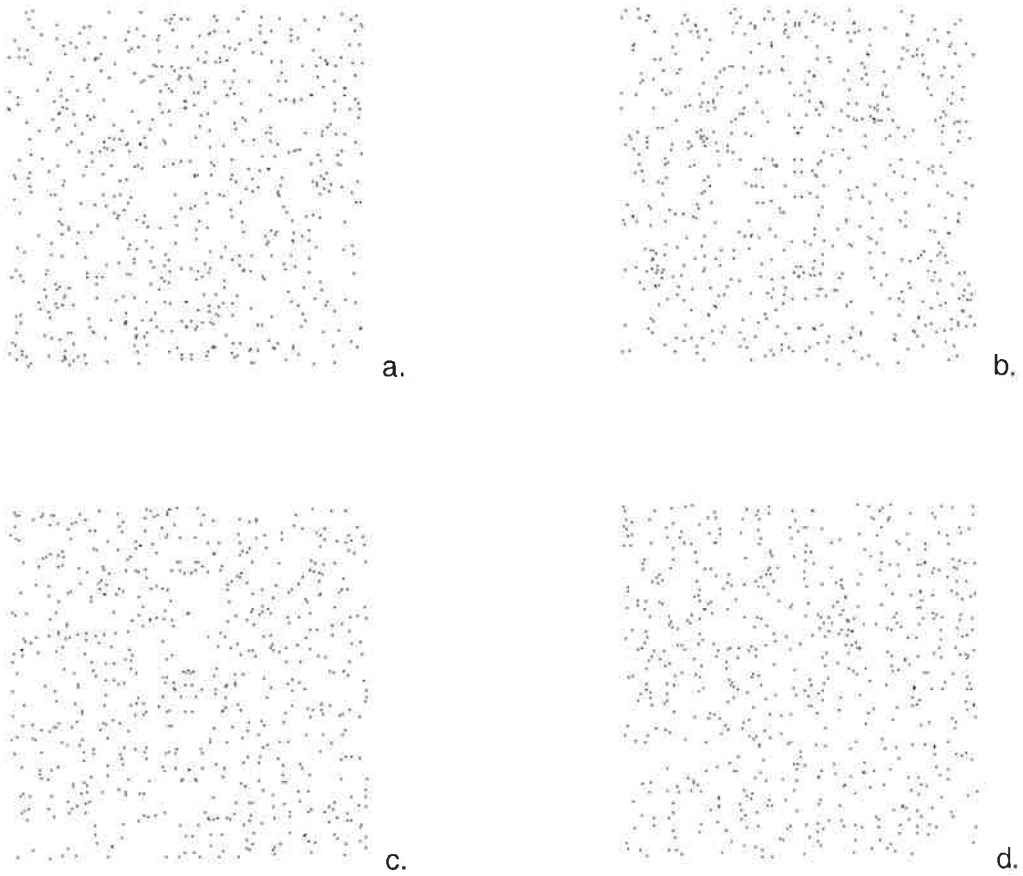

Figure 5.18 Examples of typical stimuli employed in Experiment 4. The experimental stimuli were comprised of a strip of symmetry embedded in random noise. Three symmetry strip widths are shown: 0.049 (a), 0.063 (b) and 0.077 (c). Figure 5.18d is an example of the standard stimulus (random noise pattern).

Because the figures in Jenkins (1983b) provide estimates of the variance associated with each data point, we are able to revert to the methodology applied to the Experiment 1 data. Table 5.4 presents the log marginal likelihoods and Bayes factors for the six model's fits to the Experiment 4 empirical data. The data indicate that the Voronoi model is the most likely model. However, given that the Isotropic/Alignment and Isotropic/Correlation models are at most 1.91 times less likely, based upon Jeffreys' (1961) guidelines we can conclude that no meaningful distinction can be made between the likelihoods of these three models. In contrast, the Barlow and Reeves and the two Oriented models have Bayes factors large enough to allow them to be rejected as plausible models of human performance for Experiment 4.

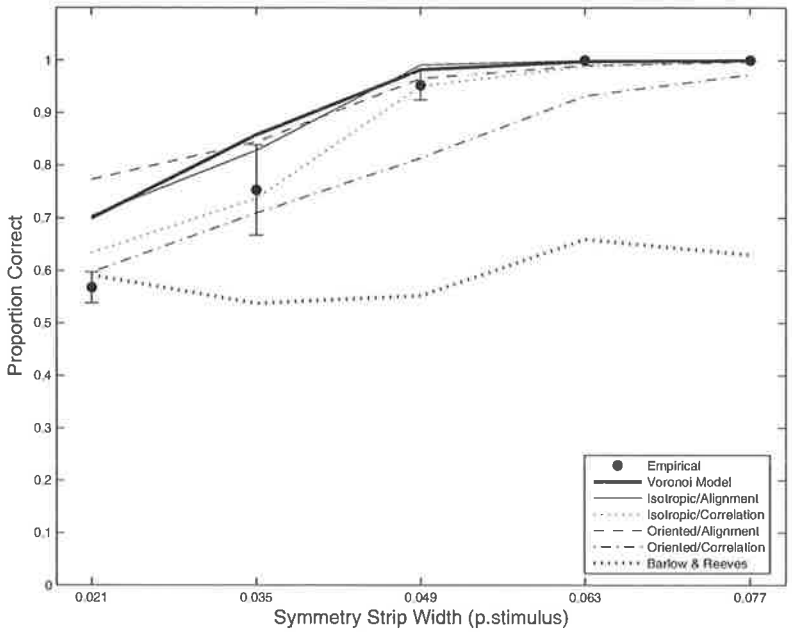

Figure 5.19 Best fits for the six models to the empirical Experiment 4 data.

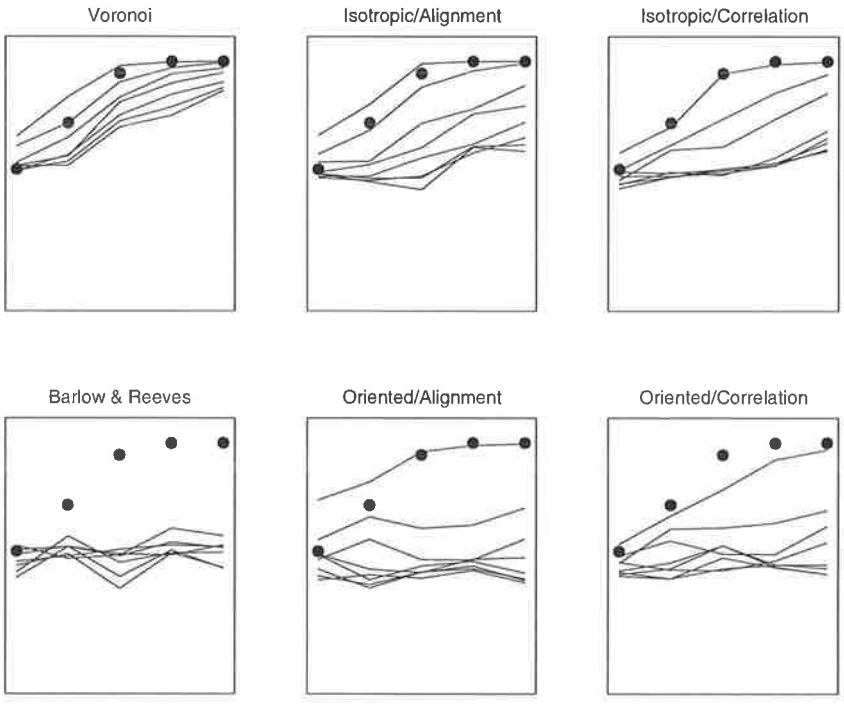

Figure 5.20 Fits of the six models across all parameterisations to the empirical Experiment 4 data.

Table 5.4 Log marginal likelihoods and Bayes Factors for the six models in Experiment 4.

Model	log marginal likelihood	Bayes Factor
Voronoi	32.52	1.00
Barlow & Reeves	-97.41	2.69×10^{56}
Isotropic/Alignment	32.12	1.49
Isotropic/Correlation	31.87	1.91
Oriented/Alignment	31.19	3.77
Oriented/Correlation	28.98	34.28

Note- Bayes Factors are taken in relation to the most likely model, which in this case is the Voronoi model.

5.10. Experiment 5- Central symmetry strip in noise vs. symmetry

Experiment 5 measured the distance from the axis at which random noise could be situated for observers to be unable to discriminate between pure symmetry and stimuli that contained a strip of symmetry flanked by noise.

5.10.1. Stimuli

The experimental stimuli employed in Experiment 5 were similar to those used in Experiment 4, but with a wider range of strip widths. The symmetry strip widths were 0.06, 0.10, 0.14, 0.17, and 0.21 (expressed as a proportion of the width of the stimulus). The standard stimuli were perfectly symmetrical stimuli. Examples of the stimuli from Experiment 5 are given in Figure 5.21.

5.10.2. Results

The results of Experiment 5 are similar to those of Experiment 3. Figure 5.22 demonstrates that model performance on this task is for the most part poor, with only two of the models (the Voronoi and the Isotropic/Alignment model) coming close to capturing the basic pattern of observer responses. Furthermore, the fits of the models across the range of their parameterisations (Figure 5.23) indicates that all of the models tend to over-perform. The similarity between the model responses to the Experiment 3 and 5 data is unsurprising: in both cases the task involves discriminating between perfect symmetry and symmetry with some form of random noise.

The log marginal likelihoods indicate that the Voronoi model provides the most likely account of the empirical responses (Table 5.5). However, the Bayes factors of the remaining models are testimony to the generally poor performance of the most likely model: despite being essentially flat functions they are only (at most) 18.89 times less likely than the best fitting model.

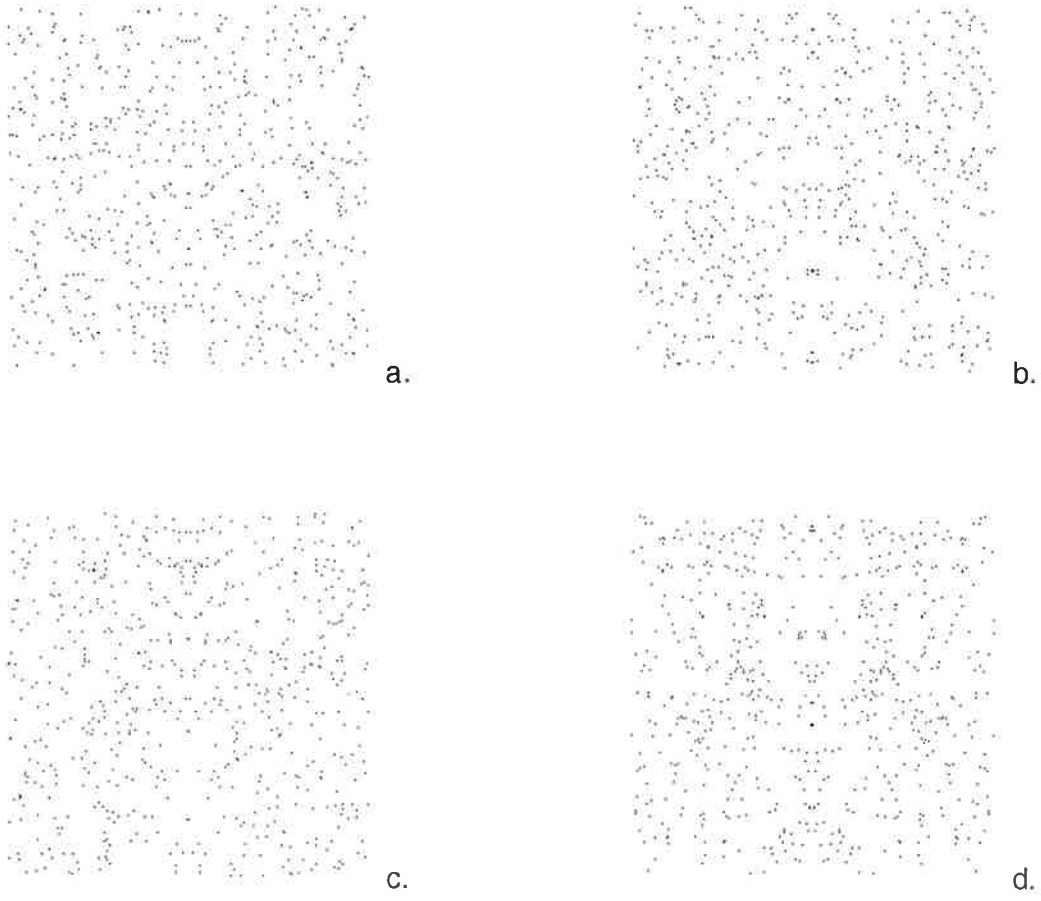

Figure 5.21 Examples of typical stimuli employed in Experiment 5. The experimental stimuli were comprised of a strip of symmetry embedded in random noise. Three symmetry strip widths are shown: 0.14 (a), 0.17 (b) and 0.21 (c). Figure 5.21d is an example of the standard stimulus (a perfectly symmetric pattern).

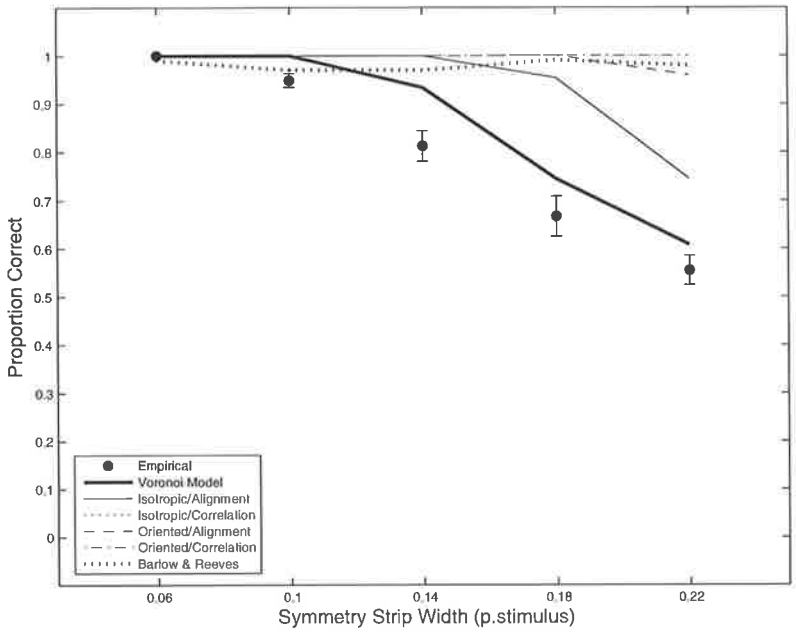

Figure 5.22 Best fits for the six models to the empirical Experiment 5 data.

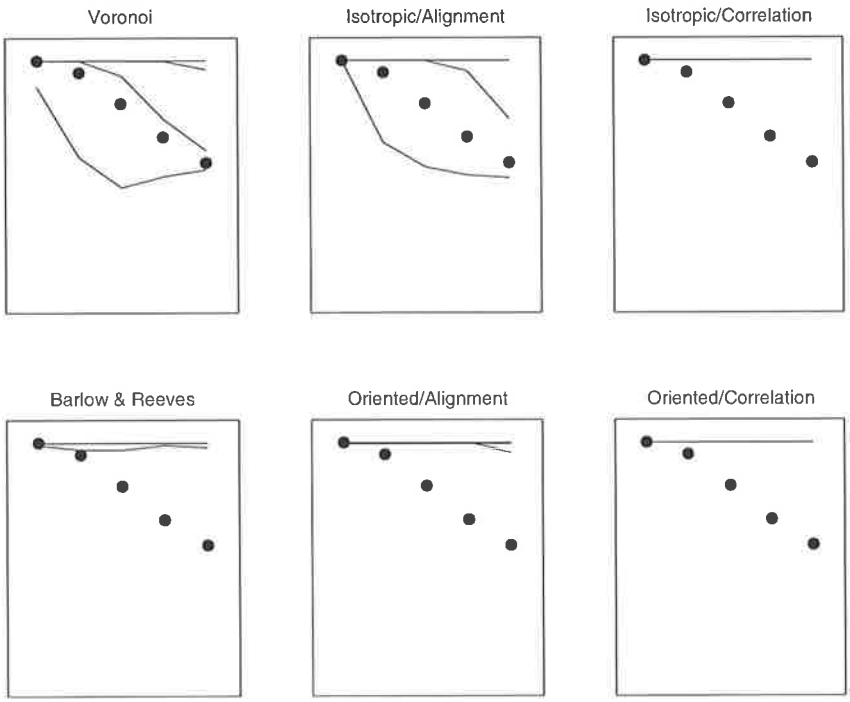

Figure 5.23 Fits of the six models across all parameterisations to the empirical Experiment 5 data.

Table 5.5 Log marginal likelihoods and Bayes Factors for the six models in Experiment 5.

Model	log marginal likelihood	Bayes Factor
Voronoi	27.54	1.00
Barlow & Reeves	24.73	16.62
Isotropic/Alignment	25.88	5.25
Isotropic/Correlation	24.63	18.39
Oriented/Alignment	24.74	16.56
Oriented/Correlation	24.63	18.39

Note- Bayes Factors are taken in relation to the most likely model, which in this case is the Voronoi model.

5.11. Experiment 6: Central noise strip in symmetry vs. noise.

Experiment 6 can be thought of as a repetition of the previous experiment, but with roles of symmetry and random noise swapped. The experiment was designed to determine the distance from the central axis at which symmetry was indistinguishable from random noise. In this case the experimental stimulus was a strip of randomly distributed points embedded in a symmetric pattern, and the standard stimulus was a randomly distributed noise pattern.

5.11.1. Stimuli

The noise strip widths were 0.06, 0.10, 0.14, 0.17, and 0.21 (expressed as a proportion of the width of the stimulus). The standard stimuli were randomly distributed noise patterns. Examples of the stimuli from Experiment 6 are given in Figure 5.24.

5.11.2. Results

Figure 5.25 indicates that best fits of the Voronoi and Oriented/Alignment models capture the observer's declining performance as width of the strip containing noise increases. However, in Figure 5.26 it can be seen that only the Oriented/Alignment model is able to approximate the empirical data across the majority of its parameterisations. Bayes factors (Table 5.6) indicate that the relative likelihoods of the Oriented/Alignment and Voronoi models (1 and 2.27 respectively) are so close that they cannot be meaningfully separated. However, in a similar manner to the previous experiment, the essentially flat functions of the two Correlation models are only 18.40 times less likely than the most likely model. Once again, the relative utility of the two most likely models should be interpreted in the light of this information.

Table 5.6 Log marginal likelihoods and Bayes Factors for the six models in Experiment 6.

Model	log marginal likelihood	Bayes Factor
Voronoi	28.74	2.27
Barlow & Reeves	26.69	17.67
Isotropic/Alignment	14.31	4.20×10^6
Isotropic/Correlation	26.65	18.40
Oriented/Alignment	29.56	1.00
Oriented/Correlation	26.65	18.40

Note- Bayes Factors are taken in relation to the most likely model, which in this case is the Oriented/Alignment model

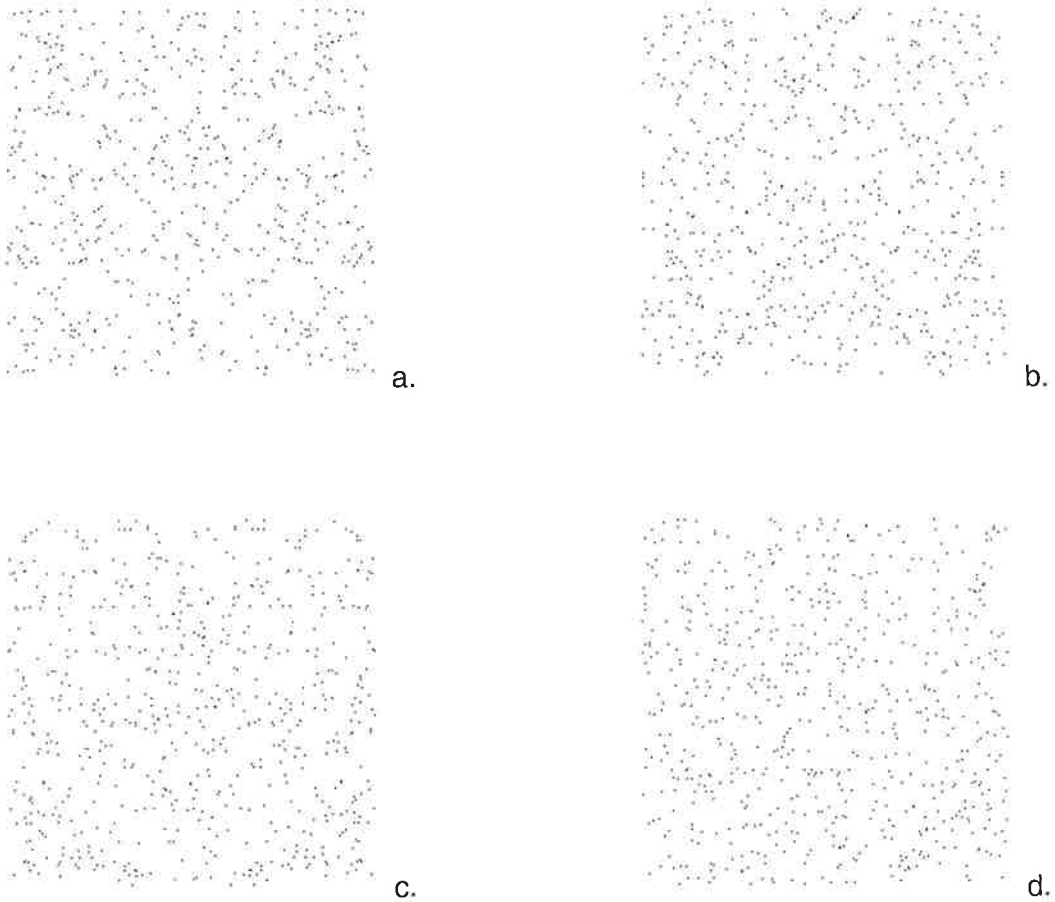

Figure 5.24 Examples of stimuli employed in Experiment 6. The experimental stimuli were comprised of a strip of noise embedded in symmetry. Three noise strip widths are shown: 0.14 (a), 0.17 (b) and 0.21 (c). Figure 5.21d is an example of the standard stimulus (a random noise pattern).

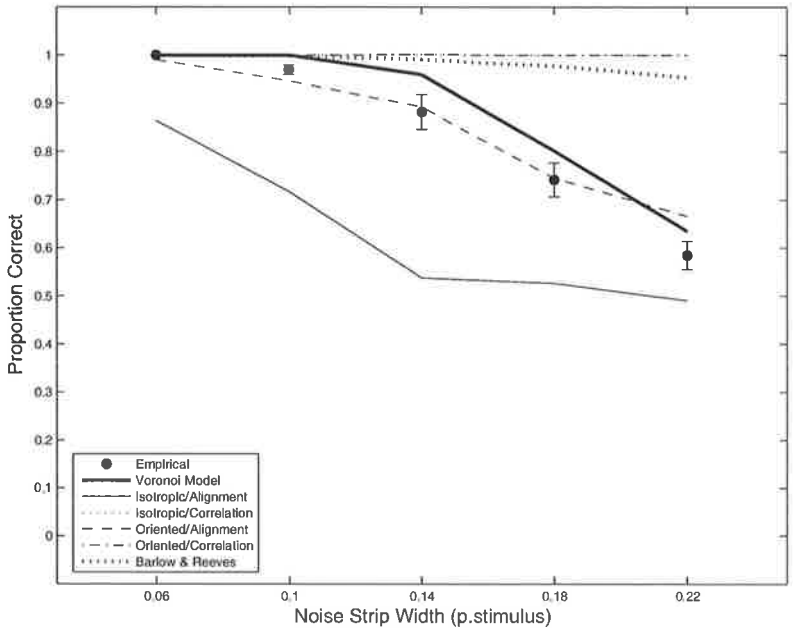

Figure 5.25 Best fits for the six models to the empirical Experiment 6 data.

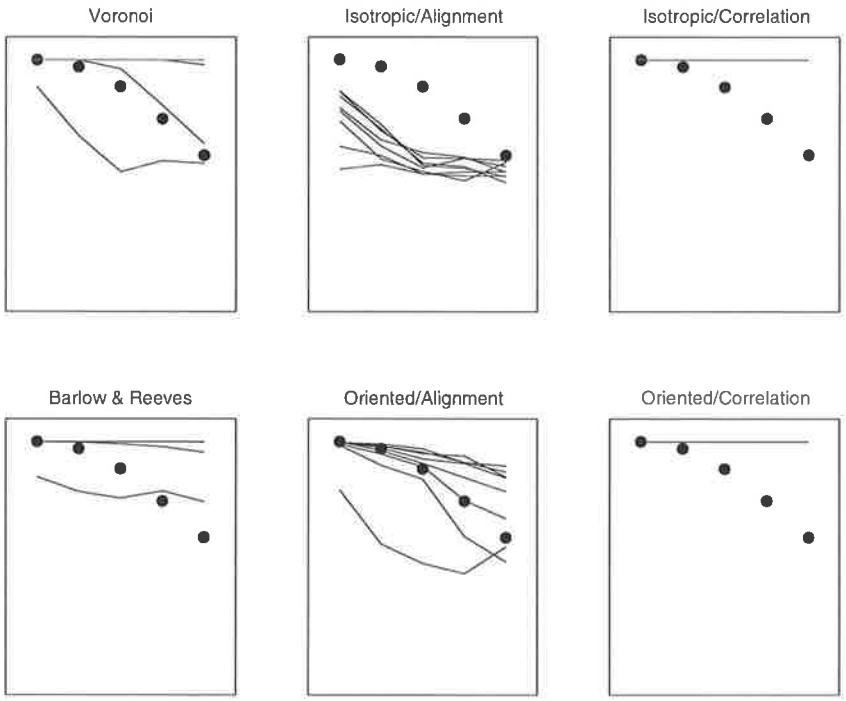

Figure 5.26 Fits of the six models across all parameterisations to the empirical Experiment 6 data.

5.12. Model evaluation overview

5.12.1. Model generality

Up until this point the analyses have been restricted to comparisons of the six models within each separate experiment. However, as has been argued in Chapter 4, the ability of a model to generalise across a number of data sets or experimental manipulations is a better test of its utility than its ability to fit a single data set. One of the strengths of Bayesian model selection is that it is able to naturally accommodate comparisons across multiple data sets by summing across each model's log marginal likelihood for each experiment.

Table 5.7 summarises the log marginal likelihoods for Experiments 1 to 6. For each experiment, the most likely model and any models with Bayes factors of less than 3.2 (i.e., Jeffreys (1961) suggested minimum criterion level for differentiating model likelihood) have been marked with an asterix. It can be seen that for five out of the six experiments the Voronoi model is either the most likely model, or has a log marginal likelihood so close to the most likely model that no meaningful distinction can be made between them. By summing the log marginal likelihoods it is possible to calculate Bayes factors for the relative likelihoods of the models across all six experiments. Unsurprisingly, the Voronoi model is the most likely model, with the next most likely model (the Oriented/Alignment model) being 3.48×10^{22} times less likely.

In Dakin and Watt's original study, based upon a qualitative comparison of the fits of the four filter-based models it was concluded that the Oriented/Alignment model was able to provide the best account of the empirical data (Dakin & Watt, 1996). Table 5.7 indicates that when the four filter-based models are considered in isolation the present study replicates Dakin and Watt's findings. This finding is important in two respects. First, the analyses confirm Dakin and Watt's conclusions and provide a quantitative account of the relative likelihood of the four models featured in their paper. Second, it is an indication of the reliability of the present study. Reproducing the results of Dakin and Watt suggests that that the four filter-based models from their study have been successfully replicated. Furthermore, it suggests that the ordering of model likelihoods is due to the different characteristics of the models, rather than any differences in the characteristics of the stimuli employed in this study and the stimuli employed in Dakin and Watt (1996).

Table 5.7 Log marginal likelihoods for Experiments 1 to 6, the sum log marginal likelihood and Bayes factors for all models and the four models from Dakin and Watt (1996) alone.

Model	log marginal likelihood							Bayes Factor	
	Exp 1	Exp 2	Exp 3	Exp 4	Exp 5	Exp 6	Sum	All Models	Filter models only
Voronoi	3.03	4.88*	-1.73*	32.52*	27.54*	28.74*	95.01	1.00	-
Barlow & Reeves	7.02*	1.16	-15.45	-97.41	24.73	26.69	-53.24	2.44×10^{64}	-
Isotropic/Alignment	-39.75	0.10	-4.88	32.12*	25.88	14.31	27.79	1.55×10^{29}	6.08×10^{3}
Isotropic/Correlation	2.06	-26.65	-73.73	31.87*	24.63	26.65	-15.15	7.01×10^{47}	2.73×10^{22}
Oriented/Alignment	-20.82	3.86*	-32.03	31.19	24.74	29.56*	36.50	2.56×10^{25}	1.00
Oriented/Correlation	5.46	2.83	-73.73	28.98	24.63	26.65	14.85	6.49×10^{34}	2.53×10^{9}

Note- Bayes Factors are taken in relation to the most likely model.
* most likely model and any models with Bayes factors less than 3.2.

Table 5.8 Comparison of the sum log marginal likelihoods, Bayes factors and order of model likelihoods for Experiments 1 to 6, and Experiments 1, 4, 5, & 6 only.

Model	Experiments 1 to 6			Experiments 1, 4, 5, & 6		
	Sum log marginal likelihood	Bayes Factor	Order of model likelihood	Sum log marginal likelihood	Bayes Factor	Order of model likelihood
Voronoi	95.01	1.00	1	91.85	1.00	1
Barlow & Reeves	-53.24	2.44×10^{64}	6	-38.95	6.46×10^{56}	6
Isotropic/Alignment	27.79	1.55×10^{29}	3	32.57	5.55×10^{25}	5
Isotropic/Correlation	-15.15	7.01×10^{47}	5	85.22	756.21	3
Oriented/Alignment	36.50	2.56×10^{25}	2	64.68	6.32×10^{11}	4
Oriented/Correlation	14.85	6.49×10^{34}	4	85.74	448.96	2

Because the log marginal likelihoods for Experiments 2 and 3 were calculated across a range of data precision assumptions, the contribution that these values add to the sum log marginal likelihoods is inherently uncertain. In order to control for this uncertainty the analyses were repeated excluding the Experiment 2 and 3 data. Table 8 provides a comparison of the two sets of analyses. As can be seen, the Bayes factors indicate that excluding Experiments 2 and 3 from the analyses has an effect on both the ordering of the model likelihoods, and on the magnitude of the differences between the models. Nonetheless, the Voronoi model remains the most likely model by a difference that is large enough to be considered decisive. It can therefore be concluded that success of the Voronoi model is not reliant on the data precision assumptions that were made in relation to the Experiment 2 and 3 data.

5.12.2. Range of parameter values

As explained in Chapter 4, log marginal likelihoods are calculated across all of a model's parameter values, not just the best fitting parameterisation. It is therefore possible that the relative likelihood of a model across all experiments is due (at least in part) to the influence of redundant parameterisations. In other words, if a model's best fits are achieved using only a limited range of parameterisations, then it is fair to question the utility of including a wider range of parameter values, particularly if the redundant parameterisations provide poor fits to the empirical data. This is particularly important in regards to the Barlow and Reeves model. Barlow and Reeves (1979) originally reported model fits for only two values (16 and 32 sub-regions); in this study, however, the model has been implemented using seven values of sub-

region numerosity. If the model only needs a restricted range of values to generalise across the range of experiments, then the redundant values may be unfairly penalising the model.

Table 5.9 lists the number of parameter values for each model, the best fitting parameterisation for each experiment, and the range of best-fitting parameterisations across all experiments. The last item can be thought of as a list of each model's non-redundant parameterisations. As can be seen, the only model showing redundancy in this respect is the Voronoi model, all other models need to utilise the full range of their parameterisations to fit the empirical data with any accuracy across all six experiments. Interestingly, if the range of parameterisations over which the Voronoi model likelihood is calculated is restricted to only 1 or 2 columns either side of the axis, the Voronoi model is still 3.16×10^{19} times more likely than the next best model. Furthermore, if the Voronoi model is restricted to only using a 2 column detection regions, it is still 1.25×10^{20} times more likely than the next best model.

Reducing the range of parameter values employed by any of the other models only serves to reduce their likelihood. It is interesting that this reduction in likelihood is the case even for the Barlow and Reeves model and the two cross-correlation models. In regards to the Barlow and Reeves model, the majority of the best-fitting parameterisations are in the range of 1-4 (between 4 and 64 subregions), suggesting that the model tends to simulate the empirical responses best when the size of the symmetrical sub-regions is reasonably large. The only exception to this is Experiment 4, when the best fit to the empirical data is achieved when the model is operating at its most sensitive level (196 subregions). However, Figure 5.20 indicates that this 'best' fit is actually reasonably poor, and does not appear to be much better than the fits achieved by the model when it employs lower sensitivity levels. The contribution that this high level of sensitivity provides to the model's overall likelihood is therefore questionable and it is possible that the fit in Experiment 4 comes at a cost of poor fits in the other experiments. As it turns out, reducing the range of parameterisations only serves to decrease the model's likelihood. When only levels 1 to 6 contribute to the likelihood, the model becomes 4.77×10^{16} times less likely, and restricting the range to levels 1 through 5 makes the model 4.75×10^{17} times less likely.

For Experiments 1 and 2, the Isotropic/Correlation and Oriented/Correlation models both achieve good fits using low frequency spatial filtering. In contrast to this, for Experiments 3, 5 and 6 these models perform equally badly across all parameterisations (see Figures 5.16, 5.23

Table 5.9 Description of model parameter, number of parameter values, best-fitting parameterisation for each experiment, and the range of best-fitting parameterisations across all experiments.

Model	Parameter	Values	Best-fitting parameterisation						Range
			Exp1	Exp2	Exp3	Exp4	Exp5	Exp6	
Voronoi	Number of columns in detection region	1 to 7 columns either side of axis	6	2	1	1	2	2	1 to 6
Barlow & Reeves	Number of subregions	7 levels (4, 16, 36, 64, 100, 144, and 196 sub-regions)	64	36	4	196	4	16	4 to 196
Isotropic/Alignment	Spread function (s) of Isotropic filter	8 levels (s = 2, 4, 6, 8, 10, 12, 14 and 16 pixels)	16	4	2	2	4	16	2 to 16
Isotropic/Correlation	"	"	12	16	-	2	-	-	2 to 16
Oriented/Alignment	Spread function (s) of Oriented filter	8 levels (s = 1.14, 2.82, 4.24, 5.65, 7.07, 8.48, 9.89 and 11.31 pixels)	11.31	2.82	1.14	1.14	1.14	4.24	1.14 to 11.31
Oriented/Correlation	"	"	11.31	11.31	-	1.14	-	-	1.14 to 11.31

Note: the Oriented/Correlation and Isotropic/Correlation models have equal fits across all parameter values for Experiments 3, 5 and 6.

and 5.26, respectively). Given that the fits achieved using high frequency spatial filtering appear to be making a positive contribution in Experiment 4 alone, it is possible that restricting the range of parameter values to only include low spatial frequency filters will increase the two models' overall likelihood values. However, as with the Barlow and Reeves model, restricting the range of parameter values has a detrimental effect only on model likelihood. For the Isotropic/Correlation model, when the likelihood is calculated using filters with s = 4 to 16 the model becomes 1.17×10^4 times less likely. Similarly, when the Oriented/Correlation model is restricted to using filters with s = 2.82 to 11.31 it becomes 1.23×10^{37} times less likely.

These results suggest that the high sum marginal log likelihood of the Voronoi model (relative to the other models) is not a result of the negative influence of redundant parameterisations. Reducing the range of parameter values available to the alternative models merely serves to decrease their likelihood. In contrast to this, the Voronoi model exhibits some redundancy in the range of parameter values necessary to successfully simulate human performance on the six tasks featured in this study. As has been demonstrated, dramatically reducing the range of Voronoi model parameterisations from the likelihood calculation does not change the order of model likelihood. However, it is possible to imagine situations in which the Voronoi model would best fit an empirical data set using the as yet unused parameterisations (for example, given that the width of the detection region appears to be related to stimulus presentation time (Dry, 2005; Tyler et al., 1995), longer presentation times would theoretically require the Voronoi model to employ more columns of cells). In this way, the 'redundant' parameterisations are an indication of the potential flexibility of the Voronoi model.

5.13. Discussion

The results summarised in Section 5.12 indicate that the Voronoi model is the most likely model of human observer symmetry perception out of the six models compared in this study. Furthermore, it has been demonstrated that the high sum log marginal likelihood of the Voronoi model is not a result of the data precision assumptions made in Experiments 2 and 3, or because the alternative models were unfairly penalised by the poor performance of redundant parameterisations.

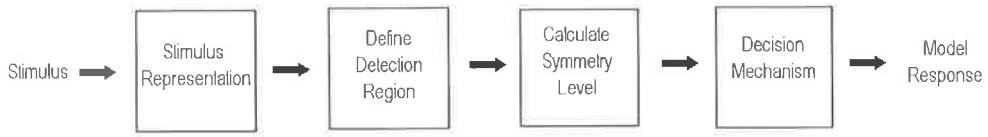

Figure 5.27 Schematic representation of the four steps involved in making a model response. All of the models share the same decision process, therefore the relative likelihoods of the models must stem from differences in the first three steps.

In order to understand the reasons for the Voronoi model's relative success we must first determine any crucial points of difference between the models. Figure 5.27 is a schematic representation of the six models under comparison in this study. The process of moving from a stimulus to a response can be thought of as a series of four discrete steps through which all of the models move. First, a stimulus representation is defined. For example, a stimulus might be transformed via some form of spatial filtering, or it might be partitioned into a series of discrete cells such as occurs with the Voronoi model and the Barlow and Reeves model. Second, the amount and location of information that it is assumed each model has access to is defined in terms of some form of detection region. For the Voronoi model and the two feature alignment models this is a spatially limited region adjacent to the central axis. For the two cross-correlation models and the Barlow and Reeves model, the region of detection encompasses the entire stimulus. Third, a value is assigned to each stimulus based upon some form of symmetry calculation (e.g., co-alignment of stimulus features, or a comparison of expected and observed dot densities). Finally, based upon the distributions of these values, a decision is made and the model discriminability is calculated.

As was discussed earlier, all of the models employ the same decision mechanism (see Section 5.4). Therefore, any differences in relative likelihood must stem from differences in the first three stages of this process. It would be tempting to suggest that the success of the Voronoi model is due to the underlying Voronoi representation and that this, in turn, is evidence of a similar representation in early vision. However, as shall be seen in the following sections, there are a number of crucial differences between the models in the middle two stages (detection region definition and symmetry level calculation) that appear to play a more immediate role in determining their relative likelihoods. First, I will discuss the effect that a

simple difference in symmetry calculation methods has upon model response to perfect symmetry. Second, I will discuss the effect of the different model assumptions relating to detection region definition.

5.13.1. Model response to perfect symmetry

Of the six experiments featured in this study, Experiment 3 produced the largest disparity in relative likelihood between the most likely model (the Voronoi model) and the remaining models. The task in this experiment involved discriminating perfectly symmetric stimuli from perturbed symmetry. The disparity between the models can be largely explained in terms of a difference in the way that the Voronoi model and the alternative models calculate the degree of symmetry within a stimulus, and the influence that this has on the resulting distributions produced by the models in response to perfect symmetry.

As was discussed in Section 5.4, for each experiment a given model produces two sets of responses, one for the experimental stimuli (i.e., the symmetric point patterns with varying levels of randomly distributed noise points) and one for the standard stimuli (i.e., the perfectly symmetric patterns). The two distributions have an associated mean and standard deviation, and the proportion of correct and incorrect decisions made by the model is determined by the degree of overlap between these distributions (Green & Swets, 1966).

For the Voronoi model, the degree of symmetry assigned to a stimulus is determined (in part) by the number of points located within the region of detection. Because the number of points within the region of detection can vary from stimulus to stimulus, the distribution of scores corresponding to the perfectly symmetric stimuli also has an associated variance. In contrast to this, each of the alternative models calculates the degree of symmetry within a stimulus in such as way that all perfectly symmetric stimuli are assigned the same value (equal to 1 for the cross-correlation and feature alignment models, and equal to 0 for the Barlow and Reeves model). Due to this, the resulting distribution of scores has no variance.

Figure 5.28 illustrates the influence that this difference in symmetry calculation method has upon the Experiment 3 data. For the Voronoi model (Figure 5.28a), there is variance associated with both the distribution corresponding to the perfectly symmetric stimuli and the distribution corresponding to the perturbed symmetry. The decision mechanism sets the

criterion at the junction of the two distributions and in many cases there will be an overlap between the two distributions and the criterion. For all of the other models (Figure 5.28b), the lack of variance in the distribution of perfectly symmetric stimuli results in the decision mechanism setting a criterion equal to the mean of this distribution (i.e., 1). Because all of the standard stimuli are equal to this criterion level, and the majority (or all) of the values in the distribution of experimental stimuli fail to overlap this criterion, model discrimination is (for the most part) perfect.

It could be argued that the response to perfect symmetry that the Barlow and Reeves, feature alignment, and cross-correlation models make is theoretically sound: a pattern is either perfectly symmetrical, or it is not. While there can be many grades of partial symmetry, only a

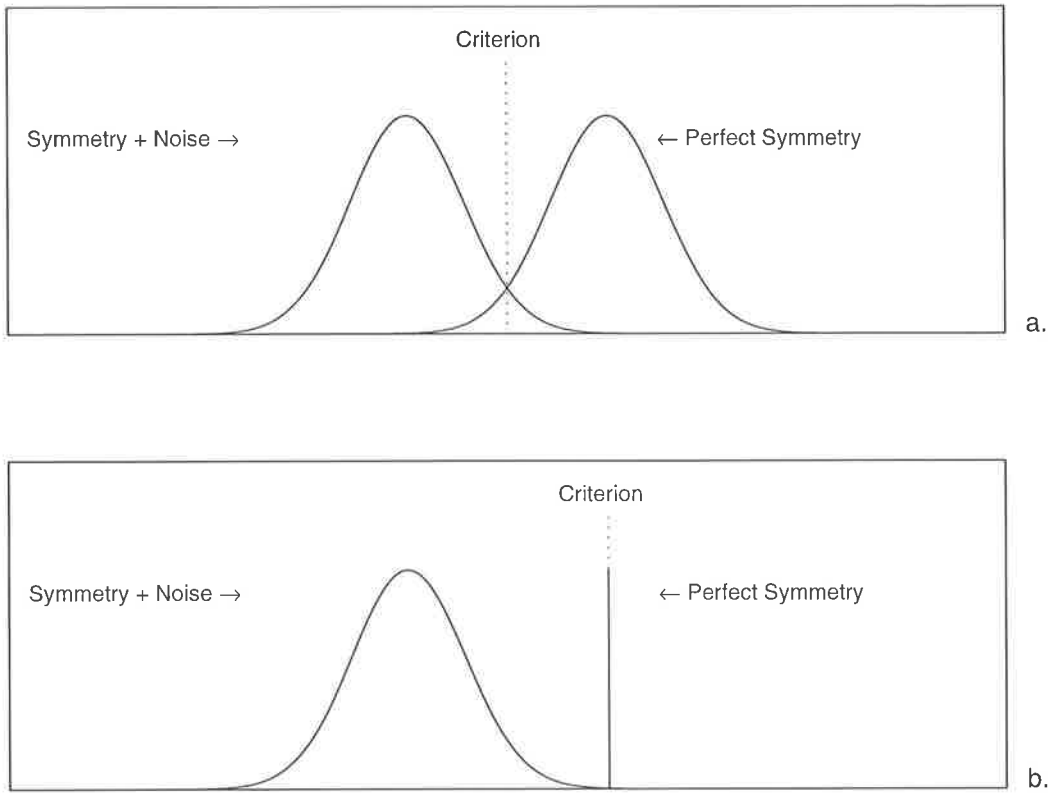

Figure 5.28 Comparison of theoretical distributions of responses to the experimental and standard stimuli in Experiment 3 for the Voronoi (5.28a) and all other models (5.28b). Note that there is no variance in the distribution of values assigned to the standard stimuli in 5.28b.

pattern that has two perfectly correlated halves can be classed as perfect symmetry. In other words, it would seem to make sense that there should be no variance associated with a model's response to these stimuli because they represent a unique state. Given this position, the response of these models is legitimate, and it is the decision mechanism that is at fault.

However, it should be remembered that the visual system only has limited access to the stimuli in this task (Barlow and Reeves (1979) reported that the stimuli in Experiment 3 were only presented for 500 milliseconds). The value that the Voronoi model assigns to a given stimulus can be thought of as being the amount of 'Symmetry' evidence accumulated by the model during this limited presentation time. It is assumed that the Voronoi model only has access to the information located within the detection region located at the stimulus midline. Due to differences in the distribution of point pairs across different perfectly symmetric stimuli, the total amount of information located within the detection region varies. Stimuli with more point-pairs within the detection region can be thought of as offering more 'Symmetry' evidence than stimuli with fewer pairs in the detection region.

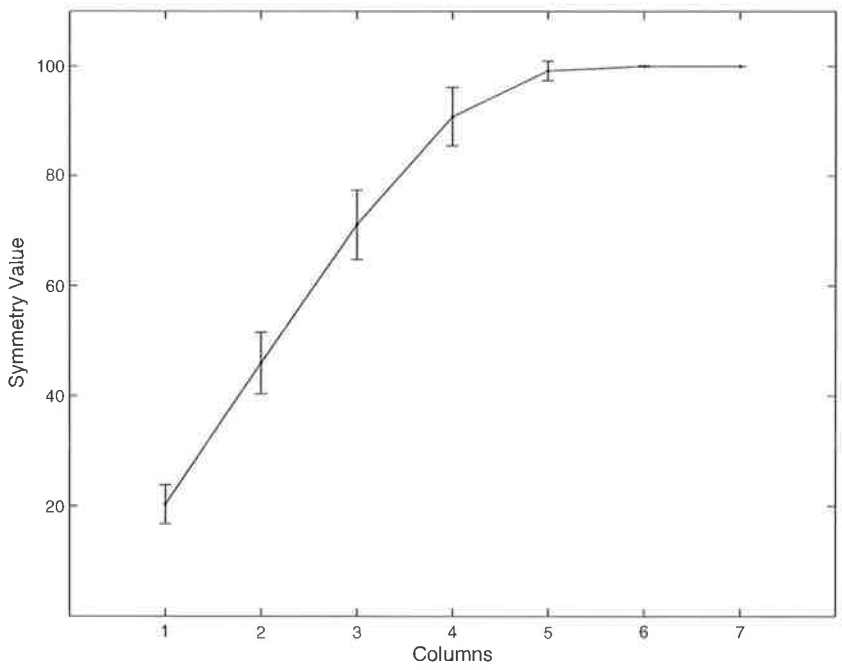

Figure 5.29 Mean and standard deviation of Voronoi model symmetry value assigned to perfectly symmetric stimuli in Experiment 3 for detection regions ranging from 1 to 7 columns either side of the stimulus midline.

It has been demonstrated that the size of the detection region for symmetry perception increases as a function of presentation time (Tyler et al., 1995). Presumably, when observers have unlimited access to stimuli the detection region would encompass the entire stimulus. Figure 5.29 presents the mean and standard deviation of the Voronoi model symmetry values assigned to perfectly symmetric stimuli in Experiment 3 across a range of detection region size assumptions. When the detection region is comprised of seven columns of cells located either side of the stimulus midline, the model has access to all of the point pairs in the stimuli. As can be seen, in this case the Voronoi model assigns the same value to all perfectly symmetric stimuli. In other words, under the assumption of unlimited stimulus presentation times the Voronoi model makes the same sort of single-value response to perfect symmetry that the alternative models make.

5.13.2. Model detection region

The second main differentiating feature of the Voronoi model is its ability to capture the limited region of detection exhibited by the human observers in Experiments 5 and 6. For example, in Experiment 6 the human observers' ability to detect the presence of symmetry approached chance levels when the central noise strip covered around 20% of the stimulus. The two cross-correlation models and the Barlow and Reeves model tend to over-perform in this experiment because they have access to all of the available information within a stimulus, and they are therefore sensitive to the presence of symmetric pairings regardless of their location. In contrast, the Voronoi model only employs information located within a spatially limited region directly adjacent to the central axis and, as a result, performance degrades in a similar manner to the performance of the human observers.

The Voronoi model is not unique in this respect: both of the feature alignment models also base their symmetry calculation upon information located within a spatially limited region. However, for Experiment 5, the two feature alignment models tend to over-perform for the same reasons as in Experiment 3. The task in Experiment 5 involved discriminating between perfect symmetry and a central strip of symmetry embedded in random noise. As with Experiment 3, the feature alignment models return the same value for the perfectly symmetric stimuli and the resulting lack of variance in the distribution of scores leads to near-perfect performance.

For Experiment 6, the Oriented/Alignment model achieves a number of good fits to the empirical data, whereas the Isotropic/Alignment model underperforms across all parameter values. This difference in performance can be explained by the existence of a relationship between the sensitivity of the feature alignment models and the size of their detection regions. For both of the feature alignment models, as filter size increases the size of the resulting detection region also increases. However, coarse filter scales tend to minimise the difference between random and symmetrical stimuli and, as a result, there is a negative relationship between model sensitivity and filter size. The detection region generated by oriented filtering was wide enough to include symmetrical information for all but the finest filter scale (s = 1.14). In contrast to this, the detection region generated by isotropic filtering only included symmetrical information when the model employed coarse filter scales and, as a result, the Isotropic/Alignment model lacked the sensitivity to discriminate between the two classes of stimuli.

For Experiments 5 and 6, the Voronoi model is either the most likely model (Experiment 5), or performs at a level such that no meaningful distinction can be made between the Voronoi and the most likely model (Experiment 6). However, Figures 5.23 and 5.26 indicate that the Voronoi model's success is due largely to the fit of a single parameterisation. When the model employs a two column detection region it is able to achieve a reasonably close fit to the empirical data, however, for all other parameterisations the model either over- or under-performs. The reason that this occurs is simple enough: of the seven detection region sizes tested for the Voronoi model, the average width of the two column region corresponds most closely with the widest value of strip width employed in Experiments 5 and 6 (i.e., strip width = 0.21, or 21% of the width of the stimulus).

Figure 5.30a shows a typical Experiment 5 stimulus with a strip width equal to 0.21. Figure 5.30b shows the detection region for one column of cells either side of the stimulus midline. Because the average width of the detection region is only 0.09, the model will not be able to discriminate at a better than chance level for stimuli with strip widths greater than this value. Figure 5.30d shows a three column detection region. In this case the average width of the detection region is 0.27 and so the model is able to detect the presence of randomly positioned points. As a result, the model performs at a level far greater than the human observers are capable of for this class of stimuli. Figure 5.30c shows a two column detection region

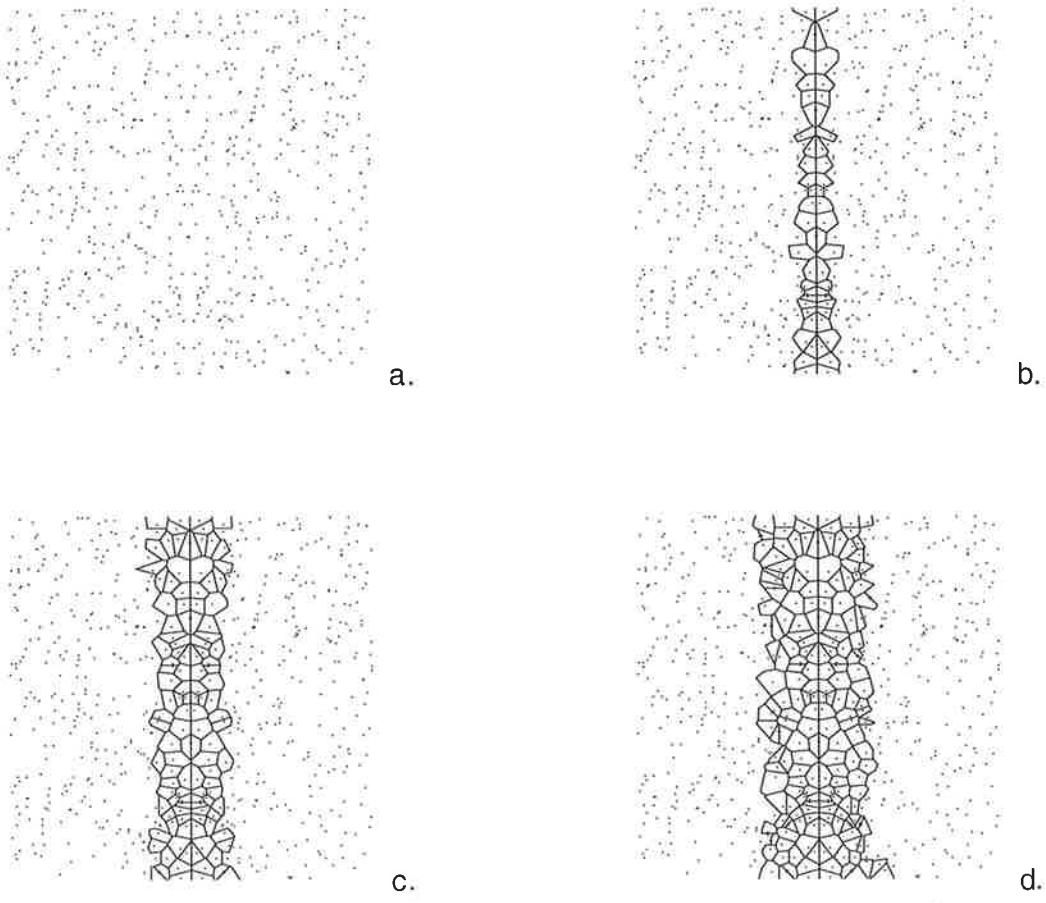

Figure 5.30 Comparison Voronoi model detection regions for a Experiment 5 stimulus comprised of central strip of symmetry strip (width = 0.21) embedded in randomly positioned points (5.30a). Three detection region sizes are shown: 1 column of cells either side of the stimulus midline (5.30b), two columns of cells (5.30c) and three columns of cells (5.30d).

(average width = 0.18). As can be seen the width of the detection region corresponds closely with the width of the central strip of symmetry, resulting in the model discriminating at a level similar to the human observers (i.e., close to chance). For stimuli with thinner symmetry strip widths the model is able to detect randomly positioned points and, as the strip width decreases, model performance improves in a similar manner to the pattern of results in the empirical data.

5.13.3. Conclusions

It has been argued that the superior performance of the Voronoi model can be largely explained by two factors. First, the Voronoi model's symmetry calculation method appears to cause the model to respond differently than the alternative models to the presence of perfect symmetry such as found in Experiments 3 and 5. For the decision mechanism to operate in a meaningful way there must be some degree of variance in the distribution of model responses to stimuli. Unlike each of the alternative models, the Voronoi model is able to return a range of scores to different perfectly symmetric stimuli. It has been demonstrated that the Voronoi model's response does not violate the notion of perfect symmetry existing as a singular state, but is a plausible reflection of the inherent uncertainty associated with limited stimulus presentation times.

Second, unlike the two cross-correlation based models and the Barlow and Reeves model, the Voronoi model is able to capture the limited region of detection demonstrated by the human observers in Experiments 5 and 6. It has been demonstrated that the Voronoi model only fits the empirical data well when it employs a two column detection region because the average width of this size of detection region corresponds most closely with the physical properties of the stimuli employed in these experiments. Furthermore, although Table 5.9 indicates that the best fitting parameterisations for the Voronoi model range from 1 to 6 columns, it has been shown that when the Voronoi model only employs a two column detection region for all experiments, the model is still 5.38×10^{19} times more likely than the next most likely model.

As Figure 5.27 indicates, for each of the models compared in this study the definition of the detection region and the calculation of stimulus symmetry levels both occur *after* the process of defining the stimulus representation. Because there are differences between the six models at these later stages, we are unable to make any conclusions about the relative utility of the various model representations. In other words, although the Voronoi model is a far more likely model of symmetry perception than the five alternative models featured in this study, we cannot conclude that Voronoi tessellation is any more likely a form of stimulus representation in early vision than a stimulus that has undergone oriented spatial filtering.

There are a number of simple modifications that can be made to the alternative models that should (theoretically) improve their performance. For example, it should be a simple matter to

restrict the size of the detection region for the Barlow and Reeves model and the two cross-correlation based models and thereby improve their performance in Experiments 5 and 6. Furthermore, it is possible that adding a small amount of random noise to the response distributions corresponding to perfectly symmetrical stimuli will improve the performance of all the alternative models in Experiments 3 and 5.

The proposed modifications may enhance the quality of the model fits to the empirical data across a number of parameterisations. However, there is also the potential that they may introduce unneeded complexity in the form of numerous parameter combinations that fail to match the empirical data. Because Bayesian model selection penalises models for complexity, it is possible that the modified models may actually end up with lower sum log marginal likelihoods than their original incarnations. One of the strengths of the Voronoi model is that the proposed modifications are integral features of the model. It is not necessary to make *ad hoc* assumptions about these features because they can be seen to change in a plausible manner in response to changes in experimental stimuli (i.e., the size of the detection region for the Voronoi model scales with dot density).

In conclusion, this study has demonstrated that the Voronoi model is a more plausible model of human symmetry perception than a number of previously published models. The Voronoi model produces good fits to empirical data, has low complexity (in a parametric sense), generalises well across a range of experimental settings and stimulus manipulations, and is based upon theoretically sound assumptions. In this sense it can be seen that the Voronoi model meets all of the requirements of a plausible model of human cognitive processing (Jacobs & Grainger, 1994; Myung & Pitt, 1997; Roberts & Pashler, 2000).

The success of the Voronoi model appears to be largely based upon two features: the spatial extent of its detection region, and the inherent variability associated with its responses to perfect symmetry. It has been suggested that there are a number of simple modifications which may possibly improve the performance of the alternative models. One of the primary aims of the next study is to compare the performance of the Voronoi model to those of a set of modified models. It is proposed that by controlling for any major differences in the models in terms of detection region definition and symmetry calculation it may be possible to draw conclusions regarding the relative utility of Voronoi tessellation as a form of stimulus representation in early vision.

6. STUDY 2

6.1. Introduction

The spatial organisation of scenes or images is one of the fundamental tasks of early vision and in this thesis it has been argued that the visual system may be employing a process similar to Voronoi tessellation to extract the spatial relations necessary for achieving this task. The Voronoi model outlined in this thesis is an example of how this process might be implemented in relation to the perception of symmetry in simple point patterns. The model operates by initially generating an underlying representation of a given stimulus based upon Voronoi tessellation and then performing a number of correspondence calculations to determine the degree to which the stimulus can be organised into a set of point pairs with uniform orientation and collinear midpoints. The Voronoi model is both theoretically motivated and physiologically plausible and it has been demonstrated that the model is capable of simulating the performance of human observers on a range of symmetry detection tasks.

In Study 1 the performance of the Voronoi model was compared to that of five previously published alternative models of symmetry detection (Barlow & Reeves, 1979; Dakin & Watt, 1996) across a range of empirical symmetry detection data sets (Barlow & Reeves, 1979; Jenkins, 1982). The results of Study 1 indicated that, of the six models under comparison, the Voronoi model was able to provide the most likely description of the processes underlying observer performance. For five out of the six tasks the Voronoi model was able to provide either the best fit to the empirical data, or a fit that was so close to that of the best-fitting model that no meaningful distinction could be made between them. Furthermore, when considered across all six tasks, Bayes factors indicated that the Voronoi model was 2.56×10^{25} times more likely than the next most likely model.

A major point of difference between the Voronoi model and each of the alternative models is the form of their underlying stimulus representation. For example, in response to evidence suggesting that perceptual grouping is the result of early spatial mechanisms filtering an image at different scales and orientations (e.g. Campbell & Robson, 1968; Hubel & Wiesel, 1962; Kuffler, 1953), four of the models derive stimulus representations via some form of spatial filtering (Dakin & Watt, 1996). Additionally, the Barlow and Reeves model obtains a representation by dividing each stimulus into a set of discrete square-shaped subregions

(Barlow & Reeves, 1979). Given these considerable differences in stimulus representation, one way to interpret the results of Study 1 is to suggest that the superior performance of the Voronoi model is evidence of the existence of a Voronoi-like representation in early vision. In other words, it could be argued that the Voronoi model is able to provide better fits to the empirical data than the alternative models because its underlying representation more closely simulates the true form of the stimulus representation in early vision.

However, there are a number of problems with this argument. As was discussed in Chapter 5, before any conclusions can be drawn about the comparative utility of the models' stimulus representations it is necessary to control for any other differences between the models. For example, it was suggested that the superior performance of the Voronoi model can be largely explained by two factors that occur subsequent to the derivation of the stimulus representation. First, the Voronoi model's symmetry calculation method leads the model to respond differently than the alternative models in the presence of perfect symmetry. Second, the Voronoi model was able to capture the spatially limited region of symmetry detection demonstrated by the human observers. There are a number of simple modifications that can be made to the alternative models that should (theoretically) improve their performance without violating any of their assumptions.

Additionally, there are two features of Study 1 that may have unfairly biased the results in favour of the Voronoi model. First, the paper from which Experiments 2 and 3 were drawn (Barlow & Reeves, 1979) did not include any indication of the variance associated with the observers' discriminability data for these tasks. As a result, the model comparisons for these two experiments were carried out using a range of data precision assumptions. In both cases the Voronoi model was found to have the highest log marginal likelihood but it is possible that this may be a result of the range of data precision assumptions that were adopted in the analyses. Second, because the spatial filter based models were only implemented using a limited number of filter scale sizes, it is possible that the probability density of these models was underestimated. As will be explained in Section 6.3.1, modifying the number of filter scale sizes over which the filter modes are implemented will provide a more accurate indication of their true log marginal likelihoods.

Given these problems, the aims of Study 2 are threefold:

1. To empirically replicate Experiments 2 and 3, as described in Chapter 5, in order to provide data sets that include variance information and thereby ensure that the associated model comparisons are not based upon potentially biased data precision assumptions.

2. To modify the models in light of the results of Study 1

3. To compare the new versions of the models across the same six tasks in Study 1 (substituting the replications of Experiments 2 and 3 for the original data sets).

In the following sections these three aims will be dealt with in turn.

6.2. Empirical replication of Experiments 2 and 3

In Study 1 the data for Experiments 1, 2 and 3 were drawn from Barlow and Reeves (1979). Unfortunately, the figures in this paper provide no indication of the variance associated with observer discriminability for Experiments 2 and 3. Following Lee (2001; 2004), the log marginal likelihoods and their associated Bayes factors for these experiments were calculated across a range of data precision assumptions and, in each case, it was demonstrated that the Voronoi model was the most likely model of observer performance.

Given that the Voronoi model's high relative likelihoods were dependant upon the range of data precision assumptions employed in the analysis of these two experiments, the sum log marginal likelihood (a measure of generalisablity across all experiments) was calculated both with and without Experiments 2 and 3. It was demonstrated that excluding Experiments 2 and 3 changed the magnitude of the differences between the models. Nonetheless, the Voronoi model remained the most likely model by a difference that was large enough to be considered decisive. It was therefore concluded that the success of the Voronoi model was not reliant on the data precision assumptions that were made in relation to the data from Experiments 2 and 3.

While this conclusion is both satisfactory and justifiable, it is preferable not to rely upon these sorts of assumptions. In the following section, empirical replications of Experiments 2 and 3 are presented; the primary aim of these were to obtain observer variance data.

6.2.1. Method

Four observers (including the author) took part in both experiments. All observers had training on each of the tasks prior to data collection.

All stimuli were presented on a Hitachi CM721F colour computer monitor with a vertical refresh rate of 85 Hz and a pixel resolution of 1280 x 1024. Pixel width was 0.26 mm. Viewing distance was set at 57 cm. Stimuli were generated and presented using MATLAB 7.0.1. All stimuli were comprised of one hundred 2 x 2 pixel black dots presented within a white circular region with a 256 pixel diameter located at the centre of the monitor. Although a range of stimulus presentation times were employed in the original experiments (0.1-0.5 seconds (Barlow & Reeves, 1979)), stimulus presentation time was held constant at 0.1 seconds for both experiments.

For Experiment 2, the standard stimuli were comprised of randomly positioned points and the experimental stimuli were comprised of a varying number of symmetric point pairs and randomly positioned points. The proportion of symmetrically positioned points in the experimental stimuli ranged from 0.2 to 1.0 in 0.2 increments. For Experiment 3, the standard stimuli were perfectly symmetric point patterns and the experimental stimuli were comprised of a varying number of symmetric point pairs and randomly positioned points. The proportion of symmetric pairings in the experimental stimuli ranged from 0.0 to 0.8 in 0.2 increments.

For both experiments, the symmetric stimuli were always vertical bilateral symmetry with an axis centred in the middle of the stimulus. For the stimuli containing random noise points, equal numbers of points were located on each side of the central axis in order to negate the potential effects of density cues. A total of 50 experimental stimuli and 50 standard stimuli were generated for each stimulus level.

Both experiments employed a single-interval, two-alternative forced-choice design. Stimuli were presented in blocks of 50 trials. The signal parameter was fixed within each set of trials, so that observers were required to discriminate between standard stimuli and a single level of experimental stimuli. The observers were informed of the level of symmetry in the experimental stimuli prior to beginning each block and were presented with multiple examples of both standard and experimental stimuli. The order of presentation for the separate

trial blocks was randomised to negate order effects. Each block contained equal numbers of standard and experimental stimuli (25 standard and 25 experimental stimuli) and, within each block, the experimental and standard stimuli were presented in a random order. Each of the data points in the following figures was derived from 800 observations.

Each trial was preceded by the presentation of a fixation point (duration = 0.5 seconds) located in the centre of the presentation window. This was followed by the presentation of a single stimulus. The observer was then required to indicate which of the two populations (standard or experimental) the stimulus was drawn from, using a mouse to click on one of two buttons presented at the bottom of the screen. The observer was then given feedback as to whether their choice was correct or not. When the observer was ready, they initiated the next trial by clicking on a button that was located equidistant from the two choice buttons.

6.2.2. Results

Observer performance was expressed as *d'*. The results of the replications are shown in Figures 6.1 and 6.2. In both cases, the mean empirical discriminability of the four observers in the current study is compared with data from Barlow and Reeves (1979). Although there are small differences between the original and replicated data sets, the two distributions correspond closely for both experiments, with the original data falling (for the most part) no more than one standard deviation away from the replicated data.

In Study 1, the log marginal likelihoods for Experiments 2 and 3 were calculated under the assumption of a data precision range determined by pilot experiments (0.1 to 0.23 and 0.09 to 0.13 standard errors respectively). The current study indicates that for Experiment 2, the standard errors range from 0.09 to 0.32. Reanalysing the Study 1 data using this range has a minimal effect on the six models' log marginal likelihoods, but does not change the rank order of relative likelihood as indicated by the Bayes factors. Similarly, the current study indicates that the range of standard errors in Experiment 3 (0.11 to 0.34) is wider than had been indicated by the pilot study, but re-analysing the Study 1 data using this range does not change the rank order of model relative likelihood.

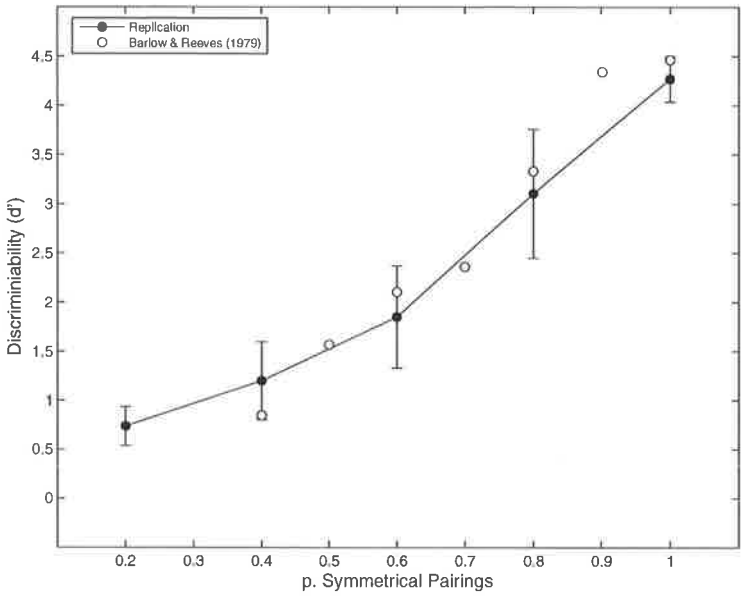

Figure 6.1 Comparison of mean empirical discriminability (d') for the replication of Experiment 2 (black circles) and Barlow and Reeves (1979) data (white circles). The error bars represent one standard deviation.

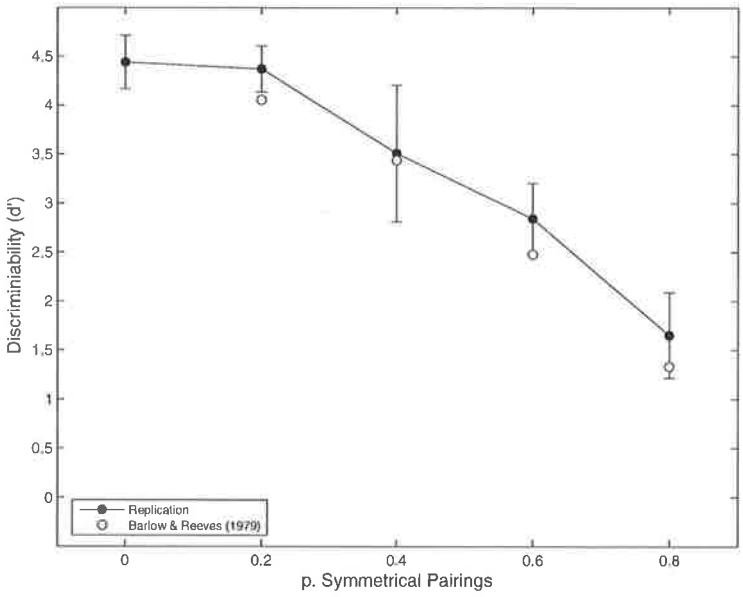

Figure 6.2 Comparison of mean empirical discriminability (d') for the replication of Experiment 3 (black circles) and Barlow and Reeves (1979) data (white circles). The error bars represent one standard deviation.

6.2.3. Summary

Comparison of the replicated data with the original data from Barlow and Reeves (1979) indicates a close correspondence between the two data sets. The data from the replications indicate a wider range of variance in observer performance than was suggested by pilot experiments. However, reanalysis of the Study 1 data using these wider ranging variance assumptions did not have a substantial effect upon the relative likelihoods of the models.

Importantly, the current experiments provide variance data for each empirical data point in Experiments 2 and 3. In the following analyses the replicated data will be used in place of the original data, thereby avoiding the need for making data precision assumptions.

6.3. Model modifications

6.3.1. Range and number of model parameter values

In Chapter 5 it was demonstrated that the range of parameter values over which a model is implemented could affect the model's relative likelihood. For example, it was shown that there was a degree of redundancy in the range of detection region sizes employed by the Voronoi model and that by reducing the model's parameter range from 1-to-7 columns in the detection region down to 1-to-5 columns it became 2.04 times more likely. None of the other models demonstrated a similar redundancy. Rather, the best-fits of the alternative models spanned the entire range of their respective parameter ranges. It is therefore plausible that better model fits could be obtained by *increasing* the range of these models parameter values.

A second factor that can potentially affect model likelihood is the number of values within a model's parameter range. For example, Figure 6.3a shows an empirical data set (circles) and the response of an imaginary model (lines) that has been implemented over 6 values of a single parameter. The model's associated likelihood function is shown in 6.3b. Figure 6.3c shows the responses of the same model but in this case it has been implemented across 11 parameter values. Here the range of the values is the same, but the increments of change are smaller. The model's new likelihood function (6.3d) shows that when the model was only implemented across six levels the model likelihood was underestimated at the centre of the distribution and slightly overestimated at the tails of the distribution.

In order to obtain a model's true marginal likelihood it is necessary to implement the model at every possible parameter value within the chosen parameter range. For models with a limited

number of discrete parameterisations (such as the Voronoi and Barlow and Reeves models) this is a simple matter to achieve. However, for the Isotropic and Oriented models it should be assumed that the underlying variable (the spread function of the filter) is either continuously, or close to continuously, distributed. In this case, obtaining model fits for every possible parameter value becomes a non-trivial computational problem.

In Bayesian model selection, likelihoods are typically sampled from a range of points in a model's parameter space using Markov Chain Mote Carlo or similar methods and, when the number of randomly sampled values is sufficiently large, the shape of the true likelihood function can be estimated with reasonable accuracy (Myung & Pitt, 1997). However, given

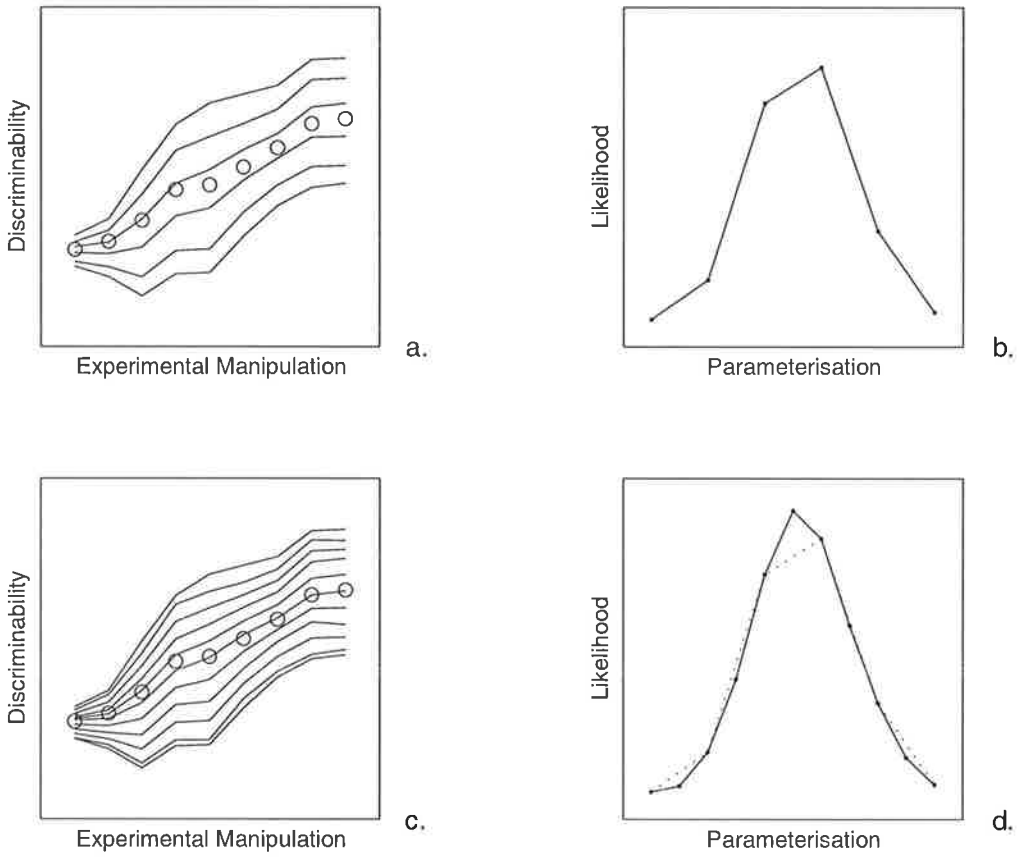

Figure 6.3 Comparison of model response and associated likelihood functions for the same model implemented across 6 parameter levels (a, b) and 11 parameter levels (c, d).

that implementing the Oriented/Alignment model across the range of values specified in Study 1 took between 18 and 24 hours for a single experiment, time constraints render this approach impractical. Although it is not ideal, a more constrained approach to estimating the filter models' likelihood functions has been adopted in the current study: in order to fill out the shape of the distribution the models were implemented using values at half-steps between the values employed in Study 1 (i.e., $s = 2, 3, 4, 5, 6$ rather than $s = 2, 4, 6$, etc).

6.3.2. Detection region

A number of experiments have demonstrated that the perception of symmetry appears to be limited to a spatially restricted region located immediately adjacent to the axis of symmetry (Barlow & Reeves, 1979; Dakin & Herbert, 1998; Jenkins, 1982; Julesz, 1971; Rainville & Kingdom, 2002). For example, in Experiment 6 the observers were required to discriminate between random noise stimuli and stimuli comprised of a strip of random noise embedded at the centre of an otherwise perfectly symmetric dot pattern. The results indicated that as the width of the central strip of noise was increased, discriminability decreased. The Voronoi model and the two feature alignment models only have access to information that is located within a limited detection region and consequently these models were able to capture the basic pattern of the empirical results. On the other hand, the two cross-correlation models and the Barlow and Reeves model all have access to information located across the entire stimulus. These models were able to detect the presence of symmetry (Experiment 6) or deviations from symmetry (Experiment 5) regardless of its location relative to the central axis and, as a result, the models tended to over-perform.

The empirical data from Experiments 5 and 6 suggest that the spatial extent of the detection region for stimuli comprised of 650 points is approximately 21% of the width of the stimulus. Theoretically, it should be possible to bring about a decrease in discriminability similar to that demonstrated by the observers by reducing the size of the area over which the cross-correlation models and the Barlow and Reeves model are able to integrate information from 100% down to around 21%. For the cross-correlation models this is a relatively simple process. Each stimulus is 256 pixels wide; therefore, the best fit to the empirical data should occur when the cross-correlation procedure is restricted to a central region with a width of approximately 54 pixels. Figure 6.4a shows a stimulus from Experiment 5 with a central

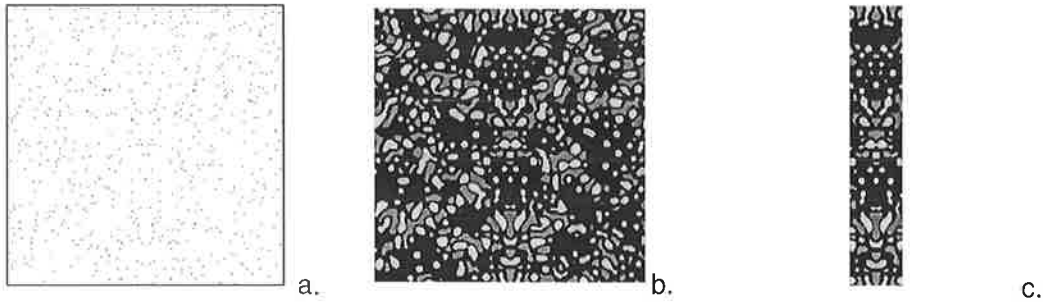

Figure 6.4 An example of a typical stimulus from Experiment 5 (a), and a comparison of the Isotropic/Correlation detection region using 256 pixels (b) and 54 pixels (c).

symmetry strip that covers 21% of the stimulus. Figures 6.4b and 6.4c show the stimulus after it has been convolved with an isotropic filter. The detection region in 6.4b encompasses the entire stimulus, whereas the detection region in 6.4c is only 54 pixels wide. As can be seen, the detection region in 6.4c only contains symmetrical information; therefore, the model would be unable to discriminate between this stimulus and a perfectly symmetric stimulus.

Although it is possible to predict the approximate width of the detection region that would maximise model likelihood in Experiments 4, 5 and 6, it does not necessarily follow that the same detection region width would maximise the likelihood in the remaining experiments. Rainville and Kingdom (2002) demonstrated that the spatial extent of the detection region scales with stimulus density. Specifically, detection region size appears to be negatively related to stimulus density, therefore we could expect that the Correlation models would need a wider detection region for Experiments 1 to 3 (dot density = 100) than for Experiments 4 to 6 (dot density = 650). Furthermore, if the models are expected to be able to generalise to very high or low density stimuli then it is necessary to allow the models detection regions to range from very thin to very wide. In order to include all possible detection region widths for stimuli 256 pixels wide and with varying dot densities, the cross-correlation models were implemented using 127 detection region column widths: from 1 pixel either side of the midline to 128 pixels in 1 pixel increments.

For the Barlow and Reeves model restricting the size of the detection region is a slightly more complex process. The Barlow and Reeves model was implemented at six levels of sub-region numerosity: 4, 16, 36, 64, 100, 144, and 196 sub-regions. The size and number of the sub-

regions dividing a stimulus places restrictions upon the range of detection region widths at which the model can be implemented. The reason for this is demonstrated in Figure 6.5, which shows a comparison of a number of different detection region sizes for the Barlow and Reeves model. For 6.5a and 6.5b the initial representation is divided into 16 sub-regions, and for 6.5c and 6.5d the initial representation is divided into 64 sub-regions. In 6.5a and 6.5c all of the sub-regions contribute to the detection region, whereas in 6.3b and 6.3d only the two centre-most columns of sub-regions are included. It should be noted that only the detection region shown in 6.3d could be expected to simulate the performance of human observers in Experiments 4 to 6: all of the other detection regions contain both symmetry and noise information, whereas the detection region in 6.3d only contains symmetric pairings.

For the Barlow and Reeves model the number of potential detection regions that an initial representation can be divided into is given by:

$$W_s = \frac{\sqrt{S}}{2},$$

where W_s is the number of detection regions widths (given as columns of sub-regions either side of the stimulus midline) and S the number of sub-regions in the initial representation. Accordingly, for a base representation comprised of 16 sub-regions, such as 6.5a, the total number of detection region sizes is 2 (i.e., either 1 or 2 columns of sub-regions either side of the midline). On the other hand, a 64 sub-region representation such as 6.5c could have detection regions comprised of 1, 2, 3 or 4 columns either side of the stimulus mid-line.

As with the cross-correlation models it is possible to predict the Barlow and Reeves model detection region widths that will generate the best fit to the empirical data in Experiments 4 to 6, but it does not follow that these would be the optimal widths for Experiments 1 to 3. In order for the model to be able to generalise to experimental situations involving either very high or very low density stimuli it is necessary to implement the model using all possible detection region widths.

Up until this point the detection region for the Voronoi model has been comprised of columns of cells spanning the entire height of the stimulus. However, as was discussed in Chapter 3, studies by Dakin and Herbert (1998) and Rainville and Kingdom (2000; 2002) have provided

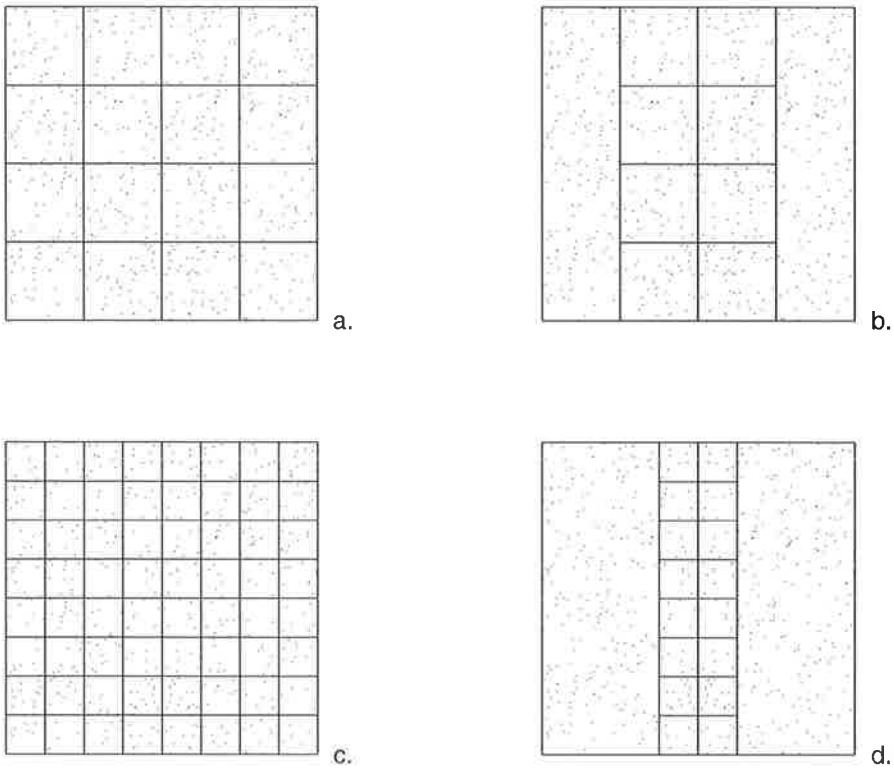

Figure 6.5 Comparison of different sized detection regions for the Barlow and Reeves model using an initial representation of 16 sub-regions (a, b) and 64 sub-regions (c, d). In 6.5 a and c the detection region encompasses the entire stimulus. In 6.5b and d the detection region is restricted to a single column of cells either side of the stimulus midline.

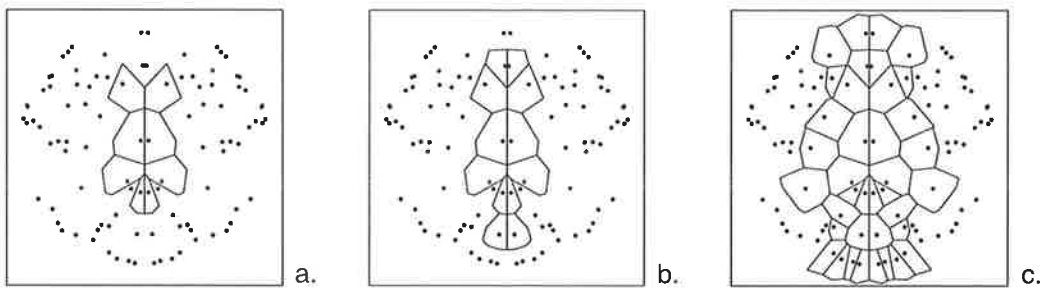

Figure 6.6 Examples of the Voronoi model detection region with varying widths and aspect ratios. 6.6a has a width of 1 column either side of the stimulus midline and an aspect ratio of 1:2. 6.6b has a width of 1 and an aspect ratio of 1:3. 6.6c has a width of 2 and an aspect ratio of 1:2.

evidence suggesting that the detection region for symmetry is limited both horizontally *and* vertically. In light of this, a more sophisticated version of the Voronoi model (which allows for limitations to both width and height of the detection region) is tested in Study 2. The empirical data from and Herbert (1998) and Rainville and Kingdom (2002) indicate that there is an approximately 1:2 width-to-height ratio for the detection region, although in Rainville and Kingdom (2000) it was demonstrated that depending on the characteristics of the stimulus the aspect ratio could be as great as 1:20. Because the actual aspect ratio is unknown, the width-to-height ratio of the Voronoi model detection region was free to vary from 1:1 to 1:7 in increments of 1.

Examples of the Voronoi model detection region with varying height to width ratios are given in Figure 6.6. As can be seen in Figure 6.6, the width of the detection region places restrictions on the number of unique height values over which the model can be implemented. The width-to-height ratio in 6.6c is only 1:2, but the detection region spans the entire height of the stimulus. Any increase in aspect ratio would return an identical model response, and it is possible that this could result in skewed log marginal likelihoods. In order to control for this only the fits of the unique detection regions were included in the final analyses.

6.3.3. Additive noise

As was discussed in Study 1, one of the major problems with the alternative models is that they each return the same value for all of the perfectly symmetric stimuli. As a result of this the models all tend to over-perform on Experiment 3 (Section 5.13.1 explains the reason for this is in detail). This problem can be addressed by assuming that there is a distribution of potential symmetry scores that a model could return for a given stimulus.

In their current form all of the models (including the Voronoi model) will return the same symmetry value for an individual stimulus. However it is reasonable to assume that the perceptual system is inherently noisy, resulting in an approximately Gaussian probability distribution of sensory values that could be attributed to a stimulus (Green & Swets, 1966). Figure 6.7 illustrates the effect of placing a Gaussian probability distribution over the values assigned by a model to each individual stimulus. The two data points in 6.7a represent the symmetry values that a model assigns to two different stimuli. As can be seen, the two values lie on either side of the criterion, and would therefore be classified as belonging to different stimulus classes (i.e., a symmetrical stimulus with random noise and a perfectly symmetric

stimulus). In 6.7b a Gaussian probability distribution has been placed over each of the scores. In this case the most likely response is to classify each stimulus in the same manner as before, but there is now a small chance that either stimulus might be misclassified.

As in Study 1, for Experiments 1 to 3 the decision criterion for each separate experimental stimulus signal level parameterisation (i) is calculated as:

$$\text{Criterion}_i = \mu_i^e + \frac{\sigma_i^e}{\sigma_i^e + \sigma^s}\left(\mu^s - \mu_i^e\right),$$

where μ and σ are the means and standard deviations of the symmetry values assigned by a model to the various standard (s) and experimental (e) stimuli. In the current study the minimum value that σ was allowed to take was equal to the standard deviation of the Gaussian probability distribution placed over each symmetry value, so even when a model assigns the same value to all of the symmetric stimuli, the criterion will no longer be equal to the mean of these values. Theoretically, this should allow the models to show a decrease in discriminability similar to that demonstrated by the human observers in Experiment 3.

For Experiments 1 to 3 the probabilities of hits *p(H)* and false alarms *p(F)* were obtained by comparing each stimulus symmetry value to the criterion and determining the proportion of the symmetry value's associated probability distribution that lies above and below the criterion. For Experiments 4 to 6 the proportion of correct responses was obtained by comparing the symmetry value for each experimental stimulus to the values of all of the standard stimuli. The proportion of the experimental stimulus probability distribution greater than the value of the standard stimulus was averaged across all comparisons to give the probability of a correct identification.

6.3.4. Voronoi model symmetry calculation

Study 2 also provides an opportunity to compare the performance of a number of alternative versions of the Voronoi model symmetry calculation procedure. This is analogous to the comparison of the Oriented/Alignment and Oriented/Correlation models, where the form of the underlying stimulus representation is identical (stimuli are convolved with horizontally aligned spatial filters), but the method of symmetry calculation is different (i.e., the alignment of midline crossing stimulus features versus a cross-correlation procedure).

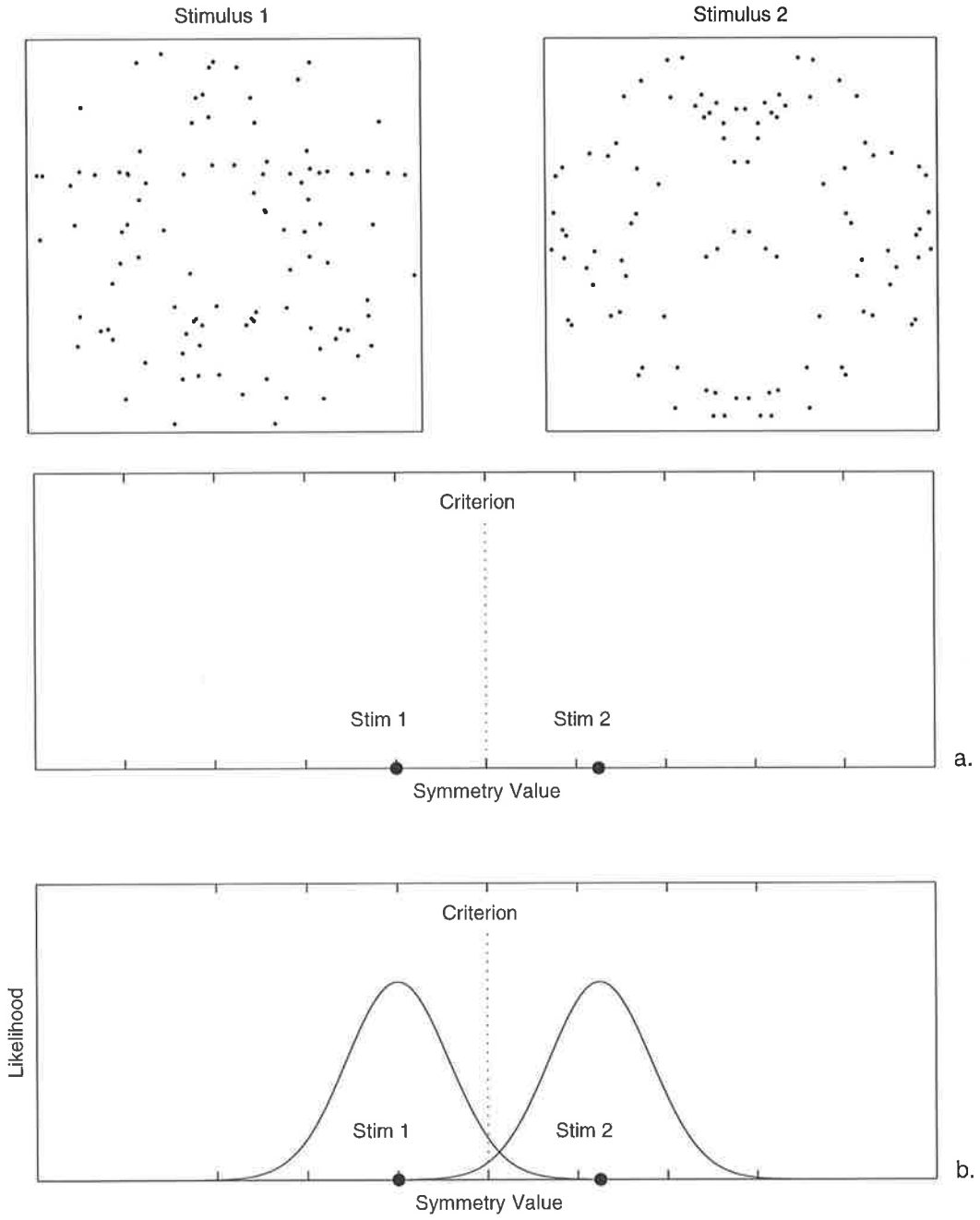

Figure 6.7 The effect of placing a Gaussian probability distribution over the value that a model assigns to each stimulus.

The Voronoi model detects symmetry by searching for pairs of points that are located in roughly the same location on opposite sides of the stimulus midline. The model achieves this by reflecting the Voronoi cell associated with each of the points in the detection region across the stimulus midline and counting the number of points that are located in the region corresponding to the reflected cell. When the observed number of points is greater than 0, the expected number of points (1) is divided by the observed number of points resulting in a value between 1 and 0. As a result, reflected cells containing a single point are given a value of 1, indicating an unambiguous point-pair match. On the other hand, reflected cells containing more than one point will be given a value of 0.5 or less. These values are then summed across all of the points in the detection region giving a total symmetry score for the stimulus.

In its current form the Voronoi model assigns a value of 0 to reflected cells that contain no points. In other words, points that have no matching pair do not contribute any value to the total symmetry score for the stimulus. However, it could be argued that points which cannot be matched are a potentially useful indicator of asymmetry: a near-perfectly symmetric stimulus will contain fewer unmatched points in comparison to a highly distorted or random stimulus. One way to utilise this information would be to penalise the total symmetry score by the number of unmatched points. Furthermore, it is possible that only either the unmatched points or the unambigously matched points are necessary to signal the presence of symmetry, and that the information provided by the ambiguous pairings is redundant.

In order to explore these possibilities four alternative versions of the Voronoi model were implemented. The Voronoi Type-1 (V1) model counts the number of points in the detection region that have a single point located in its reflected cell. The Voronoi Type-2 (V2) model has a symmetry calculation procedure identical to that employed in Study 1: for each point in the detection region the expected number of points in the reflected cell is divided by the observed number. Points with no match are assigned a value of 0. The Voronoi Type-3 (V3) model is identical to the V1 model with the exception that unmatched points are more heavily penalised: in this case unmatched points are assigned a value of -1. Finally, the Voronoi Type-4 (V4) model only counts the number of points in the detection region that have no matching point.

Although these models do not exhaust the universe of potential calculation methods that the Voronoi model could employ they are reasonably representative and can be thought of as

defining a continuum. The V1 model measures only perfect or near-perfect symmetry, the V2 and V3 models penalise this symmetry measure in terms of the level of asymmetry in a stimulus (with the V3 model weighting asymmetry more heavily than the V2 model), and the V4 model only takes into account the level of asymmetry within a stimulus.

6.4. Symmetry model comparison

This section presents a comparison of model performance across the six empirical data sets. Initially I will present the fits of the five alternative models (the Barlow and Reeves, Isotropic/Correlation, Isotropic/Alignment, Oriented/Correlation and Oriented/Alignment models) paying particular attention to the effect of any modifications on model performance. This will be followed by a comparison of the four different approaches to Voronoi model symmetry calculation.

6.4.1. The alternative models

The Barlow and Reeves model has three parameters: number of sub-regions in the base representation (S), number of columns of sub-regions included in the detection region (W_s), and level of added Gaussian noise (n_σ). In Study 1 the range of sub-region numerosity values was 4, 16, 36, 64, 100, 144, and 196. The results of Experiment 4 in Study 1 indicated that the best fit of the Barlow and Reeves model to the empirical data was achieved using 196 sub-regions. It is possible that better fits may be achieved with larger numbers of sub-regions, therefore, in Study 2 the range of sub-region numerosity values was increased to encompass 4, 16, 36, 64, 100, 144, 196, 256, 324, and 400. As was discussed previously the number of potential detection region widths for each level of sub-region numerosity is given by:

$$W_s = \frac{\sqrt{S}}{2}.$$

The level of added noise (n_σ) ranged from 0 to 0.3 in increments of 0.05, where each value represents the standard deviation of the Gaussian probability distribution placed over each symmetry value.

The Isotropic/Alignment and Oriented/Alignment models both have two parameters: filter size (s) and level of added Gaussian noise (n_σ). In order to obtain a more accurate estimate of the true model likelihood functions the models employed a wider range of filter sizes, with smaller increments of change than employed in Study 1. The spread function of the Isotropic

filter ranged from 1 to 17 pixels in 1 pixel steps. For the Oriented filter the spread function ranged from 0.70 to 11.02 in 0.70 steps. The level of added noise (n_σ) ranged from 0 to 0.3 in increments of 0.05.

The Isotropic/Correlation and Oriented/Correlation models both have three parameters: filter size (s), level of added Gaussian noise (n_σ), and detection region width (W). The filter size and Gaussian noise assumptions were identical to those of the feature alignment models. The width of the detection region ranged from 1 column of pixels either side of the stimulus midline to 128 columns, in 1 pixel increments.

Each model was implemented using the full range of parameter values indicated above. Fits were then examined to determine the range of parameter values that maximised each model's likelihood across all six experiments. For example, Figure 6.8 shows contour plots of the Isotropic/Alignment model fit for the six experiments across the full range of its parameterisations. As can be seen, for the majority of the experiments the best fits are obtained under the assumption of a small amount of additive noise. This indicates that there is a degree of redundancy in the range of values that the additive noise parameter needs to take in order for the model to adequately fit the empirical data and suggests that reducing the range of parameter values over which the model fits are calculated would increase the model's overall likelihood.

For the spatial filter (s), additive noise (n_σ), and subregion numerosity (S) parameters, the range of parameter values was reduced to the upper bound of the best fitting values across all six empirical data sets. In regards to the detection region width parameter values, unlike the Voronoi models (which have detection regions that scale automatically with changes in stimulus density), in order that the Barlow and Reeves and cross-correlation models be able to generalise across a range of varying stimulus densities it was necessary to include all of their potential detection region widths when calculating the log marginal likelihoods. In other words, none of the values of W_p or W_s was considered to be redundant.

Figures 6.9 to 6.13 show the non-redundant fits of each of the alternative models to the empirical data from Experiments 1 to 6. For each experiment the best fitting parameter combination is shown by a black line, and the corresponding parameter values are indicated at

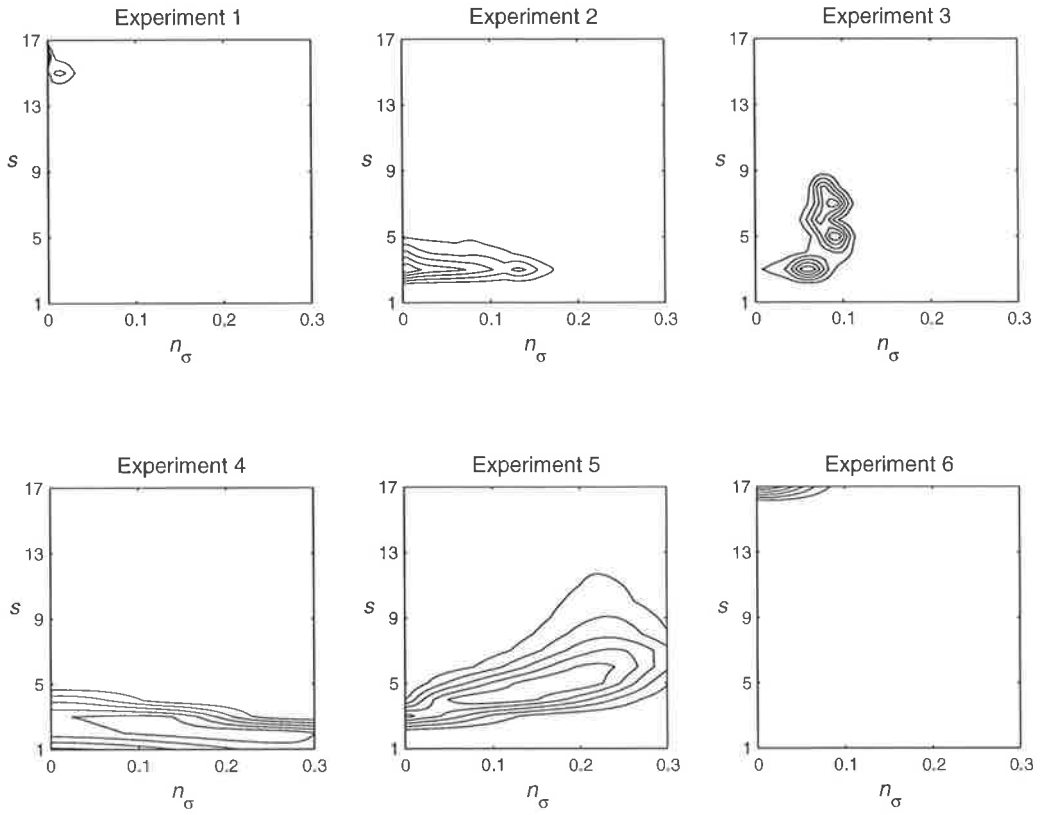

Figure 6.8 Contour plots showing the Isotropic/Alignment model level of fit for the six experiments across a parameter space ranging from 1 to 17 for the filter size parameter (s) *and 0 to 0.30 for the additive noise parameter* (n_σ).

the bottom of each panel. A qualitative comparison of the data fits reveals two immediately obvious features. First, for the majority of the models the best-fitting parameterisations are able to provide reasonably good fits to the pattern of the empirical responses across all six experiments. In particular, model fits are no longer uniform and close to perfect for Experiments 3 and 5, suggesting that the modifications that were made to the models contribute significantly to improving their ability to simulate the performance of the human observers.

The second obvious feature revealed by Figures 6.9 to 6.13 is that all of the models also produce numerous response patterns that are not consistent with the empirical data. This

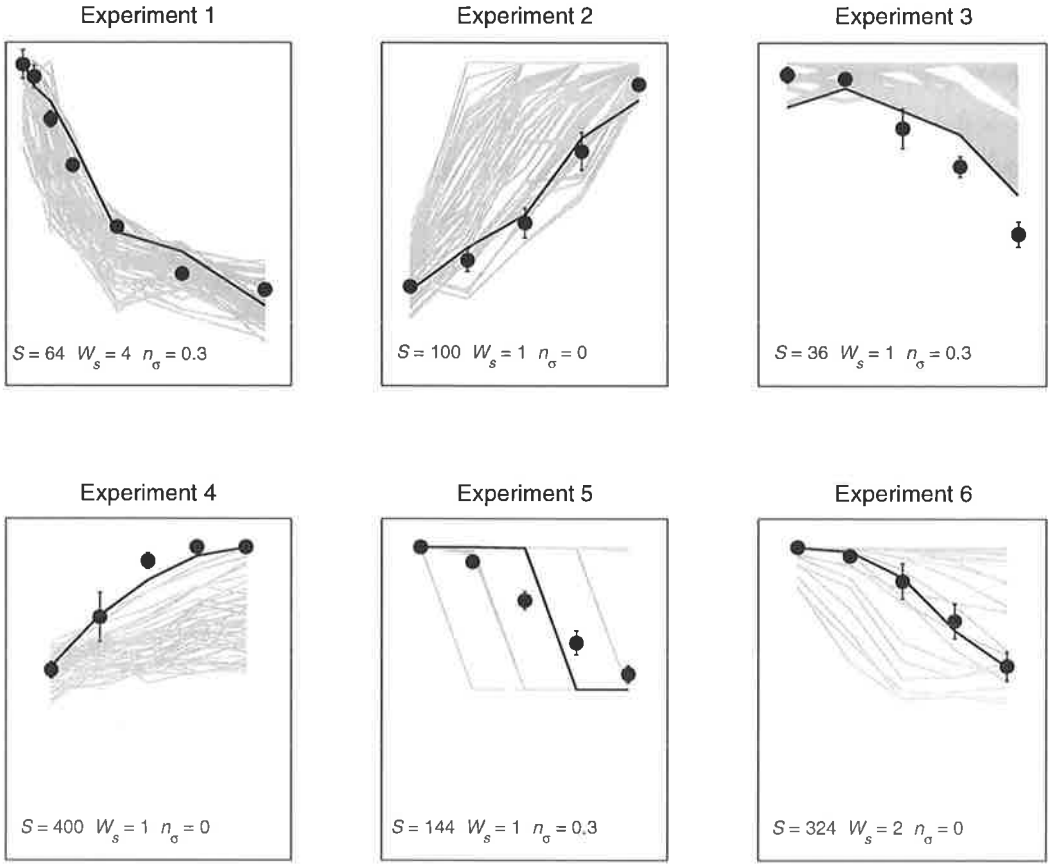

Figure 6.9 Barlow and Reeves model fits for Experiments 1 to 6 using a sub-region (S) parameter range of 36 to 400, detection region width (W_s) parameter range of 1 to 10 sub-regions and an additive noise (n_σ) parameter range of 0 to 0.3. For each experiment the best fitting parameter combination is shown by a black line. The parameter values of the best fit are indicated in the bottom left corner of each panel.

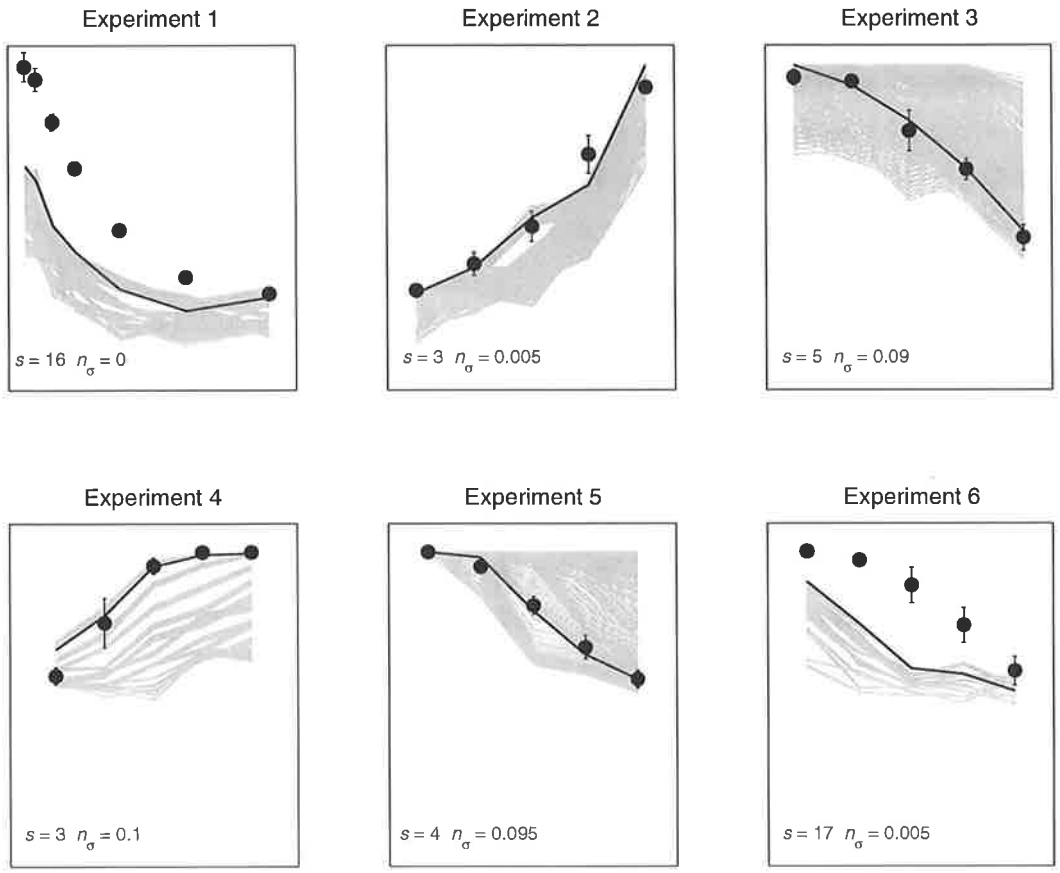

Figure 6.10 Isotropic/Alignment model fits for Experiments 1 to 6 using a spatial filter (s) parameter range of 3 to 17 pixels and an additive noise (n_σ) parameter range of 0 to 0.1. For each experiment the best fitting parameter combination is shown by a black line. The parameter values of the best fit are indicated in the bottom left corner of each panel.

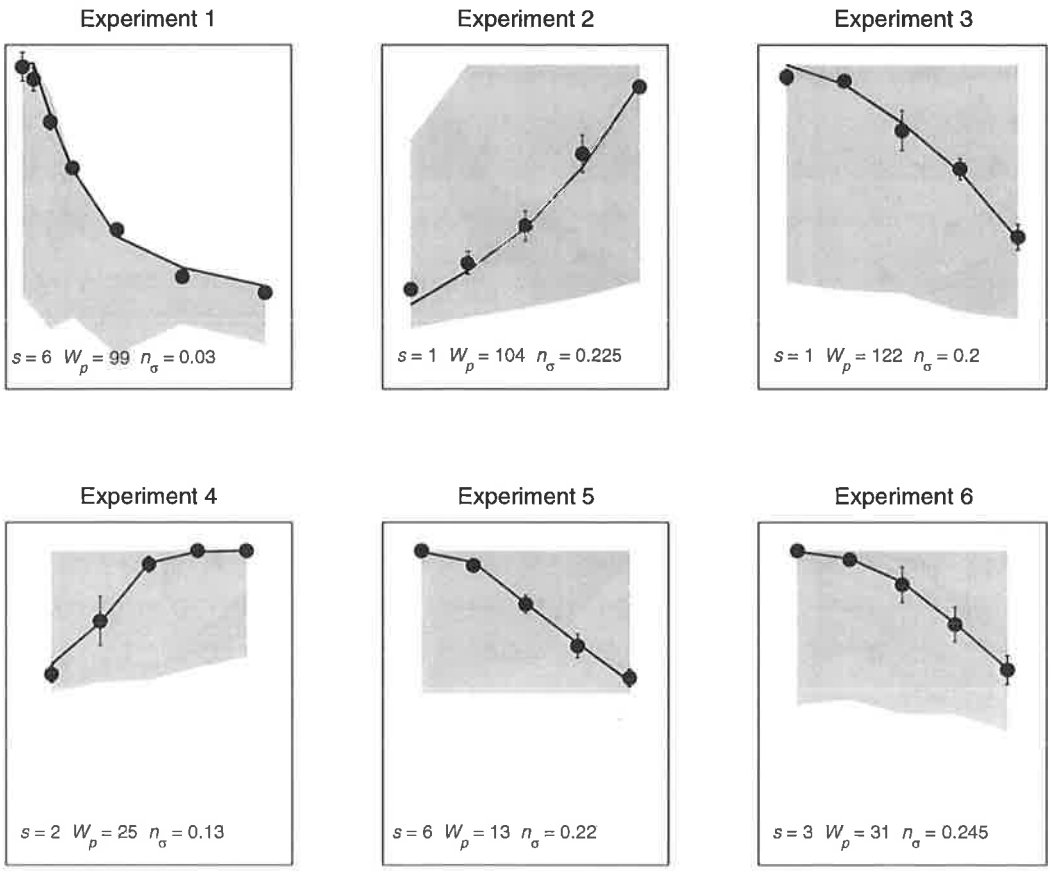

Figure 6.11 Isotropic/Correlation model fits for Experiments 1 to 6 using a spatial filter (s) parameter range of 1 to 6 pixels, detection region width (W_p) parameter range of 1 to 128 pixels and an additive noise (n_σ) parameter range of 0.03 to 0.245. For each experiment the best fitting parameter combination is shown by a black line. The parameter values of the best fit are indicated in the bottom left corner of each panel.

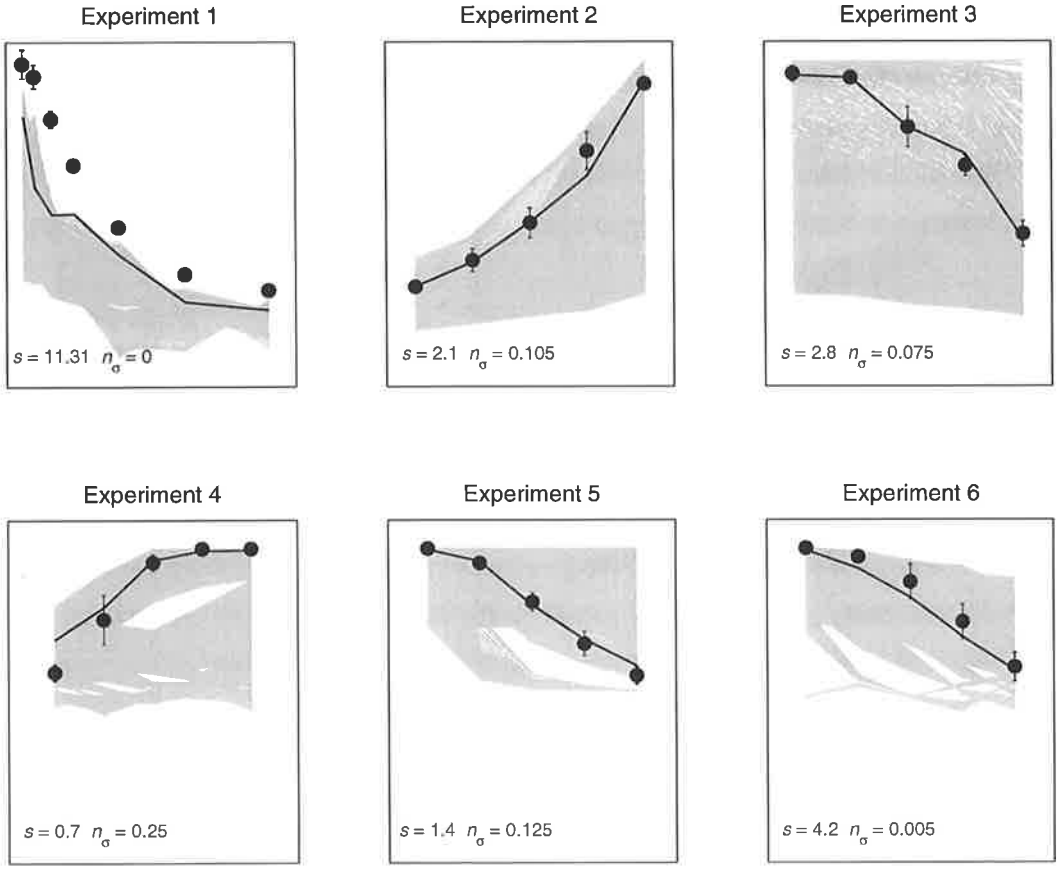

Figure 6.12 Oriented/Alignment model fits for Experiments 1 to 6 using a spatial filter (s) parameter range of 0.70 to 11.31 pixels and an additive noise (n_σ) parameter range of 0 to 0.25. For each experiment the best fitting parameter combination is shown by a black line. The parameter values of the best fit are indicated in the bottom left corner of each panel.

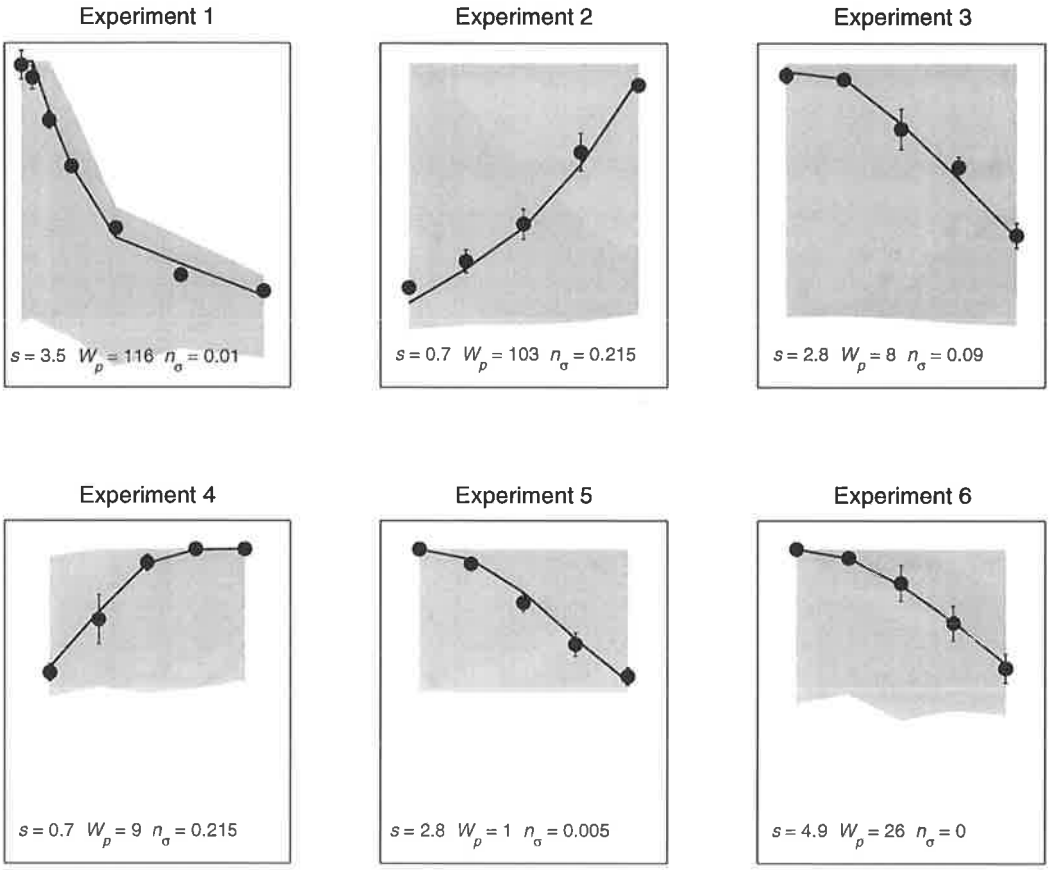

Figure 6.13 Oriented/Correlation model fits for Experiments 1 to 6 using a spatial filter (s) parameter range of 0.70 to 4.9 pixels, detection region width (W_p) parameter range of 1 to 128 pixels and an additive noise (n_σ) parameter range of 0 to 0.215. For each experiment the best fitting parameter combination is shown by a black line. The parameter values of the best fit are indicated in the bottom left corner of each panel.

Table 6.1 Log marginal likelihoods for Experiments 1 to 6, the sum log marginal likelihood and Bayes factors for the alternative models. The Bayes factors are given for pairwise comparisons of the new and the original models, and across all five new models.

Model		log marginal likelihood							Bayes Factor	
		Exp 1	Exp 2	Exp 3	Exp 4	Exp 5	Exp 6	Sum	New vs. Original	New Only
Barlow and Reeves	New	5.33	1.05	-5.14	29.70	26.27	24.66	81.89	1.00	7.47×10^{4}
	Original	7.16	0.26	-22.51	-136.00	24.75	23.95	-102.36	1.05×10^{80}	
Isotropic/Alignment	New	-41.95	0.91	1.64	31.49	27.78	15.35	35.23	1.00	1.37×10^{25}
	Original	-40.41	1.09	-0.91	32.01	27.16	15.79	34.74	1.62	
Isotropic/Correlation	New	4.62	2.20	1.74	32.67	27.15	24.70	93.11	1.00	1.00
	Original	3.25	-7.59	-27.80	31.97	24.63	24.02	48.49	2.39×10^{19}	
Oriented/Alignment	New	-25.27	0.67	0.02	31.01	26.86	26.86	57.78	1.00	2.19×10^{15}
	Original	-21.54	1.46	-8.11	30.99	25.25	25.48	53.53	70.21	
Oriented/Correlation	New	4.43	1.92	1.40	31.98	26.80	24.86	91.42	1.00	5.43
	Original	5.61	0.23	-27.80	31.12	24.63	24.02	57.83	3.85×10^{14}	

suggests that the models' good fits to the empirical data come at the cost of high complexity. In other words, the models are able to replicate the empirical data because they are able to replicate almost any pattern of results. This is particularly evident in regards to the Isotropic/Correlation model in Experiment 5 (Figure 6.11).

Log marginal likelihoods were calculated using the weighted sum square data fits of each model to the empirical data from Experiments 1 to 6. In order to determine the effect that the modification of the models had upon their ability to fit the empirical data, likelihoods were calculated for both the new and the original versions of each of the five alternative models. Table 6.1 summarises the log marginal likelihoods for each experiment, and the sum log marginal likelihoods and Bayes factors for all six experiments. Bayes factors were calculated for a pairwise comparison of the new and the original versions of each model, and across all of the new models.

The Bayes factors for the pairwise comparisons indicate that in each case the modified models are more likely than their original form, although the magnitude of the difference ranges from negligible to very large. The largest change in relative likelihood occurs for the models that were modified in terms of both the addition of random noise and restrictions placed upon the detection region (i.e., the Barlow and Reeves model and the two cross-correlation models). In contrast to this, the two feature alignment models (which already featured a restricted detection region) show very little improvement at all. This result is partly due to the fact that the cross-correlation models and the Barlow and Reeves model had more potential for improvement, being able to benefit from two modifications rather than just one. It is also due to the direction of the effect that the addition of noise has upon model discriminability: it is only able to degrade model performance, never enhance it. This means that models which tended to over-perform in Study 1 are now able to provide better fits to the empirical data, whereas models that underperformed (such as the two feature alignment models) can only be shifted further away from the distribution of empirical responses. The effect of this process can be clearly seen in Figure 6.10 for Experiments 1, 2, 4 and 6.

The Bayes factors for the five new versions of the alternative models reveal a strikingly different pattern of results to those found in Study 1. Study 1 replicated Dakin and Watt's (1996) finding that the Oriented/Alignment model was able to provide the best account of the empirical data. In the current study, the Bayes factors indicate that the Isotropic/Correlation

model is the most likely of the five alternative models, followed by the Oriented/Correlation and Barlow and Reeves models. Given the high complexity of the Isotropic/Correlation model (as evidenced by the large number of poorly fitting parameterisations in Figure 6.11) this result seems somewhat surprising. However, it should be remembered that Bayesian model selection balances the competing effects of model fit, complexity and generalisability. Although the Isotropic/Correlation model provides many fits which bear little or no resemblance to the pattern of empirical results, this is balanced by the number of fits that match the empirical data very well. Furthermore, the Isotropic/Correlation model is able to generalise well across all six experimental data sets. In contrast to this, the quality of the best-fitting parameterisations for the two feature alignment models is comparatively poor and in the cases of Experiments 1 and 6 for the Isotropic/Alignment model and Experiment 1 for the Oriented/Alignment model, they fail to closely match the empirical data.

6.4.2. Voronoi model

All four versions of the Voronoi model had the same three parameters: detection region width (W), detection region height (H), and level of added noise (n_σ). Study 1 indicated that the optimal detection region width for the Voronoi model is two columns of cells either side of the stimulus midline. Increasing the width of the detection beyond two columns resulted in discrimination levels that were far greater than those demonstrated by the empirical data in Experiments 5 and 6. Because these two experiments are effectively measuring the spatial limits of symmetry detection, it is implausible to suggest that the Voronoi model should be implemented using values of W greater than 2. There is no theoretically interesting reason for limiting either the range of detection region heights or the level of added Gaussian noise; therefore these parameters were free to vary. Detection region height (H) ranged from 2 to 14 cells above and below the stimulus horizontal midline. The Voronoi model symmetry values were standardised to lie between 0 and 1 and the level of added noise (n_σ) ranged from 0 to 0.3 in 0.05 increments. Following the same approach that was applied to the alternative models, the model fits were examined to determine the range of non-redundant parameter values and log marginal likelihoods were calculated from the parameter values that allowed the models to best generalise across all six experiments.

Figures 6.14 to 6.17 show the fits of the four Voronoi model symmetry calculation variants. For each model the best fitting parameter combination is shown by a black line and the corresponding parameter values are indicated at the bottom of each panel. As can be seen, the

different forms of the Voronoi model all tend to make fewer predictions than the alternative models, but the quality of the best-fitting parameterisations are generally poorer in relation to the more complex models such as the Isotropic/Correlation model. Table 6.2 shows the log marginal likelihoods of each variant of the Voronoi model for Experiments 1 to 6. The sum log marginal likelihoods indicate that the best account of the empirical data is given by the Type-4 model, with Bayes factors suggesting that it is at least 174.70 times more likely than the other versions of the Voronoi model. In other words, the most likely symmetry calculation method for the Voronoi model is one based purely upon the level of *asymmetry* within a stimulus.

The Type-4 model's success appears to be largely due to its ability to fit the Experiment 1 data and if this experiment is ignored, the other Voronoi models are at least 5.17 times more likely. Interestingly, Figures 6.14 to 6.17 demonstrate that all of the Voronoi model variants tend to under-perform in Experiment 1 and the reason for this is as follows: As has been explained previously, the Voronoi model attempts to match each point in a stimulus with a point (or points) located in roughly the same location on the opposite side of the stimulus midline. Specifically, any points that are located within a tolerance region defined by a reflection of a point's Voronoi cell are considered to be a potential symmetrical match. For Experiments 2 to 6, which employ stimuli comprised of strips of pure symmetry or symmetry that has been perturbed via the addition of random noise, there will always be *at least one* point in the reflected cell corresponding to a symmetrical point. However, for Experiment 1, in which symmetry is perturbed by randomly relocating one half of each pairing by a given distance, this is not necessarily the case.

Theoretically, the furthest distance that one half of a symmetrical pairing can be moved and still fall within the region corresponding to its partner's reflected cell is equal to the distance between the point and the most distant vertex of the cell. For stimuli with a dot density of 100 (the dot density employed in Experiment 1) the average distance between each point and its most distant vertex is around 0.10 (where distance is expressed as a proportion of the width of the stimulus). This suggests that as the level of distortion in the Experiment 1 stimuli is increased beyond 0.10 the probability of the Voronoi model being able to make any correct pairings will be very low and performance will drop. The data support this hypothesis: all of

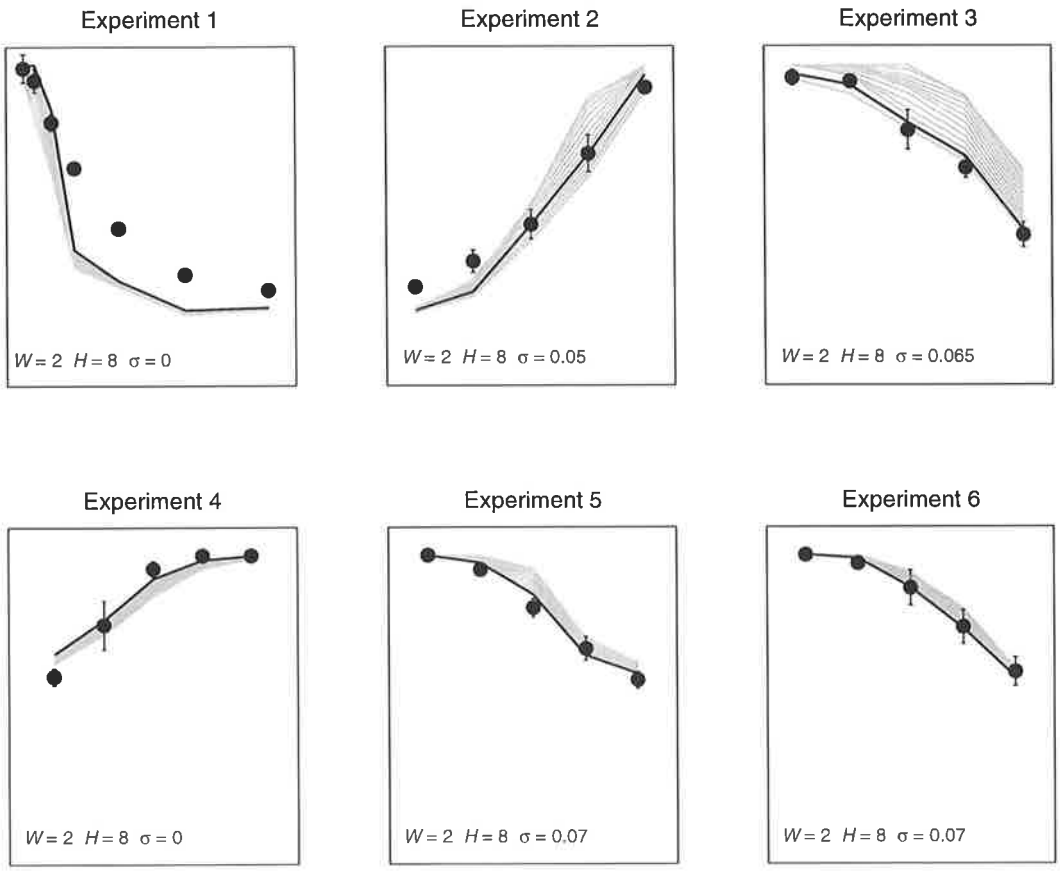

Figure 6.14 Voronoi Type-1 model fits for Experiments 1 to 6 using a detection region width (W) = 2 columns of cells, detection region height (H) = 8 rows of cells, and an additive noise (n_σ) parameter range of 0 to 0.095. For each experiment the best fitting parameter combination is shown by a black line. The parameter values of the best fit are indicated in the bottom left corner of each panel.

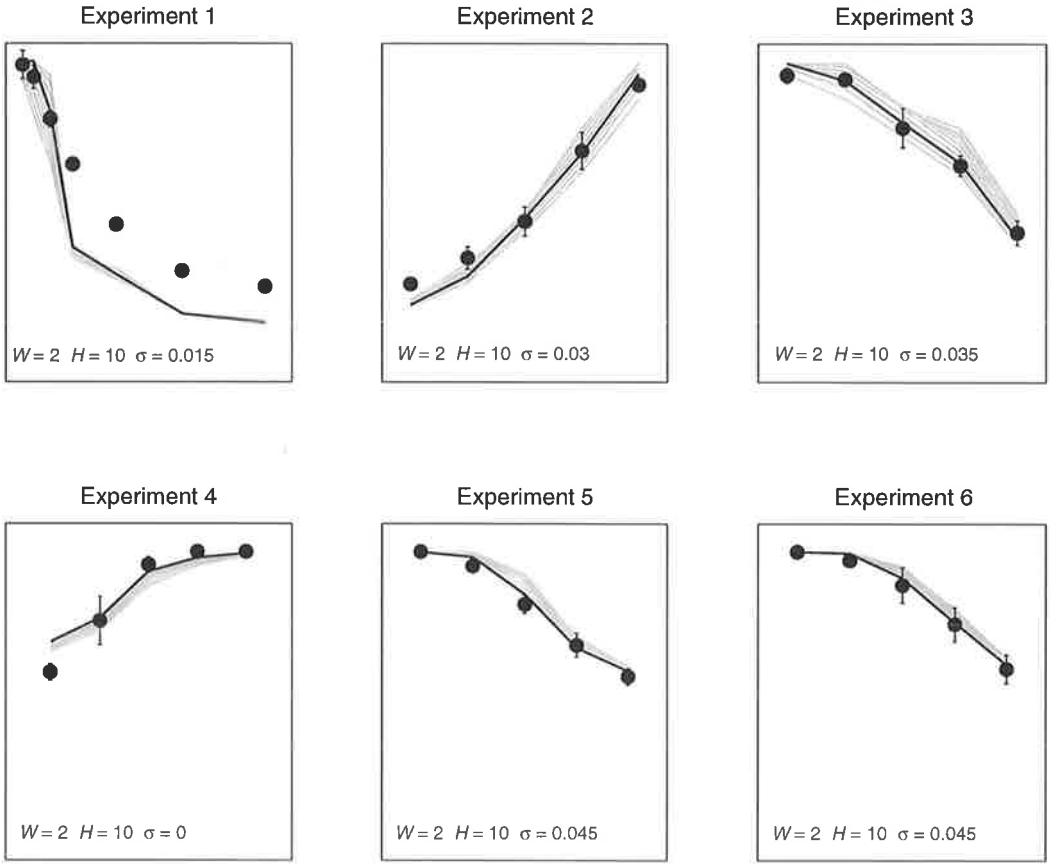

Figure 6.15 Voronoi Type-2 model fits for Experiments 1 to 6 using detection region width (W) = 2 columns of cells, detection region height (H) = 10 rows of cells, and an additive noise (n_σ) parameter range of 0 to 0.045. For each experiment the best fitting parameter combination is shown by a black line. The parameter values of the best fit are indicated in the bottom left corner of each panel.

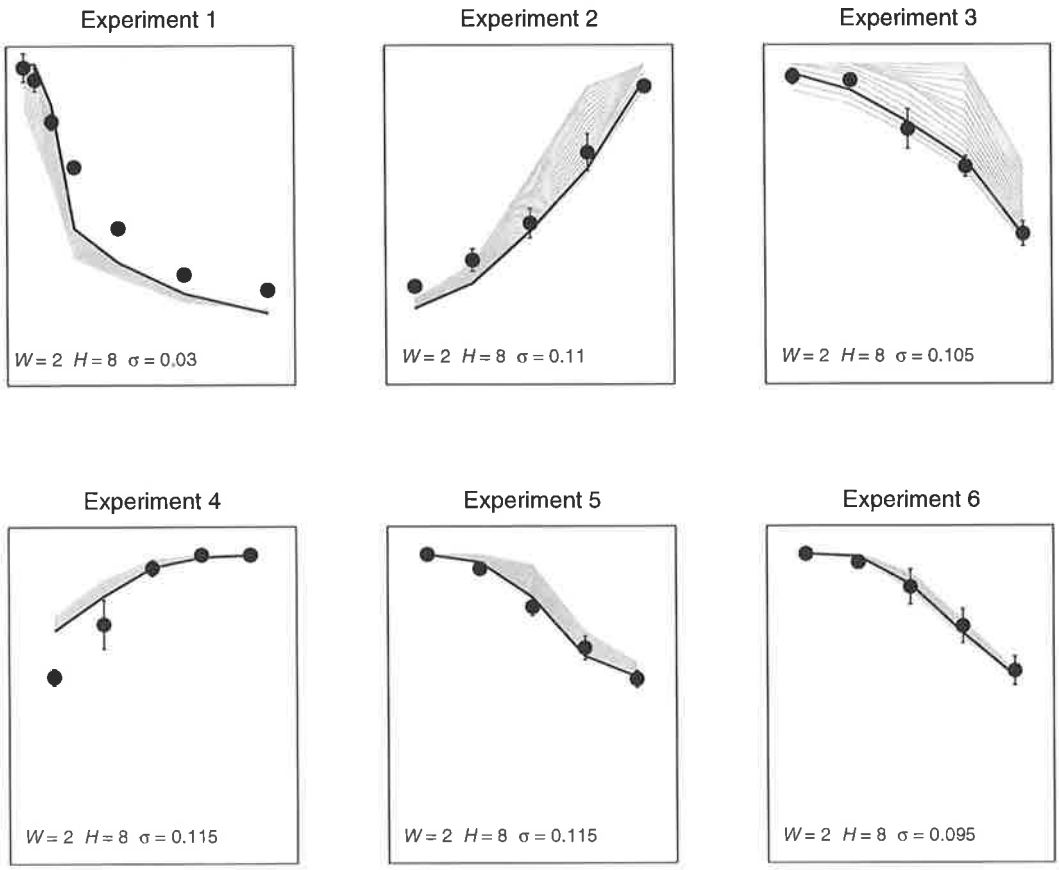

Figure 6.16 Voronoi Type-3 model fits for Experiments 1 to 6 using a detection region width (W) = 2 columns of cells, detection region height (H) = 8 rows of cells, and an additive noise (n_σ) parameter range of 0.03 to 0.13. For each experiment the best fitting parameter combination is shown by a black line. The parameter values of the best fit are indicated in the bottom left corner of each panel.

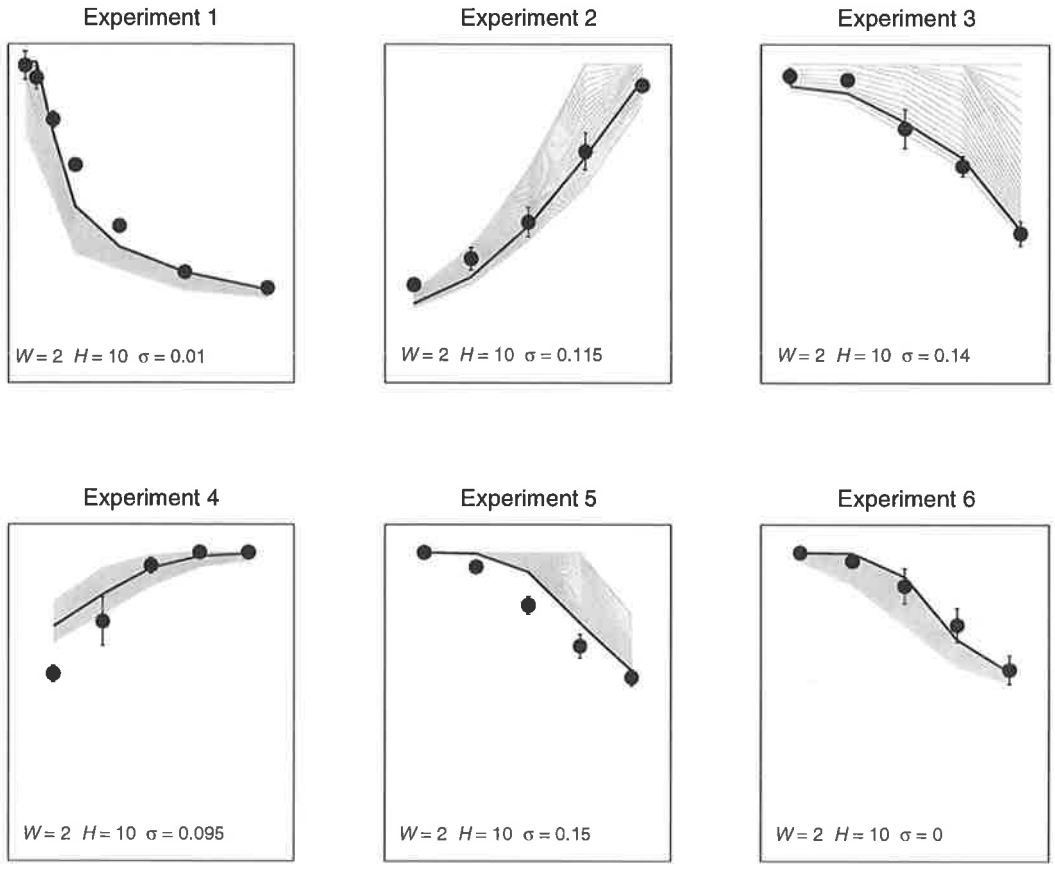

Figure 6.17 Voronoi Type-4 model fits for Experiments 1 to 6 using a detection region width (W) = 2 columns of cells, detection region height (H) = 10 rows of cells, and an additive noise (n_σ) parameter range of 0 to 0.15. For each experiment the best fitting parameter combination is shown by a black line. The parameter values of the best fit are indicated in the bottom left corner of each panel.

Table 6.2 Log marginal likelihoods for Experiments 1 to 6, the sum log marginal likelihood and Bayes factors for the four Voronoi model symmetry calculation variants (V1 to V4), and the modified versions of the alternative models.

Model		log marginal likelihood							Bayes Factor	
		Exp 1	Exp 2	Exp 3	Exp 4	Exp 5	Exp 6	Sum	Voronoi Only	All Models
Voronoi	Type-1	-12.52	2.07*	2.47*	33.56*	29.59*	26.37*	81.56	6.89×10^{4}	6.89×10^{4}
	Type-2	-14.17	3.02*	3.35*	33.33*	29.69*	26.35*	81.58	6.81×10^{6}	6.81×10^{6}
	Type-3	-2.25	2.48*	2.53*	33.34*	29.63*	26.63*	92.15	174.70	174.70
	Type-4	4.86*	2.46*	1.96	33.07*	28.72*	26.21*	97.31	1.00	1.00
Barlow and Reeves								81.89		4.99×10^{6}
Isotropic/Alignment								35.23		9.15×10^{26}
Isotropic/Correlation								93.11		66.79
Oriented/Alignment								57.78		1.46×10^{17}
Oriented/Correlation								91.42		362.74

Note- Bayes Factors are taken in relation to the most likely model.
* denotes most likely model or models with Bayes factors of less than 3.2

the Voronoi variants are able to match the empirical data for the three lowest distortion levels (0.03, 0.04 and 0.07). However, when the level of distortion rises above 0.11, model performance drops dramatically.

The reason that performance of the Type-4 model does not degrade as rapidly as the other Voronoi models appears to be related to the type of evidence that the model employs to make classification decisions. Unlike the other symmetry calculation variants the Type-4 model does not count the number of matching points but rather the number of points which *cannot* be matched because there are no points located in the region corresponding to a reflection of their Voronoi cell. It appears that the number of unmatched points in the distorted stimuli does not approach the level found in random stimuli as rapidly as the number of matching points declines towards the number of spurious matches that occur in the random stimuli.

While this method of symmetry calculation benefits the Type-4 model in Experiment 1, it causes problems for it elsewhere. In particular, for Experiments 3 and 5 (which both involve perfectly symmetric standard stimuli), the Type-4 model suffers from the same difficulties experienced by the alternative models: it returns the same value for all of the perfectly symmetric stimuli because the number of non-matching points in a perfectly symmetric stimulus will always be equal to zero. As a result the model is able to discriminate between the two classes of stimuli at a close to perfect level. Figure 6.17 demonstrates that this problem can be addressed through the use of additive noise but that the high levels of noise needed to achieve reasonable fits in Experiments 3 and 5 result in numerous poor fits for the remaining experiments.

All of the Voronoi models appear to benefit from the addition of random noise, although the range of σ-values needed to obtain good fits to the empirical data varies considerably across the four models. It should be noted that the plausibility of any additive noise assumptions should not be based upon the *amount* of noise needed to fit the data, but the *range* of noise levels needed to fit the data across all six experiments. It is reasonable to assume that there is a moderately high level of noise associated with the decision process but that this level of noise should be stable across the different experimental settings. Importantly, all of the Voronoi models require a range of n_σ-values that is smaller than is required by all the alternative models (with the exception of the Isotropic/Alignment model).

Aside from the addition of random noise and the use of alternative symmetry calculation methods, the Voronoi models in the current study also differed from the Study 1 model in regards to the vertical extent of their detection region. In each case the models' best fit to the empirical data was obtained with a single detection region height parameter value: $H = 8$ for the Type-1 and Type-3 models and $H = 10$ for the Type-2 and Type-4 models, indicating that the optimal detection region width to height ratio is in the range of 1:4 to 1:5. Both of these aspect ratios can be interpreted in much the same way. For stimuli with a dot density of 100 points (i.e., Experiments 1 to 3) the detection region is comprised of two columns of cells spanning a stimulus from top to bottom. However, for the higher density stimuli (650 points) the detection region does not include information at the very top and bottom edges of a stimulus. While this is interesting in that it indicates that the Voronoi-based models can be implemented using an arguably more plausible detection region, it should be noted that limiting the vertical aspect of the detection region in this way only had a very small effect on the models' overall likelihoods. When the models were implemented with no limit placed upon the vertical extent of the detection region (in other words, identical to the Voronoi detection region employed in Study 1) the models were at most only 1.8 times less likely.

Table 6.2 also includes the sum log marginal likelihoods for the five new versions of the alternative models. Bayes factors indicate that the Voronoi Type-4 model is 66.79 times more likely than the most likely alternative model (the Isotropic/Correlation model). According to Jeffrey's (1961) guidelines this can be considered to be strong evidence in favour of the Voronoi Type-4 model. A qualitative comparison of the Isotropic/Correlation and Voronoi Type-4 best fits clearly indicates that the Isotropic/Correlation model is far more successful at capturing the pattern of empirical results. Nonetheless, as has been suggested, these fits appear to come at the price of complexity, with the Isotropic/Correlation model making numerous predictions that do not resemble the empirical data at all. In contrast the Type-4 model (like all of the Voronoi model variants) makes relatively few predictions, the majority of which fit the basic pattern of observer responses.

6.5. Discussion

The results of Study 2 suggest that the most likely model of symmetry perception is a form of the Voronoi model (the Type-4 model) which differentiates between stimuli in regards to the level of asymmetry in the stimulus. Study 1 also found that a model based upon Voronoi tessellation was able to provide the best account of the empirical symmetry perception data.

However, it is possible that the results of Study 1 were due to two factors that may have unfairly favoured the Voronoi model: the variance assumptions associated with Experiments 2 and 3, and the limited range and number of parameter values over which the alternative models were implemented. In Study 2 these potential confounds were addressed and, as a result, we can be confident that the Voronoi model's high relative likelihood is due to its ability to closely simulate empirical performance rather than any methodological artefacts.

Aside from the nature of its underlying representation, the Voronoi model differed from the original forms of the alternative models in two important respects. In the first place, all of the alternative models returned the same value for perfectly symmetrical stimuli, resulting in a tendency to over-perform in Experiment 3. Second, the two cross-correlation models and the Barlow and Reeves model lacked the ability to place limitations on the spatial extent of the detection region. In Study 2, the models were modified to allow these features and it was demonstrated that in each case the model's ability to simulate the empirical data improved. In some cases the improvement was considerable. It should also be noted that the order of likelihood for the four models from Dakin and Watt (1996) differed from that of the original paper and from Study 1: in the current study the two modified Correlation models were found to be more likely than the alignment-based models by a difference that can be considered decisive.

It was suggested that by controlling for these differences between the Voronoi and alternative models we might be able to make an assessment of the relative utility of the underlying stimulus representations. In other words, if the models were identical in all respects except for the form of stimulus representation then any difference in likelihood must be due to the nature of the representation. Following this logic we could assume that the visual system appears to be detecting symmetry using a stimulus representation based upon Voronoi tessellation rather than some form of spatial filtering or a division of the visual field into discrete square-shaped sub-regions. However, there is a major problem with this argument: the models also differ considerably in terms of the methods that are employed to calculate the actual level of symmetry within a stimulus (i.e., cross-correlation versus the degree of alignment of stimulus features versus the comparative dot density of symmetrically positioned regions, etc). As was demonstrated in the current study, changing the form of the symmetry calculation method had a substantial effect on the likelihood of the Voronoi model, with the Type-4 model being at least 174.70 times more likely than the other Voronoi model variants. It would appear,

therefore, that differences between the models in terms of symmetry calculation method are potentially just as important as any other model feature. Unfortunately, there is no way that the different models can be made equivalent in respect to the method of symmetry calculation and, in light of this, it would seem that we cannot make any conclusions about the relative utility of the stimulus representations.

Despite the fact that the current study cannot tell us anything specific about the relative utility of the various stimulus representations, it is still highly informative in regards to a number of other factors, such as the assumptions regarding the shape and extent of the detection region, and the additive noise assumption. In the following sections these will be discussed in turn.

6.5.1. The spatial extent of the detection region.

In the first place, the results suggest that in order to fit the Experiment 4, 5 and 6 data it is necessary to assume that the visual system's detection region is spatially limited. Modifying the Barlow and Reeves model and the two cross-correlation based models to allow for a spatial limitation along the horizontal plane had a dramatic effect on the quality of the model fits to the empirical data, increasing their relative likelihoods to a level approaching that of the most likely version of the Voronoi model.

Nevertheless, it should be noted that even if a model feature is able to improve the model's likelihood it is still of little value if it cannot be plausibly interpreted. For example, the data indicate that in Experiment 5 the Oriented/Correlation obtains its best-fit to the empirical data using a detection region that has a width of just 1 pixel either side of the stimulus midline. Given that the empirical data from Experiments 5 and 6 suggest that the optimal width of the detection region for stimuli with a density of 650 points is approximately 27 pixels either side of the stimulus midline, this value seems unreasonably low. Similarly, it is difficult to predict what the optimal detection width should be for Experiments 1 to 3 but it can be safely assumed that because the dot density of the stimuli in these experiments is lower than that used in Experiments 4 to 6 (100 points vs. 650 points), the detection region should be wider. Furthermore, because dot density is held constant across these three experiments the width of the detection region should reflect similar constancy. The data suggest that this is not the case for either the Barlow and Reeves model (the best fitting parameter values correspond to detection region widths of 128, 25.6 and 42.66 pixels for Experiments 1 to 3, respectively), or the Oriented/Correlation model (W_p = 116, 103 and 8 for Experiments 1 to 3, respectively). Of

the three models, only the Isotropic/Correlation model has best-fitting values that are high and reasonably stable for Experiments 1 to 3 (W_p = 99, 104 and 122, respectively) and reasonably close to 27 pixels for Experiments 4 to 6 (W_p = 25, 13 and 31, respectively). Nonetheless, even for the Isotropic/Correlation model the values for Experiments 3 and 5 are stretching the bounds of interpretability.

Study 2 also demonstrated that it is possible to limit the vertical extent of the Voronoi model detection region, and that this results in a minor, positive effect on model likelihood. While this modification is of little benefit in the current study (aside from the slight improvement the quality of the data fits) it suggests that the model should be able to generalise to experimental situations that require a vertically limited detection region (e.g., Rainville and Kingdom (2002)). This finding also raises the possibility that the alternative model's relative likelihoods might be increased by a similar modification. Due to time constraints this was not attempted in the current study; but, theoretically, there is no reason to suggest that vertically limited versions of the models could not be implemented. Nonetheless, given the already high complexity of the Correlation models in particular, and the small effect that this modification had on the Voronoi models, it is unlikely that any significant gain would result. Finally, even if the four Voronoi models' detection regions were similarly constrained to a spatial limitation only along the horizontal plane, the Type-4 model would still 59.41 times more likely than the Isotropic/Correlation model. It is therefore reasonable to conclude that the alternative models were not unfairly discriminated against in this regard.

6.5.2. The addition of noise to the decision-making process

The second main finding of interest in Study 2 is in regards to the presence of noise in the perceptual decision making process. It was demonstrated that all of the models benefited from the addition of random noise, with the greatest improvement in quality of fit being found for the models that tended to over-perform in the tasks. In contrast to this, when the models underperformed (such as the feature alignment or Voronoi models in Experiment 1), the model response patterns were always shifted further away from the empirical distribution. This is a logical and interpretable pattern of results because the addition of noise to a system should only ever serve to degrade performance. Furthermore, it should be noted that for the Voronoi Type-4 model and each of the alternative models, the addition of noise is an integral assumption of the models' ability to fit the Experiment 3 and 5 data. Each of these models returns the same value for all of the perfectly symmetric stimuli and, because of this, there is

little or no overlap between the distributions of scores for the two classes of stimuli being discriminated, resulting in near-perfect performance. The addition of noise to the decision process adds variance to the distribution of values assigned to the perfectly symmetric stimuli, and, as a result, the models are more capable of reproducing the pattern of empirical responses.

There is a potential problem associated with this important assumption: for each of these models there are also a number of experimental settings in which little or no benefit is derived from the addition of noise. In other words, the range of n_σ-values required by the models is relatively wide. The use of additive noise as a means of inducing variability in model responses is relatively uncontroversial and is an integral assumption of numerous models of perceptual and cognitive processes. The interpretation of what should be considered a plausible range of added noise is more problematic. It is reasonable to assume that any decision-making process is inherently noisy; nonetheless, the level of this interference should be reasonably stable across similar tasks and experimental settings. This suggests that models requiring a limited range of additive noise values to successfully fit the empirical data across all six experiments should be considered more plausible than models requiring a wide range of values. In this respect the Type-4 model compares favourably with the majority of the other noise-reliant models but compared to the other Voronoi model variants it does quite badly.

There are a number of factors that might be implicated as potential sources of noise, such as differences in experimental setting and method and individual differences in ability and attention. Furthermore, it is reasonable to assume that the proportion of variance in the empirical responses that is accounted for by the additive noise could vary considerably across different observers and experimental settings. Given that the empirical data employed in this study were drawn from three different sources (Experiment 1 from Barlow and Reeves (1979), Experiments 2 and 3 from the empirical replications of Barlow and Reeves' original studies reported in section 6.2, and Experiments 4 to 6 from Jenkins (1982)), it could be argued that we cannot rule out the possibility that the wide range of n_σ-values required by the models is a credible reflection of variance across the different observers and settings. However, if this were the case we could expect that the range of n_σ-values required to fit the data well should not vary highly *within* a given experimental setting or set of observers. A comparison of the range of values required to fit the three Jenkins (1982) data sets indicates that the only model that meets this condition is the Voronoi model Type-3; all of the other

models require n_σ-values that span at least half of the total range needed to fit the data across all six experiments. Consequently, we must begin to question the explanatory value of the additive noise assumption in relation to the majority of the models.

6.5.3. Goodness-of-fit, complexity and generalisability

The results of the current study are also informative regarding the competing demands of goodness-of-fit, complexity and generalisability. A good model of symmetry perception should be able to fit an empirical data set well, without producing numerous predictions that bear little-or-no resemblance to the empirical responses. Furthermore, it should be able to produce good fits across a range of different experimental manipulations. None of the models under comparison in this study was able to meet all three of these demands with the same level of success.

The two most likely models (Voronoi Type-4 and Isotropic/Correlation) provide an interesting contrast in respect to goodness-of-fit and complexity. The Type-4 model produces few predictions, the majority of which capture the basic pattern of empirical responses. However, the overall quality of the best-fitting parameterisations is relatively poor, particularly in Experiments 1, 4 and 5 where the model fails to make predictions that lie within the bounds of credible inter-observer variance indicated by the error-bars for one or more levels of stimulus manipulation. In contrast, the best-fitting responses produced by the Isotropic/Correlation model tend to pass through or very close to each of the empirical data points. However, as Figure 6.11 indicates, the model also makes a large number of responses that do not resemble the empirical data. If the models' likelihoods were based purely upon the best-fitting parameterisation, the Isotropic/Correlation model would be 18.69 times more likely than the Type-4 Voronoi model. However, because Bayesian model selection averages a model's likelihood across all of its responses, the Isotropic/Correlation model is heavily penalised for its numerous unlikely responses.

The results suggest that the Type-4 Voronoi model's main strength is its ability to generalise across the six data sets using a limited range of parameter values. In particular, all of the Voronoi models (like the two feature alignment models) are able to automatically scale the size of their detection region with changes in stimulus density. However, in the current study only two different stimulus densities were employed, and density was never varied within a single experimental setting. Theoretically, the Type-4 model should be able to perform

equally well across a wide range of stimulus density manipulations. However, how well the data actually match these theoretical predictions is yet to be determined.

6.6. Summary

Study 2 compared the performance of modified versions of the models of symmetry perception that were introduced in Study 1. It was demonstrated that placing limitations on the spatial extent of the detection regions of the two cross-correlation models and the Barlow and Reeves model enabled them to simulate the performance of the human observers in Experiments 4 to 6. However, for at least two of the models (Oriented/Correlation and Barlow and Reeves) the values of the best-fitting parameter values were questionable. For these two models the addition of a spatially limited detection region contributes little to our understanding of the processes underlying symmetry perception. In contrast to this, the best-fitting detection region width parameter values for all of the Voronoi model variants, and (for the most part) the Isotropic/Correlation model, were plausibly interpretable. For these models the detection region appears to be operating in a credible manner across changes in stimulus density and, in this way, the presence of a spatially limited detection region can be seen to be more than just an *ad hoc* feature that merely serves to increase model-fit.

It was also demonstrated that adding variability to the decision-making process by placing a Gaussian distribution over each of the values assigned by a model to a given stimulus was able to increase the quality of the fits to empirical data for all of the models under comparison. This modification was particularly important for the five alternative models and the Type-4 Voronoi model. These models assign the same value to all of the perfectly symmetric stimuli and this causes the models to over-perform in Experiments 3 and 5. The addition of noise to the decision process degrades the models' ability to discriminate between the two classes of stimuli and consequently shifts the models' response distributions closer to the empirical data. However, for each of these models the level of added noise needed to fit the Experiment 3 and 5 data tended to be far greater than the level needed to fit the remaining data sets. It was suggested that this difference could be attributed to differences in observer ability and task requirements across the six experiments. In this case it could be expected that variability in n_o-values should be low for the three experiments taken from Jenkins (1982) which shared the same three observers and similar task requirements. Nonetheless, for all of the models (with the exception of the Type-3 model) there was a higher degree of variance in the necessary level of added noise than could be reasonably expected.

The sum log marginal likelihoods indicated that of the models compared in Study 2, the most likely model of symmetry perception was the Type-4 Voronoi model. One of the strengths of all of the Voronoi model variants is their ability to scale with changes in stimulus density. This in turn makes a number of strong predictions about the behaviour of the Voronoi models in response to a stimulus density manipulation. In the following study these predictions will be examined.

7. STUDY 3

7.1. Introduction

Study 2 compared the performance of a number of models of symmetry perception across six empirical symmetry perception data sets. The two most likely models were a version of the Voronoi model based upon the level of asymmetry in a stimulus (the Type-4 model), and a modified version of the Isotropic/Correlation model from Dakin and Watt (1996). Although the best-fitting predictions of the Isotropic/Correlation model were closer to the empirical responses than those of the Voronoi model, its high complexity resulted in the Isotropic/Correlation model being 66.79 times less likely.

The high complexity of the Isotropic/Correlation model stems largely from the number of detection region width parameter values over which the model must be implemented. It has been demonstrated that changes in stimulus density affect the spatial extent of the region within which symmetry can be detected (Rainville & Kingdom, 2002). Because the Isotropic/Correlation model has no means of adjusting the width of its detection region in response to changes in stimulus density, in order for the model to respond in a plausible manner to all potential stimulus densities it must be implemented using all possible detection region widths. In contrast to this, the Voronoi representation-based models are implemented using a single detection region width parameter value. The Voronoi model detection region width (W) is based upon the number of columns of Voronoi cells located either side of the stimulus midline that are included in the symmetry calculation procedure. Because the size of a Voronoi cell automatically scales in response to changes in stimulus density, the spatial extent of the detection region also automatically scales with stimulus density. As a result, the Voronoi models were able to fit all six empirical data sets with a detection region comprised of two columns of cells either side of the stimulus midline (i.e., $W = 2$).

The Voronoi model makes a number of other strong predictions in relation to changes in stimulus density. For example, it has been demonstrated that perturbing symmetric stimuli by adding distortion in the form of a random relocation of one half of each symmetric pairing has an effect on discriminability. Specifically, as the extent of the random relocation is increased, discriminability declines (Barlow & Reeves, 1979). Because the degree of tolerance to distortion afforded to the Voronoi model is defined by the size of each point's Voronoi cell and the size of each cell scales with stimulus density, in theory the amount of distortion that

the model is able to tolerate should decline as stimulus density is increased. It follows that if the Voronoi model is a plausible model of human symmetry perception then the degree of distortion that human observers can tolerate should be similarly affected by stimulus density.

An alternative approach to perturbing symmetry is to replace symmetric point pairs with increasing numbers of randomly located points; it has been demonstrated that as the proportion of symmetric points pairs within a stimulus declines, discriminability also declines (Barlow & Reeves, 1979). In this case the Voronoi model should be unaffected by changes in stimulus density, because the chance of a random noise point falling within the region corresponding to a symmetric points reflected Voronoi cell (or, for the Type-4 model, of a random point having no matching point) is equal across different densities as long as the signal-to-noise ratio is kept constant. As a result we expect that the performance of human observers should be similarly uniform across changes in stimulus density.

The following study sets out to test these two hypotheses. The performance of human observers on a symmetry-with-distortion discrimination task and on a symmetry-with-added-noise discrimination task is tested across four stimulus density levels. Following this, the performance of the human observers is compared to that of the four Voronoi model variants and the Isotropic/Correlation and Oriented/Correlation models.

7.2. The effect of stimulus density on symmetry perception

Although the effect of stimulus density on distortion tolerance has been demonstrated in previous papers (Rainville & Kingdom, 2002), the author is not aware of any demonstration of this phenomenon in which only stimulus density and distortion level have been manipulated. Similarly, there appears to be no previously published data exploring the effect of density on symmetry perception in the presence of added random noise points. The following experiments present an opportunity to explore these two phenomena. Experiment 1 explores the effect of density on distortion tolerance; Experiment 2 explores the effect of density on added random noise tolerance. In both cases stimulus density is manipulated across four levels: 10 points, 50 points, 100 points and 200 points per stimulus.

For Experiment 1 it is hypothesised that there will be a clear ordering of discriminability in regards to stimulus density. Specifically, observers should be able to tolerate higher levels of distortion in the 10 point stimuli, than in the 50 point stimuli, and higher levels in the 50 point

stimuli than the 100 point stimuli. Distortion tolerance should be lowest in the 200 point stimuli. In contrast to this, it is hypothesised that there should be no ordering of discriminability in terms of density for the Experiment 2 stimuli.

7.2.1. Method

Two observers (one was the author) took part in both experiments. Both observers had training on each of the tasks prior to data collection.

All stimuli were presented on a Hitachi CM721F colour computer monitor with a vertical refresh rate of 85 Hz and a resolution of 1280 x 1024. Pixel width was 0.26 mm. Viewing distance was set at 57 cm. Stimuli were generated and presented using MATLAB 7.0.4. All stimuli were comprised of 2 x 2 pixel black dots presented within a white circular region with a 256 pixel diameter located at the centre of the monitor. Stimulus presentation time was held constant at 0.1 seconds for both experiments.

For Experiment 1 the standard stimuli were comprised of randomly positioned points and the experimental stimuli were symmetric stimuli that had been subjected to varying degrees of positional distortion. The symmetrical stimuli were distorted by repositioning one half of each symmetrical pair a random distance within a given horizontal and vertical 'tolerance area' centred on the point's original position. The size of the tolerance area was manipulated across seven levels: 0.0, 0.05, 0.1, 0.15, 0.2, 0.25 and 0.3 of the diameter of the stimulus. For Experiment 2 the standard stimuli were comprised of randomly positioned points and the experimental stimuli were comprised of both symmetric point pairs and randomly positioned points. The proportion of symmetrically positioned points in the experimental stimuli ranged from 0.2 to 1.0 in 0.2 increments.

For both experiments the symmetric stimuli were always vertical bilateral symmetry with an axis centred in the middle of the stimulus. All of the stimuli had equal numbers of points located on either side of the stimulus midline. In both experiments the stimulus density was manipulated across four levels: 10, 50, 100 and 200 points. A total of 50 experimental stimuli and 50 standard stimuli were generated for each density/distortion (Experiment 1) or density/signal-to-noise (Experiment 2) stimulus level.

Both experiments employed a single-interval, two-alternative forced-choice design. Stimuli were presented in blocks of 50 trials. The signal parameter was fixed within each set of trials, so that observers were required to discriminate between standard stimuli and a single level of experimental stimuli. The observers were presented with multiple examples of both standard and experimental stimuli prior to beginning each block. The order of presentation for the separate trial blocks was randomised to negate order effects. Each block contained equal numbers of standard and experimental stimuli (25 standard and 25 experimental stimuli) and, within each block, the experimental and standard stimuli were presented in a random order.

Each trial was preceded by the 0.5 second presentation of a fixation point located in the centre of the presentation window. This was followed by the brief presentation of a single stimulus. The observer was then required to indicate from which of the two populations (standard or experimental) the stimulus was drawn, using a mouse to click on one of two buttons presented at the bottom of the screen. The observer was then given feedback as to whether their choice was correct or not. When the observer was ready, they initiated the next trial by clicking on a button that was located equidistant from the two choice buttons.

7.2.2. Results

Observer performance was expressed as d'. The results of Experiment 1 are shown in Figure 7.1. As can be seen, the order of discriminability across the four stimulus densities is (for the most part) as predicted: observers were able to tolerate a greater degree of positional distortion in the low density stimuli than in the high density stimuli. The results of Experiment 2 are also as predicted: when discriminating between random noise stimuli and stimuli comprised of symmetry with varying levels of added noise points, the observers' performance is relatively uniform across changes in density (Figure 7.2). Although the data suggest that there may be slight advantage for the 50 and 200 point stimuli, it important to note that there is no clear ordering of discriminablity in terms of stimulus density such as is found in the Experiment 1 data (i.e., $d'_{10} > d'_{50} > d'_{100} > d'_{200}$).

According to the experimental hypotheses, the Experiment 1 data should be best fit by four separate functions (i.e., one for each level of stimulus density), whereas it should be possible to account for the Experiment 2 data with a single function. Bayesian statistical inference (Kass & Raftery, 1995) was employed to test these hypotheses. A similar use of function-

fitting to test experimental hypotheses has been previously applied in a number of papers (for example: Dry, 2005; Dry et al., in press; Vickers, Lee et al., 2003).

According to this approach a number of potential experimental outcomes are defined. For example, that the data is best fit by a single function (i.e., $d'_{10} = d'_{50} = d'_{100} = d'_{200}$), or two functions (i.e., $d'_{10} = d'_{50} \neq d'_{100} = d'_{200}$), or four separate functions (i.e., $d'_{10} \neq d'_{50} \neq d'_{100} \neq d'_{200}$), etc. Each of these outcomes becomes a rival model, and Bayesian model selection is employed to determine their relative likelihood. For both the Experiment 1 and Experiment 2 data there are 15 models that can be used to describe the full range of potential density effects. At one extreme the data is best fit by a single function, and at the other extreme four functions are needed to account for the empirical data. Additionally, each of the 13 possible two and three function combinations were also considered.

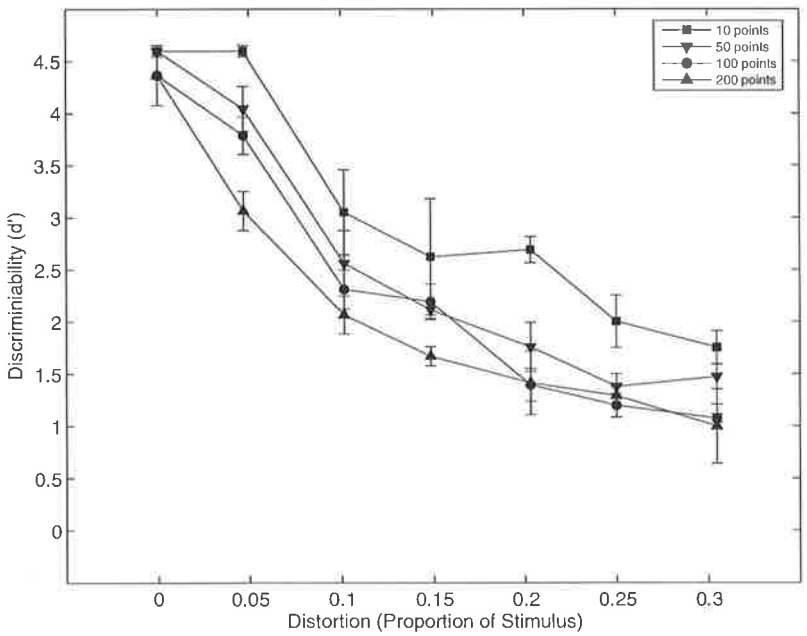

Figure 7.1 Comparison of empirical discriminability across the four stimulus density levels for symmetry with varying degrees of distortion in Experiment 1. The error bars represent one standard error.

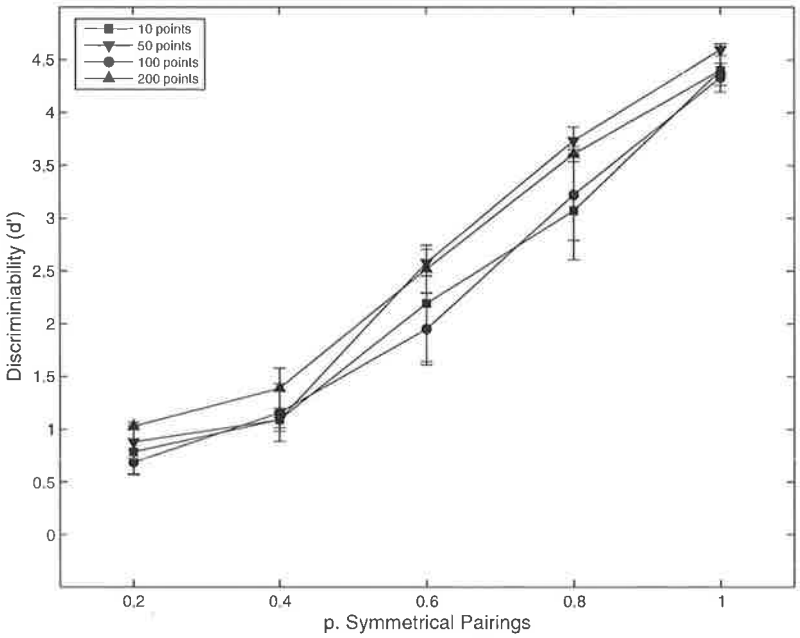

Figure 7.2 Comparison of empirical discriminability across the four stimulus density levels for stimuli containing varying levels of symmetric pairings in Experiment 2. The error bars represent one standard error.

The maximum likelihood fits to the empirical Experiment 1 and 2 data were calculated for all fifteen models. In each case the maximum likelihood fits were made under the assumption of a Gaussian likelihood function. Given these data fits it is possible to calculate the Bayesian information criterion (BIC) for each model (Schwarz, 1978) using:

$$\mathrm{BIC} = \mathrm{WSSE} + P\ln(N),$$

where WSSE is the weighted sum-squared error, P is equal to the number of model parameters, and N represents the number of data points to which the model is being fitted (i.e., for Experiment 1 $N = 7$, and for Experiment 2 $N = 5$). In this case P is equal to the number of functions that each model predicts will be needed to account for the empirical data. The relative likelihood of each model can then be determined by calculating Bayes factors. Table 7.1 summarises the results of these analyses, showing the maximum likelihood fit of the functions predicted by the models, their parametric complexity and the associated BIC and Bayes factors for both the Experiment 1 and 2 data.

Table 7.1 Summary of models, and weighted sum squared error data fit (WSSE), Bayesian information criterion (BIC) and Bayes factors for the Experiment 1 and 2 data.

Model Summary			Experiment 1			Experiment 2		
No.	Prediction	Parameters	WSSE	BIC	Bayes Factor	WSSE	BIC	Bayes Factor
1	$d'_{10} = d'_{50} = d'_{100} = d'_{200}$	1	178.70	202.02	6.20×10^{26}	23.11	38.09	2.81*
2	$d'_{10} = d'_{50} \neq d'_{100} = d'_{200}$	2	43.35	90.00	293.79	15.08	45.04	90.69
3	$d'_{10} = d'_{100} \neq d'_{50} = d'_{200}$	2	118.52	165.17	6.18×10^{18}	6.06	36.02	1.00*
4	$d'_{10} = d'_{200} \neq d'_{50} = d'_{100}$	2	102.05	148.70	1.63×10^{15}	13.41	43.37	39.45
5	$d'_{10} \neq d'_{50} = d'_{100} = d'_{200}$	2	38.75	85.40	29.44	19.38	49.34	780.23
6	$d'_{10} = d'_{50} = d'_{100} \neq d'_{200}$	2	99.73	146.38	5.13×10^{14}	11.73	41.69	16.98
7	$d'_{10} = d'_{50} = d'_{200} \neq d'_{100}$	2	119.43	166.08	9.74×10^{18}	10.07	40.03	7.41
8	$d'_{10} = d'_{100} = d'_{200} \neq d'_{50}$	2	163.73	210.37	4.02×10^{28}	14.70	44.66	75.00
9	$d'_{10} \neq d'_{50} \neq d'_{100} = d'_{200}$	3	16.13	86.10	41.75	10.91	55.84	2.01×10^{4}
10	$d'_{10} = d'_{50} \neq d'_{100} \neq d'_{200}$	3	27.22	97.20	1.07×10^{4}	4.17	49.10	692.87
11	$d'_{10} \neq d'_{50} = d'_{100} \neq d'_{200}$	3	8.66	78.64	1.00*	8.60	53.54	6.37×10^{3}
12	$d'_{10} = d'_{100} \neq d'_{50} = d'_{200}$	3	89.52	159.49	3.60×10^{17}	0.53	45.47	112.60
13	$d'_{10} = d'_{200} \neq d'_{100} \neq d'_{200}$	3	93.38	163.36	2.49×10^{18}	4.81	49.74	953.16
14	$d'_{10} \neq d'_{100} \neq d'_{200} = d'_{50}$	3	29.00	98.98	2.61×10^{4}	5.53	50.46	1.36×10^{3}
15	$d'_{10} \neq d'_{50} \neq d'_{100} \neq d'_{200}$	4	0.00	93.30	1.51×10^{3}	0.00	59.91	1.53×10^{5}

The results indicate that for Experiment 1 the data is best accounted for by Model 11 which has three separate functions: one for the 10 point data, one for the 50 and 100 point data and one for the 200 point data. Although this is not as strong an effect as was initially hypothesised, it does indicate an effect of stimulus density on discrimination for distorted symmetry. In other words, the observers have greater distortion tolerance for the 10 point stimuli compared to the 50 and 100 point stimuli, and show the lowest tolerance to distortion for the 200 point stimuli.

For the Experiment 2 data the results indicate that the most likely model is Model 3 which has two functions: one for the 10 and 100 point stimuli, and one for the 50 and 200 point stimuli. However, the Bayes factors indicate that Model 1, which corresponds to the experimental hypothesis of a single function for all four distributions, is only 2.81 times less likely. We can therefore conclude that there is no meaningful difference between the relative likelihoods of these two models. Importantly, although Model 11 indicates that the data is best accounted for by two separate functions, it does not predict an ordering of discriminability in terms of stimulus density.

In conclusion, although the effect was not as strong as was initially hypothesised, the data suggest that there is relationship between stimulus density and tolerance for distortion. Specifically, as stimulus density increases distortion tolerance decreases. Furthermore, density does not appear to have a similar effect on discrimination for symmetric stimuli that have been perturbed via the addition of random noise points. The results of the Bayesian model fitting indicated that the Experiment 2 data was best fit by either a single function or two functions (one for the 50 and 200 point data, and one for the 10 and 100 point data). However, there does not appear to be a similar ordering of tolerance to noise in regards to density.

7.3. Symmetry model comparison

The results of Experiments 1 and 2 were predicted by the theoretical behaviour of the Voronoi model. Given that the experimental hypotheses were supported by the empirical data (albeit not as strongly as predicted) we assume that the Voronoi model would have little trouble simulating the results of Experiments 1 and 2. In the following section this assumption is put to the test by comparing the performance of the two most likely models from Study 2 (the Voronoi model Type-4, and the Isotropic/Correlation model) on the Experiment 1 and 2 tasks.

In order to determine the range of model parameter values the model fits were considered across the eight new data sets from the current study as well as the six data sets from Study 2. It is important to do this for two reasons, in the first place it is possible that the new data may be well fit using a more limited range of values than was needed for the Study 2 data. However, even if this were the case it would still be necessary to include the fits across the full range of parameter values when calculating the log marginal likelihoods. Second, and more likely, it possible that the range of values employed in Study 2 underestimates the range needed to fit the new data. In this case the Study 2 log marginal likelihoods need to be re-calculated in order to include the contribution of the extra parameterisations.

For each model the log marginal likelihoods were calculated under two different assumptions. The first assumption is that of a constrained range of potential detection region widths. For the Voronoi Type-4 model this is identical to the values employed in Study 2: a width of 2 columns of cells, and a height of 10 cells. For the Isotropic/Correlation model constraining the range of potential detection region widths is more difficult, because (as has been discussed) the model has no means of automatically scaling the extent of its detection region in response to changes in density. The Jenkins (1982) data suggest that the optimal width for Experiments 4 to 6 in Study 2 is approximately 27 pixels. Arriving at an optimal width for the Study 2 Experiment 1 to 3 data is more difficult but the model fits appear to suggest that it is around 100 pixels or greater. Although it is necessarily *ad hoc*, the detection region widths (W_p) for the Isotropic/Correlation model were set as follows: for 650 point stimuli W_p = 20 to 30, for the 200 point stimuli, W_p = 50 to 100, and for stimuli with 100 points or less W_p = 100 to 128. The second assumption was that of unrestrained detection region widths. For the Voronoi model Type-4 this means the detection region width (W) is free to vary between 1 and 7 and Height (H) is free to vary between 2 and 14. For the Isotropic/Correlation model this is identical to the values employed in Study 2, with W_p ranging from 1 to 128.

There are a number of reasons for comparing the models across these two different sets of assumptions. In the first place, it is interesting to see if the performance of the Isotropic/Correlation could be improved if it had some means of scaling detection region size with stimulus density. Second, although the Voronoi model's constrained detection region is theoretically motivated, it is interesting to determine the degree to which this contributes to (or possibly detracts from) the model's overall likelihood.

Figures 7.3 to 7.6 show the fits of the Isotropic/Correlation model to the Experiment 1 and 2 data. The data indicate that in order to account for the empirical data the model must employ a wider range of spatial filter sizes than was necessary in Study 2. In Study 2 the spatial filter sizes ranged from s = 1 to 6, whereas in the current study the unconstrained version of the Isotropic/Correlation model obtains the best fit to the 200 point stimuli in Experiment 1 using a filter with a spread function of 17 pixels (Figure 7.3) and the constrained version of the model obtains a best fit to the 200 point stimuli in Experiment 2 using a filter with a spread function of 15 pixels (Figure 7.6). In regards to the additive noise parameter, the data indicate that both the constrained and unconstrained versions of the model require a slightly wider range of n_σ-values than was employed in Study 2. In this case n_σ ranged from 0.0 to 0.245.

The effect of constraining the range of Isotropic/Correlation model detection region width values can be clearly seen in Figures 7.3 to 7.6. A comparison of these figures indicates that qualitatively there is very little difference in the best fits of the constrained and unconstrained versions of the Isotropic/Correlation model. However, for the constrained version of the model, all of the predictions tend to be more tightly clustered around the empirical data. As a result, the sum log marginal likelihood of the constrained width Isotropic/Correlation model is greater than that of the unconstrained model when compared across both the combined Study 2 and Study 3 data and the Study 3 data alone (Table 7.2). This result highlights the importance of enabling the detection region to scale with changes in stimulus density. When the detection region is not matched to stimulus density, all of the model's predictions (good and bad) must be included in the calculation of the model likelihood. On the other hand, when the detection region width is set to a range that is more appropriately matched to the density of the stimuli a large number of poorly fitting predictions are excluded from the model's likelihood calculation, resulting in a higher overall likelihood.

The fits of the constrained and unconstrained versions of the Voronoi Type-4 model to the Experiment 1 and 2 data are shown in Figures 7.7 to 7.10. In this case the pattern of results is quite different to those of the Isotropic/Correlation model. When the Voronoi Type-4 model's detection region is free to vary (i.e., the unconstrained model) the quality of the best-fitting parameterisations are generally of a higher quality than those of the constrained model. Furthermore, although the unconstrained model makes a number of predictions that are

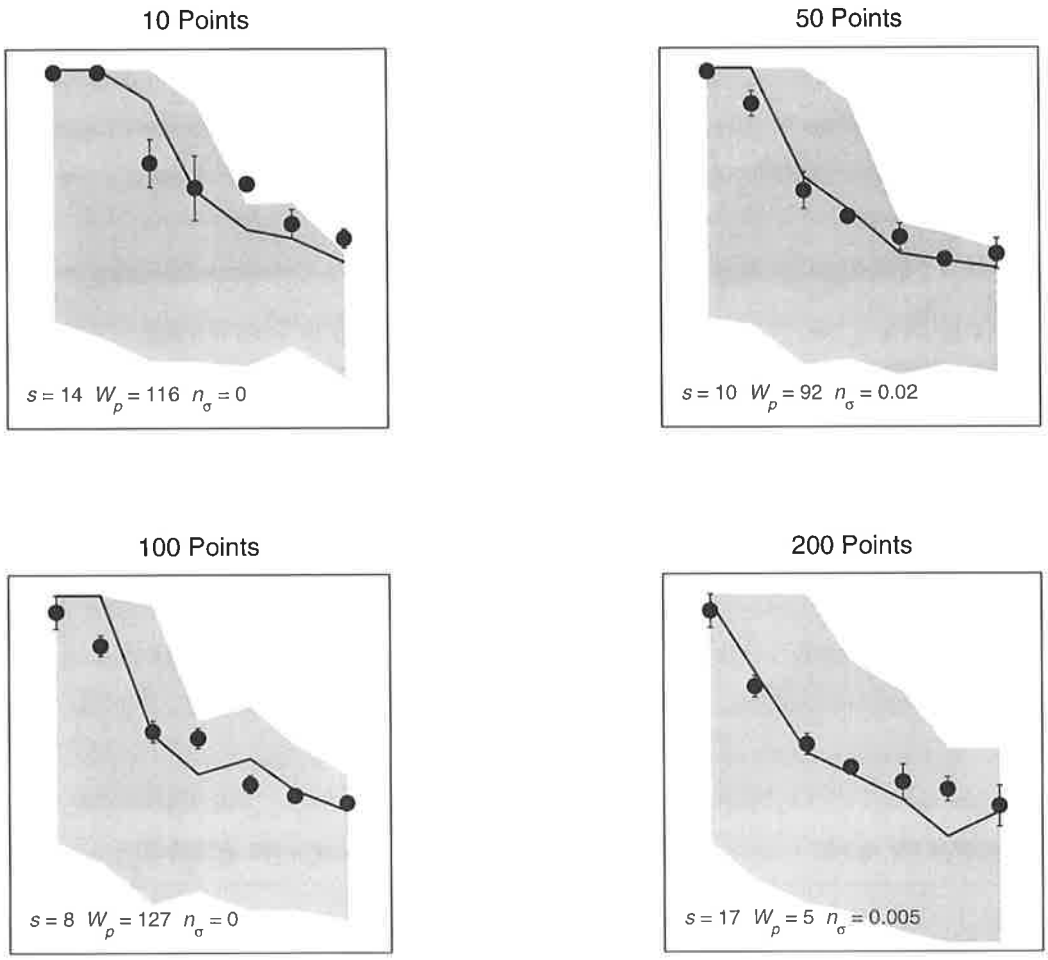

Figure 7.3 Comparison of the Isotropic/Correlation model fits to the Experiment 1 data under the assumption of an unconstrained range of detection region values. Spatial filter values ranged from s = *1-17, detection region width* (W_p) = *1-128 pixels, and the additive noise* (n_σ) *parameter range = 0.0-0.245. The best fitting parameter combination is shown by a black line, and the parameter values of the best fit are indicated in the bottom left corner of each panel.*

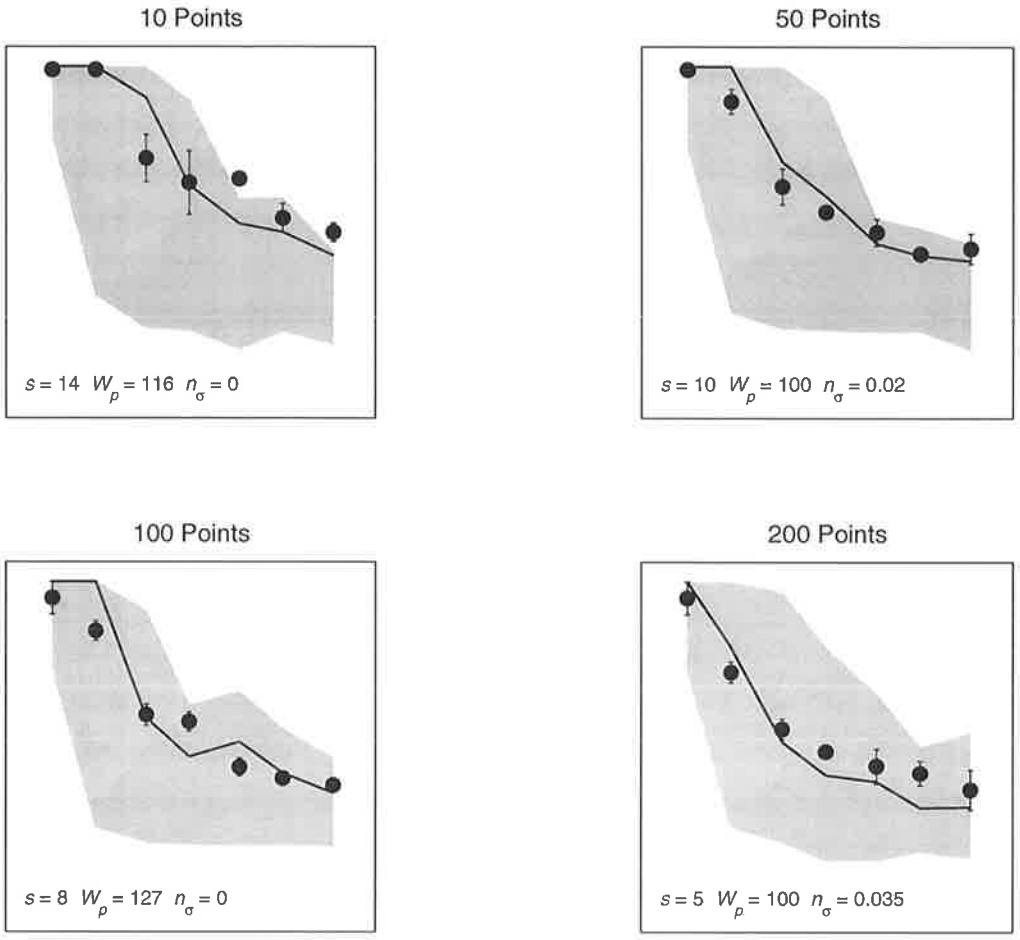

Figure 7.4 Comparison of the Isotropic/Correlation model fits to the Experiment 1 data under the assumption of a constrained range of detection region values. Spatial filter values ranged from s = *1-15, and the additive noise* (n_σ) *parameter range = 0.0-0.245. The detection region width values ranged from* W_p = *50-100 for the 200 point stimuli, and* W_p = *100-128 for all other stimuli. The best fitting parameter combination is shown by a black line, and the parameter values of the best fit are indicated in the bottom left corner of each panel.*

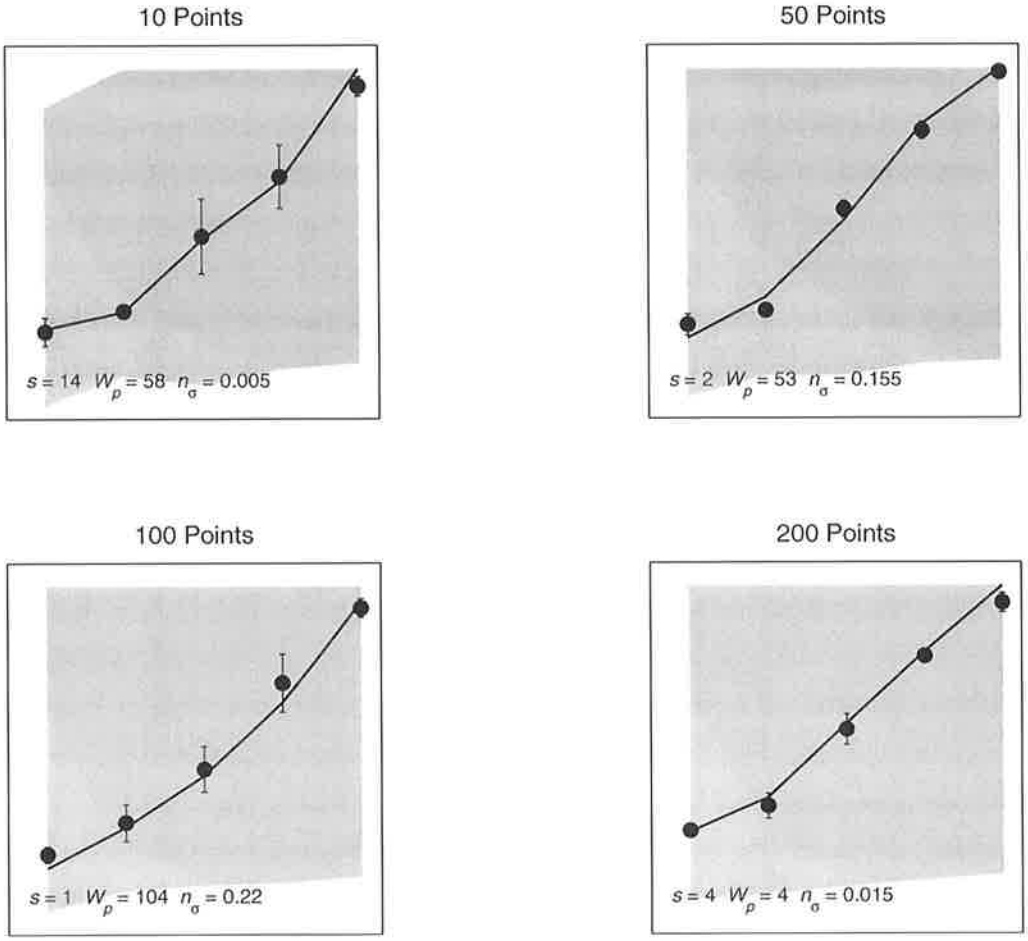

Figure 7.5 Comparison of the Isotropic/Correlation model fits to the Experiment 2 data under the assumption of an unconstrained range of detection region values. Spatial filter values ranged from s = *1-17, detection region width* (W_p) = *1-128 pixels, and the additive noise* (n_σ) *parameter range = 0.0-0.245. The best fitting parameter combination is shown by a black line, and the parameter values of the best fit are indicated in the bottom left corner of each panel.*

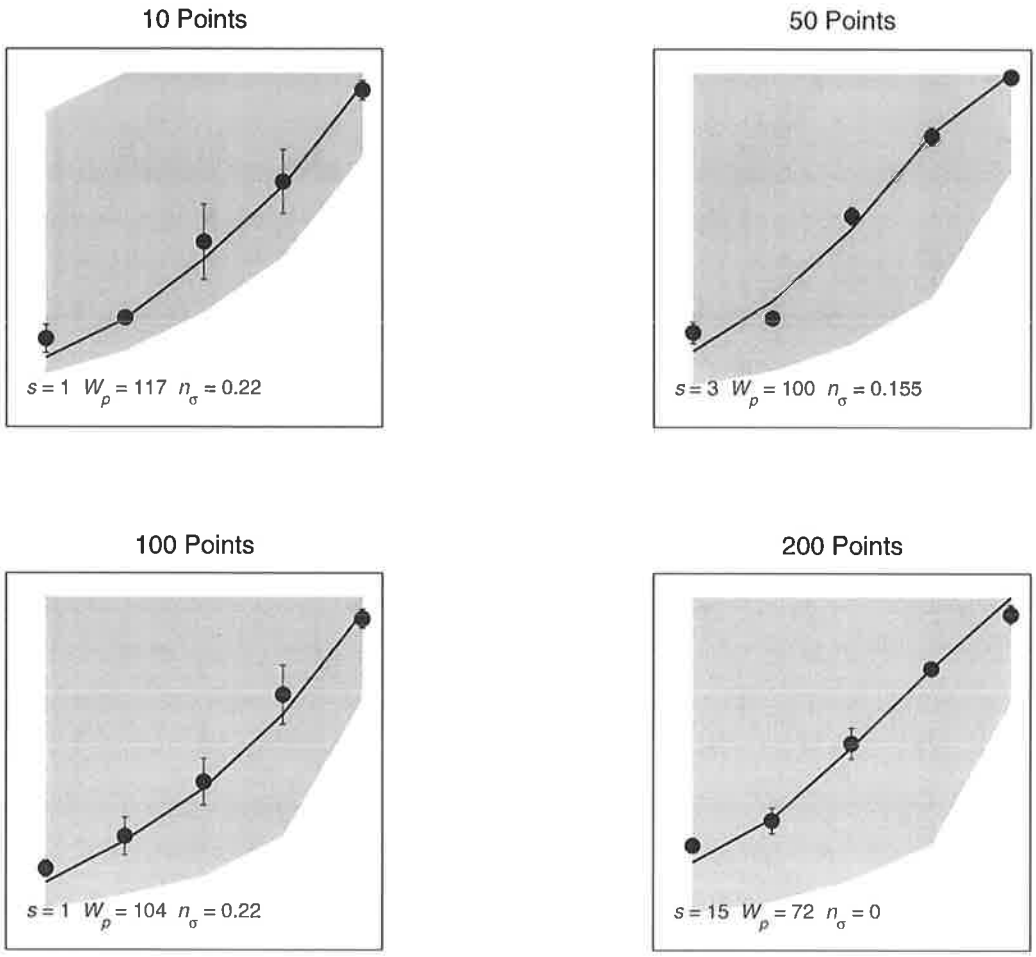

Figure 7.6 Comparison of the Isotropic/Correlation model fits to the Experiment 2 data under the assumption of a constrained range of detection region values. Spatial filter values ranged from s = *1-15, and the additive noise* (n_σ) *parameter range = 0.0-0.245. The detection region width values ranged from* W_p = *50-100 for the 200 point stimuli, and* W_p = *100-128 for all other stimuli. The best fitting parameter combination is shown by a black line, and the parameter values of the best fit are indicated in the bottom left corner of each panel.*

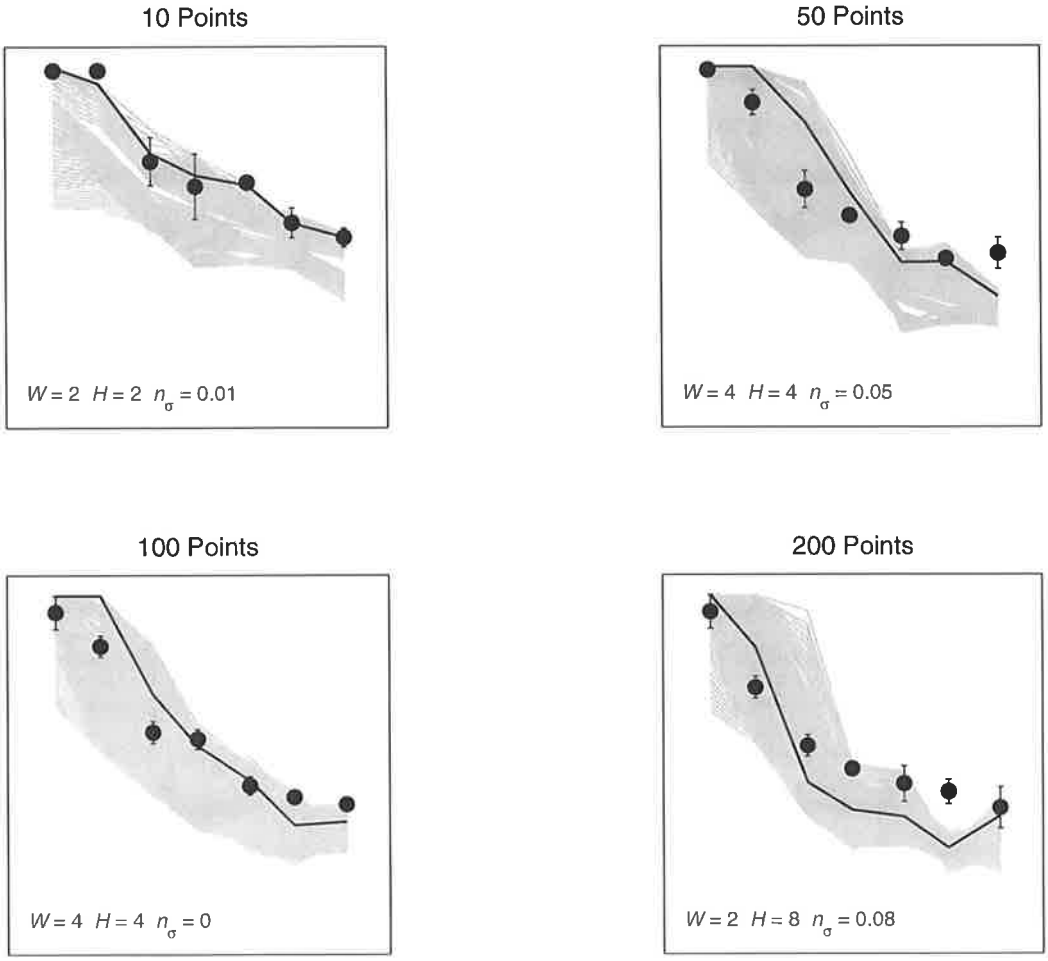

*Figure 7.7 Comparison of the Voronoi Type-4 model fits to the Experiment 1 data under the assumption of an unconstrained range of detection region values. Detection region width (*W*) = 1:7, detection region height (*H*) = 2:14, and the additive noise (n_σ) parameter range = 0.0-0.115. The best fitting parameter combination is shown by a black line, and the parameter values of the best fit are indicated in the bottom left corner of each panel.*

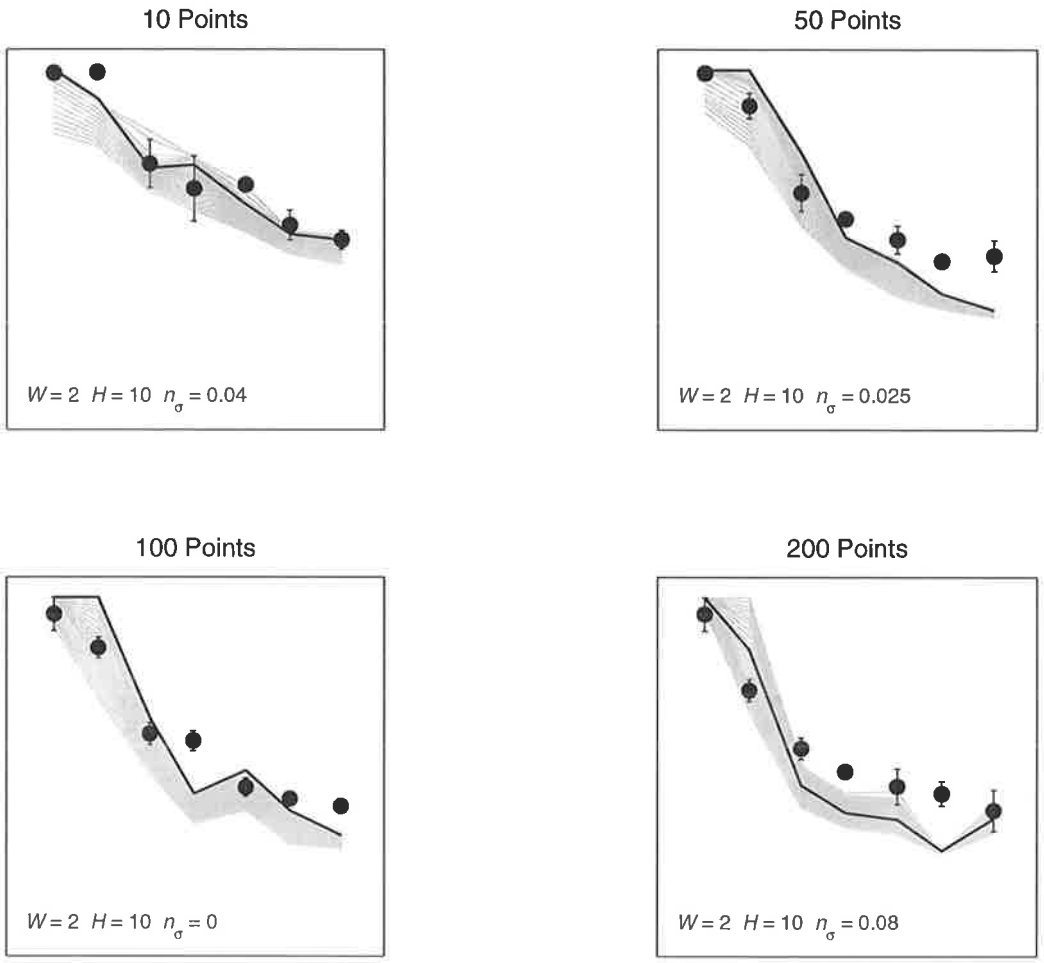

Figure 7.8 Comparison of the Voronoi Type-4 model fits to the Experiment 1 data under the assumption of a constrained range of detection region values. Detection region width (W) = 2, detection region height (H) = 10, and the additive noise (n_σ) parameter range = 0.0-0.115. The best fitting parameter combination is shown by a black line, and the parameter values of the best fit are indicated in the bottom left corner of each panel.

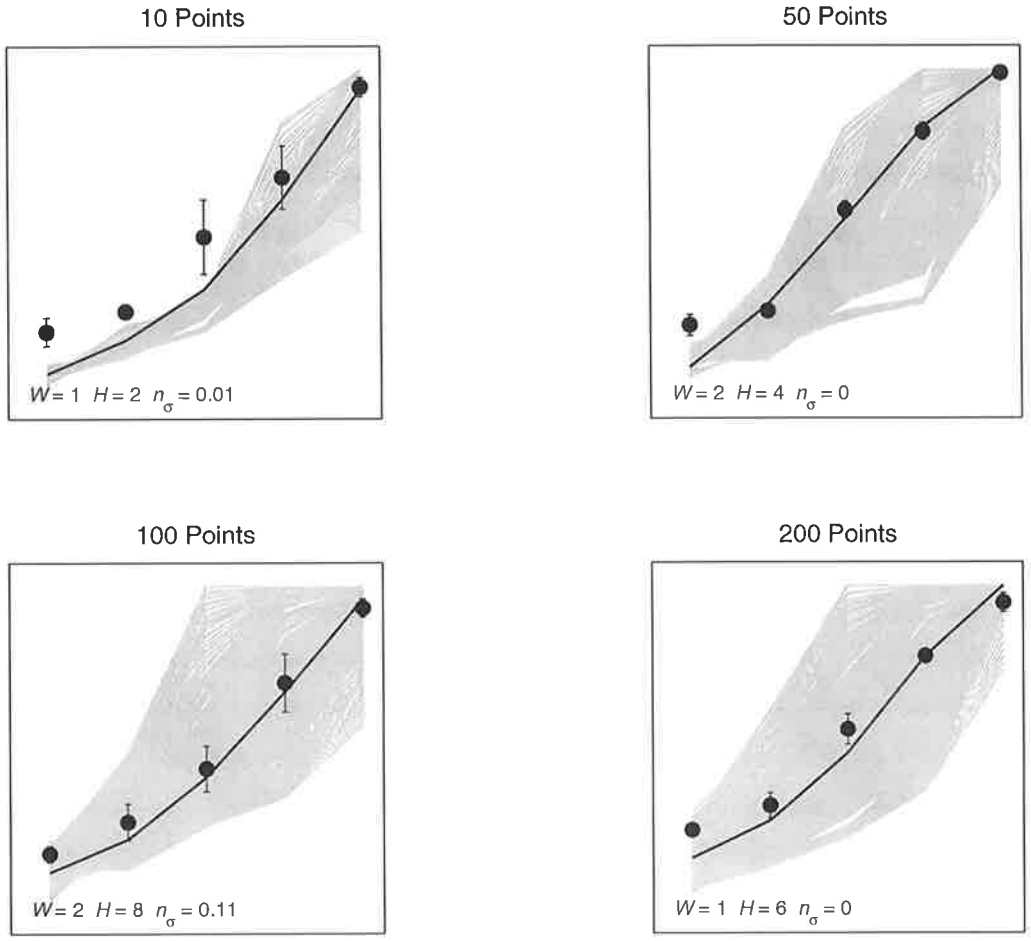

*Figure 7.9 Comparison of the Voronoi Type-4 model fits to the Experiment 2 data under the assumption of an unconstrained range of detection region values. Detection region width (*W*) = 1:7, detection region height (*H*) = 2:14, and the additive noise (n_σ) parameter range = 0.0-0.115. The best fitting parameter combination is shown by a black line, and the parameter values of the best fit are indicated in the bottom left corner of each panel.*

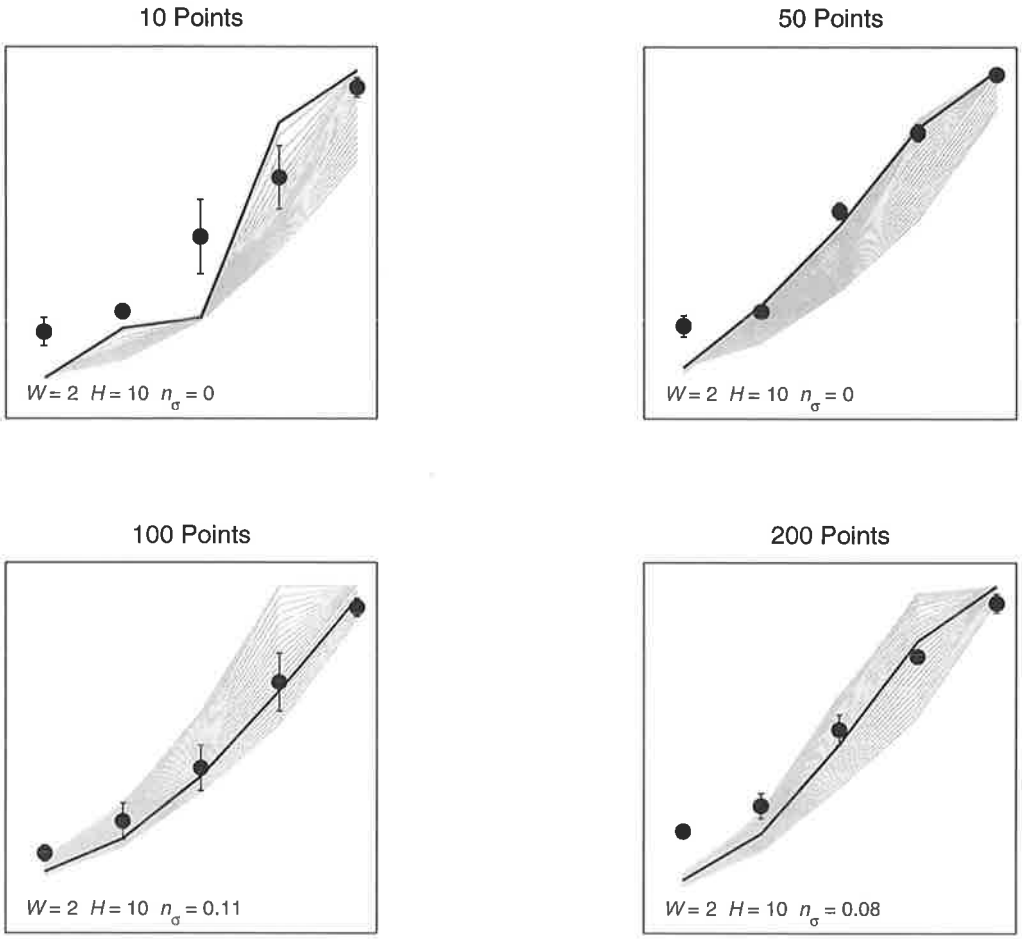

Figure 7.10 Comparison of the Voronoi Type-4 model fits to the Experiment 2 data under the assumption of a constrained range of detection region values. Detection region width (W) = 2, detection region height (H) = 5, and the additive noise (n_σ) parameter range = 0.0-0.115. The best fitting parameter combination is shown by a black line, and the parameter values of the best fit are indicated in the bottom left corner of each panel.

Table 7.2 Log marginal likelihoods for Experiments 1 and 2, the sum log marginal likelihood and Bayes factors for the Voronoi Type-4 and the Isotropic/Correlation models under the assumption of constrained and unconstrained detection region widths.

Model	log marginal likelihood										Bayes Factor	
	Experiment 1				Experiment 2				Sum		Study 3	Study 2 & Study 3
	10 Points	50 Points	100 Points	200 Points	10 Points	50 Points	100 Points	200 Points	Study 3	Study 2 & Study 3		
Isotropic/Correlation (unconstrained)	-2.87	6.09	-0.40	-4.96	-0.14	6.62	-0.29	3.35	7.39	97.40	21.11	523.21
Isotropic/Correlation (constrained)	-1.74	6.78	0.69	-5.45	-0.09	6.86	-0.36	3.10	9.79	102.78	1.91*	2.41*
Voronoi Type-4 (unconstrained)	4.35	5.08	1.39	-6.97	-2.49	6.83	-0.13	2.38	10.44	101.40	1.00*	9.58
Voronoi Type-4 (constrained, $W = 2$)	4.12	-12.03	-0.30	-5.28	-2.91	7.22	0.65	-0.51	-9.06	88.04	2.94×10^{8}	6.07×10^{6}
Voronoi Type-4 (constrained, $W = 2$-3)	4.04	4.76	2.06	-5.66	-2.83	6.95	0.47	-4.75	9.32	103.66	3.06*	1.00*

Note- Bayes Factors are taken in relation to the most likely model.
* denotes most likely model or models with Bayes factors of less than 3.2

further away from the observer's responses than those of the constrained model, the sum log marginal likelihood is considerably greater when considered across the Study 3 data alone, and across both the Study 2 and Study 3 data (Table 6.2). This appears to be mainly due to the constrained model's inability to fit the 100 point stimuli in Experiment 1 (Figure 7.8).

One possible reason for the difference between the Isotropic/correlation and Voronoi Type-4 models in regards to the relative likelihoods of the constrained and unconstrained model variants may be the extent to which the two models are constrained. For the Voronoi model the limitation of the detection region is quite specific: $W = 2$ and $H = 10$. In contrast, for the constrained Isotropic/Correlation model there is still a fair degree of variance in the W_p-values that the model is allowed to take (i.e., $W_p = 50\text{-}100$ for the 200 point stimuli and $W_p = 100\text{-}128$ for the 10, 50 and 100 point stimuli). Allowing the Voronoi Type-4 model even a small degree of flexibility in its detection region width range ($W = 2\text{-}3$, $H = 5$) increases the quality of the model's fit to the 100 point stimuli in Experiment 1, and raises the model's overall likelihood when considered across both the Study 2 and Study 3 data (Table 7.2).

The sum log marginal likelihoods of the models when considered across both the Study 2 and Study 3 data are particularly informative because they provide an indication of each model's ability to generalise across the six data sets from Study 2 and the eight data sets from Study 3. As can be seen in Table 7.2, the best account of the empirical symmetry perception data is provided by the Voronoi Type-4 model when employing a detection region constrained to W-values of 2-to-3, and a H-value of 5. Importantly, however, the Bayes factors indicate that the constrained-width Isotropic/Correlation model is only 2.41 times less likely, suggesting that no meaningful distinction can be made between these two models.

7.4. Discussion

The results of Study 3 are interesting for a number of reasons. In the first place, it was demonstrated that varying stimulus density has a differential effect on discrimination for distorted symmetry (Experiment 1) compared to discrimination for symmetry that has been perturbed via the addition of randomly positioned noise points (Experiment 2). In the case of the distorted stimuli, it was shown that as stimulus density increases, the level of distortion that the observers can tolerate tends to decrease. In contrast to this, for stimuli comprised of symmetry with added noise, there was no ordering of discriminability in terms of stimulus density.

As was discussed in the introduction to this Chapter, this pattern of results was predicted by the theoretical behaviour of the Voronoi model. However, the data from the model comparison indicates that the actual performance of the Voronoi model tends to fall short of its theoretical prediction. In particular, Figures 7.9 and 7.10 demonstrate that unlike the human observers, the performance of the Voronoi model (constrained or otherwise) differs in response to changes in stimulus density for the Experiment 2 data. This finding is interesting not just because it suggests that the Voronoi model may be an inadequate model of human symmetry perception, it also tells us something about model comparison methodology. It is reasonably common for empirical studies to be based upon the theoretical behaviour of a model and, if the empirical data match the theoretical predictions, then this is taken as evidence of the utility of the model in question. Examples of research such as this from the symmetry perception literature include Pashler (1990), Royer (1981) and Palmer and Hemenway (1978) among others. However, this study has demonstrated that in practise the actual behaviour of a model does not always match its theoretical predictions. This highlights the need for the quantitative comparison of formal models rather than relying upon empirical data to support or reject potentially inaccurate theoretical predictions.

The main finding of the Study 3 symmetry model comparison relates to the nature of the detection region for symmetry perception. It was demonstrated that in order for a model to have any utility as a potential explanation for the processes underlying symmetry perception it needs to have not only a spatially limited detection region but it must also be able to scale the detection region in response to changes in stimulus density. The unconstrained version of the Isotropic/Correlation model was able to provide good fits to the empirical data but it also made numerous unlikely predictions. Constraining the range of the model's detection region widths removed a large number of the poor fits and increased its overall likelihood. However, it should be remembered that constraining the model in this way relied upon hand-tuning the parameter values: in its current form the Isotropic/Correlation model has no means of automatically adjusting the detection region width to match stimulus density.

In contrast to this, the Voronoi models are able to automatically scale the size of the detection region in response to changes in stimulus density. In this sense at least, the Voronoi Type-4 model can be considered the more plausible of the two models. However, this plausibility is severely undermined by its inability to adequately simulate the performance of the human

observers using a single width/height parameter combination. As has been discussed, the empirical data from Jenkins (1982) indicate that the optimal detection region width for the Voronoi model is equal to 2 columns of cells either side of the stimulus midline. In theory, the Voronoi model should be able fit each of the data sets using this single parameter value. However, the results of the model comparison suggest that the Voronoi Type-4 model was only able to provide the most likely account of the empirical data when W was free to vary between 2 and 3.

It is worthwhile considering whether or not it is reasonable to assume a small degree of flexibility in Voronoi model W-value. On the one hand, the Jenkins experiments were effectively measuring the spatial extent of symmetry detection. If the data suggest that the limit of the Voronoi detection region in these experiments is defined by a W-value of 2, then it is implausible to suggest that the detection region should encompass more columns of cells in other experimental settings. In other words, if the visual system were able to behave in this manner then, presumably, it would have done so in the Jenkins experiments too. Following this logic it is implausible to assume any flexibility in regards to Voronoi model W-values.

However, it is possible that the apparent need for a small degree of flexibility might be explained by the results of Tyler, Hardage and Miller (1995). In this paper, the Jenkins (1982) experiments were replicated under two different conditions. The first condition used dynamic textures such as were employed in the original study and the second condition employed static textures, with the results suggesting that the detection regions for the static textures were slightly wider than those of the dynamic textures. Given that the empirical data from Experiments 1 to 3 in Study 2 and Experiments 1 and 2 in the current study are based upon static textures, whereas the empirical data from Experiments 4 to 6 in Study 2 (i.e., the Jenkins experiments) are based upon dynamic textures, Tyler et al.'s findings suggest that the Voronoi-model would best fit the data using two separate W-values. When W is set to 2 for the Jenkins experiments and 3 for all of the other experiments the sum log marginal likelihood is equal to 102.35, a value that is 3.69 times less likely than when the model is free to vary across W = 2-to-3, but 1.63×10^6 times more likely than when the width is held constant at W = 2.

Regardless of the degree to which the Voronoi model detection region is constrained, the current study highlights a number of fundamental weaknesses in the model. As has already

been discussed, there is theoretically no reason the performance of the model should vary dramatically with changes in stimulus density for the Experiment 2 data and yet Figures 7.9 and 7.10 indicate that both the constrained and unconstrained model under-perform in regards to the 10 point stimuli. Furthermore, both versions of the model tend to under-perform in response to high levels of distortion in Experiment 1.

It seems fitting that the Bayes factors indicate that no distinction can be made between the relative likelihoods of the two most likely models under comparison in Study 3. On the one hand the best fitting parameterisations of the constrained Isotropic/Correlation model are able to closely simulate the observer's behaviour but the model also makes numerous unlikely responses. On the other hand, the 'semi-constrained' Voronoi Type-4 model (i.e., with $W = 2$-to-3) makes relatively few predictions but the quality of the best-fitting predictions is not high. Furthermore, both of the models test the bounds of plausibility by relying upon what could be considered *ad hoc* assumptions: for the Isotropic/Correlation model it is the hand-tuning of the detection region widths and for the Voronoi model it is the assumption of a small degree of flexibility in W-values. From this we could rightly conclude that they are not very good models of symmetry perception. However, it should be remembered that these two models are still a great deal more likely than any of the other models that have been compared in this thesis. In many ways it is unsurprising that the models in question have failed to adequately capture the pattern of empirical results. They are (after all) only models, or simplified versions of what is an undeniably complex and well-calibrated system. Despite their flaws, these models have provided a number of interesting insights into the processes underlying the perception of symmetry, and in the following chapter these findings will be reviewed and discussed.

8. DISCUSSION

In Chapter 1 it was indicated that the aims of this thesis were threefold. First, to develop a theoretically plausible model of symmetry perception based on the relational structure revealed by Voronoi tessellation. Second, to introduce Bayesian model selection as a methodology for comparing models of visual processes. And third, to compare the performance of the Voronoi based model of symmetry perception to the performance of previously published models across a range of empirical symmetry detection tasks. Given that each of these goals has been met, I will now turn to a discussion of the main findings revealed by this research.

I will first discuss a number of findings relating to the Voronoi model, paying particular attention to both the possible implications that a Voronoi representation might have for our understanding of other models of symmetry perception and also to the limitations of a Voronoi based approach. This will be followed by a discussion of the utility of the alternative models in elucidating the processes underlying symmetry perception in the human visual system. I will then briefly discuss a number of issues relating to the decision mechanism employed by all of the models. Finally, given the reasonably novel status of the Bayesian model comparison methodology employed in this thesis, it is important to spend some time reviewing both its strengths and weaknesses.

8.1. The Voronoi model

One of the most interesting findings of this research was the indication that a Voronoi based model is at least as plausible as the spatial filter based models. Studies 1 to 3 provided quantified evidence indicating that a model of symmetry perception based upon a Voronoi representation is able to match and, for the most part, out-performed the spatial filter based models across a number of different empirical symmetry perception tasks. Importantly, this quantified evidence is not only based upon the model's best-fits to the empirical data, but takes into account all of the potential responses that the model is able to make. Furthermore, the evidence drawn from a wide range of neurophysiological and psychophysical studies that was presented in Chapter 2 suggests that a Voronoi based representation of relational structure is both theoretically motivated and physiologically plausible.

In this sense the Voronoi model can be seen to be meeting (albeit with varying success) all four of Jacob and Grainger's (1994) criteria for assessing model utility (i.e., descriptive adequacy, low complexity, generality, and explanatory adequacy). Meeting these criteria is highly important. For example, if the Barlow and Reeves model were able to match the most likely models in terms of goodness-of-fit, complexity and generality, a good case could still be made against the utility of their model on the grounds of its explanatory adequacy, ostensibly because the model's underlying representation (the visual field divided into equal-sized square sub-regions) is implausible.

Although there have been a number of previous models of symmetry perception that have been based on relational structure (i.e., Stevens, 1978; Wagemans et al., 1993), the Voronoi model is unique in the respect that an attempt has been made to specifically link the model to neural mechanisms. Admittedly, the body of evidence suggesting that a spreading activation or wavefront process (Kovacs et al., 1998; Lee et al., 1998) might be responsible for generating a medial axis or Voronoi-like representation in early vision is nowhere near as extensive as the body of evidence upon which the spatial filter models are based. Nonetheless, it is argued that linking the Voronoi model to this evidence goes at least some of the way towards meeting Rainville and Kingdom's challenge by demonstrating that a model of symmetry perception that is not based specifically upon spatial filtering can be "... psychophysically, physiologically, and theoretically motivated ..." (Rainville & Kingdom, 2000, p.2638).

8.1.1. The Voronoi representation and other models of symmetry perception

This finding raises the possibility that features of a number of previously published models of symmetry perception might be explained in terms of a Voronoi representation. For example, in some ways the Voronoi model can be thought of as a more plausible version of the Barlow and Reeves model. Both of the models calculate symmetry on the basis of the comparison of the dot density of symmetrically located stimulus regions; however, in the case of the Barlow and Reeves model, the underlying representation is highly implausible. It should be remembered that Barlow and Reeves were not actually suggesting that the visual system would calculate symmetry in this way. Rather, their model was meant to represent the kind of process that might be carried out by a neuron with a fixed receptive field. The results of the model comparisons suggest that a Voronoi representation forms the basis of another potential explanation for how the visual system might perform this process.

Furthermore, Jenkins (1983a) suggested that the perception of symmetry may arise as a consequence of the visual system detecting pairs of points that are uniformly orientated and have collinear midpoints. As has been demonstrated, it is possible to obtain this relational information from a Voronoi representation. It should be noted that the Voronoi model is not unique in this respect: for example, Dakin and Watt (1996) indicated that the Oriented/Alignment model was partly inspired by Jenkins' suggestion that the search for uniformly aligned and collinear point pairs could be achieved by orientationally tuned neurons in V1. Nonetheless, the model comparisons suggest that the Voronoi model was by far and away the more successful of the two approaches in terms of simulating the observers' responses.

In Chapter 2 it was suggested that the spreading activation or wavefront process that is proposed to be generating the Voronoi representation could also reveal the structure represented by Delaunay triangulation, the dual graph of Voronoi tessellation. If this is the case then it may be possible to explain Wagemanns et al.'s (1993) 'bootstrap' model of symmetry perception in terms of a Voronoi based representation. In a series of experiments (Wagemans, Van Gool, & d'Ydewalle, 1990; Wagemans, Vangool, & Dydewalle, 1992, etc) it was demonstrated that the perception of symmetry breaks down as stimuli are subjected to increasing amounts of skew. Figure 8.1 demonstrates this phenomenon, comparing an unskewed symmetric dot pattern (8.1a) with patterns that have been skewed by 15 degrees (8.1b) and 30 degrees (8.1c), respectively. It should be noted that in each case the stimuli

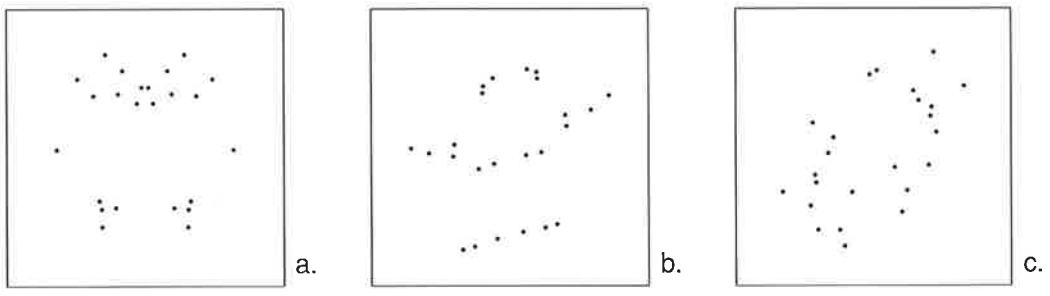

Figure 8.1 Comparison of an unskewed symmetric dot pattern (8.1a) with patterns that have been skewed by 15 degrees (8.1b) and 30 degrees (8.1c).

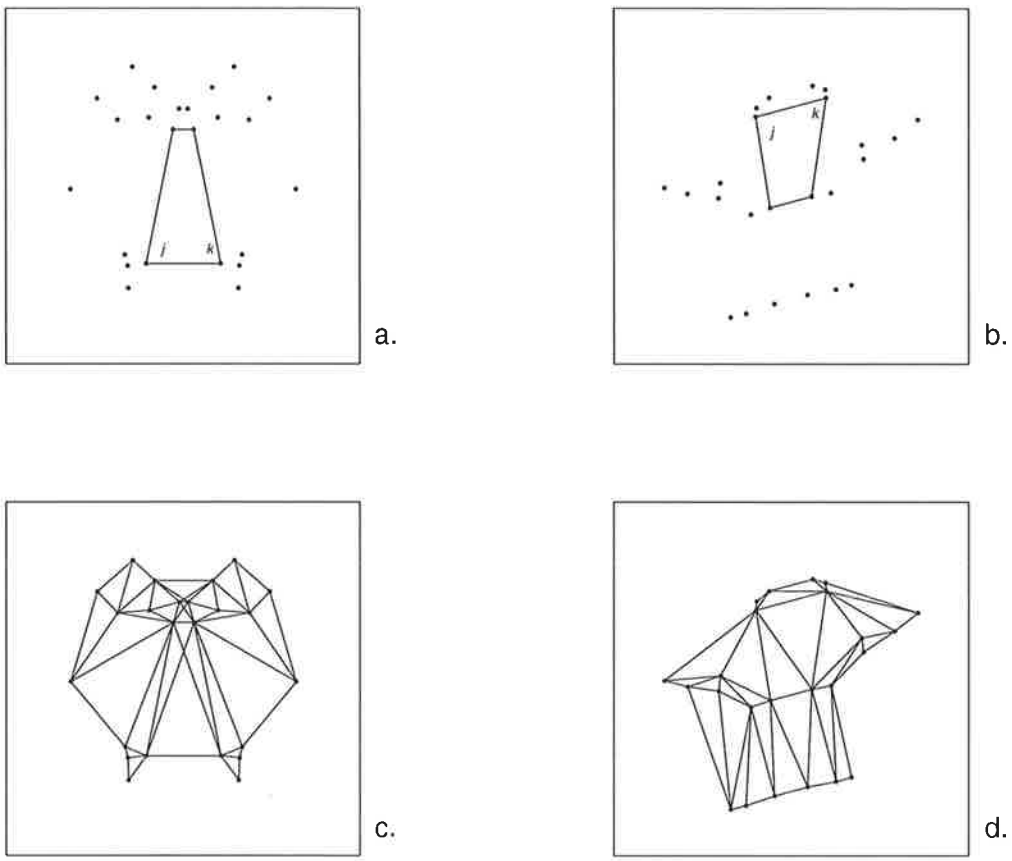

Figure 8.2 Examples of the difference in correlation between the inner angles (j, k) *formed between pairs of point pairs in an unskewed (8.2a) and skewed (8.2b) symmetric dot pattern. The corresponding Delaunay triangulations of the two point sets are shown in 8.2c and 8.2d.*

retain the characteristics of symmetric dot patterns as defined by the Jenkins (1983a) model (i.e., they are comprised of points pairs with uniform orientation and collinear midpoints); nonetheless, symmetry is more perceptible in 8.1a than in 8.1c.

Wagemanns et al. (1993) argued that the Jenkins model was insufficient to describe this phenomenon and suggested that, in order to extract symmetry relations, the visual system might also be relying on correlations between the angles formed by pairs of point pairs. For example, the inner angles (*j, k*) formed between two sets of point pairs in an unskewed symmetric pattern are equal (Figure 8.2a), whereas this is not the case for a skewed pattern (Figure 8.2b). Wagemanns et al. implemented a model that was able to capture the basic pattern of empirical responses in a skewed symmetry versus random noise detection task.

However, they did not explicitly specify how the visual system might go about obtaining this relative orientation information. Importantly, in regards to the current discussion, as can be seen in Figures 8.2c and 8.2d much of this relative orientation information is carried by the Delaunay triangulation of the point set.

8.1.2. Limitations of the Voronoi model

Despite the success of the Voronoi model (relative to the other models under comparison), the results of the simulations indicated that are a number of problems with the Voronoi model in its current form. In the following section I will discuss these problems, and also highlight a number of other limitations to the utility of the Voronoi model.

In Study 2 four different variants of the Voronoi model symmetry calculation procedure were compared with the results suggesting that the empirical data were best predicted by a version of the model that differentiated between stimuli in terms of their degree of asymmetry (the Voronoi Type-4 model). In other words, rather than searching for pairs of symmetrically located points, the Type-4 model searched for points for which no match could be found. While adopting this approach served the purpose of fitting the empirical data more closely, its underlying logic seems a little counter-intuitive. It would seem reasonable to assume that the model is able to detect both symmetry and asymmetry but that the demands of the experimental context were such that asymmetry provided the better cue for discrimination (i.e., it would explain how the visual system might detect both symmetry and deviations from symmetry). However, the model comparisons indicated that for a number of the experiments, the other three Voronoi model variants (which all based their symmetry calculation method at least partly upon the number matching or symmetric points) actually discriminated at a closer to optimal level than the Type-4 model. If we assume that the visual system has access to this information, then it seems implausible that it would ignore it in favour of a sub-optimal strategy.

A second major problem is the Voronoi model's tendency to under-perform in tasks involving the discrimination of distorted symmetry and random noise stimuli. In Study 2 it was suggested that this underperformance was a result of the point pairs in the experimental stimuli being randomly relocated beyond the bounds of the Voronoi model's tolerance region. For example, for stimuli with a dot density of 100 points the average distance between each

point and the most distant vertex of that point's associated Voronoi cell was approximately 0.10 (where distance is expressed as a proportion of the width of the stimulus). This represents the maximum distance (on average) that one half of a symmetrical pairing can be moved and still fall within the region corresponding to its partner's reflected cell. In other words, this value is an indication of the degree of distortion that the Voronoi model can tolerate. Importantly, the empirical data suggest that human observers can tolerate levels of distortion that exceed this value and, because the model's distortion tolerance range is defined by a physical property of its underlying representation, this could be taken as a strong challenge to the assumption of a Voronoi-like representation in symmetry perception.

Study 3 revealed a third major problem with the Voronoi model: on the basis of the theoretical behaviour of the model it was predicted that there would be an effect of stimulus density on tolerance for distortion, but not on tolerance for added noise. The results of the empirical research in Study 3 supported these predictions. However, the model simulations indicated that the actual behaviour of the Voronoi model did not match its theoretical predictions. Specifically, for the experiment involving symmetry with added noise (Experiment 2), the model demonstrated that stimulus density affects discriminability in the form of under-performance in the low density stimulus condition.

Finally, the data indicate that there are a number of problems with the assumptions relating to the Voronoi model detection region. The Jenkins (1982) data suggest that the optimal detection region width for the Voronoi model should be equal to two columns of cells either side of the stimulus midline. However, the Voronoi model fits indicate that for a number of the other experiments, the data are best predicted when the width of the detection region is free to vary between 2-to-3 columns of cells. As was discussed in Study 3, it is possible that this may be attributed to differences between the characteristics of the stimuli used in the Jenkins experiments and the stimuli employed in the remaining experiments (Tyler et al., 1995). Unfortunately, the methodology employed in the Tyler et al. study means that the empirical data do not lend themselves easily to the type of comparison necessary to test these assumptions in the Voronoi model (i.e., the data in the Tyler et al. study are threshold estimates for discrimination performance at different durations). In the absence of further data comparisons we must necessarily consider the assumption of a small degree of variability in the Voronoi model detection region width as *ad hoc*. Nonetheless, it is an assumption that could easily be tested in a future study.

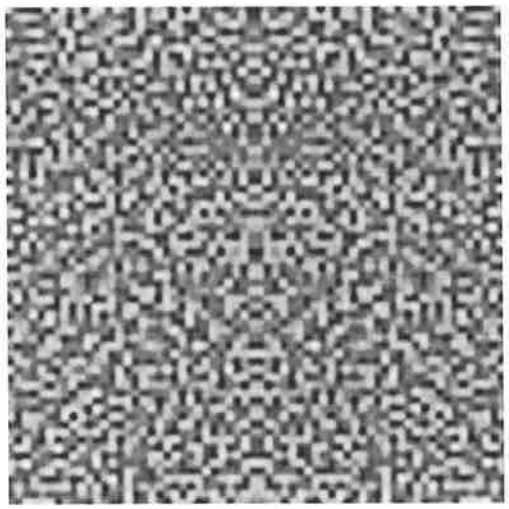

Figure 8.3 Example of the type of band-pass filtered noise textures employed in studies such as Dakin and Hess (1997) and Rainville and Kingdom (1999).

Aside from the problems specifically revealed by the results of the simulations, the Voronoi model is limited in a number of other important ways. In the first place, it has been demonstrated that spatial filter based models are able to account for symmetry detection in a wide range of different stimuli. For example, Figure 8.3 shows an example of the kind of high-density, band-pass filtered stimuli employed in studies by Dakin and Hess (1997) and Rainville and Kingdom (1999). In its current form the Voronoi model relies upon the tessellation produced by MATLAB 7.0.4, and is therefore only able to deal with dot patterns. As was indicated in Chapter 2, it is theoretically possible to obtain a Voronoi or medial axis representation from any form of stimulus input; however, it is difficult to envision how it would be possible to use this information to detect symmetry in stimuli such as Figure 8.3, in a manner that is analogous to the Voronoi model tested in this thesis.

This thesis has only focused on a limited sub-set of the many different types of symmetry perception phenomena that have previously been studied. For example, a number of studies (e.g., Barlow & Reeves, 1979; Pashler, 1990; Wagemans et al., 1992; Wenderoth, 1994) have demonstrated that the perception of vertical symmetry (Figure 8.4a) is more salient than either horizontal symmetry (Figure 8.4b), or symmetry that is presented at any other

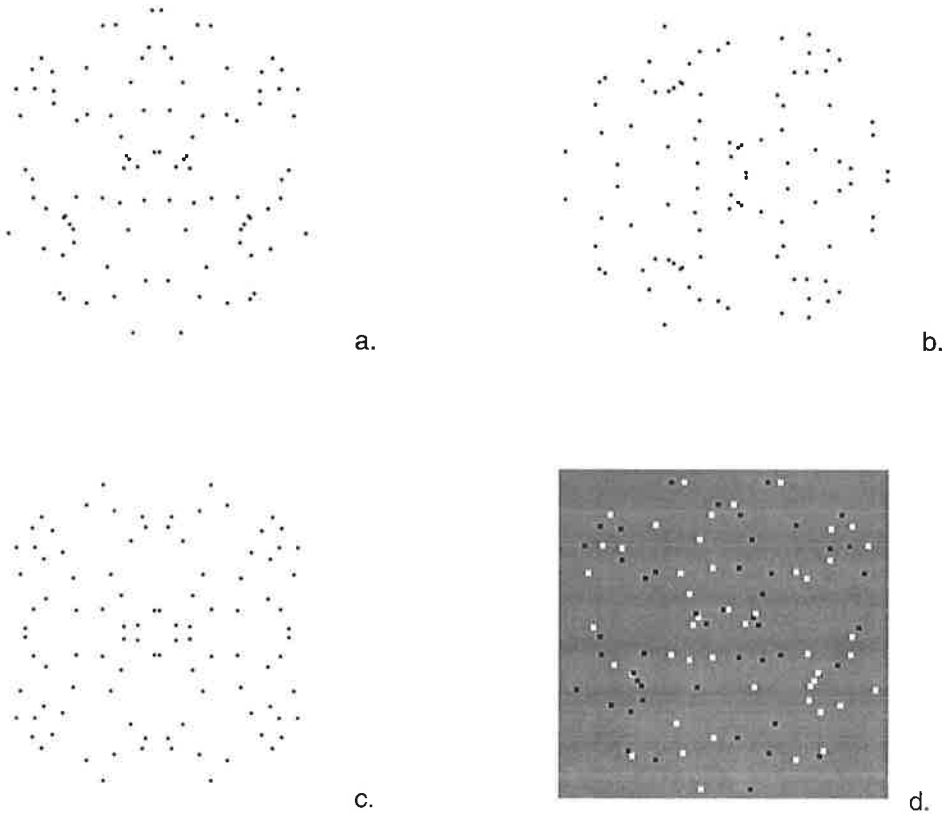

Figure 8.4 Examples of vertical symmetry (a), horizontal symmetry (b), multiple symmetry (c), and a pattern comprised of symmetrically located points with opposite contrast polarity (d).

orientation, and that stimuli with multiple symmetry axes (Figure 8.4c) are more salient than stimuli with a single axis (Palmer & Hemenway, 1978; Wagemans, Van Gool, & d'Ydewalle, 1991; Wenderoth, 1997). Additionally, it has been demonstrated that manipulating the contrast polarity of symmetric point pairs (Figure 8.4d) disrupts (but does not destroy) the saliency of the symmetric structure (Wenderoth, 1996b). It should be noted that the Voronoi model is unable to provide any insight into these phenomena but, by the same token, none of the other models under comparison in this thesis is particularly enlightening in this regard either.

These limitations pose some strong challenges to the Voronoi model. However, they do not necessarily rule out the possibility of a Voronoi-based representation in early vision, because it appears that a number of the weaknesses revealed by the results of the simulations may be plausibly attributed to problems associated with the model's symmetry calculation method. For example, as was discussed above, changing the assumptions relating to symmetry calculation method had a non-trivial effect on the model's relative likelihood and one instance in which this was quite apparent was in the distorted symmetry versus noise experiment in Study 2. Although the best fitting model in this case (the Type-4 model) is arguably implausible, it is possible that a future version of the Voronoi model might be able to overcome the limited distortion tolerance of the other models by drawing upon additional information that is available in the Voronoi representation (such as the inter-angle correlations between pairs of point-pairs shown in Figure 8.2). In other words, it is possible that a Voronoi representation is plausible, but the way that that the Voronoi model(s) currently use the information made available by this representation is insufficient to adequately describe the actual processes occurring in these sorts of symmetry detection tasks. The challenge of future research is therefore to elucidate a Voronoi-based model of symmetry perception that is able to overcome these weaknesses.

8.2. The alternative models

At this point I would like to shift the discussion away from the Voronoi model and discuss some of the findings revealed by the results of the model comparisons in relation to the five alternative models (i.e., the Barlow and Reeves, Isotropic/Alignment, Oriented/Alignment, Isotropic/Correlation, and Oriented/Correlation models). I will first discuss the two feature alignment models, followed by the Barlow and Reeves model, and finally the two cross-correlation models.

8.2.1. The feature alignment models

One of the most surprising outcomes of Study 2 was the indication that the feature alignment models were found to be the least likely out of all of the of the models under comparison. This finding is unexpected because these two models share a number of qualities that suggest that they should have done quite well simulating the empirical response patterns. First place, they are based on a spatially filtered representation which means that they should be have been reasonably tolerant of any positional distortion or added noise in the symmetrical stimuli (i.e., the filtering smooths out the image and removes any abrupt changes in luminance).

Second, one of the built-in assumptions of both of these models is a spatially restricted detection region which suggests that the models should have been able to capture the pattern of empirical results in the Jenkins (1982) data. Additionally (although this has little bearing on predicting the quality of the data fits), the feature alignment models both share a degree of theoretical plausibility: as has been discussed, the feature alignment models can be thought of as a means of formally implementing the Jenkins (1983a) model and are based upon known neural mechanisms.

There appear to be two main factors which are responsible the poor performance of the alignment models. The first factor relates to the general inadequacy of the alignment model symmetry calculation procedure. Briefly; the feature alignment models calculate the degree of symmetry in a stimulus based upon the collinearity of the axis-crossing stimulus features (see Section 5.3.3 for a more detailed description). However, the data suggest that this approach penalises positionally distorted symmetric stimuli too heavily, resulting in underperformance on the distorted symmetry versus random noise discrimination task (i.e., see Figures 6.10 or 6.12).

Second, there are a number of problems with the symmetry detection region of the feature alignment models. The biggest problem appears to be the fact that the spatial extent of the detection region is largely dictated by the scale of the filter employed to derive the initial stimulus representation. In other words, convolving the stimulus with a small filter will result in a highly restricted detection region, whereas using a large filter will result in a much wider detection region. This is particularly problematic for the Isotropic/Alignment model in Experiment 6 (see Figure 5.26 for Study 1, or Figure 6.10 for Study 2). In this experiment the symmetric point pairs are separated by a centrally located strip of noise points. In order for the detection region to be wide enough to include this symmetric information, the stimulus must be convolved with a coarse scale filter (i.e., $s = 16$ or 17). However, this inhibits the model's ability to distinguish between random noise and symmetric structure and results in the model underperforming on the task.

Additionally, there is a further problem with the detection region of the feature alignment models that argues strongly against their plausibility: because only stimulus features that cross the midline are considered to fall into the model's detection region, the alignment models

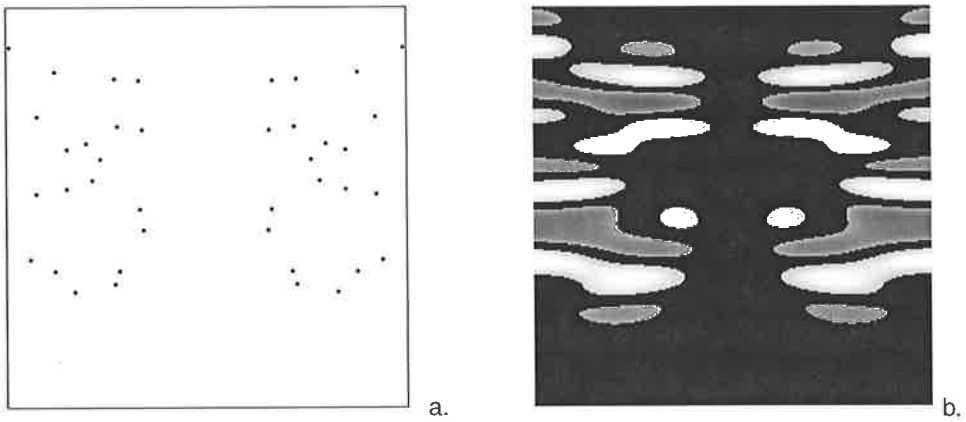

Figure 8.5 A symmetric dot pattern and the resulting stimulus representation derived by the Oriented/Alignment model using a spread function of s = *7.07*

would not be able to detect any symmetry in stimuli such as Figure 8.5a where there is a large gap between the two pattern halves. For example, Figure 8.5b shows the stimulus representation derived by the Oriented/Alignment model for the symmetric pattern in Figure 8.5a using a spread function of s = 7.07. Because none of the stimulus features cross the midline the models detection region contains no information and, therefore, the symmetry cannot be detected. This is obviously a serious problem for these models, because the visual system would surely detect the symmetry in Figure 8.5a, even if it were only presented for a very brief duration. Furthermore, it is important to note that none of the other models under comparison would have a similar problem with these sorts of stimuli. One potential solution to this problem is to increase the size of the filter employed to derive the initial representation. However, as has been discussed, this tends to impede the model's ability to distinguish between symmetry and random noise.

8.2.2. The Barlow and Reeves model

In many ways it was not necessary to perform any quantified model comparison to determine that the Barlow and Reeves model was an implausible explanation for the processes underlying symmetry perception. Specifically, the nature of the model's representation (the visual field divided into equal sized square sub-regions) is sufficient to rule out the model on the grounds of neurophysiological implausibility. Nonetheless, for both Study 1 and Study 2 the Barlow and Reeves model was found to be more likely than any of the other models in

regards to the symmetry with distortion versus random noise discrimination task. In this respect the Barlow and Reeves model is still potentially informative: by examining the reason for the models success in this task it may be possible to learn something about the way the visual system deals with distorted symmetry.

In Study 2 the next most likely model on the distorted symmetry task was a variant of the Voronoi model. A common feature of these two models is that symmetry is calculated on the basis of some form of dot density comparison. However, for the Voronoi model this is performed over relatively small regions (the reflection of each point's Voronoi cell) compared to the Barlow and Reeves model. In this sense the Barlow and Reeves model can be thought of as comparing clusters of points rather than just individual points (Wenderoth, 1996b). There is potential utility in assuming that the visual system may be detecting symmetry via some form of cluster comparison. First, it is possible that this could explain the observers' reasonably high tolerance to distortion, in that points can be moved outside of the tolerance region defined by a Voronoi cell or spatial filter and yet still remain part of a cluster of points. Second, a number of researchers have suggested that clusters of points forming salient sub-structures (such as the two semi-circular clusters in Figure 8.6) may play an important role in symmetry detection (Wagemans, 1999; Wenderoth, 1996c). Third, there is empirical evidence suggesting that this type of grouping or clustering of stimulus elements may actually precede symmetry detection (Labonte et al., 1995; Locher & Wagemans, 1993).

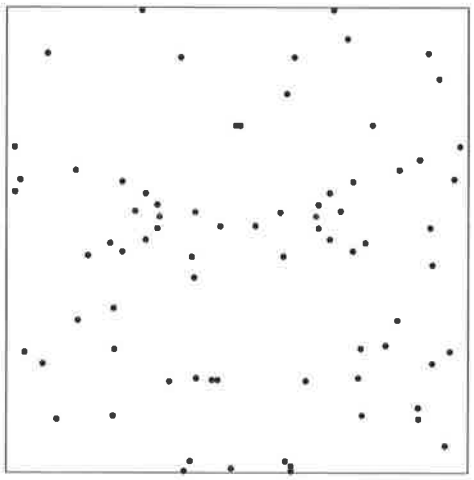

Figure 8.6 Example of a symmetric dot pattern with salient sub-structures.

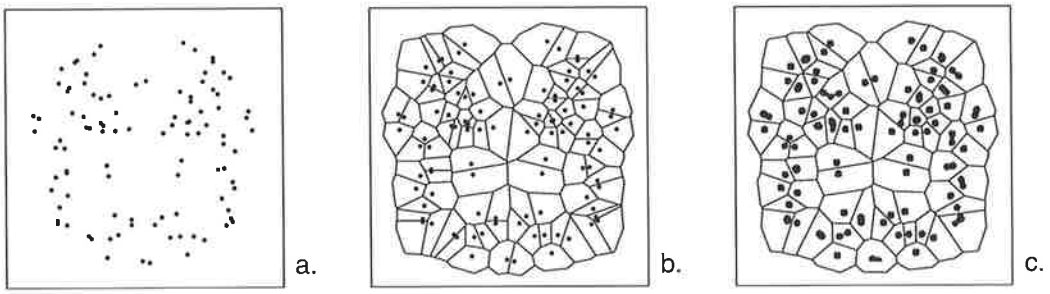

*Figure 8.7 A distorted symmetric stimulus (a), and its associated Voronoi tessellation (b). The tessellation obtained after initially convolving the stimulus with a fine-scale isotropic filter (*s = *4) is shown in 8.7c.*

Of course, it must be stressed that the implication of cluster-comparison as a potential explanation for the success of the Barlow and Reeves model in this task is purely speculative. Nonetheless, it does point toward a potential line of development for future models of symmetry detection. For example, under the assumption of a Voronoi representation it might be possible to compare symmetrically located clusters of points defined by a sub-set of the Delaunay triangulation such as the nearest neighbour clusters. Alternatively, if we assume that symmetry perception is based on a spatial filter representation, the cluster comparison process might be better explained by comparing the relative size and location of the blobs or features in a convolved stimulus. This in turn suggests the possibility that the data might be best explained by some form of a hybrid-model, in which the tessellation of a stimulus occurs after an initial filtering stage. For example, Figure 8.7 compares the Voronoi tessellation of a distorted symmetric stimulus before filtering (8.7b) and after filtering with a fine-scale isotropic filter (8.7c). As can be seen, in 8.7c a number of the individual points have been clustered together or fused into larger stimulus features. As a result, there are fewer cells in the tessellation, and the average size of the cells is larger than in 8.7b. It is possible that this increase in cell size may result in greater distortion tolerance.

8.2.3. The cross-correlation models

One of the most common approaches that researchers have taken in implementing models of symmetry detection is to employ some form of cross-correlation procedure (e.g., Dakin & Watt, 1996; Gurnsey et al., 1998; Pintsov, 1989; Rainville & Kingdom, 1999; Rainville & Kingdom, 2002; Tjan & Liu, 2005). As was demonstrated in Study 1, when cross-correlation is applied to an entire image, the models are unable to account for data indicating that the

perception of symmetry is limited to a spatially restricted region directly adjacent to the axis of symmetry (e.g., Jenkins, 1982). However, in Study 2 it was demonstrated that this weakness could be overcome by simply adding a spatially limited detection region to the cross-correlation models. This result is important because it suggests that other models that do not specifically include a spatially limited detection region might also benefit from this simple modification.

The simulation results indicated that the best-fitting parameterisations of both of the cross-correlation models were generally able to closely replicate the empirical response patterns. However, this high descriptive adequacy came at the cost of high complexity, with the models also being able to produce responses ranging from perfect performance down to chance performance for nearly all of the experiments (e.g., see Figures 6.11 and 6.13). Clearly, the utility of such a model as a potential explanation for the processes underlying symmetry perception is limited. The model may be able to simulate the behaviour of the visual system in response to symmetry, but in order to do so it must also be capable of behaving in ways that are clearly implausible.

The high complexity of the cross-correlation models stems largely from the high number of detection region width parameter values over which they must be implemented. It has been demonstrated that changes in stimulus density affect the spatial extent of the region within which symmetry can be detected (Rainville & Kingdom, 2002). Because the cross-correlation models have no means of adjusting the width of the detection region in response to changes in stimulus density, in order to respond in a plausible manner to all potential stimulus densities they must be implemented using all possible detection region widths. In a future model it may be possible to reduce this complexity by specifically matching detection region width to stimulus density, in a similar manner to Rainville and Kingdom's (2002) suggestion that spatial scale selection could be determined by stimulus density. However, although Rainville and Kingdom suggest that this should be possible, they do not offer any indication of how it might be plausibly achieved.

An additional problem with the cross-correlation models that has been mentioned by a number of researchers (Dakin & Hess, 1997; Dakin & Watt, 1996; Jenkins, 1982, 1983a; Rainville & Kingdom, 2000) is that it is difficult to imagine how the visual system might actually carry out a cross-correlation process. The underlying representation employed by the

two cross-correlation models in this thesis is relatively uncontroversial (i.e., a spatially filtered image). However, the actual correlation process implies that the visual system is comparing luminance values across an enormous number of different locations in the visual field. While it is possible that this kind of process could be achieved at a neuronal level, the degree of interconnectivity that would be required to detect symmetry at different axis orientations and at different locations in the visual field seems fairly improbable (Dakin & Hess, 1997; Rainville & Kingdom, 2000). Of course, it should be remembered that a similar criticism could be directed at the Voronoi model: specifically, it is one thing to demonstrate how the visual system might plausibly obtain relational information via a Voronoi-like processes, but it is another thing to demonstrate how the visual system might actually use this information at a neuronal level.

8.3. The decision mechanism

Up until this point I have avoided discussing one of the other major findings of the model comparisons: the apparent need for variability in the decision making process. In the following section this finding will be discussed, along with other issues relating to the decision mechanism employed by the models.

In Study 1, all of the models (with the exception of the Voronoi model) assigned the same value to the perfectly symmetric stimuli. Because of this lack of variance there was little or no overlap between the distributions of scores for the two classes of stimuli (i.e., perfect symmetry and symmetry with added noise) in Experiments 3 and 5 and, as a result, the model's discrimination performance tended to be implausibly high. In Study 2, this problem was addressed by placing a Gaussian distribution over each of the values assigned by a model to a given stimulus. It was demonstrated that this modification was able to increase the quality of the fits to empirical data for all of the models under comparison. However, it was noted for the majority of the models there was a higher degree of variance in the level of added noise needed to fit the models across all of the data sets than could be reasonably expected.

Although the assumption of inherent noise in the decision-making process is uncontroversial (Green & Swets, 1966) the data suggest that the above-mentioned approach to accounting for this variance is insufficient. Perhaps one of the problems with this approach is that it was based upon the assumption that the error rates for both of the classes of stimuli were equal. In other words, no attempt was made to model bias in the decision process, or the degree to

which the decision threshold is shifted away from the mean of one of the two distributions. For the majority of the data sets employed in the model comparisons no information was available to suggest that the observers may be demonstrating bias. However, for both of the replicated experiments in Study 2 the relative proportions of hits and false alarms suggest that there may be a bias towards categorising stimuli as symmetric rather than random. Furthermore, there is evidence from other studies that also suggests that this might be occurring (e.g., Wenderoth, 1997). In light of this, it appears that modelling observer bias might be worth pursuing in future model comparisons.

These problems highlight one of the major difficulties associated with modelling cognitive or perceptual processes. Because the brain is arguably a massively interconnected and inter-dependant set of systems and processes, it is very difficult (and maybe even impossible) to simulate a single process in isolation. This thesis is ostensibly concerned with models of symmetry perception. However, in order to implement these models strong assumptions needed to be made about the nature of the decision-making process associated with the empirical tasks. As has been demonstrated, these assumptions can have a non-trivial affect upon the results of the model comparisons. It is quite possible that if different assumptions were made about the nature of the decision process that the results of the model comparisons could be quite different to those that have been reported.

This is clearly quite problematic and, unfortunately, there does not appear to be an easy solution. One approach to dealing with this difficulty would be to employ more sophisticated decision-making processes. For example, although the signal detection theory (Green & Swets, 1966) approach that was adopted in this thesis is widely accepted and applied within the visual perception literature, there is a growing body of evidence to suggest that it is overly simplistic and that decision-making is more sufficiently described by sequential sampling models such random walk or accumulator processes (see Ratcliff & Smith, 2004 for a recent review of these, and related, models). However, these models tend to be computationally demanding to implement and there is little agreement over which of these approaches has the greater utility. Furthermore, there is also the added danger that adopting a more sophisticated decision-making process may add unwarranted complexity to the models and, as a result, make it more difficult to learn anything about the actual process of interest.

Regardless of the form of the decision mechanism, it is important to ensure that all of the associated assumptions are both theoretically motivated and interpretable. In this sense the current decision-making mechanism has some claim to utility: because the decision process was reasonably simple, it was easy to see that the assumptions regarding additive noise were implausible. In the case of a multi-parameter decision process this may not have been so easy to determine.

8.4. Bayesian model selection

As was stated in the introduction, one of the three primary aims of this thesis was to demonstrate Bayesian model selection as a means of comparing formal models of visual phenomena. Given that this approach is relatively novel (at least within the visual perception literature) it is important to spend some time discussing both the strengths and weaknesses of this approach.

I will begin by discussing a number of the strengths of Bayesian model selection, paying particular attention to the methodology's ability to address the problems associated with previous research involving models of symmetry perception. One of the greatest problems with previous studies has been the lack of direct comparison between rival models. As has been demonstrated, all of the models under comparison in this thesis were able to capture the general pattern of observer responses in most of the empirical data sets. If the models were only ever considered in isolation, then any of them could be taken as a potentially useful explanation for the processes underlying symmetry perception. In other words, we would not be able to narrow down the range of potential explanations for this phenomenon. However, by directly comparing the models we are able to obtain a rank them in terms of their utility. Furthermore, Bayesian model selection offers an interpretable means of quantifying this utility in the form of a measure of relative likelihood.

Another issue associated with modelling psychological processes that has gained much attention in recent years is the general over-reliance upon using the quality of a model's best-fitting parameter combination as an indication of its utility (Lee, 2004; Myung & Pitt, 1997; Pitt & Myung, 2002; Roberts & Pashler, 2000). One of the strengths of Bayesian model selection is that it is able to take into account not only the descriptive adequacy of a model, but also its complexity and generality. However, this raises an important question when applied in relation to psychological models: Given that the human brain is arguably a highly

complex system is it feasible to penalise models for demonstrating similarly high levels of complexity? The answer to this question lies in defining what is meant by complexity. A number of other model selection techniques (e.g., the AIC (Akaike, 1983), BIC (Schwarz, 1978) and RMSEA (Steiger, 1990)) also penalise models in terms of complexity. However, in the case of these alternative approaches, complexity refers exclusively to parametric complexity, or the number of parameters that a model has. In Bayesian model selection on the other hand, complexity is better thought of as an index of the range of responses that a model is able to make across each of its parameter combinations. In this case it is possible for a model to have a high number of parameters but still have a low complexity, as long as each of the responses that the model makes corresponds closely with the empirical responses.

The processes underlying human symmetry perception may be complicated (in the general sense of the word) but they are also highly constrained. The empirical data indicate that there tends to be little variability in the range of responses produced by the human observers. In contrast to this, a number of the models were able to produce a wide range of responses, many of which bore little-or-no resemblance to the empirical response patterns. For example, the Isotropic/Correlation model was able to provide close fits to the empirical data in Study 2 (see Figure 6.11). However, it also produced responses ranging from chance performance up to perfect performance for most of the experiments. In light of the limited range of responses produced by the human observers it is fair to suggest that this model is implausible.

It should be noted that the quantification of model likelihood provided by the Bayesian analyses should never be taken as the be-all and end-all of any model comparison. There will always be an element of qualitative interpretation involved in any of the analyses. Bayes factors provide an indication of the relative likelihood of a set of models; however, a high relative likelihood is not necessarily an indication of model utility. For example, in Experiment 4 from Study 1, the Isotropic/Alignment model was found to be 2.28×10^5 times less likely than the Isotropic/Correlation model, but a qualitative comparison of the data fits in Figure 5.26 indicates that neither model is particularly informative. Furthermore, Bayesian model selection cannot provide any indication of the explanatory adequacy of a model. For example, in Study 2 the Barlow and Reeves model was found to be more likely than the two feature alignment models. Nonetheless, given the implausible assumptions of the Barlow and Reeves model it could be justifiably argued that the alignment models actually have the greater utility.

The last two points should not be interpreted as limitations so much as caveats (albeit highly important ones). However, the analyses did indicate that there were a number of issues related to the Bayesian approach that could be considered to be potentially problematic. The first of these issues is the difficulty associated with determining each model's necessary range of parameter values. In Study 1 it was demonstrated that limiting the range of parameter values over which a model is implemented could reduce model complexity and, following this, in Study 2 the model likelihoods were only calculated using the non-redundant parameter values. However, limiting parameter values in this way is potentially problematic in regards to generalisability, or how well a model is able to replicate empirical responses across a range of different experimental situations and data sets. For example, in Study 2 the Isotropic/Correlation model was implemented using a spatial filter parameter range of $s = 1$ to 6. In contrast to this, for Study 3 the data suggested that in order to adequately account for the empirical responses the model needed to be implemented using a spatial filter parameter range of $s = 1$ to 17. In other words, the initial range of parameter values badly underestimated the range needed by the model to generalise to new data sets.

One possible solution to this would be to implement all of the models across as wide a range of parameter values as possible. However, because there is no uniformity across models in terms of type or number of parameters there is no way to ensure that the models will be equally penalised by an approach such as this. Furthermore, adding complexity to the models in this way would most probably severely limit the interpretability of the simulation results. An alternative (and arguably preferable) approach would be to ensure that the models are compared across as wide a range of data sets as is practically possible. This of course does not rule out the possibility that the range of parameter values would need to be adjusted in light of any subsequent comparisons. Nonetheless, the larger the pool of empirical data sets that are included in the comparison process, the more confident we can be that the range of parameter values is an accurate reflection of the model's true non-redundant parameter range.

The second methodological issue relates to the high computational demands of Bayesian model selection. As was discussed in Study 2, in order to obtain an accurate indication of a model's true marginal likelihood it is necessary to implement the model at every possible parameter value within the chosen parameter range. For models with a limited number of discrete parameterisations (such as the Voronoi, and Barlow and Reeves models) this is a

simple matter to achieve. However, for the Isotropic and Oriented models it could be assumed that the underlying variable (i.e., the spread function of the filter) is either continuously or close to continuously distributed. In this case, obtaining model fits for every possible parameter value becomes a non-trivial computational problem. This problem is made even more difficult when the implementation of particular models is in itself computationally demanding, for example the Oriented/Alignment model could take more than 24 hours to complete a single experiment. Although an attempt was made to increase the number of incremental steps over which the models were implemented the possibility still exists that the marginal likelihoods were under- or overestimated. This is highly problematic, especially when the margin of difference between the two most likely models is as slim as was the case in Study 3.

These last two issues pose some difficulties for Bayesian model selection as a methodology. However, on the whole, the positive benefits of this approach appear to out-weigh these difficulties. At the very least it is certainly preferable to purely qualitative comparisons, which are overly reliant upon subjective interpretation. Furthermore, given the number of models and experiments featured across the three studies, it is difficult to imagine how a qualitative approach could possibly keep track (and make sense) of all of the comparisons. Finally, it should be noted that while the current model comparisons were solely concerned with models of symmetry detection, there is no reason why a similar approach could not be applied to the comparison of models of other cognitive and perceptual phenomena.

8.5. Conclusions

We can safely conclude that none of the models under comparison in this thesis represents the true process underlying symmetry perception in the human visual system. This is not really surprising given that the models are simplified versions of what is arguably a complicated (but well calibrated) system. However, as a number of researchers have suggested, the true utility of a model should not be defined by how well it fits the data, but by what it can reveal about the phenomenon of interest (Hintzman, 1991, etc; Jacobs & Grainger, 1994; Lewandowsky, 1993). In this sense at least, the models have proven to be useful in a number of ways.

In the first place, past empirical research has indicated that there are various characteristics associated with symmetry perception (e.g., a reasonably high tolerance to distortion or

interference from added noise), and the models represent specific theories seeking to explain these characteristics. It has been demonstrated that none of the models was able to provide sufficient explanations for all of the characteristics of interest that were focussed on in Studies 1 to 3. Although the model comparisons could not indicate what the processes underlying symmetry perception *are*, we now have a better idea of what they *are not*. This may seem to be of little help, but it must be remembered that the models are (for the most part) theoretically motivated and represent plausible potential explanations. In other words, the models are not just arbitrary, but are based on our current understanding of visual processing. Demonstrating that specific models are insufficient helps to constrain the pool of potential explanations, and therefore adds to this understanding.

Second, although the current models are insufficient, in some cases their specific failings point towards potential future models. For example, adding a spatially limited detection region to the Isotropic/Correlation model improved its descriptive adequacy and generality. Furthermore, the results of Study 3 suggest that future modifications should focus upon reducing the model's high complexity by using stimulus density to constrain the range of its detection region widths.

Third, it has been argued that the characteristics of certain models may be informative in regards to other models of symmetry perception. For example, Figure 8.2 demonstrates that it may be possible to explain the bootstrap model (Wagemans et al., 1993) in terms of a Voronoi representation. Similarly, the success of the Barlow and Reeves model in regards to distorted symmetry suggests that the visual system may precede symmetry detection with some form of clustering process. It is possible that a future version of the Voronoi model might benefit from an initial filtering stage to obtain clusters of points such as was demonstrated in Figure 8.7.

Fourth, the results of the model comparisons are informative in regards to indicating the need for specific new empirical studies. For example, the two experiments in Study 3 were designed as a test of the theoretical behaviour of the Voronoi model in regards to tolerance for distortion or interference from added noise points across different stimulus densities. Similarly, in a future study it would be useful to replicate the Jenkins (1982) experiments across a range of stimulus densities as this would provide useful insight into the relationship between stimulus density and the detection region, and provide a direct test of the Voronoi model's apparent scale invariance.

It must be remembered that the models under comparison in this thesis are only a small sub-set of all of the previously published models of symmetry perception. A limitation of the current research is that any conclusions that are drawn in regards to the likelihood of a particular model can only be interpreted relative to the other models under comparison. Obviously, therefore, in future studies it would be highly useful (and interesting) to expand the set of models included in the comparisons.

In conclusion, it is clear that the current research falls far short of the goal of elucidating the processes underlying symmetry perception in the human visual system. Nonetheless, it has arguably been successful in demonstrating the utility of Bayesian model selection as a theoretically sound and practically achievable means of quantifying the relative utility of rival models of symmetry perception. Furthermore, the results of the model comparisons have increased our understanding of symmetry perception and indicated future directions for the development of new models and new empirical studies. Although we are still a long way off from a full understanding of this phenomenon, it is hoped that the research presented in this thesis has moved us a little closer towards this goal.

References

Ahuja, N. (1982). Dot Pattern Processing Using Voronoi Neighborhoods. *Ieee Transactions on Pattern Analysis and Machine Intelligence, 4*, 336-343.

Ahuja, N., & Tuceryan, M. (1989). Extraction of Early Perceptual Structure in Dot Patterns - Integrating Region, Boundary, and Component Gestalt. *Computer Vision Graphics and Image Processing, 48*, 304-356.

Akaike, H. (1983). Information measures and model selection. *Bulletin of the International Statistical Institute, 50*, 277-290.

Aurenhammer, F. (1991). Voronoi diagrams - A survey of a fundamental geometric data structure. *ACM Computing Surveys, 23*, 345-405.

Barlow, H. (2001). The exploitation of regularities in the environment by the brain. *Behavioral and Brain Sciences, 24*, 602-607.

Barlow, H., & Reeves, B. C. (1979). The versatility and absolute efficiency of detecting mirror symmetry in random dot displays. *Vision Research, 19*, 783-793.

Barrett, B. T., Whitaker, D., McGraw, P. V., & Herbert, A. M. (1999). Discriminating mirror symmetry in foveal and extra-foveal vision. *Vision Research, 39*, 3737-3744.

Baylis, G. C., & Driver, J. (1994). Parallel computation of symmetry but not repetition within single visual shapes. *Visual Cognition, 1*, 377-400.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review, 94*, 115-147.

Blum, H. (1973). Biological shape and visual science. *Journal of Theoretical Biology, 38*, 205-287.

Bonneh, Y., Reisfeld, D., & Yeshurun, Y. (1996). Quantification of local symmetry: application to texture discrimination. In C. W. Tyler (Ed.), *Human Symmetry Perception* (pp. 303-318). Utrecht: VSP.

Broadbent, D. (1987). Simple models for experimental situations. In P. Morris (Ed.), *Modeling Cognition* (pp. 169-185). London: Wiley.

Browne, M. W., & Cudeck, R. C. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*, 230-258.

Burbeck, C. A., & Pizer, S. M. (1995). Object representation by Cores: Identifying and representing primitive spatial regions. *Vision Research, 35*, 1917-1930.

Burbeck, C. A., Pizer, S. M., Morse, B. S., & Ariely, D. (1996). Linking object boundaries at scale: a common mechanism for size and shape judgements. *Vision Research, 36*, 361-372.

Caelli, T. M. (1981). Some Psychophysical Determinants of Discrete Moire Patterns. *Biological Cybernetics, 39*, 97-103.

Caelli, T. M., & Julesz, B. (1978). On perceptual analysers underlying visual texture discrimination: Part 1. *Biological Cybernetics, 28*, 167-175.

Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology, 197*, 551-566.

Carlin, B. P., Kass, R. E., Lerch, F. J., & Hugenard, B. R. (1992). Predicting working memory failure: A subjective Bayesian approach to model selection. *Journal of the American Statistical Association, 87*, 319-327.

Carmody, D. P., Nodine, C. F., & Locher, P. J. (1974). Global detection of symmetry. *Perceptual Motor Skills, 45*, 1267-1273.

Chater, N., & Brown, G. D. A. (1999). Scale-invariance as a unifying principle. *Cognition, 69*, 17-24.

Compton, B. J., & Logan, G. D. (1993). Evaluating a computational model of perceptual grouping by proximity. *Perception & Psychophysics, 53*, 403-421.

Dakin, S. C. (1997). The Detection of Structure in Glass Patterns: Psychophysics and Computational Models. *Vision Research, 37*, 2227-2246.

Dakin, S. C. (1999). Orientation variance as a quantifier of structure in texture. *Spatial Vision, 12*, 1-30.

Dakin, S. C., & Herbert, A. M. (1998). The spatial region of integration for visual symmetry detection. *Proceedings of the Royal Society of London Series B-Biological Sciences, 265*, 2455-2455.

Dakin, S. C., & Hess, R. F. (1997). The spatial mechanisms mediating symmetry perception. *Vision Research, 37*, 2915-2930.

Dakin, S. C., & Watt, R. J. (1996). Detection of bilateral symmetry using spatial filters. In C.W.Tyler (Ed.), *Human Symmetry Perception and its Computational Analysis* (pp. 187-207). Utrecht: VSP.

Driver, J., Baylis, G. C., & Rafal, R. D. (1992). Preserved figure-ground segregation and symmetry in visual neglect. *Nature, 360*, 73-75.

Dry, M. (2005). *Mirror Symmetry Perception: Effect of Proximity to the Central Axis.* Paper presented at the XXXVII Annual Conference of the Cognitive Science Society, Stresa, Italy.

Dry, M., Lee, M. D., Vickers, D., & Hughes, P. (in press). Human Performance on Visually Presented Traveling Salesperson Problems with Varying Numbers of Nodes. *Journal of Problem Solving.*

Dry, M., Vickers, D., Lee, M., D., & Hughes, P. (submitted). The role of nearest neighbors in the perception of Glass patterns.

Earle, D. C. (1991). Some observations on the perception of Marroquin patterns. *Perception, 20*, 727-731.

Elliot, P. B. (1964). Appendix 1-Tables of d'. In J. A. Swets (Ed.), *Signal Detection and Recognition by Human Observers* (pp. 201-220). New York: John Wiley & Sons Inc.

Fabbri, R., Estrozi, L. F., & Costa, L. F. (2002). On Voronoi Diagrams and Medial Axes. *Journal of Mathematical Imaging and Vision, 17*, 27-40.

Gabriel, K. R., & Sokal, R. R. (1969). A new statistical approach to geographic variation analysis. *Systematic Zoology, 18*, 259-278.

Glass, L. (1969). Moire Effect from Random Dots. *Nature, 223*, 578-580.

Glass, L., & Perez, R. (1973). Perception of Random Dot Interference Patterns. *Nature, 246*, 360-362.

Graham, S. M., Joshi, A., & Pizlo, Z. (2000). The traveling salesman problem: A hierarchical model. *Memory & Cognition, 28*, 1191-1204.

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: John Wiley and Sons, Inc.

Gurnsey, R., Herbert, A. M., & Kenemy, J. (1998). Bilateral symmetry embedded in noise is detected accurately only at fixation. *Vision Research, 38*, 3795-3803.

Herbert, A. M., & Humphery, G. K. (1996). Bilateral symmetry detection: testing a 'callosal' hypothesis. *Perception, 25*, 463-480.

Hintzman, D. L. (1991). Why are formal models useful in psychology? In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. MurdocK.* Hillsdale: NJ: Lawrence Erlbaum Associates.

Hong, S., & Pavel, M. (1996). Determinants of symmetry perception. In C. W. Tyler (Ed.), *Human Symmetry Perception* (pp. 135-155). Utrecht: VSP.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's striate cortex. *Journal of Physiology, 160*, 106-154.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology, 195*, 215-243.

Jacobs, A. M., & Grainger, J. (1994). Models of Visual Word Recognition- Sampling the State of the Art. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 1311-1334.

Jarvis, R. A. (1972). Clustering Using a Similarity Based on Nearest Neighbours: Visual Image Experiments. In J. F. O'Callaghan (Ed.), *Pictorial Organisation and Shape* (pp. 90-97). Canberra: CSIRO Division of of Computing Research.

Jeffreys, H. (1961). *Theory of Probability*. Oxford, UK: Oxford University Press.

Jenkins, B. (1982). Redundancy in the perception of bilateral symmetry in dot textures. *Perception & Psychophysics, 32*, 171-177.

Jenkins, B. (1983a). Component processes in the perception of bilaterally symmetric dot textures. *Perception & Psychophysics, 34*, 433-440.

Jenkins, B. (1983b). Spatial Limits to the Detection of Transpositional Symmetry in Dynamic Dot Textures. *Journal of Experimental Psychology, 9*, 258-269.

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

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*, 773-796.

Kimia, B. B. (2003a). On the medial geometry in human vision. *Journal of Physiology- Paris, 97*, 155-190.

Kimia, B. B. (2003b). On the role of medial geometry in human vision. *Journal of Physiology- Paris, 97*, 155-190.

Koehler, W. (1929). *Gestalt Psychology*. New York: Liveright.

Koffka, K. (1935). *Principles of Gestalt Psychology*. New York: Harcourt.

Kovacs, I., Feher, A., & Julesz, B. (1998). Medial-point description of shape: a representation for action coding and its psychophysical correlates. *Vision Research, 38*, 2323-2333.

Kovacs, I., & Julesz, B. (1993). A closed curve is much more than an incomplete one. *Proceedings of the National Academy of Sciences, 90*, 7495-7497.

Kovacs, I., & Julesz, B. (1994). Perceptual sensitivity maps within globally defined shapes. *Nature, 370*, 644-646.

Kubovy, M., Holcombe, A., & Wagemans, J. (1998). On the lawfulness of grouping by proximity. *Cognitive Psychology, 35*, 71-98.

Kuffler, S. W. (1953). Discharge patterns and functional organisation of mammalian retina. *Journal of Neurophysiology, 16*, 37-68.

Kurbat, M. A. (1996). A network model for generating differential symmetry axes of shapes via receptive fields. In C. W. Tyler (Ed.), *Human Symmetry Perception* (pp. 227-236). Utrecht: VSP.

Labonte, F., Shapira, Y., Cohen, P., & Faubert, J. (1995). A Model for Global Symmetry Detection in Dense Images. *Spatial Vision, 9*, 33-55.

Lee, M. D. (2001). On the Complexity of Additive Clustering Models. *Journal of Mathematical Psychology, 45*, 131-148.

Lee, M. D. (2004). A Bayesian analysis of retention functions. *Journal of Mathematical Psychology, 48*, 310-321.

Lee, M. D., & Webb, M. R. (2005). Modeling individual difference in cognition. *Psychonomic Bulletin & Review, 12*, 605-621.

Lee, T., Mumford, D., Romero, R., & Lamme, V. A. F. (1998). The role of the primary cortex in higher level vision. *Vision Research, 38*, 2429-2454.

Lewandowsky, S. (1993). The Rewards and Hazards of Computer Simulations. *Psychological Science, 4*, 236-243.

Leyton, M. (1992). *Symmetry, Causality, Mind.* Cambridge: Bradford Books.

Li, W., & Westheimer, G. (1997). Human Discrimination of the Implicit Orientation of Simple Symmetrical Patterns. *Vision Research, 37*, 565-572.

Locher, P. J., & Wagemans, J. (1993). Effects of element type and spatial grouping on symmetry detection. *Perception, 22*, 565-587.

Logan, G. D. (1996). The CODE Theory of Visual Attention: An Integration of Space-Based and Object-Based Attention. *Psychological Review, 103*, 603-649.

MacGregor, J. N., Chronicle, E. P., & Ormerod, T. C. (2004). Convex hull or crossing avoidance? Solution heuristics in the traveling salesperson problem. *Memory & Cognition, 32*, 260-270.

MacGregor, J. N., & Ormerod, T. (1996). Human performance on the traveling salesman problem. *Perception & Psychophysics, 58*, 527-539.

MacGregor, J. N., Ormerod, T. C., & Chronicle, E. P. (2000). A model of human performance on the traveling salesperson problem. *Memory & Cognition, 28*, 1183-1190.

Mach, E. (1886/1959). *The analysis of sensations and the relation to the psychical.* New York: Dover.

Maloney, R. K., Mitchison, G. J., & Barlow, H. B. (1987). Limit to the detection of Glass patterns in the presence of noise. *Journal of the Optical Society of America, 4*, 2336-2341.

Marr, D. (1982). *Vision*. San Fransisco: W.H. Freeman and Company.

Marroquin, J. L. (1976). *Human Visual Perception of Structure.* Unpublished M.S. thesis, MIT.

Moller, A. P. (1993). Patterns of fluctuating asymmetry in sexual ornaments predict female choice. *Journal of Evolutionary Biology, 6*, 481-491.

Moller, A. P. (1995). Bumblebee preference for symmetrical flowers. *Proceedings of the national Academy of Sciences of the United States of America, 92*, 2288-2292.

Moore, D. J. H., & Parker, D. J. (1974). Analysis of global pattern features. *Pattern Recognition, 6*, 149-164.

Moore, E. F. (1956). Gedanken-experiments on sequential machines. In C. E. Shannon & J. McCarthy (Eds.), *Automata studies*. Princeton, NJ: Princeton University Press.

Myung, I. J. (2003). Tutorial On Maximum Likelihood Estimation. *Journal of Mathematical Psychology, 47*, 90-100.

Myung, I. J., Karabatsos, G., & Iverson, G. J. (2005). A Bayesian approach to testing decision making axioms. *Journal of Mathematical Psychology, 49*, 205-225.

Myung, I. J., & Pitt, M. A. (1997). Applying Occams razor in modelling cognition: A Bayesian approach. *Psychonomic Bulletin and Review, 4*, 79-95.

Navarro, D. J., & Lee, M. (2004). Common and distinctive features in stimulus representation: A modified version of the contrast model. *Psychonomic Bulletin & Review, 11*, 961-974.

Oka, s., van Tonder, G. J., & Ejima, Y. (2001). A VEP study on visual processing of figural geometry. *Vision Research, 41*, 3791-3803.

Osorio, D. (1996). Symmetry detection by categorisation of spatial phase, a model. *Proceedings of the Royal Society of London Series B-Biological Sciences, 263*, 105-110.

Palmer, S. E. (1985). The Role of Symmetry in Shape Perception. *Acta Psychologica, 59*, 67-90.

Palmer, S. E., & Hemenway, K. (1978). Orientation and Symmetry: Effects of Multiple, Rotational, and Near Symmetry. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 691-702.

Palmer, S. E., & Rock, I. (1994). On the nature and order of organizational processing: A reply to Peterson. *Psychonomic Bulletin & Review, 1*, 515-519.

Papathomas, T. V., Kovacs, I., Gorea, A., & Julesz, B. (1995). A unified approach to the perception of motion, stereo, and static-flow patterns. *Behavior Research Methods, Instruments, & Computers, 27*, 419-432.

Pashler, H. (1990). Coordinate Frame for Symmetry Detection and Object Recognition. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 150-163.

Pintsov, D. (1989). Invariant pattern recognition, symmetry, and Radon transforms. *Journal of the Optical Society of America, 6*, 1544-1554.

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *TRENDS in Cognitive Sciences, 6*, 421-425.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Towards a method of Selecting Among Computational Models of Cognition. *Psychological Review, 109*, 472-491.

Pizlo, Z., Salach-Golyska, M., & Rosenfeld, A. (1997). Curve Detection in a Noisy Image. *Vision Research, 37*, 1217-1241.

Pomerantz, J. R. (1981). Perceptual Organisation in Information Processing. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual Organisation* (pp. 141-180). NJ: Lawrence Erlbaum.

Rainville, S. J. M., & Kingdom, F. A. A. (1999). Spatial scale contribution to the detection of mirror symmetry in fractal noise. *Journal of the Optical Society of America A, 16*, 2112-2123.

Rainville, S. J. M., & Kingdom, F. A. A. (2000). The functional role of oriented spatial filters in the perception of mirror symmetry - psychophysics and modeling. *Vision Research, 40*, 2621-2644.

Rainville, S. J. M., & Kingdom, F. A. A. (2002). Scale invariance is driven by stimulus density. *Vision Research, 42*, 351-367.

Ratcliff, R., & Smith, P. L. (2004). A Comparison of Sequential Sampling Models for Two-Choice Reaction Time. *Psychological Review, 111*, 333-367.

Roberts, S., & Pashler, H. (2000). How Persuasive Is a Good Fit? A Comment on Theory Testing. *Psychological Review, 107*, 358-367.

Rouder, J. N., Lu, J., Speckman, P., Sun, D. H., & Jiang, Y. (2005). A hierarchical model for estimating response time distributions. *Psychonomic Bulletin & Review, 12*, 195-223.

Royer, F. L. (1981). Detection of Symmetry. *Journal of Experimental Psychology-Human Perception and Performance, 7*, 1186-1210.

Sally, S., & Gurnsey, R. (2001). Symmetry detection across the visual field. *Spatial Vision, 14*, 217-234.

Sasaki, Y., Vanduffel, W., Knutsen, T., Tyler, C. W., & Tootell, R. (2005). Symmetry activates extrastriate visual cortex in human and non-human primates. *Proceedings of the national Academy of Sciences of the United States of America, 102*, 3159-3163.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461-464.

Scognamillo, R., Rhodes, G., Morrone, C., & Burr, D. C. (2003). A feature-based model of symmetry detection. *Proceedings of the Royal Society of London. B., 27-0*, 1723-1733.

Siddiqi, K., Kimia, B. B., Tannenbaum, A., & Zucker, S. W. (2001). On the psychophysics of the shape triangle. *Vision Research, 41*, 1153-1178.

Smits, J. T. S., & Vos, P. G. (1986). A Model for the Perception of Curves in Dot Figures: The Role of Local Salience of "Virtual Lines". *Biological Cybernetics, 54*, 407-416.

Stanislow, H., & Todorov, N. (1999). Calculation of signal detection measures. *Behavior Research Methods, Instruments, & Computers, 31*, 137-149.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*, 173-180.

Stevens, K. A. (1978). Computation of Locally Parallel Structure. *Biological Cybernetics, 29*, 19-28.

Stevens, K. A., & Brookes, A. (1987). Detecting Structure by Symbolic Constructions of Tokens. *Computer Vision, Graphics, And Image Processing, 37*, 238-260.

Tek, H., & Kimia, B. B. (2003). Symmetry Maps of Free-Form Curve Segments via Wave Propagation. *International Journal of Computer Vision, 54*, 35-81.

Tjan, B. S., & Liu, Z. (2005). Symmetry impedes symmetry discrimination. *Journal of Vision, 5*, 888-900.

Toussaint, G. T. (1980). The Relative Neighborhood Graph of a Finite Planar Set. *Pattern Recognition, 12*, 261-268.

Tripathy, S. P., Mussap, A. J., & Barlow, H. B. (1999). Detecting collinear dots in noise. *Vision Research, 39*, 4161-4171.

Tummers, B., van der Laan, J., & Huyser, K. (2005). Data Thief III (Version 1.0).

Tyler, C. W., & Hardage, L. (1996). Mirror symmetry detection: predominance of second-order processing throughout the visual field. In C. W. Tyler (Ed.), *Human Symmetry Perception and its Computational Analysis* (pp. 151-171). Utrecht: VSP.

Tyler, C. W., Hardage, L., & Miller, R. T. (1995). Multiple Mechanisms for the Detection of Mirror Symmetry. *Spatial Vision, 9*, 79-100.

Ullman, S. (1984). Visual Routines. *Cognition, 18*, 97-159.

Uttal, W. R. (1990). On some two-way barriers between models and mechanisms. *Perception & Psychophysics, 48*, 188-203.

van der Helm, P. A., & Leeuwenberg, E. L. J. (1996). Goodness of visual regularities: A nontransformational approach. *Psychological Review, 103*, 429-456.

van Oeffelen, M. P., & Vos, P. G. (1982). Configurational effects on the enumeration of dots: Counting by groups. *Memory & Cognition, 10*, 396-404.

van Oeffelen, M. P., & Vos, P. G. (1983). An Algorithm for Pattern Discrimination on the Level of Relative Proximity. *Pattern Recognition, 16*, 341-348.

van Rooij, I., Stege, U., & Schactman, A. (2003). Convex hull and tour crossings in the Euclidean traveling salesperson problem: Implications for human performance studies. *Memory & Cognition, 31*, 215-220.

van Tonder, G. J., & Ejima, Y. (2000a). From image segmentation to anti-textons. *Perception, 29*, 1231-1247.

van Tonder, G. J., & Ejima, Y. (2000b). The patchwork engine: image segmentation from shape symmetries. *Neural Networks, 13*, 291-303.

Vickers, D., Bovet, P., Lee, M. D., & Hughes, P. (2003). The perception of minimal structures: Performance on open and closed versions of visually presented Euclidean travelling salesperson problems. *Perception, 32*, 871-886.

Vickers, D., Lee, M. D., Dry, M., & Hughes, P. (2003). The roles of the convex hull and the number of potential intersections in performance on visually presented traveling salesperson problems. *Memory & Cognition, 31*, 1094-1104.

Vickers, D., Lee, M. D., Dry, M., Hughes, P., & McMahon, J. A. (2006). The aesthetic appeal of minimal structures: judging the attractiveness of solutions to traveling salesperson problems. *Perception & Psychophysics, 68*, 32-42.

Vickers, D., Mayo, T., Heitmann, M., Lee, M., & Hughes, P. (2004). Intelligence and individual differences in performance on three types of visually presented optimisation problems. *Personality and Individual Differences, 36*, 1059-1071.

Vickers, D., Preiss, K., & Hughes, P. (submitted). The role of nearest neighbours in the perception of structure and motion in dot patterns.

Wagemans, J. (1993). Skewed Symmetry - a Nonaccidental Property Used to Perceive Visual Forms. *Journal of Experimental Psychology-Human Perception and Performance, 19*, 364-380.

Wagemans, J. (1996). Detection of visual symmetries. In C. W. Tyler (Ed.), *Human Symmetry Perception and its Computational Analyses* (pp. 25-48). Utrecht: VSP.

Wagemans, J. (1997). Characteristics and models of human symmetry detection. *Trends in Cognitive Sciences, 1*, 346-352.

Wagemans, J. (1999). Toward a better approach to goodness: Comments on Van der Helm and Leeuwenberg (1996). *Psychological Review, 106*, 610-621.

Wagemans, J., Van Gool, L., & d'Ydewalle, G. (1990). Visual search in dot-patterns with bilateral and skewed symmetry. In D. Brogan (Ed.), *Visual Search*. London: Taylor & Francis.

Wagemans, J., Van Gool, L., & d'Ydewalle, G. (1991). Detection of symmetry in tachiscopically presented dot patterns: Effects of multiple axes and skewing. *Perception & Psychophysics, 50*, 413-427.

Wagemans, J., Van Gool, L., Swinnen, V., & Van Horebeek, J. (1993). Higher-Order Structure in Regularity Detection. *Vision Research, 33*, 1067-1088.

Wagemans, J., Vangool, L., & Dydewalle, G. (1992). Orientational Effects and Component Processes in Symmetry Detection. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 44*, 475-508.

Watson, P. J., & Thornhill, R. (1994). Fluctuating asymmetry and sexual selection. *Trends in Ecology and Evolution, 9*, 21-25.

Watt, R. J. (1988). *Visual Processing: Computational, Psychophysical and Cognitive Research*. London: Lawrence Erlbaum Associates.

Watt, R. J. (1991). *Understanding Vision*. London: Academic Press.

Wenderoth, P. (1994). The Salience of Vertical Symmetry. *Perception, 23*, 221-236.

Wenderoth, P. (1995). The Role of Pattern Outline in Bilateral Symmetry Detection with Briefly Flashed Dot Patterns. *Spatial Vision, 9*, 57-77.

Wenderoth, P. (1996a). The effects of dot pattern parameters and constraints on the relative salience of vertical bilateral symmetry. *Vision Research, 36*, 2311-2320.

Wenderoth, P. (1996b). The effects of the contrast polarity of dot-pair partners on the detection of bilateral symmetry. *Perception, 25*, 757-771.

Wenderoth, P. (1996c). The role of pattern outline in bilateral symmetry detection with briefly flashed dot patterns. In C.W.Tyler (Ed.), *Human Symmetry Perception Perception and its Computational Analysis* (pp. 49-69). Utrecht: VSP.

Wenderoth, P. (1997). The effects on bilateral-symmetry detection of multiple symmetry, near symmetry, and axis orientation. *Perception, 26*, 891-904.

Wertheimer, M. (1938). Laws of organisation in perceptual forms. In W. Ellis (Ed.), *A Source Book of Gestalt Psychology*. New York: Harcourt.

Westheimer, G. (1981). Visual hyperacuity. In H. Autrum (Ed.), *Progress in sensory physiology* (Vol. 1, pp. 1-30). New York: Springer.

Wilkinson, D., & Halligan, P. W. (2003). Stimulus symmetry affects the bisection of figures but not lines: evidence from event-related fMRI. *Neuroimage, 20*, 1756-1764.

Wilson, H. R., & Wilkinson, F. (1998). Detection of global structure in Glass patterns: implications for form vision. *Vision Research, 38*, 2933-2947.

Wilson, H. R., & Wilkinson, F. (2003). Further evidence for global orientation processing in circular Glass patterns. *Vision Research, 43*, 563-564.

Wilson, H. R., Wilkinson, F., & Assad, W. (1997). Rapid Communication: Concentric orientation summation in human form vision. *Vision Research, 37*, 2325-2330.

Zahn, C. (1971). Graph-Theoretical methods for Detecting and Describing Gestalt Structures. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS., 20*, 68-86.

Zhu, S. C., & Yuille, A. L. (1996). FORMS: A Flexible Object Recognition and Modelling Systems. *International Journal of Computer Vision, 20*, 187-212.