<table>
<thead>
<tr>
<th>Title</th>
<th>Basic aesthetic features and their influence on attention and performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Mulcahy, Paul</td>
</tr>
<tr>
<td>Publication Date</td>
<td>2015-01-23</td>
</tr>
<tr>
<td>Item record</td>
<td><a href="http://hdl.handle.net/10379/4908">http://hdl.handle.net/10379/4908</a></td>
</tr>
</tbody>
</table>
BASIC AESTHETIC FEATURES IN VISUAL PROCESSING
AND THEIR INFLUENCE ON
ATTENTION AND PERFORMANCE

Thesis submitted for the Degree of Doctor of Philosophy

Paul Mulcahy, BA (Psychology)

School of Psychology
National University of Ireland, Galway
Galway
February 2015
# TABLE OF CONTENTS

Abstract ........................................................................................................ iv
DeclarationAcknowledgments ...................................................................... vii
Acknowledgments ...................................................................................... viii
List of Works ................................................................................................ ix
List of Figures ............................................................................................... x
List of Tables ................................................................................................ xi
Chapter 1: Perceptual Issues in Philosophical and Empirical Aesthetics ........ 1
  1.1 Introduction .......................................................................................... 1
  1.2 Defining Aesthetics ............................................................................ 3
  1.3 Historical/Philosophical Approaches .................................................. 5
  1.4 Empirical Aesthetics ......................................................................... 14
  1.5 Summary ............................................................................................ 23
Chapter 2: The Golden Section as a Perceptual Phenomenon ..................... 25
  2.1 Introduction ....................................................................................... 25
  2.2 The Golden Section in Nature ............................................................ 26
  2.3 The Golden Section in Art and Design .............................................. 28
  2.4 Experimental aesthetics .................................................................... 30
  2.5 Theoretical Approaches .................................................................... 37
  2.6 Present Study .................................................................................... 43
  2.7 Experiment 2.1: Target Section Localisation in Mondrian-type Grids ... 45
  2.8 Experiment 2.2: Paired Comparisons of Ratio-Sectioned Mondrian Grids ... 55
  2.9 Experiment 2.3: The Influence of an Aesthetic Attitude on Paired
                   Comparisons of Ratio-Sectioned Mondrians .................................. 60
  2.10 General Discussion ......................................................................... 70
Chapter 3: The Golden Section and Perceptual Processing ......................... 73
  3.1 Introduction ....................................................................................... 73
  3.2 Experiment 3.1: Target Localisation in Equiluminant Mondrians ........ 73
  3.3 Natural Image Statistics as Constraints on Visual Processing .......... 80
  3.4 Experiment 3.2: The Influence of Retinal Spatial Frequency and Task
                   Specification on Target Localisation .............................................. 87
  3.5 Experiment 3.3: Golden Sectioning and Spatial Frequency Coding ...... 95
  3.6 General Discussion ......................................................................... 101
Chapter 4: Visual Balance and Search Performance ..................................... 105
  4.1 Introduction ....................................................................................... 105
  4.2 Physical and Phenomenal Balance ..................................................... 108
  4.3 Balance, Emergence, and Gestalt ....................................................... 108
  4.4 Balance and Perceptual Processing ................................................... 110
  4.5 Experiment 4.1: Visual search in an Axially-aligned Distracter Array .... 112
  4.6 Experiment 4.2: Visual Search in an Axially-misaligned Distracter Array . 126
  4.7 Experiment 4.3: Paired comparisons of Stimuli from Experiment 4.1 .... 137
  4.8 Discussion of Experiment 4.3 ............................................................ 139
  4.9 General Discussion ......................................................................... 139
Chapter 5: Visual Balance as an Influence on Orienting Preferences in Domestic
          Chicks (Gallus gallus domesticus). ...................................................... 143
  5.1 Introduction ....................................................................................... 143
  5.2 Evolution and Divergence of Amniote Vision .................................... 144
  5.3 Avian Aesthetics .............................................................................. 145
  5.4 Comparative Psychology and Visual Experience .................................. 146
Abstract

Introduction. Within experimental aesthetics it is often claimed that artists can exploit the normal activity of the perceptual system in order to achieve aesthetic effect. The present research extends this proposal by providing evidence to suggest that the elementary units of aesthetic experience are intrinsic to the normal functioning of the perceptual system. The practical implication of this proposal is that aesthetic relevance can be studied via its effects on perceptual performance instead of reliance on preference measures which may poorly reflect real-world experience. It also underscores the importance of research in experimental aesthetics as it proposes that we experience the world in a fundamentally qualitative manner. The strategy of the thesis was to select two phenomena with extensive previous research literature and avowed aesthetic relevance, namely visual balance and golden ratio sectioning, and apply visual search paradigms to investigate their effects on performance. Furthermore, and in order to follow up on results from findings relating to visual balance, a study involving domestic chicks (Gallus gallus domesticus) was conducted in order to test the proposal that the observer effects of this phenomenon in the human studies were due to perceptual organisation rather than the specific neural architecture of human observers.

Methods. The main experiments involved were based on visual search procedures, with the participant aiming to locate a target region within a randomised display of distracter elements of increasing set size. Response time was measured against set size in order to determine how the phenomenon of interest affects performance in relation to the attentional bottleneck. Preference measures were also implemented by using 2AFC comparisons of pairs of stimuli which allowed a unidimensional scale of preference to be established for the range of stimulus
categories. Further investigations were conducted by referring to the parameters of the stimulus arrangements; spatial anisotropy and inter-item structure in visual balance experiments, and fractal dimension and spatial frequency for golden sectioning. The domestic chick study was conducted by measuring orientation behaviour for balanced or imbalanced stimuli during the generalisation phase following visual category training.

**Results.** Detailed in Chapters 2 and 3, the results of the golden-section experiments revealed no special preference for golden-section stimuli, but the performance data revealed that perceptual processing of these stimuli appeared to be slowed relative to an otherwise linear function relating RT to ratio. Further experiments sought to determine the locus of the effect, and showed that the spatial frequency of the stimuli sectioned according to the golden ratio exhibited a lack of spatial frequency information in the band corresponding to highest sensitivity in the normal human contrast sensitivity function. In Chapter 4, two experiments are reported which suggest that visual balance affects the efficiency of target localisation in search, but balance itself depends on the complexity of the display; more populated displays define local structure via the arrangement of inhabiting items, while less populated displays affect performance via target alignment with the structural axes of the frame. Chapter 5 presents evidence from chicks which suggests that imbalance may affect stimulus encoding or the deployment of attention during visually-guided orientation.

**Conclusions.** The present research offers evidence from two main strands which demonstrates that if aesthetic experience is considered as an aspect of normal perception, then research paradigms normally associated with measuring the efficiency of processing can demonstrate effects with the potential for aesthetic value.
A lack of corollary effects in measures of preferences underscores the importance for empirical aesthetics to expand the repertoire of experimental procedures being used to examine putative aesthetic phenomena. Performance results are interpreted in terms of two frameworks; perceptual fluency theory and constraints on visual processing from natural scene statistics. The results reported here have relevance to visual aesthetics and to issues concerning perceptual organisation, particularly as it deals with the deployment of attention.
Declaration

The author certifies that this thesis is his own work, and he has not previously received a degree on the basis of this work at NUI, Galway or any other University.
Acknowledgments

This research received funding support from the Irish Research Council for the humanities and social sciences.

A huge amount of gratitude is due to Mark Elliott, whose support and knowledge were never far away, no matter where in the world he might have been.

I would like to thank my thesis committee members Caroline Heary and Ian Stewart for their help in keeping the degree on-track. I also want to add my name to the long list of postgraduate students who feel owe a debt of gratitude to Declan Coogan for his help in technical matters. I have met many wonderful people in the postgraduate community of the School of Psychology, but I must give special mention to Chris, Rob, Heike, Mark and Caithriona. For their help at various stages I would like to thank Sinéad, Julie, Dee, and Naomi. In Padua, many people helped and welcomed me, but I especially thank Lucia Regolin and Orsola Rosa-Salva for their combination of friendliness and scientific expertise.

I am hugely thankful to my family, especially Marie, Karen, and Séan, for helping me along the way. My children Kaelan and Luke helped keep me committed and inspired, and made me smile when I needed it.

More than anyone else, I am grateful to my wife Marion, whose encouragement, love, and support are the reasons this thesis exists.
List of Works

The following are presentations and previous publications which arose from work presented in this thesis:

Publications


Conference Presentations


Seminar Presentations

List of Figures

Figure 2.1 A logarithmic spiral in geometry and nature ................................................................. 27
Figure 2.2 Examples of stimuli used in Experiment 2.1 ................................................................. 48
Figure 2.3 Exponents of mean log-transformed RTs by set size and ratio for Experiment 2.1 ........ 52
Figure 2.4 Mean error percentages for each ratio for Experiment 1 ............................................. 53
Figure 2.5 Preference rating results for Experiment 2.2 ................................................................. 57
Figure 2.6 Example BTL score plots for selected participants in Experiment 2.2 ........... 59
Figure 2.7 Preference rating results for Experiment 2.3 ................................................................. 64
Figure 2.8 Exponents of mean log-transformed RT distributions by set size and ratio for Experiment 2.4 ......................................................................................................................... 67
Figure 2.9 Mean error percentages for each ratio for Experiment 2.4 ........................................ 69
Figure 3.1 Exponents of mean log-transformed RT distributions by set size and ratio for Experiment 3.1 .............................................................................................................................................. 77
Figure 3.2 Mean error percentages for each ratio for Experiment 3.1 ....................................... 78
Figure 3.3 Box counting dimension for ratio- sectioned Mondrian stimuli .................. 82
Figure 3.4 Power spectra of experimental stimuli resulting from spatial frequency analysis ................................................................................................................................................... 87
Figure 3.5 Exponents of mean log-transformed RTs by set size and ratio for Experiment 3.1 .............................................................................................................................................. 90
Figure 3.6 Mean error percentages for each ratio and distance for Experiment 3.2 .... 92
Figure 3.7 Example of noise-masked stimuli used in Experiment 3.3 ....................... 96
Figure 3.8 Exponents of mean log-transformed RTs by set size and ratio for Experiment 3.2 .............................................................................................................................................. 97
Figure 3.9 Mean error percentages for each ratio for Experiment 3.3 ......................... 98
Figure 4.1 Arnheim’s (1954) structural skeleton and some examples ................................................ 107
Figure 4.2 An example of stimuli used in Experiment 4.1 ......................................................... 113
Figure 4.3 Exponents of mean log-transformed RTs by set size and target position for Experiment 4.1 .............................................................................................................................................. 115
Figure 4.4 Exponents of the means of the log-transformed RTs by set size, position, and location in Experiment 4.1 ......................................................................................................................... 117
Figure 4.5 An example of stimuli used in Experiment 4.2 ......................................................... 127
Figure 4.6 Exponents of mean log-transformed RTs by set size and target position for Experiment 4.2 .............................................................................................................................................. 129
Figure 4.7 Exponents of mean log-transformed RTs by set size, position, and location for Experiment 4.2 .............................................................................................................................................. 131
Figure 4.9 Arcsine-transformed preference proportions by set size and target position for Experiment 4.3 .............................................................................................................................................. 139
Figure 5.1 Principal and medial axes of four-sided polygons ............................................... 152
Figure 5.2 Suzume Odori-zu (Dancing Sparrows) by Hokusai ...................................................... 154
Figure 5.3 Experimental stimuli used during (a) shaping, (b) training, and (c) testing phases in Experiment 5.1 .............................................................................................................................................. 155
Figure 5.4 Schematic diagram of the experimental apparatus for all experiments ........... 157
Figure 5.5 Boxplot of scores for each training category in Experiment 5.1 ................... 159
Figure 5.6 Experimental stimuli used during (a) shaping, (b) training, and (c) testing phases in Experiment 5.2 .............................................................................................................................................. 160
Figure 5.7 Boxplot of scores for each training category in Experiment 5.2 ................... 161
List of Tables

Table 2.1  Linear and curvilinear regression results for log-transformed means in Experiment 2.1 for the full range of ratios, and for the range excluding the golden ratio ................................................................. 51
Table 2.2  Coefficient of consistency (ζ) and mean circular triads (CTM) by participant for Experiment 2.2 ................................................................. 58
Table 2.3  Coefficients of consistency (ζ) and mean circular triads (CTM) by participant in Experiment 2.3 ................................................................. 64
Table 2.4  Linear and curvilinear regression results for log-transformed means in Experiment 2.3 for the full range of ratios and for the range excluding the golden ratio ................................................................. 68
Table 3.1  Linear and curvilinear regression results for log transformed mean RTs in Experiment 3.1 for the full range of ratios and for the range excluding the golden ratio ................................................................. 77
Table 3.2  Linear and curvilinear regression results for log-transformed mean RTs in Experiment 3.2 for both sessions ................................................................. 94
Table 3.3  Linear and curvilinear regression results for log-transformed RTs in Experiment 3.3 for the full range of ratios and for the range excluding the golden ratio ................................................................. 100
Table 4.1  Fixed and random effect estimates for response times as a function of Palmer-Guidi ratings in Experiment 4.1 ................................................................. 120
Table 4.2  Fixed and random effect estimates for response times as a function of mean inter-dot distance from target to distracters in Experiment 4.1 ................................................................. 123
Table 4.3  Model fit diagnostics for anti-logs of mean log-transformed response time by inter-dot distance from the target to distractor items for each set size in Experiment 4.1 ................................................................. 124
Table 4.4  Fixed and random effect estimates for response times as a function of Palmer-Guidi ratings in Experiment 4.2 ................................................................. 133
Table 4.5  Fixed and random effect estimates for response times as a function of inter-dot distance in Experiment 4.2 ................................................................. 135
Table 4.6  Model fit diagnostics for anti-logs of mean log-transformed response time by inter-dot distance from the target to distractor items for each set size in Experiment 4.2 ................................................................. 135
Chapter 1: Perceptual Issues in Philosophical and Empirical Aesthetics

1.1 Introduction

Artists have long exploited the fact that artworks can mimic the normal activity of the perceptual system to elicit an aesthetic effect (e.g. Mammasian, 2008). For example, artistic effect has been achieved by exploiting the way the visual system processes luminance and spatial scale in the work of Monet and Da Vinci (Livingstone, 2002), line orientations in the work of Mondrian and Malevich (Zeki, 1999), and surface reflectance in the work of Rubens (Cavanagh, Chao, & Wang, 2008). However, as general aesthetic principles these ideas have been largely speculative and often limited to specific artworks, so it remains an open question as to what perceptual-cognitive mechanisms or processes are employed in aesthetic experience more generally. According to Gregory (1995), aesthetic experience arises from the brain’s attempts to make probability statements (inferences) about the environment, an aptitude that is so intricate and evolutionarily advanced that Turner (2006) suggests that it represents a discontinuity between humans and all other life. This thesis provides evidence to contradict these claims, and advances the proposal that localising aesthetic experience entirely to higher-level cognition modularises the phenomenon in a way that is philosophically untenable, and at odds with empirical data.

In the present thesis, it is claimed that aesthetic experience is an intrinsic aspect of all perception, imbuing everything that is perceived with the potential for aesthetic evaluation. This is proposed to be the reason why aesthetic experience sometimes seems ineffable or deeply personal; its elementary units often exert their influence prior to attention or conscious awareness. The strategy reported here is to
assess the extent to which geometric features of a visual display that have putative aesthetic relevance may affect normal perceptual processing. Aesthetics considered in this way is taken to be a fundamental aspect of visual perception, and not due to the specific neural architecture of humans but to the nature of the visual process. While Gregory (1995) echoes Turner’s sentiment and suggests that “how sight-for-survival became adapted to seeing and creating beauty is anyone’s guess”, the research presented in this thesis provides some evidence for an adaptive relevance for visual features that affect early perceptual organisation prior to attention, the deployment of which is proposed to be an important aspect of aesthetic experience (e.g. Leder, Bär, & Topolinski, 2012; Leder, Belke, Oeberst, & Augustin, 2004; Molnar, 1987; Muth & Carbon, 2013; Tinio, 2013).

The present chapter presents some historical and theoretical information relating to the empirical study of visual aesthetics. Some of the major frameworks for philosophical aesthetics are briefly discussed in terms of the dichotomy between aesthetic and non-aesthetic experience, in order to situate the thesis as a challenge to this division. Following this, some discussion of the findings and theories from empirical aesthetics is given, with the aim of establishing and clarifying the research strategy followed in subsequent chapters. Chapters 2 and 3 examine the golden section as a geometric arrangement of proportions with a long history of being held as especially relevant to aesthetic experience. In partial support of these claims, evidence is presented wherein this proportion is shown to uniquely affect perceptual processing in comparison with other ratio sections. Chapter 4 follows on from this by implementing a similar methodology in the assessment of visual balance as a basic aesthetic feature. Chapter 5 provides an extension of these results in terms of non-human (avian) orientation behaviour, with the aim of investigating the extent to which
visual balance relates to perceptual organisation rather than inferential cognitive processes. In both cases, evidence is presented which supports the proposal that visual balance may be an intrinsic aspect of normal perceptual organisation which can influence attention. The final chapter offers a summary, interprets these findings in relation to philosophical and empirical aesthetics, and visual perception more generally, and describes some directions for future research. One of the most persistent objections to an empirical aesthetics approach is that there is not enough to conceptually differentiate normal perception or cognition (or the related neurology) from aesthetically-relevant psychological functions (e.g. Seeley, 2009). The outcomes from the research presented here offer some support for the proposal that aesthetic experience can derive from normal visual perception via the relationship between stimulus features and the visual system.

1.2 Defining Aesthetics

The fact that the term “aesthetic” is a nebulous one within experimental or philosophical aesthetics is not especially insightful, nor is it useful or optimistic for the scientific analysis of aesthetic phenomena. However, people seem to be able to easily and consistently identify a common referent phenomenology; in other words, we know it when we feel it (van Tonder & Spehar, 2013). Baumgarten (1750, 1758) introduced the word aesthetica, translated as aesthetic, to refer to a distinction between things that we know and things that we perceive (Benjafield, 2010). The clearest explanation of what this entails is to consider the etymological antonym anaesthetic – aesthetic events are those which are strikingly felt, while those that are experienced in a dulled manner are anaesthetic, and this is best considered as a continuous variable (van Tonder & Spehar, 2013). It is now often used to refer to a judgment, property, attitude, experience, pleasure, or value (Goldman, 2005). In this
thesis, most of these are treated as different aspects of aesthetic experience that are defined by their usage and context. An aesthetic experience is one in which the interaction of an observer with an object or a property of an object (or event) enables a hedonic response for the observer that is signed by a phenomenal quality. This experience may give pleasure to the observer, and allow them to make a judgement that ascribes some value to the object of the experience, or it may guide the behaviour of the observer in a non-conscious way that is nonetheless based on qualitative aspects of the visual scene.

The following sections present some historical discussion of philosophical aesthetics, the aim of which is provide background for the empirically-oriented operational definition adopted in the studies which follow. This definition is largely in accordance with a preference-based conceptualisation of aesthetic experience, grounded in a functional relationship between an aspect of the environment and the observer (Fechner, 1876). While much of the writing on aesthetics has focused on art (because art involves a deliberate attempt to provide aesthetic experience), aesthetics and art are not treated as synonyms because people may have aesthetic experiences in situations outside of the viewing of an artwork. Similarly, an artwork may enable an experience that is labelled as aesthetic by one person, but no such experience might be available for another person in connection with the same artwork. Shimamura (2012) suggests that we consider an aesthetic experience as any hedonic preference response to a sensory experience. While the phenomenon can therefore apply to any event, artworks are those objects or events that are solely intended to provoke a hedonic experience.

---

1 The experimental aesthetics approach, and particularly Fechner’s foundational methodology, is not a theory in its own right, but problems arise when it is mistaken for one. Accordingly, this thesis does not propose that all of aesthetic experience is explained by the kinds of phenomena that can be analysed in an experimental situation (i.e. that social or cultural context has no influence on aesthetic experience). This strawman depiction of the experimental approach is as obviously invalid as the hard claim that there are no invariants in aesthetic experience (a claim that has been contradicted by over a century of empirical data).
response. In a major review of experimental aesthetics, Palmer, Schloss and Sammartino (2013) suggest that every experience has an aesthetic aspect, but some situational factors bring it into attentional awareness and subsequently affect behaviour in the way described. These circumstances include the extremity of the aesthetic response itself (seeing something very beautiful), the context (aesthetic attitude), or by explicit awareness of the potential for such responses (being told to look for beauty in a stimulus in an experiment). This definition allows that preferences are subjective and an entirely normative definition is not possible, but the circumstances may be manipulated to facilitate measurement of aesthetically-relevant responses. By situating aesthetic experience as a continuous variable, the difficulty of defining it relative to normal experience is then obviated in favour of something like a psychophysical threshold (Fechner, 1876). This problem, differentiating aesthetic from quotidian experience, is not a trivial one historically. However, this thesis argues that the distinction is unnecessary.

1.3 Historical/Philosophical Approaches

The empirical aesthetics approach is sometimes characterised as being ahistorical, or that it eschews sociocultural factors in formulating hypotheses and theories (e.g. Bullot & Reber, 2013). However, a brief survey of some of the philosophical perspectives may show that such a criticism fails to account for the divisiveness in philosophical aesthetics, or for the extent to which empirical aesthetics is a conscious and deliberate attempt to address some of these issues. Philosophical approaches may themselves to be too delimited in their approach; non-art or non-Western art styles are seldom included, and this ignores qualitative similarities between these realms. The empirical aesthetics approach may appear ahistorical because it doesn’t usually integrate with a culturally-specified set of variables.
However, it can also be seen to be the result of some tightly constraining theoretical commitments. As mentioned, the following sections outline the conceptual development of one of the major objections to perceptually-based aesthetics approaches – its supposed failure to account for the dichotomy between normal and aesthetic perceptual experience (Seeley, 2006). In many of the following sections, examples of aesthetic experience derive from art. This is reflective of a general emphasis in the literature. As mentioned in subsequent discussion, the emphasis on art is taken to undermine the usefulness of any theories so constrained.

1.3.1 Mimesis

The mimetic approach contends that art is a mirror of reality, and the success of an artwork depends on its proximity to realistic depiction. This idea is attributed in its earliest known form to Plato (in *Phaedrus*), who held art to be an impoverished and imperfect reflection of nature. Ideal forms cannot be realised by humans, so the arts are distant from ideals. For example, a carpenter’s bed is a copy of an ideal, and a painting of it is even worse than a copy of this copy, as it lacks the functionality of being a bed. Plato furthermore felt that art stirs the emotions, which is deleterious to rationality, which obligated him and his followers to adopt an antagonistic attitude to art. Aristotle (in *The Poetics*) disagreed with Plato’s interpretation, taking the stance that art is mimetic but also possesses a beneficial didactic role, and represents a natural and real form of pleasure. He claimed that art can depict certain essentials of human experience and that we can learn from the mistakes and lessons of fictional people. The Western tradition of philosophy of aesthetics took the concept of mimesis as a normative principle on board without its ethical implications by adopting the Aristotelian modification of Plato’s approach (Shimamura, 2012).
The mimetic approach dominated the philosophy of aesthetics for much of the time between Plato and the 19th Century, and gained much of its prominence during the Renaissance, with the advent of technical skills that allowed convincingly real depictions of the world. For instance, artists studied the behaviour of light and how it reflects in the environment, formulated new and more realistic pigments and painting materials, and developed chiaroscuro ("light/dark") to enhance three-dimensionality through the use of shading. The parallels between these techniques and visual processing mechanisms have been noted (Cavanagh, 2005, Livingstone, 2002; Zeki, 1999), and the psychology of colour owes its early formulation to theories developed in this era. However, while it is clear that the development of artistic techniques and materials has allowed for greater verisimilitude, this has not been the sole aim of art, especially since the 20th Century (or arguably ever, outside of the European context), and has limited applicability outside of art. Artistic use is nonetheless the critical factor which determines aesthetic relevance according to this theory, and this demarcation means that the mimetic approach cannot adequately account for general aesthetic experience as described in this thesis.

1.3.2 Expressivism

The expressivist approach takes the view that the primary goal of art is emotional arousal. Francis Hutcheson (1725) described art as instilling a sense of beauty and pleasure in the observer. This happens independently of whatever the art object’s function is. In *What is Art?* Tolstoy (1897) argued that the aim of art is to communicate a feeling from artist to observer. The experience is essentially subjective, and exhibits some variability amongst people, but is held to adhere to absolute standards. Hume’s *Of the Standard of Taste* (1757) argued that the subjectivity of aesthetic experience is real, but that universal standards are essential
nonetheless. Good taste is taken to be a matter of knowledge and training, and ideally should be divorced from the influence of other people. In more modern theorising, the expressivist approach is most prominently described by Croce (1965) and Collingwood (1938/1975). According to Croce (1965), aesthetic experience is characterised by an intuitive apprehension of artworks as symbolic representations of intense feelings. One obvious argument to this is that we seem able to acknowledge the aesthetic value of an artwork without feeling whatever emotion inspired it, or is literally depicted in it. Like Croce, Collingwood (1938/1975) attests to the role of emotiveness in artworks, but also claims that imagination plays a central role. In other words, there need be no pre-existing emotion in the creator or resultant emotion in the observer, but an artwork might nonetheless provide some knowledge about that emotion – the distinction is encapsulated in the comparison of “expression of” versus “expressive of” (Graham, 2005). The function of art is therefore in providing information about emotion or some “truth” of consciousness (Graham, 2005).

Expressivism has in common with mimesis the intuitively appealing proposition that aesthetic experience can be qualified by its success as a vector of some pre-determined quality – i.e. how realistically it portrays its subject, or how well it conveys the emotions of the artist (or how it facilitates emotional awareness). Both of these, however, depend on some familiarity with a definite representational or emotional aspect to the aesthetic object, and none may exist. For instance, a person may have an aesthetic experience that relates to natural objects that is assuredly of the same kind as an experience that derives from an artwork (what Edmund Burke [1771] calls the sublime), yet there is no artist to emote through the object, nor can verisimilitude sensibly apply in this situation. Like mimesis, the expressivist approach is too circumscribed to be acceptable as a serious philosophical framework for
aesthetics as it cannot account for aesthetic experience arising from abstract art, non-Western art (such as Arabic tiling patterns), or natural objects.

1.3.3 Kantian Aesthetics

Kant (1781) argued that there is more to knowledge than what is given by the senses, namely that there is an interaction between knowledge and experience. Humans share some fundamental mental categories and concepts (Verstandesbegriffe) that shape sensory input, and the manner in which aesthetic judgements are made is the result of a special case of this process. In the Critique of Judgement (1790) he says that judgements of taste are evaluations of pleasure, and are mostly subjective. The appreciation of beauty goes beyond the physical pleasures of food, shelter, etc. - aesthetic judgements are made in a disinterested manner, that is, without relation to the function of the referent object. In such a state the observer’s cognitive faculties are operating fully, but without the constraint of a corporeal need or imperative that defines them. Kant caused a shift from objective features determining beauty towards the idea that the form of the experience of an aesthetic object is what determines beauty. Certain objects allow us to become aware of the purposiveness of perception, isolated from nature’s determininistic imperatives (McMahon, 1999). According to Mothersill (1984), there are no pre-specifications for the types of objects that allow this to occur, but the outcome is the same - we derive meta-cognitive awareness of efficient perception and experience pleasure. Art can accentuate - instead of merely represent - reality by this process of “disinterested interest” (or disinterestedness).

Fechner considered his aesthetics-from-below to be in opposition to the Kantian approach, and this thesis argues that the concept of higher-level mental

---

2 Kant’s depiction of beauty is not the same as the normal definition, which equates it to pleasing physical form. Instead, it involves an apprehension of the Zweckmassigkeit, or match between a set of elements and an ordered or unified whole, similar to the concept of rightness discussed in later sections (see Sammartino-Gardner, 2011).
faculties guiding aesthetic experience is not entirely supportable given experimental evidence. Furthermore, while disinterestedness may carry some value as a way to characterise aesthetic experience, it is more difficult to apply it as a necessary condition. However the proposal that aesthetic experience may arise from awareness of normal cognitive mechanisms has been applied to perceptual processing by McMahon (1999) who suggests that stimuli which challenge or exemplify the way the perceptual system operates can facilitate aesthetic experience. This has relevance to the thesis proposal, and will be discussed in more detail in the concluding chapter.

1.3.3.1 Aesthetic Attitude

While disinterest is a fundamental component of aesthetic attitude theories, many approaches similar to Kant’s have been offered which focus on the observer’s intentional stance. While Kant suggested that awareness of normal cognitive processes are what characterise aesthetic experience, aesthetic attitude theories effectively propose that the processes themselves are of a different kind when a particular stance is taken. There are similarities, however. Stolnitz (1960) suggested that attitudes guide perception, so aesthetic attitudes are essentially disinterested in nature. Bullough (1912) claimed that such attitudes are emotionally uninvested (cf. expressivism). As with the Kantian approach, it does appear likely that some kind of volitional receptiveness may facilitate aesthetic experience, but the question arises – receptiveness to what? If there are no constraints on the objects of an aesthetic attitude, then it is permissible that all experience can be aesthetic if the correct stance is taken, and this seems unlikely to be correct, so some properties of objects would appear to have special relevance for aesthetic experience. The idea of universal aesthetic properties that address a fundamental human capacity across cultures – what Kant called a sensus communis – is persistent (e.g. Dutton, 2009), but is unlikely to be
due to an attitude or stance. Dickie (1964) proposed that disinterest is impossible; what appears to be a disinterested cognitive stance is simply interest in something else, so what is posited as an experiential difference is actually a motivational one. Kemp (1999) argues that these factors are not as discrete as Dickie suggested. Perceptual orientation is largely a matter of motivation or purpose, and a non-pragmatic attentional motivation is not an incoherent concept. However, it is difficult to understand what might characterise the type of processing that is somehow made different by adopting the appropriate attitude.

Aesthetic attitude theories have fallen from favour in philosophical aesthetics, largely as a result of criticisms issued by Dickie (1964; Kemp, 1999). The present thesis offers the speculative claim that an aesthetic attitude can be redefined as the result of aesthetic experience rather than its cause. The formal aspects of the stimulus are posited to be an integral part of this process, and may enable aesthetic experience by enhancing the perceptual process itself. Disinterest is therefore taken to be an orientation towards the processes involved in perceptual organisation. As such, disinterestedness is not suggested to be the necessary precursor to aesthetic experience, but a part of the dynamics of the interactions between an observer and the object of their aesthetic experience. While traditional aesthetic attitudes were proposed as a requirement for aesthetic experience, in this thesis stimulus features may facilitate a non-pragmatic evaluative orientation by influencing the deployment of attentional resources in the absence of an immediate need or requirement.

1.3.4 Formalism

In the second half of the 19th Century, representational accuracy was captured in more or less perfect form by photography (Shimimura, 2012), so many artists turned to an attempt to capture the essential features of form and structure by reducing
the amount of representational detail. Modernists, impressionists, and post-
impressionists tried to approach their subjects by emphasising what they considered to be their essential characteristics. Bell (1914) suggested that the essence of aesthetic experience is the perception of “significant form”. Formalism asserts that an artwork should be evaluated on the basis of sensory features such as line, colour, and shape. Importantly, this idea does not limit aesthetic experience to art or a particular type of processing. The present thesis is largely in agreement with this approach, specifically advocating the possibility that perceptual organisation is influenced by the formal aspects of a stimulus, and therefore some features may be more aesthetically relevant than others.

1.3.5 Cultural and Structural Approaches

The following approaches are largely related, by virtue of their dependence on a structure extraneous to a perceiver and the object of their aesthetic experience. Similar factors make these approaches largely unsuitable for empirical work, as can be seen in the following two examples.

1.3.6 Conceptualism

According to the conceptual approach to aesthetics, art is essentially about conveying meaning rather than a picture of reality or providing some sense of pleasure in formal aspects. Prehistoric painting seems to be about a concept or message rather than attempting to recreate reality, or to express emotion. Art in the past half-century has been mainly focused on the representation of concepts. An early example is Duchamp’s Fountain (1917) – a men’s urinal turned on its back and signed “R. Mutt”. It was submitted for exhibition but never shown. The artwork’s aesthetic value is difficult to gauge - it doesn’t appear to be created by an artist, it isn’t intended to be beautiful, and it (supposedly) doesn’t elicit a sense of significant form.
The intention was surely to make people question the nature of Art, meaning that the only way to assign it aesthetic relevance is conceptually. To interpret these types of work, the observer needs to understand what Goodman (1976) calls the “languages of art”; i.e. they must have knowledge about art history and the changing definitions of art. Some of the central claims of this approach may not be valid. The appeal to primitive art as being inherently conceptual rather than representational is intuitively appealing, but some evidence contradicts this. According to Reznikoff and Dauvois (1988; Reznikoff, 2008), cave drawings tend to be found in echoic cave chambers. The suggestion is that these early artworks were directed towards enhancing storytelling or musical performance by providing a multi-sensory depiction so that the perceiver can attain deeper mimetic appreciation (Scarre, 1989). The formal qualities of the artworks are therefore very important, even if the depictions themselves do not seem realistic in isolation. Similarly, it is difficult to say with any assurance that aesthetic experience can be entirely conceptual, as some physical object will necessarily operate as the vector for the concept. Finally, in this thesis, the extreme form of the conceptual approach (positing that art is entirely conceptual) is taken to be invalid for all aesthetic experience as it can’t account for untutored aesthetic phenomena or for non-art aesthetic events.

1.3.7 Institutionalism

Some theorists have denied the possibility of defining art in terms of properties, and have suggested that invariant features may be non-perceptual, relational ones (Davies, 1990). According to Davies (1990), these approaches can be either functional or procedural. In the functional approach, art is defined largely by its purpose, as mimesis or expressivism. This has the problem of accounting for those art objects that don't appear to communicate or express some artists' intent, for example Marcel
Duchamp's *Fountain*. In the procedural approach, art gets its meaning from the creation and reception process (the artworld; Danto, 1964; Dickie, 1964). This approach can be seen as an extension of formalism, but one in which the significant form is not instantiated in the artwork itself, but in the community that creates, experiences, and discusses it. One problem with this idea is that it doesn't account for early art; there surely wasn’t a pre-existing artworld waiting for art to come along (Dutton, 2009). Dickie (1964) tackled this issue by suggesting a second way in which things become art according to the artworld. Besides similarity to previous art, artworks can gain their status via artefactuality. In this view, an artefact is a material that is being recruited to fit some purpose, and specific to the artworld, some artistic purpose. Stecker (1986) claimed that this revision removes the need for an institutional framework at all, so is actually detrimental to the validity of the approach. Also, the concept of the artworld may rest on a circular definition - if the distinction between the artworld and regular social practices is to be made, it surely requires some definition of its central concern (art) beyond the institution that supports it. Otherwise the artworld is like any other community discourse. It appears difficult to justify a purely institutional approach, and reducing the artworld to a less deterministic status offers no conceptual advantage over traditional art criticism which focuses on form and content.

### 1.4 Empirical Aesthetics

In the late 19th century, a scientific turn in the psychology of aesthetics led by Fechner aimed to establish an empirical approach to the field by identifying the features of objects that allowed them have aesthetic effects. Figures such as Helmholtz (1954), Kulpe (1901), Lipps (1962), and Stumpf (1883) provided work under this framework. This research represented a move away from an introspective
discursive approach in which the value of an aesthetic theory was due to its philosophical rigour, or how convincingly it was expressed. The experimental approach did not represent a narrowing of the scope of aesthetics to a purely physicalist framework – instead it allowed aesthetic phenomena that were difficult to verbalise to be systematically studied via their effects on perception and behaviour. Fechner’s aesthetics introduced some important research tools for experiments: in the method of choice subjects are presented with objects and rate them for pleasingness. In the method of production, participants produce an object that conforms as closely as possible to their personal taste. The method of use involves analysing existing highly regarded artworks for consistent characteristics. Procedures have become more varied since this, and Palmer et al. (2013) provide a review of some of the popular methods, but these techniques have remained popular in the last century of research (see Chapter 2 for a more detailed discussion of these methods).

Fechner’s aesthetics-from-below offered a way to investigate aesthetic experience via its effect on participant preferences as an indicator of aesthetic experience, and one which is not necessarily dependent on introspection. This is an important advance, as artistic creation and reception are often characterised by inaccessibility of the relevant processes. For instance, Gombrich wrote:

Ideas about beauty and expression are rarely mentioned by artists…What an artist worries about as he plans his picture, makes his sketches, or wonders whether he has completed his canvas, is something much more difficult to put into words. Perhaps he would say he worries about whether he has got it ‘right’. Now it is only when we understand what he means by that modest little word ‘right’ that we begin to understand what artists are really after (1972, pp. 13-14).
He goes on to say that we all have these moments where we look for something to be ‘right’, even if that rightness is ineffable (Gombrich, 1972). This is what Yeats (1935) was referring to when he wrote: “A poem comes right with a click like a closing box”.

Recently, Locher (2003, 2006) investigated the rightness theory, specifically as it relates to structural organisation of features in visual art. He found that people who were untrained in visual art were faster to discriminate between original and structurally-altered artworks when the alterations were severe. Performance was best when people were focused on the style and form of abstract artworks and when they focused on content and realism in representational artworks. While Locher, Stappers, and Overbeeke (1998) found that people glean a sense of rightness and inevitability from structurally coherent displays, they claim that there is no support for the hard view of rightness theory, that there is only one such organisation that supports rightness. According to Locher (2003), compositional rightness might achieve its aesthetic effect by exhibiting a structure that facilitates the expression of an artwork’s content. However, the concept itself may be too vague to be employed with appropriate scientific rigour.

One way to operationalise rightness is to relate it to perceptual processing, particularly in the formation of perceptually-organised holistic units. Bell’s (1914) concept of significant form is superficially similar to that of aesthetic Gestalten, in that it arises from the organisation of formal features rather than from the features themselves. Bell’s theory suggests that as long as something has significant form, it has aesthetic potential. According to Carroll (1999), it is easy to ascertain such an idea from instrumental musical pieces; it is clear that the aesthetic value of this artform does not depend on representation, but almost exclusively on “the temporal play of aural form (Carroll, 1999)”. If artworks have a common denominator, it is form, so
the formalist argument goes that there is something in the *type* of form possessed by the objects of aesthetic experience that gives them their value. Carroll (1999) suggests that this shifts the locus of ambiguity from ‘art’ or ‘aesthetic’ to ‘significant form’. However, it is not entirely obvious that there is no added constraint in shifting the focus in this manner; relating the concept to *gute Gestalten* offers the possibility of experimental analysis (see Chapters 4 and 5).

### 1.4.1 Gestalt Psychology and Aesthetics

Ogden (1937) advocated a return to the perceptually-driven aesthetics as defined by Baumgarten, and claimed that aesthetic experience derives from the level of the Gestalt. His book *The Psychology of Art* (1938) derived much of its influence from Hambridge’s (1920; 1923) idea of dynamic symmetry. This notion was placed in opposition to static symmetry exemplified by the regularity of figures such as a square or triangle. Dynamic symmetry, in contrast, relates to more irrational proportions of which the golden section provides a canonical example, and is thought to be inherently more interesting (see Chapters 2 and 3). Le Corbusier (1954/2004, 1958/2004) was heavily influenced by this idea (Benjafield, 2010). Ogden sought to accommodate this principle into a Gestalt framework, relating it to figure-ground and part-whole segregation. Deriving from the minimum principle, percepts were held to assume simple geometric shapes with are more pleasing when their parts bear some reliable and easily apprehended relationship to each other. This is done without the explicit awareness of the observer. Furthermore, Ogden (1937) suggested that *all* visual perception therefore has “geometric implications (p. 198)”, and that the perception of beauty is fundamentally geometrical.

Arnheim (1954) described the ‘hidden structure’ of a regularly proportioned frame, which allows observers to immediately infer internal asymmetry. The effect is
so pronounced that it obviates the need for verification by measurement, and amply
demonstrates our visual system’s capability of performing a relational computation
that isn’t done in piecemeal object-by-object fashion, but as an implicit property of
the whole visual field. This phenomenon provides us with information about the
relations between distal structures, and gives rise to a semantic component – the
central disc appears to be in a state of ‘tension’, moving away from or toward the
centre. Arnheim concludes that the whole of the visual array is the interplay of
directed tensions, which he terms ‘psychological forces’. The loci of particular forces
are specific to each stimulus configuration, and are induced in similar fashion to
Kant’s synthetic a priori. Visual dynamics are proposed to be “a property inherent in
shapes, colors, and locomotion, not something added to the percept by the
imagination of the observer” (Arnheim, 1954).

1.4.2 Cognitive Psychology and Aesthetics

Fechner’s (1876) model of empirical aesthetics proposed that low-level
perceptual features could be assessed in terms of consistent and reliable effects on
aesthetic experience. Cognitively-oriented process models of aesthetic experience
(e.g. Belke, Leder, & Augustin, 2006; Carbon, 2011; Carbon & Leder, 2005; Leder, et
al. 2004; Solso, 1996; Tinio, 2013) usually integrate the Fechnerian contribution
within an initial perceptual stage which organises the information for later cognitive
processing based on mnemonic factors relating to issues such as style, artistic intent,
emotion, etc. This thesis presents the claim that, in agreement with Fechner and
Arnheim, the perceptual stage is critical, and that theories aiming to codify the
entirety of aesthetic experience are premature. While experimental aesthetics is often
classified as lacking progress (e.g. van Tonder & Spehar, 2013), recently there
have been a number of attempts to account for mechanisms related to aesthetic
experience in perception. Two of these with relevance to the present thesis are described in the following sections.

1.4.3 Perceptual Fluency Approach

The perceptual fluency approach has been described by Palmer et al. (2013) as “the single most general explanation of aesthetic preference”. The idea proposed is that aesthetic experience is a function of the perceiver’s processing efficiency; the more fluently or easily the perceiver can process an object, the more positive his or her aesthetic response (Reber, Schwarz, & Winkielman, 2004; Reber, Winkielman, & Schwarz, 1998). Fluency is hedonically marked such that high fluency is experienced positively, and people take this affective information into account when making aesthetic judgements. Importantly, this theory provides an account of unexpected or surprising aesthetic experiences. The impact of fluency is moderated by expectations, and has higher impact when fluent processing is a surprise.

Rather than simply assigning an affective component to a representation, fluency may affect the process of categorisation and change the semantic implications of an object. Oppenheimer and Frank (2008) specifically looked at the idea that processing fluency impacts on category judgments by asking participants to judge whether a word (e.g. “robin”) was a good category member of another supraordinate word concept (“bird”). They manipulated fluency by altering the font of the word displays, and found that when people have trouble with perceptual processing due to poorer font legibility, they ascribe this lack of fluency to their judgment of the word item’s category membership. Furthermore, Oppenheimer (2008) found that participants assigned a low intelligence score to a writer when they read a writing sample with complex vocabulary, grammar, and poor font legibility. However, when they simulated a toner depletion pattern to the printed pages, the participants gave the
writer a higher intelligence rating. The author concluded that this is because they experienced low perceptual fluency, but attributed their metacognitive evaluation to the quality of the print.

Muth and Carbon (2013) note that there seems to be some disagreement between fluency theories and the observation that we like novelty or innovation; some artworks actively inhibit processing efficiency, yet are liked. Similarly, Van de Cruys and Wagemans (2011a, 2011b) claim that artists use familiar patterns along with some deviation from familiarity to produce a conflict between an expectation and what is given. Furthermore, the evolutionary value of either fluency or novelty can be convincingly hypothesised; familiar stimuli would be easier to process (and if encountered many times are probably harmless), but searching for novel features in the environment might also be adaptive and intrinsically rewarding.

The finding that an intermediate level of ambiguity is optimally preferred (Berlyne, 1974; Jakesch & Leder, 2009), has led to the suggestion that fluency and novelty may be independent rather than parts of a single continuum. Muth and Carbon (2013) have shown that people do not tend to increase in liking for objects regardless of their initial evaluation. Blijlevens, Carbon, Mugge, and Schoormans (2012) found that arousal and typicality both contribute to aesthetic appreciation. The authors define typicality as the degree of exemplification of a category, or goodness-of-example. Typical stimuli have been found to result in positive appraisals of product designs, but atypicality also seems to be attractive (e.g. Schoormans & Robben, 1997). The latter may involve some kind of problem-solving whose solution results in positive affect. Crucian et al. (2000) suggested that explicit awareness of arousal being attributed to a stimulus is necessary for aesthetic experience. Blijlevens et al. (2012) found that typicality had a curvilinear relationship with aesthetic ratings, and
arousal had a positive linear function related to ratings. These relationships did not appear to interact, so it is possible that fluency theory cannot provide an adequate account of both simultaneously. Evidence presented in Chapter 5 appears to support this idea.

Carbon (2010) emphasised the need for aesthetic experience to be considered as a dynamic process involving variability in the observer’s state. After an object is evaluated according to perceptual and cognitive elaboration, preferences along with perception itself are altered. When stimuli are perceptually challenging, the observer adopts a stance of continuous elaboration, and pleasure is gained when information or meaning is extracted from previously challenging stimuli (Muth & Carbon, 2013). In summary, there would appear to be good reason for rejecting the harder version of perceptual fluency theory. It might well be true that the efficiency of perceptual processing affects its evaluation, but this is not the limit of aesthetic experience. Ambiguity and novelty are often part of such experiences, and it may be more enlightening to focus on the ways in which visual exploration is affected by perceptual organisation – a reasonable amount of perceptual difficulty or lack of fluency can enhance aesthetic experience when considered as a dynamic process involving elaborative mechanisms.

1.4.4 Natural Scene Structure and Aesthetic Experience

Related to perceptual fluency is the claim that images are preferred when they exhibit the same kinds of structure as natural scenes; we prefer those statistical structures that match our evolutionary environments (Graham & Redies, 2010). This is proposed to be a result of the underlying ecological constraints on perceptual processing (Field, 1987). Horizontal and vertical lines are preferred to obliques, and these line orientations are more common in nature, but also in art (Appelle, 1972;
Latto, 1995; Latto, Brain & Kelly, 2000). People also appear to prefer lines with curved, rather than sharp, contours, and this holds for abstract shapes (Bar & Neta, 2006; Silvia & Barona, 2009). Some recent evidence suggests that there are natural predispositions to favour certain spatial frequencies. While line orientation preference is linked with population differences in visual cortical encoding, psychophysical evidence suggests that the early visual system also performs a Fourier analysis of incoming visual information via local spatial frequency filters (Campbell & Robson, 1968; Shapley & Lennie, 1985). Spectral analysis of paintings and natural scenes have shown similar amplitude spectra for both, highest for low spatial frequencies and decreasing in linear fashion with the log of spatial frequency in the spectrum of $1/F$ noise (Redies, Hänisch, Blickhan, & Denzler, 2007; Redies, Hasenstein, & Denzer, 2008). More natural looking paintings have the $1/f$ power spectrum, and increases correlate with visual discomfort (Graham & Redies, 2010).

Object shapes may also exert an influence – the golden section appears to offer some possibility that proportion ratios influence preference (Green, 1995). Stimulus complexity has also been investigated as a source of aesthetic value - Birkhoff (1933) proposed that preference varies with the number of elements and inversely with complexity (expressed as the number of non-collinear sides). There has been little empirical support for this idea, however, possibly because it predicts linear increases in preference with complexity, when an intermediate complexity seems to be most preferred (Berlyne, 1971). In general, people who are familiar with simpler stimuli seem to prefer greater complexity and vice versa (Tonio & Leder, 2009).

The way in which objects are positioned relative to each other and the frame, or how they are distributed around the centre, also influences aesthetic preferences. Arnheim (1954) speculated on the structural skeleton, with dynamic internal forces,
and this has been supported by experimental work (Palmer, 1991; Palmer & Griscom, 2012; Palmer & Guidi, 2011; see Chapters 4 and 5). Early studies on visual balance used a fixed line of some length, width and colour, with the task being to adjust the position of a test line to balance the entire display (Pierce, 1894; Puffer, 1903). These tended to be placed farther from the fixed line if they were shorter than it, and closer if they were longer or thicker. In studies where artworks are modified to influence balance and composition, more balanced images are preferred and when subjects are asked to place shapes in an empty frame, physical weight tends to be clustered at the centre. For spatial arrangement of meaningful objects, Palmer, et al. (2008) found strong tendencies for symmetrical objects that faced forward to be preferred to be located at centre, and those facing left to be located on the right, and vice versa. An inward bias has also been shown in paintings of animals (Bertamini, Bennett, & Bode, 2011). Sammartino and Palmer (2012) found similar preferences for vertical position – if the object is symmetrical around a horizontal axis, it is centred, and if it faces upwards or downward there was an inward bias. Palmer et al. (2013) suggest that ecological salience is the determining factor for these kinds of objects. The present thesis claims that natural scene structure influences and constrains perceptual processing and this can be seen in the special status accorded to certain stimulus configurations over others, namely in the areal contrast of golden sectioned objects (Chapters 2 and 3) and the visual balance of items relative to a surrounding frame (Chapters 4 and 5).

1.5 Summary

According to van Tonder and Spehar (2013), the field of experimental aesthetics has been bewilderingly unproductive, especially considering its claim as one of the first areas to receive systematic focus in experimental psychology. They
attribute this lack of progress to a focus on mechanistic formulations of aesthetic experience that are overly reliant on the concept of pleasure and simple stimuli far removed from the actual kinds of objects that cause such experiences in the real world. On the other hand, it appears that process models have failed to provide adequately testable hypotheses, and perpetuate the idea that there is a discontinuity between aesthetic and non-aesthetic processing. In this thesis, the preferences that people can express for visual attributes are assumed to have the same phenomenal quality as more pronounced aesthetic experiences. Rather than relying on a dichotomised version of aesthetic/non-aesthetic (which introduces an unnecessary hard problem), every experience is taken to have a quality that arises from normal perceptual processing. As suggested by van Tonder and Spehar:

…every segment of the visual field appears to be imbued with some quality. Whatever appears has the potential to repulse or attract. From the smallest fragments to the most holistic ensembles in perception, the forces of attraction and repulsion combine into what is experienced as the aesthetic appeal of reality (2013, p. 395).

The goal of the following chapters is to investigate phenomena with avowed aesthetic relevance (proportion, balance) under the assumption that their effects can be demonstrated in perceptual performance.
Chapter 2: The Golden Section as a Perceptual Phenomenon

2.1 Introduction

This chapter describes an empirical investigation of the golden section hypothesis, which proposes that the presence of the golden section in a visual pattern can produce distinct effects on perception and cognition that might contribute to aesthetic experience (Adams-Webber, 1978; Benjafield & Adams-Webber, 1976; Plug, 1980). The golden section describes the ratio of two unequal subdivisions of any whole where the smaller section (a) is to the larger as the larger (b) is to the sum of both sections. In simple notation, \( a/b = b/(a+b) \), the solution of which is \( (1 + \sqrt{5}) / 2 \).

This numerical relationship is called the golden ratio, and the present work deals mostly with its instantiation in physical geometry. The positive expansion of this is the irrational \( 1.61803\ldots \) and the negative is \(-0.61803\ldots\), often interchangeably referred to as phi (\( \phi \))\(^3\), and belonging to a group of famous mathematical abstractions such as \( e \) and \( \pi \) that have provoked intense interest since at least Pythagoras (Green, 1995). Johannes Kepler described the golden section as a “treasure of geometry” (Schoot, 2001). Numbers like phi that cannot be expressed as whole integer ratios are called incommensurable, and Green (1995) has suggested that there may be a pantheon of these numerical relations (labelled ‘Pythagorean’ due to the numerical mysticism associated with Pythagoras and followers) that may have aesthetic relevance\(^4\).

---

\(^3\) Strictly speaking, these are not treated interchangeably. Most commentators choose one or the other value, and adopt the negative of the negative solution when that value is chosen. Green (1995) is a notable exception within the psychologically-oriented literature. His notation is that \( \phi = -1.618 \), \( -\phi' \) is the negative solution (\(- -0.618\)), and its negative is \( \phi' \) (\(-0.618\)). For the sake of clarity, this dissertation will not differentiate in each case during discussion of precedents in the literature because the properties of the solutions are the same in each case. This equivalent treatment comes from the fact that \( 1.618:1 = 1:0.618 \), so that ratio subdivisions are equal for each value of \( \phi \).

\(^4\) Aside from phi, Green (1995) suggests that both rational and irrational root subdivisions (\( \sqrt{2}, \sqrt{3}, \sqrt{4} \)) are also aesthetically relevant (see also Davis, 1933). For example, in Islamic architecture, \( \sqrt{2} \) is considered to be an important ideal ratio (Kak, 2011).
The name “golden section” is relatively recent, and comes from Martin Ohm in 1835 as *goldener Schnitt* (Fowler, 1982). However, codification of its geometric properties begins in earnest with Euclid, who referred to a golden section as the division into “extreme and mean ratio”, and suggested methods for its construction. Luca Pacioli (1509) developed the first major thesis on phi (*De Divina Proportione*), which listed many supposed instances of its appearance in art and architecture. Leonardo da Vinci provided illustrations for this volume, the thesis of which was that the golden ratio referred to God’s beauty permeating nature. The most divine, and therefore beautiful, artworks were considered to be those that prominently displayed phi, making the golden section a naturally relevant proportion to all observers. This view is an early definition of the golden ratio as a normative and naturally occurring phenomenon, the depiction of which provides consonant feelings of pleasure in the observer.

### 2.2 The Golden Section in Nature

The golden section is closely related to the Fibonacci sequence, devised by Leonardo of Pisa - the author of *Liber Abaci* (1202) - who proposed it to explain generativity in rabbits. Each generation in rabbit reproduction is the addition of the previous two generations, or more abstractly, each number in the sequence is equal to the addition of the previous two terms. A relationship between Fibonacci sequences and phi was discovered by Robert Simson in 1753 (Wells, 1986) who found that the ratio of each term to its predecessor approaches the golden ratio as the sequence increases \((x_n = x_{(n-1)} + x_{(n-2)})\). This is true regardless of the integers that provide the starting values of the sequence. Such a pattern is often seen in nature, in the “golden-angled” (137.5°) spiral configurations of pineapples, sunflowers, pinecones, cauliflowers, etc. This is explained by the recurrence of natural growth sequences,
and the maximising of the exposure of iterative elements (e.g. petals) to the environment \(\textit{phyllotaxis}\); Thompson, 1945; Green, 1995). A logarithmic spiral can trace the corners of subdivisions of a rectangle into a pattern which is seen in the growth pattern of the nautilus shell (Figure 2.1 [a] and [b]) and the predatory flight descent of the peregrine falcon (Tucker, 2000). Such phenomena are not limited to bio-organisms; Douady and Couder (1992) found evidence for a golden spiral arrangement in the equilibrious state of magnetic fluid suspended in silicone oil. A possible implication of this for the aesthetic potential of the golden section is that it therefore has especial phylogenetic relevance to humans as adaptive organisms (e.g. McManus & Weatherby, 1997).

![Figure 2.1 A logarithmic spiral in geometry and nature.](image)

(a) represents a logarithmic spiral inscribed in a golden rectangle. Image is adapted from “Logspyr 4” by user Homk (2012). (b) represents a cross-sectioned view the interior chambers of the nautilus shell. Image is “Nautilus cutaway logarithmic spiral” by user Chris73 (2004). Both (a) and (b) are Wikipedia and Wikimedia Commons images used under the creative commons cc-by-sa 3.0 license, retrieved 2014.

These phenomena are supportive of a privileged natural status for phi, but attempts to relate them to a general standard of beauty in nature according to adaptive constraints have proven to be less convincing. For example, some commentators
(Ghyka, 1946; Hambridge, 1923) have proposed that an ideal human form is in the golden ratio in such pairings as overall height to navel height; sub- to supra-navel height; and adjacent finger bones. Such proposals do not receive empirical support; Pallett, Link, and Lee (2010) found that an optimal alteration of facial proportions occurs when the pupil-to-pupil distance is 46% of the width of the face, and the distance from the eyeline to the mouth is 36% of the height of the face. Other research has suggested that ratios of height to width are not related to attractiveness at all, and that symmetry or a higher-order property relating to sex-specific norms are more important (e.g. Grammer, Fink, Møller, & Thornhill, 2003; Penton-Voak et al., 2001; Rhodes, 2006; Scheib, Gangestad, & Thornhill, 1999; Thornhill & Gangestad, 1993).

2.3 The Golden Section in Art and Design

There have many attempts to generalise the putative attractivity of phi to non-naturally occurring phenomena, with the aim of establishing a correlation between works of cultural significance and golden sectioning. Gowlett (2011) argues that Acheulean hand-axe construction in Kilombe in Kenya has breadth/length relationships in the golden ratio, as do bifaces from Syria and Boxgrove (Pope, Russel, & Watson, 2006). Lynch and Hathaway (1993) found a tendency for proportions of 6,000 year-old flint arrowheads to cluster around the golden ratio which appears to be a definite and deliberate design choice unrelated to production constraints. This might imply that the artefacts were fashioned within a set of codified design principles. Stevick (1994; 1998; 1999; 2004) proposes that many examples of western European religious artefacts from the 6th – 8th century are examples of coherent or constructive geometry – a pre-mathematical combination of aesthetic concerns and geometric constraints that often led craft-makers to fashion objects of widely-accepted beauty that exhibit strict geometrical relationships, one of which is
frequently the golden ratio. Stevick (1999) analysed a well-preserved example of an Irish high cross from Durrow Castle, Offaly, and found that all ratios within the subdivisions of the cross could be approximated by the use of three numbers: 1, 2, and $\phi$. The Tara brooch, one of the most important relics in Ireland, is also suggested to demonstrate the golden section in the embedding of its arcs (Stevick, 1998).

Zeising (1854, 1884) suggested that the golden ratio is intrinsically beautiful to behold, ubiquitous in nature, and that this has been reflected in notable architectural works - he claimed that the Parthenon and the Great Pyramid provide clear examples of deployment of the ratio (see also Ghyka, 1946; Ogden, 1937; but see Benjafield, 2010; Livio, 2002; and Rossi & Tout, 2002, for dissenting opinions on the validity of these examples). Ideas regarding phi in artefact design and architecture tend to imply that there may be some totemistic significance to particular geometric proportions, but it is also possible that ergonomic constraints in the construction of these artefacts may have had an influence. In the absence of a written record of the designers’ intentions, these findings are mostly anecdotal. Similarly, for many of the supposed instances of phi in art (Da Vinci, Seurat, Mondrian), there is no definite evidence of the artist’s intent to utilise the golden section (Fischler, 1981; Livio, 2002; Markowsky, 1992). While “absence of evidence is not evidence of absence” (Livio, 2002; p. 46), attempts to establish its presence in pre-modern art especially may be too subject to bias to be informative (Green, 1995; Nickerson, 1998). As such, the golden section effect would appear to be best approached by the methods of experimental aesthetics.
2.4 Experimental aesthetics

2.4.1 Early Research: Identifying an Effect

The first scientific examination of the golden section and its supposed attractiveness was conducted by Fechner (1871, 1876), who assessed preferences for rectangular stimuli of various proportions. He found that when people selected their most and least preferred stimulus there was a high degree of interindividual variability, but the golden section rectangle was the most preferred (chosen by 35% of people), and nobody selected this stimulus as their least preferred. Some commentators have suggested that averaging rankings in this manner might bias towards the middle of the range (Godkewitsch, 1974; McManus, 1980). However, in the case of Fechner’s study, participants did not rank order the stimuli; they simply reported their favourite and least favourite.\(^5\) Witmer (1894) asked subjects to judge the pleasantness of simple geometric objects (triangles, etc.) in a number of proportions. He found a tendency for preference to increase for equal ratios (1:1) and another trend somewhere between 2:3 and 1:2, the average of which closely matched the golden section (B/L ratio of 1:1.651; Green, 1995). Lalo (1908) presented 10 horizontally-oriented rectangles of various dimension ratios, but unlike previous experiments, all stimuli were shown simultaneously in order to ameliorate mnemonic effects. Fechner’s (1871) results were replicated, with preference for the golden rectangle, and less pronounced tendency towards the centre of the range. Like

\(^5\) McManus and Weatherby (1997) suggest that Fechner’s mystical leanings may have provided too much of a conflict of interest for his work on the golden section to be free from experimenter bias. Livio (2002, p.180) further suggests that Fechner may have unknowingly produced experimental situations that supported his ideas of unifying spirit and matter. He supports this claim with the fact that Fechner did not publish research that failed to find an effect in “golden ellipses”. However, the negative findings in the unpublished study demonstrate that Fechner’s methodology was robust enough to withstand experimenter bias if it did exist, and the lack of publication may have been an example of publication bias, a persistent and by no means new problem in scientific research (Dickersin & Min, 1993; Petticrew, 1998). Furthermore, the results demonstrated that while the 1.5:1 B/L ratio was preferred (by 42%), the golden ratio was the second most preferred (at 16.7%) and all other ratios were far less preferred than either of these (see Green, 1995).
Fechner's study, Lalo's also showed that no participant had the golden rectangle as their least preferred. C.O. Weber (1931) used the method of paired comparison, developed by Thurstone (1927), to find a preferred rectangle shape ratio, and found that people tended to prefer the rectangle of ratio 1.871:1, but this tended towards the golden ratio on a second test two weeks later.

2.4.2 Later Research: Methodological Challenges

These early findings received criticism from Thorndike (1917), who proposed that preference for the golden section was too variable to be convincing. He asked subjects to rank geometric shapes with various proportions in order of preference, and found that none of the objects was always ranked first, and that the order was seldom consistent. However, as Green (1995) notes, Thorndike failed to control for potential confounds such as size or presentation frequency. Woodworth (1938) subsequently re-analysed Thorndike’s data and presented some evidence for a bimodal distribution of preference scores, with one of the modes being around the golden ratio. His interpretation was that symmetry is liked because of our own bilateral physical symmetry, but that there is no such basis for our preference for the golden section. Thorndike’s criticism nonetheless had an influence on subsequent research, with issues of methodology being prioritised over interpretation of the effect when it was shown. Fechner describes a number of methods for experimental aesthetics still in use in modern research, and while many of them have produced results to support a golden section effect, it is worth describing the methodological criticisms that have been forwarded in each case in order to explore the proposal that a combination of preference and performance measures may provide some advance on previous research.
2.4.2.1 Experimental procedures

One of the most popular early methods for determining preferences is the method of choice, where people might simply select their favourite stimulus from a range that varies by the dimension under investigation. This approach has received some criticism. In rank-ordering choice procedures people may choose an extreme stimulus as their first preference, and attempt to equally space successive choices in increasing distance from the first (Godkewitsch, 1974). According to Godkewitsch (1974), if people tend to give random first-choices in a range of shapes with different dimension ratios, and then progressively move away from this in subsequent choices, the stimulus with the highest mean preference will be in the middle of the range, where he claimed is usually the place experimenters put the golden section alternative. In other words, people actually select the extremes of the range in their first choice, then use this as an anchor for subsequent ratings. This results in overall mean preferences to tend towards the middle of the range. Plug (1980) suggests that the first choice is more reflective of actual preference, and the preferred stimulus may lie outside of the presented range, so the experiment is overly-constrained by the choice of stimuli. As noted by Green (1995) however, the prevalence of this methodological error is overstated; while it is true of some experiments which found a golden section effect (Berlyne, 1970; Haines & Davies, 1904; Thompson, 1946; Thorndike, 1917), it is not true of others (Fechner, 1876; Lalo, 1908; see Plug, 1980 for a review).

In the method of production, subjects are required to draw what they consider to be the most pleasing example of a geometric object, and the ratio of its dimensions is measured by the experimenter. This approach ameliorates experimenter effects and does not limit the participant to a given range of stimulus dimensions (Plug, 1980).
Konečni (2004) suggests that a golden section effect is subtle and elusive, and best captured by an actively engaged task such as production, rather than detached evaluation. Davis (1933) used this method and found some support for the golden section hypothesis. McManus and Weatherby (1997) found that the mean placement of an object on the horizontal plane during production dissected the frame at the golden section. Konečni (2003; 2004) asked professional painters to recreate a series of images, and found the highest accuracy for those stimuli which featured the golden section. The method of production is not without confounds, however. Ogden (1937) suggests that the dimensions of the drawing surface and the page may have an influence. Furthermore, it may be unrealistic to assume that participants can produce their preferred geometric proportions with total accuracy based on some preliminary introspection.

In the method of adjustment, subjects are able to change stimulus dimensions until they produce the proportions that they prefer. In an early experiment of this kind, Pierce (1894) showed participants three lines (10cm long, 5cm wide) two of which were at a fixed distance from each other, and asked them to move the third line to the place between the others that allowed for the most pleasing effect. There was a marked tendency to divide the line according to the golden section. Pierce's aim was to investigate a preference for equality, and he found that when he increased the number of fixed vertical lines, the moveable line was placed in a position increasingly equidistant from the two adjacent lines. In other words, given lines at 0cm, 20cm, and 60cm relative to the left border of the available, people tended to maximise equality in the display by placing the line at 40cm. It appears likely however that the obvious blank space elicited a demand characteristic in the response – i.e. a suggestion of where the item should be placed (Green, 1995). Angier (1903), and Svennson (1977)
also used this method and found some support for the golden section hypothesis. The methods of adjustment and production share an important drawback, however; neither allow simultaneous comparison with alternative dimensions, so rely on a mnemonic representation of other possible configurations.

In the method of paired comparisons, first suggested by Fechner (David, 1969), and refined by Thurstone (1927), the entire stimulus range is entered into two-alternative forced choice (2AFC) experimental trials, and a mathematical model is applied to construct a unidimensional scale of preference scores. Piehl (1978) used the method of paired comparison and found a pronounced preference for golden rectangles for controlled area rectangles. McManus (1980) used this procedure and found peak preference ratings in the 1.5 – 2.1 ratio range. As Green (1995) notes, there were no golden rectangles in that experiment, so it is difficult to interpret this information, but the range of values is suggestive of a tendency towards the golden section. The paired comparison procedure is attractive to experimental aesthetics researchers because it has a psychophysically-oriented theoretical background, allows for the precise specification of stimulus geometry prior to testing, and produces a unidimensional scale for analysis (see Appendix 1 for more detailed discussion).

2.4.2.2 Statistical analysis

Derived preference statistics, by whatever method, are a persistent source of controversy. Angier (1903) raised the point that averaging across participant preference scores creates a parametric value that describes the sample, but often doesn't reflect the choices of any individual. This leads to the proposition that golden section preferences may be a mathematical artefact rather than a real phenomenon. This however, is a well-known issue with parametric analysis generally, and is not specific to the golden section. In any case, modal values often coincide with the mean
in golden section research (Green, 1995). Most of the research on ratio preferences has indicated some tendency for the 1.5 – 2.0 range (Plug, 1980). However, inter-individual variability may be too great to establish more precision (Davis, 1933; McManus, 1980). Green (1995) suggests that aesthetic effects are fragile and variability may be a defining characteristic, so cannot be used as a refutation of the effects. Thurstone (1927) states that aesthetic preference is not anchored to a physical variable, so is not best characterised in a deterministic way. A measure of central tendency is taken to be a corrective of normal response stochasticity, with the observer drawing from a probability distribution of possible responses to a stimulus from trial to trial. Population averaging performs a similar function in order to determine normative aesthetic trends.

Bimodality in preference scores has also presented interpretative challenges. Boselie (1992, 1997) suggests that if the golden ratio is not chosen unequivocally, then there is no golden section effect. However, Green (1995) points out that many supposed findings purportedly refuting a golden section effect (Boselie, 1992; Davis & Jahnke, 1991; Nakajima & Ohta, 1989) are those in which a preference for unity ratio divisions was found, but golden-sectioned stimuli followed closely. Green argues that this does not contradict the “Pythagorean” idea of the golden section as one of many interesting ratios, so that all that has been shown in such studies is that unity divisions are sometimes more interesting than golden ratios, but this does not mean that the latter are themselves uninteresting. Different aesthetic principles might even underlie the attractivity of each of these ratios (Woodworth, 1938).

2.4.2.3 Stimulus selection

There have been a variety of stimulus types used in experiments on phi. Line length division or simple polygonal proportioning has tended to show a preference for
golden sectioning, as well as symmetry (e.g. Benjafield, 1976; Benjafield, Pomeroy, & Saunders, 1980; Macrossan & Strachan, 1997; Schiffman & Bobko, 1978; Svennson, 1977). Konečni (2005) criticised these kinds of studies for being too abstracted from real-world preferences. He used positioning of vases on a mantelpiece and found that people preferred to place the object in the middle. The biggest problem with procedures of this kind is that the use of real-world stimuli introduces a problem of association; people may have placed the vase in the middle because it appears less likely to fall over the edge, or they may simply have seen more vases placed at the centre. Furthermore, it is questionable whether vase placement is a more quotidian aesthetic activity than judging polygons for preferred dimensions.

Another approach is to use existing artworks and manipulate their interior dimensions. For example, some research on preference for classical sculpture has produced results that indicate a golden section effect (De Dio, Macaluso, & Rizolatti, 2007). When the stimuli are observed, however, it is clear that the experimental manipulations are obvious deformations, and the effect may be due to selecting an existing artwork that already looked pleasing and happened to possess golden ratio proportions. Green (1995) also claims that using real artworks makes it difficult to isolate a golden section effect because so many other factors are dynamically arranged in a composition. Considering the care exercised in their creation, existing artworks cannot be assumed to be malleable enough to withstand manipulation of this sort. In summary, stimulus selection is important, and a balance must be struck between experimental manipulation and ecological validity. In the experiments presented in this chapter, Mondrian-type grids were used because of their simplicity, the ease of manipulation, and their resemblance to artworks with established aesthetic effect.
2.4.3 Summary of Methodological Issues

Opponents of the golden section hypothesis, or experimental aesthetics procedures more generally, have forwarded some valid, but not insurmountable, criticisms. In many cases, however, the objections appear to dismiss the empirical findings by appealing to a straw-man depiction of the hypothesis, characterised by Arnheim as the dichotomy: “the golden section is always preferred” versus “the golden section is never preferred” (1981, p.60), by demanding an unrealistic level of agreement between participants and precision in responding, or dismissing the effect if occurs as part of a bimodal preference trend along with symmetry (e.g. Höge, 1995, 1996, 1997a, 1997b). If a golden section effect is being investigated, it seems likely that there will be a moderate to high degree of variability due to the sensitivity of the effect, and a possible multimodal distribution of preferences. To address the former, it seems prudent to adopt an approach which minimises demand characteristics or over-reliance on subjective report, which may have been the reason for a lack of consensus in the literature. The latter requires careful consideration of the impact of the golden section (or its absence) on the pattern of results. Ultimately, and as suggested by Green (1995), the increase in methodological rigour brought about by the criticisms detailed here have failed to make the golden section effect go away.

2.5 Theoretical Approaches

If the trend in recent research on the golden section hypothesis has been to focus on methodology, there have nonetheless been some proposals that might provide a theoretical framework for the effect itself.

2.5.1 Basic Vision: Spatial Frequency and Fractal Dimension

Recent research has advanced the proposal that the basic structures of visual art are registered in mechanisms sensitive to spatial frequency in the retina, lateral
geniculate nucleus, and primary visual cortex (DeValois & DeValois, 1980; see also Molnar, 1987 and reviews by Mamassian, 2008 and Graham & Redies, 2010). Artists are shown to produce works with spatial frequency structure matching that of natural scenes, suggesting that regularities in the power spectrum of visual art contribute to aesthetic value (Graham & Field, 2008; Molnar, 1987). This suggests the possibility that the elementary units of aesthetic value are processed prior to the involvement of cognitive processing. Furthermore, the statistical structure of natural scenes is fractal in that it generally exhibits scale invariance with little change in the relative power of high and low frequencies in the Fourier spectrum when one zooms in and out of the image. Similarly, many artworks, including abstract (non-naturalistic) compositions are fractal in that they exhibit scale invariance (Graham & Field, 2008; Redies, Hänisch, Blickhan, & Denzler, 2007; Redies, Hasenstein & Denzler, 2008; Taylor, Micolich & Jonas, 1999a & 1999b). Spectral analysis of paintings and natural scenes have shown similar amplitude spectra, highest for low spatial frequencies and decreasing in linear fashion with the log of spatial frequency in the spectrum of 1/F noise. More natural looking paintings have the 1/f power spectrum, and increases correlate with visual discomfort (Palmer, Schloss, & Sammartino, 2012). According to Hagerhall, Purcell, & Taylor (2004), fractal dimension might provide part of the explanation for why ‘naturalness’ itself is an aesthetic predictor. People tend to prefer horizontal and vertical lines to obliques (Latto, Brain, & Kelly, 2000), and these are more common in paintings (Latto & Russell-Duff, 2002), and this matches natural scenes at higher spatial frequencies. Golden sectioning is also a form of scale invariance, because the ratio repeats, leading to the possibility that irrespective of its judged aesthetic quality, it may be processed differently in early visual mechanisms to other, naturally occurring fractal patterns. This possibility has been raised by recent
research identifying golden sectioning as a relation inhibiting cross-frequency synchronization in the EEG, an effect that may be of consequence for processes such as the temporal binding of neurons which is held to be one important physiological mechanism for perceptual organization (Pletzer, Kerschbaum, & Klimesch, 2010). High-frequency bands (beta and gamma) are associated with the integration of attributes in a complex visual object, and phase synchrony in these bands is shown to be enhanced in artists (versus non-artists) during visual cognition related to artworks (Bhattacharya & Petsche, 2002).

2.5.2 Perceptual Organisation

According to a Gestalt approach to experimental aesthetics, the golden section’s status as an aesthetic phenomenon (assuming this to be the case) depends on a Gestalt quality that emerges from the configuration of elements. Arnheim (1954) claims that the golden ratio (and other preferred ratios) are liked because they are optimally balanced between simplicity and ambiguity (or “unity and dynamic variety”; p.71). Benjafield et al. (1980) claim that the golden section might amount to a figure-ground relation which allows figure (minor portion) to be positioned with optimal contrast against a background (major portion; see also Berlyne, 1971, p.232), therefore providing a good Gestalt. McWhinnie (1987) suggested that the golden section represents harmony between opposing organisational principles; it allows two gestalt tendencies – part-whole relationships (grouping), and figure-ground separation of the object as a whole - to be equally evident. The former principle provides some qualitative image aspects, while the latter relates these aspects to the spatial layout of the scene (Wagemans, et al., 2012a). Assuming that this is the case, the question remains: “why is this proportion, and not some other, the ideal proportion in these situations?” (Arnheim, 1954; p.74). Arnheim (1954; also McManus & Weatherby,
1997) suggests that any aesthetic effects of the golden section arise from its place as a non-transitional item along the spectrum of possible shape dimensions. In other words, it functions as a clear conceptual exemplar of its shape category. In modern Gestalt psychology, states of equilibrium can be characterised in terms of dynamical systems, with maximised prototypicality being an attractor state in brain activity (a dynamical cycle structure), particularly in patterns of synchrony and desynchrony (Wagemans, et al., 2012). Stable periods exhibit lower frequency rates, and coherence intervals such as these appear to be longer when stimuli are less ambiguous (van Leeuwen, 2007).

2.5.3 Visual Perimetry and Cognitive Associationism

Stone and Collins (1965) proposed that people prefer proportions that resemble visual field dimensions – the “perimetric hypothesis”. Stone and Collins (1965) drew a rectangle that was just outside the visual field of participants, which was found to have a height to width ratio of 0.768, then another just at the visible limit, and this had a ratio of 0.565. The average of these is 0.665, which is close to the golden ratio. The authors suggested that inter-cultural differences in ratio preference could be due to physiological differences between ethnic groups. The hypothesis was supported by the observation that people tend to prefer horizontal arrangements over verticals or diagonals (Schiffman, 1966). However, the same author did not find production of preferred rectangles to be at or near the golden section. Furthermore, in a later experiment, Schiffman (1969) found that people drew vertically-oriented rectangles slightly more often (57%). Hintz and Nelson (1970; 1971) found that there was no significant correlation between subjects’ visual field and dimension ratio preference, but their participants did prefer rectangles slightly closer to the golden

---

**6** Harrington (1964) reported the visual field on average as subtending 200 laterally, and 130 vertically, the Height/Width ratio of which is 0.65, a figure close to Stone and Collins’s.
section than in Schiffman's studies. Specifically, participants were instructed to rank a series and through successive approximations a central tendency was reached. The median score for chosen preference was 0.558, and the mode was 0.6. Subjects were also requested to produce their favoured rectangle, and the results were further from the golden section (median = 0.545, mode = 0.57). However, it should be noted that the dimensions of the paper used for presentation and for production were different in each case (8x10 inch vs. 10x12 inch), so may have biased results in the production experiment (a flaw in many production experiments; Ogden, 1937). Plug (1976) found that there was no difference in the types of dimension ratios preferred in horizontal and vertical arrangements, but again preferences were closer to the golden section than Schiffman found.

Green (1995) argues that few studies have been performed that provide evidence on the psychological validity of the golden ratio in terms of the innate or cognitive functions of form representation. Such a research programme would appear to have relevance based on the fact that golden sectioning appears in human artefact design for at least seven thousand years (Gowlett, 2011; Stevick, 2004; Lynch & Hathaway, 1993), which is a notable robust aesthetic feature given changes in socially-derived definitions of “aesthetic”. This implies that if psychological validity is to be found, it may be fundamentally cognitive (Hambidge, 1967; Solso, 1996) rather than found in the cultural principles sometimes held to decide aesthetic value. Solso (1996), proposes that artworks function by the implementation of art schemata; inference-based knowledge sets that are acquired from experience regarding styles and modes of expression. He suggested that aesthetic experience, and by extension any trend towards the golden section, is due to experiential factors, specifically the cultural and personal milieu of the observer. In support of this, Fechner (1876) failed
to find golden section effects in children, and Thompson (1946) found that people’s ratio preferences become more pronounced with age. Berlyne (1970) compared children's preferences from Canadian and Japanese schools, and found that the latter preferred squares, while the former exhibited a preference for a 1.5:1 rectangle. Personality may also be involved, with introverts tending to prefer golden rectangles more strongly than extroverts (Eysenck & Tunstall, 1968). People also tend to prefer aspect ratios based on some semantic associations – invitations for a serious event are preferred to be in “golden” proportions in B/L measurements, while a children’s party is preferred to be roughly square (see Raghubir & Greenleaf, 2006). According to Palmer et al. (2013), the implication is that we have developed an emotional category for events such as these, and we easily assign an appropriate shape based on that category due to its ecological valence.

2.5.4 Processing Fluency

Processing fluency theory[^7] claims that aesthetic preference is a function of how easily the visual system can process information related to a stimulus (see Chapter One for a more detailed account). There may be some disagreement between perceptual fluency accounts and the observation that we enjoy novelty, or aesthetic objects that might actively inhibit processing efficiency (Berlyne, 1974; Muth & Carbon, 2013; Palmer et al., 2013). Van de Cruys and Wagemans (2011a, 2011b) have suggested that artists intentionally produce a conflict between familiar or expected patterns in order to exploit this property. Blijevens, Carbon, Megge, and Schoormans (2012) found that arousal due to novelty - and processing efficiency due to typicality - contribute independently to aesthetic appreciation. Carbon (2010) emphasises the need for aesthetic experience to be considered as a dynamic process.

[^7]: Also known as perceptual fluency when referring to the experiences associated with expedited early processing in sensory systems. Processing fluency refers to the more general effect that can also be characterised by efficiency in higher-level cognitive operations.
involving variability in the observer’s state. He argues that after we evaluate a work according to perceptual and cognitive ‘elaboration’, preferences and perception itself are altered. Elaboration is proposed as an alternative to passive engagement with stimuli, and involves a range of dynamic processes. Muth and Carbon (2013) propose that viewers adopt a stance of continuous elaboration of challenging stimuli. Pleasure is gained when information or meaning is extracted in this manner from challenging stimuli. There is some evidence that more challenging aesthetic images are rewarding due to their facilitation of elaboration (Belke, Leder, & Augustin, 2006; Carbon & Leder, 2005). The dynamic aspect of elaboration means that processing fluency has an influence, but not uniformly throughout elaboration. Ramachandran and Hirstein (1999) propose that synchronisation of disparate activity patterns resulting from ambiguous stimulation might be intrinsically rewarding. Muth and Carbon (2013) argue that the creation of perceptual Gestalten while viewing ambiguous stimuli is taken to be an experimentally-observable instance of insight during elaboration. They found that people tended to have higher preference ratings for ambiguous images after they had a gestalt ‘insight’, i.e. when they had perceived a unified object amidst the visual pattern. In summary, processing fluency may help explain a golden section effect, but it would appear to be a complex interaction of ease of processing and tendency for exploration. As such, the direction of influence of the golden section on response performance must be determined empirically.

2.6 Present Study

The present research sought to investigate whether the golden section effect, if present, exerts its influence at the level of perceptual organisation. Both a preference task and a performance task were employed, the first of which elicited reaction time (RT) data in a visual search task, the second of which used a paired comparison
procedure. While preference measurements are sought in line with the prediction that golden-sectioned stimuli would be more pleasing, the adoption of a performance-based paradigm serves to minimise the sensitivity of aesthetic effects to demand characteristics, as well as the nebulousness of terms associated with aesthetic preference. Furthermore, as Makin, Pecchinenda, and Bertamimi (2012) have remarked, there may be both implicit and explicit components to aesthetic experience, and a maximally sensitive experimental approach should provide some account of both of these. Perceptual organization may be less efficient as a result of more difficult synchronisation of neural assemblies underlying object representation, and some evidence suggests that phase sequences in the golden ratio may exhibit the most asynchronous interphasic sequences possible (Pletzer et al., 2012). RT measures are an indirect measure presumed to be sensitive to a variety of preattentive perceptual operations, which include operations requiring binding, such as multiple item comparison and visual scene segmentation (e.g. Driver, Davis, Russell, Turatto & Freeman, 2001; Duncan & Humphreys, 1989). As Boselie (1997) and Koneční (1997) have pointed out, line sections and golden-proportioned shape outlines may not be appropriate as simplified versions of aesthetic stimuli, so the present study used Mondrian-type grid patterns with a sectioning scheme across a range of ratios. These stimuli were chosen for their resemblance to established artworks as well as the rationale that a golden section effect may be due to population coding of areal ratios. We predicted that a general tendency for target localisation RTs to increase when smallest area subsections were similar in size would be disrupted in the golden-sectioned stimuli.
2.7 Experiment 2.1: Target Section Localisation in Mondrian-type Grids

Method

Participants

Twelve participants (7 female, 5 male, mean age of 25.27 years +/- 3.41 years, all with normal or corrected to normal vision) took part. Based on normal colour/form conjunction RTs and SDs (Quinlan & Humphreys, 1987; Wolfe, 1998; Wolfe, Cave, & Franzel, 1989), the effect was predicted to be large, and this sample was taken to be sufficient to detect the effect at 80% power at the 0.05 significance level (Champely, 2012).

Apparatus and Stimuli

Stimuli were generated using E-prime 2.0 Professional (Psychology Software Tools Inc.) on a Pentium 4 PC running Windows XP. Stimulus control and presentation were programmed using native E-prime macro scripts with system integrity checked by means of an E-prime refresh clock test which gave a diagnostic classification of ‘good’ (measurement error +/- 1 ms). Stimuli were presented on a 19” Magic Displays monitor, model CPD-4402 Trinitron with resolution 1024x768 and refresh rate 75 Hz. Responses were recorded using a two key Ergodex® DX1™ Input System.

Stimuli were rectangular, horizontally oriented Mondrian-like grids with 8-, 16- or 32 subareas; examples of which are shown in Figure 1. Each grid contained a set of paired area sections with self-similar dimensions. The ratios of the paired sections examined were derived from the relations of the whole to the major sections: 1:0.468, 1:0.518, 1:0.568, 1:0.618*, 1:0.668 (*0.618 is the golden ratio, or more precisely the inverse golden ratio, but in this case the two are interchangeable; see § 2.1, fn.3). These ratio intervals and sizes were chosen because they allowed equal
increments that would not cause large discontinuities in the appearance of the stimuli across ratios. They also avoid the possibility of grids producing overly small targets when sectioned beyond a certain point. The position of the golden ratio within the range of ratios employed sought to avoid any biased responding that may arise from a central position in the stimulus range (a criticism levelled at Fechner’s original findings by both Godkewitsch, 1974, and McManus, 1980; although see Green, 1995 for a rebuttal).

The sectioning procedure generates the stimuli in each case by first drawing the rectangular boundary of the whole grid irrespective of ratio, the dimensions of the grid match those of the screen to avoid biasing responses towards ratios that more closely match the screen dimensions (in other words we did not rule out the possibility that self-similarity or the effects of fractal complexity could extend beyond the stimulus). The grid was overall 0.5 times the area of the screen and centrally placed. The colour of the grid’s interior was randomly assigned to one of either of the potential target colours (RGB co-ordinates “192, 192, 192” and “75, 75, 75”), which correspond to “light gray” and “gray” in VBA/E-Basic designations and with luminance levels of 71 and 120 cd/m², respectively. The first sectioning was achieved by multiplying the length along the x-axis to achieve the subdivisions of the relevant ratios. After establishing the first section, the sectioning algorithm then performed the same operation on one of these sub-sectioned areas, with the section matching the target ratio being assigned to the smallest of the two areas thus established. The operation proceeded recursively and with the number of sectioning operations determined by the required-Set Size (in this instance there were 8-, 16 and 32 subsection grids). The two subsections of each paired section were shaded light and dark grey with equal probability across the overall set of grids.
This procedure was the same for all ratios in order to establish a consistent layout, and to make sure that the target appears in approximately the same location. Larger set sizes follow the same procedure as longer series of repetitions. There were 8 different arrangements of paired sections, which were the same for each ratio, meaning that there were slight differences in the overall size of the smallest section as a function of ratio. To account for this, overall stimulus size was also varied slightly leading to 4 different stimulus sizes which varied between 1 - 2° of visual angle.

To ensure that each target subtended the same degree of visual angle, the mean target area was taken to be the one corresponding to the 1:0.568 ratio, and other ratios were either increased or decreased in size until the target size matched the mean, by using image scaling in gnu Image Manipulation Program (GIMP, version 2.6 for Windows; http://downloads.sourceforge.net/gimp-win), an open source image retouching application. Resizing involved increasing/decreasing the width of the image, and increasing/decreasing the height by the same amount to maintain the same proportions within the image. When the image’s target matched the desired target area, the alteration was saved. Consequently, because the ratios differed in how unequal a portion is sectioned into two parts, the target would be progressively smaller as the ratio increased, and therefore the correction led to a larger overall image. Any image artefacts caused by rescaling were corrected using the same application. All grids were presented at the centre of the monitor screen. See Figure 2.2 for an example of images used as stimuli in this experiment.

**Design and Procedure**

Experiment 1 used a within subjects design and a 2AFC speeded response task in which participants were asked to respond as rapidly and accurately as possible to the shading (dark – right key, light – left key) of the overall smallest grid section.
Each trial started with the presentation of a central fixation cross for 500 milliseconds (ms). The fixation cross was then immediately replaced by the search grid, to which observers responded. Grids remained on screen until a response was recorded. In case of an erroneous response or a time-out (i.e., after a period of 2,500 ms without response), feedback was given by a computer-generated tone and an alert was presented for 500 ms at the centre of the screen. Each trial was separated from the next by a variable interval between 500 and 1,000 ms. Following a 20-trial practice session participants completed 1200 trials in one 20-block session (with 60 trials per block) ensuring 80 trials per experimental condition (5 Ratios x 3 Set Sizes). Grid presentation was fully randomised across all blocks and separately for each participant. The experiment was carried out in a sound proof booth with low ambient lighting with a chin rest used to ensure distance between participants and monitor was kept constant at 55 cm.

*Figure 2.2 Examples of stimuli used in Experiment 2.1*
Analysis and Results

Trials with response errors (15% of all trials) were removed from the data prior to subsequent analyses. Error reaction times (RTs) were closely matched to correct RTs, and an analysis of the probability correct by RT revealed no significant negative correlation between error proportions and RTs, arguing against the correct RT data being contaminated by accuracy-speed trade-offs. Examination of the correct RTs revealed non-normal distribution with pronounced positive skew. Shapiro-Wilk tests showed the RT distributions for each condition, and on aggregate, to be significantly non-normal (all p’s < .001) and on this basis subsequent analyses were conducted on the anti-logs of the means of the log-transformed RT distributions (for supporting ideas see Box & Cox, 1964, 1982).

Repeated-measure ANOVAs were carried out on the log-transformed mean RTs with the factors ratio and set size (8, 16, or 32 sub-areas). Mauchly’s test for sphericity indicated that the assumptions for sphericity had been violated for the main effect of ratio and the interaction of set size and ratio, so degrees of freedom were corrected using Greenhouse-Geisser estimates for sphericity. There was a significant main effect for ratio (F2.1, 23.0 = 41.24, p < .001, ηG2 = 0.17), and set size (F2, 22 = 16.9, p < .001, ηG2 = 0.11) and their interaction was also significant (F3.04, 33.46 = 13.25, p < .001, ηG2 = .10). Generalised eta-squared values (ηG2; see Bakeman, 2005) indicated that these effects were small to medium based on Cohen’s (1988) guidelines where .02 is small, .13 is medium, and .26 is large.

Figure 2.3 (a) shows that, apart from Ratio 5, there was no linear increase in RT as a function of set size, which would have been the result if participants were engaged in serial search for the target. The 16-item grid produced slower RTs than either of the other set sizes, likely explaining the main effect of set size. Figure 2.3 (b)
and (c) show that across set sizes, RTs tended to decrease overall as ratio increased, but that the nature of this decrease implies different models in each case. Furthermore, the golden section stimuli appear to exhibit a discontinuity in the pattern of RT decrease by ratio for the 8 and 16-item sets. Based on these graphs and the significant interaction of ratio and set size, it may be interpreted that the gradual slowing of RT by ratio is a linear or curvilinear relationship that changes across set size, and that this relationship may be furthermore sensitive to a golden section effect. Table 2.1 shows that RTs decreased nonlinearly with increasing ratio for the 32-item grids, with a quadratic function explaining a significant amount of variance in RT by ratio. The 16-item grid was explained by a near perfect linear function only following removal of the golden ratio RT, and not on the basis of analysis of the full range of ratio RTs. The 8-item set showed an improvement in fit for a quadratic function following excision of the golden section, but this did not reach significance. As mentioned, RTs did not increase linearly with increasing set size, so it seems reasonable to attribute the effect of golden sectioning to local target discrimination rather than search. Based on this, it appears that the effect on discrimination at the target location relates to the ratio-based compositional structure of the grid.
Table 2.1

Linear and curvilinear regression results for log-transformed means in Experiment 2.1 for the full range of ratios, and for the range excluding the golden ratio

<table>
<thead>
<tr>
<th>Set Size</th>
<th>Including Ratio 4 (+1:0.618)</th>
<th>Excluding Ratio 4 (-1:0.618)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Linear R^2</strong></td>
<td><strong>Logarithmic R^2</strong></td>
</tr>
<tr>
<td>8</td>
<td>.584 (F = .215)</td>
<td>.752 (F = 9.07)</td>
</tr>
<tr>
<td>16</td>
<td>.448 (F = 2.43)</td>
<td>.407 (F = 2.06)</td>
</tr>
<tr>
<td>32</td>
<td>.174 (F = .633)</td>
<td>.370 (F = 1.76)*</td>
</tr>
</tbody>
</table>

Note: *p < .05 **p < .01 ***p < .001
Figure 2.3 Exponents of mean log-transformed RTs by set size and ratio for Experiment 2.1.

(a) Exponents of mean log-transformed RT distributions by set size for each of the five ratios, (b) exponents of mean log-transformed RT distributions by ratio for the three set sizes, and (c) Exponents of mean log-transformed RT distributions by ratio, collapsed across set sizes.
Figure 2.4 Mean error percentages for each ratio for Experiment 1.
Error bars represent standard errors

Error Analysis

An ANOVA of arcsine-square transformed error probability data (with the same main terms as RT data ANOVA) revealed similar effects to those found from analysis of the mean RT data: there were main effects for ratio (Greenhouse-Geisser adjusted; $F_{2.1, 23.1} = 39.88, p < .001, \eta_G^2 = 0.23$), set size ($F_{2, 22} = 117.31, p < .001, \eta_G^2 = 0.28$) and a significant ratio by set size interaction ($F_{8, 88} = 23.66, p < .001, \eta_G^2 = 0.26$).

Simple main effects analyses were conducted to assess the two-way interaction effect. Errors for the 16-item set were shown to be significantly higher than either of the other sets (both $p$’s < .001), while there was no significant difference for error production between the 8 or 32-item sets ($p$ = n.s.). There was a decrease in errors as ratio increased for the 8 and 16 item sets, indicating that ratio sections closer to unity were more difficult to discriminate. This effect was not shown in the largest set, where errors stayed similar across ratios. The slope of the overall reduction in errors by ratio was more extreme in the 16-item ($\beta = -0.09$) than 8-item set ($\beta = -0.04$). For the 8-area set, Ratio 1 (1:0.468) was significantly different to
Ratio 3 (1:0.568; p < .01) and Ratio 5 (1:0.668; p < .01). None of the other comparisons for the set showed a significant difference. For the 16-item set, the golden-sectioned stimuli produced significantly higher errors in comparison to neighbouring ratios (p < .05), but produced lower errors than Ratio 2 (1:0.518; p < .01) and no significant difference to Ratio 1 errors (p = n.s.). For the 32-item set, there was no significant difference for any of the comparisons (all p’s = n.s.). Figure 2.4 shows two peaks in an otherwise downward trend; one at Ratio 2 and one at Ratio 4.

2.7.1 Discussion of Experiment 2.1

Our results showed that the relationship between RT and ratio was complex, and appeared to be characterised by different functions for each set size (albeit all exhibiting negative slopes), suggesting that different factors were influential in each case. In the middle set size, the relationship appeared to be sensitive to a golden section effect; namely, a significant curvilinear function relating RT to ratio was apparent only when the golden-sectioned RTs were excluded from the data. This ratio produced a discontinuity in the functional relationship that might otherwise have characterised responding. While the middle set size exhibited more errors than the other sets, unlike RT analysis, the golden-sectioned grids in this set did not produce the greatest number of errors overall. While the golden-sectioned grids produced more errors than neighbouring ratios, there was no significant difference between golden ratio errors and 1:0.468 errors, and a significant reduction in errors when compared with 1:0.518 errors. This implies that the relative slowing of golden-sectioned RTs is unlikely to be due to an effect that produces both high RTs and error rates, whereas this kind of general relationship (likely the discriminability of the target area from its neighbour) was shown for ratios close to unity sectioning. The lack of a linear increase in RT for set size increases meant that the results could not be related to
attentional processing in search. As such, it appeared unlikely that the influence of golden sectioning occurred at the level of the global display as a scene to be processed via attentional deployment. Instead, the effect is due to processing at the target section.

2.8 Experiment 2.2: Paired Comparisons of Ratio-Sectioned Mondrian Grids

Experiment 1 demonstrated a possible effect of golden-sectioned grids in the middle set size (and a similar but non-significant effect in the 8-item set) which produced extended RTs, reflecting a slowing in target localisation and responding. As the golden ratio is a normative proposition, we investigated the possibility that the effect on RT would be related to conscious aesthetic experience as expressed in a preference for golden-sectioned grids over other ratios.

Method

Experiment 2.2 was conducted in the same session with the same participants as Experiment 2.1 (N = 12)\(^8\), and used a paired comparisons procedure to investigate whether or not a significant aesthetic preference existed for the golden ratio relative to the other ratios used in Experiment 2.1. Each stimulus grid employed in Experiment 1 was presented alongside every other grid on 5 separate occasions (presentation orders were fully randomized. There were 40 trials per condition leading to 600 trials overall). Participants were asked to report which grid of the two they thought was the most aesthetically pleasing. Aesthetic was defined to the subject in simple terms of which of the two grids the participant preferred. The stimuli, stimulus presentation, and experimental conditions were identical to those employed in Experiment 2.1 (note: the running order of Experiments 2.1 and 2.2 were varied such that Experiment 2.1 was conducted first for 50% of participants).

\(^8\) A simulation study by Rajae-Joordens & Engel (2005) determined z-score differences differences via Thurstone scaling (see Appendix 1), and indicated that s sample of 12 participants would enable a medium-sized effect to be determined with 80% probability.
Analysis and Results

Paired comparison data were analysed by establishing an \( m \times m \) preference matrix, representing the grouped preferences for each stimulus comparison. Scale scores were derived by the Bradley-Terry-Luce (BTL) model of preference scaling (Bradley & Terry, 1952; Luce, 1959) using the *eba* package in R (version 1.7-1; Wickelmaier, 2012). The BTL model scores, unlike proportion preference scores, have roughly equal variance across the scale (unlike e.g. the Thurstone-Mosteller model; see Appendix 1). Therefore the standard assumptions for ANOVA are sufficiently satisfied, and repeated-measures analysis of variance was conducted on scale scores by ratio and set size (Palmer, Gardner, & Wickens, 2008). Higher BTL scores represent a greater degree of preference for a stimulus.

Mauchly’s test of sphericity was significant for all effects, so Greenhouse-Geisser estimates were adopted. There was no significant main effect of ratio, set size, or their interaction (all \( p \)’s = n.s.). BTL scores showed that Ratio 4, contrary to expectations, had the lowest BTL value of any stimulus (for the 8-item set). When averaged across set sizes, the middle ratio was the least preferred, with the golden-sectioned grids being second least preferred. Figure 2.5 (b) shows that the preference scores followed a u-shaped function, suggesting that participants tended to select the more extreme ratios, rather than the middle of the range (as predicted by Godkewitch, 1979), or unity-sectioned (e.g. Davis & Jahnke, 1991) or golden-sectioned stimuli.
Based on the results of the ANOVA, a lack of main effects may indicate that participants are responding randomly or with poor consistency across trials. A statistical test of this, where the null hypothesis predicts no consistency in responding, adopts the method of Kendell & Babington-Smith (1940). This procedure assumes that inconsistency is captured by the number of circular triads within the response matrix, which are breaks in transitivity across preferences (e.g. A>B, B>C, but C>A). This is calculated by:

$$\zeta = 1 - \frac{C}{C_{\text{MAX}}}$$

Where zeta ($\zeta$) is the coefficient of consistency, C is the number of circular triads observed, and $C_{\text{MAX}}$ is the maximum possible number of circular triads (see Appendix 1 for further details on the derivation of this value). This measure requires the observation of all possible pairs, and it varies between 1 (no circular triads) to 0 (maximum) and was computed by the eba package in R. Because the procedure runs on binary choice data rather than preference counts, and summation might produce indifference responses where none were actually available to the participant, the data were split into sequential complete runs of the binary comparison matrix, enabling the
calculation of $\zeta$ for each run, giving an average coefficient for each subject (see Table 2.2). All but three of the subject coefficients are above 0.5, implying a moderate level of transitivity.

Kendall & Babington-Smith showed that circular triads are closely modelled by a chi-square distribution (1940; see also Alway, 1962; Iida, 2009; Knezek, Wallace, & Dunn-Rankin, 1998). However, for $n < 8$ objects, chi-squared and significance values are not accurate (Iida, 2009), so will not be employed in the present analysis. Investigation of the data shown in Table 2.2 reveals that two of the participants (P6 and P7) averaged 3 circular triads per comparison cycle (where 5 is the maximum). When the data were re-analysed by excluding the two participants with the highest number of circular triads, there was still no significant effect of set size, ratio, or their interaction.

Table 2.2

<table>
<thead>
<tr>
<th>Subject</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\zeta$</td>
<td>0.54</td>
<td>0.79</td>
<td>0.58</td>
<td>0.71</td>
<td>0.63</td>
<td>0.49</td>
<td>0.41</td>
<td>0.53</td>
<td>0.76</td>
<td>0.54</td>
<td>0.59</td>
<td>0.54</td>
</tr>
<tr>
<td>$CT_M$</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

While the aggregate preferences do not demonstrate a preference for the golden section (or stimuli close to unity-sectioning), there may nonetheless have been some tendency for a preference in some of the individual participants across the experiment. As such, BTL scores for each participant were plotted against ratio, and the plots are shown in Appendix 2. Some representative examples are shown in Figure 2.6. The figures do not tend to exhibit clear preferences across participants or for the same participants across set sizes.
2.8.1 Discussion of Experiment 2.2

The results of this experiment suggest that people did not tend to prefer the appearance of the golden-sectioned grids in comparison with the other ratios. The presence of circular triads in preferences confirms that people were sometimes unable to consistently choose a preferred stimulus, and there was a large degree of inter-individual difference in which stimuli were chosen by the scaling method. This would seem to suggest that the effect of slowing on RT for golden-sectioned grids does not automatically translate to aesthetic experience. This finding argues against the proposal that more perceptually fluent stimuli would predict higher preference ratings, or that less fluent images would be least liked. While the middle set size had elevated RTs, it produced a similar pattern of preferences to the other ratios. Furthermore, while it should be noted that the lowest average BTL score received by any of the stimuli was the 8-item golden-sectioned grid, in the absence of a more general RT/BTL relationship it is difficult to qualify this result as an aspect of processing fluency in the golden-sectioned grids due to variability across participants, inconsistent responding within subjects, and the absence of a similar effect in the 16-item scores. Kendall and Babington-Smith (1940) suggested that breaks in transitivity have a number of implications: observers may themselves be inconsistent judges, the
stimuli may be difficult to distinguish, the property under investigation may not be amenable to linear scaling, or there may be combination of all of these. Most of our respondents demonstrated some intransitivity, so the lack of preference for one clear winning ratio across set sizes, or an aggregated winning ratio across participants may be due to their inability to assess the stimuli for genuine preference, rather than arbitrary choosing. Another possibility is that the instructions provided an unclear definition of the acceptance criterion for each comparison pair.

2.9 Experiment 2.3: The Influence of an Aesthetic Attitude on Paired Comparisons of Ratio-Sectioned Mondrians

In the previous experiment, there appeared to be little tendency for participants to like or dislike the golden-sectioned stimuli or systematically prefer any of the other ratios. Some intransitivity occurred, suggesting that participants may have been selecting the winning stimuli in each pair arbitrarily in some cases. Nevertheless, the possibility remains that participants were unsure of the basis of choice required by the experiment. Their instructions in Experiment 2.2 requested that they respond on the basis of which they found more aesthetically pleasing, but this terminology may have been too vague to allow the respondents to establish a consistent response criterion. The aim of the current experiment was to operationalise the aesthetic according to a number of terminological classifiers. We used 3 instructional definitions corresponding with some putative function or characterisation of aesthetic value, and a fourth group category in which no attempt was made to formally define a response criterion. On the basis of the presence of a moderate degree of transitivity in the previous experiment, we predicted that one or more of the groups might exhibit more systematically targeted responding or less variability in selecting a winning ratio.
Method

Participants

24 Participants completed Experiment 2.3 (12 Male, 12 Female, mean age = 27.11, SD = 2.27 years).

Stimuli and Procedure

Stimuli and procedures in Experiment 2.3 were identical to those of Experiment 2.2, but subjects were split into 4 groups, each receiving a slight difference in instructions on how to complete the task. Group 1 was instructed to choose which of the stimuli in each pair was the most “interesting”, for Group 2 the defining characteristic was which of the stimuli was “attractive”, for Group 3 it was “pleasing form”, and Group 4 chose based on “any reason”. Again, each participant completed 600 trials in total.

Results and Analysis

As in the previous experiment, paired comparisons were analysed by deriving scale scores via the Bradley-Terry-Luce model of unidimensional scaling using the eba package in R (version 1.7.1; Wickelmaier, 2012). Based on the rationale for Experiment 2.2 (and supported by Kolmogorov-Smirnov tests for all conditions), the data were taken to satisfy the normality assumptions for mixed factorial ANOVA (conducted in R), with ratio and set size as the repeated-measures factors, and group (1 – 4) as the between-groups factor. None of the repeated measures main effects or interactions were significant (all p’s = n.s.). The between subject effect of Group was also non-significant (p = n.s.). Figure 2.7 (a) and (b) indicates that there is no easily discernible tendency to like or dislike the golden-sectioned stimuli for groups 1 – 3. In group 4, however, there appears to be a noticeable tendency for the golden-sectioned grids to be at the nadir in overall preference score distributions, regardless of set size.
Paired comparisons demonstrated that this ratio was significantly different to ratio 1 (p < .05), but not any of the other sectioning factors.

As in Experiment 2.2, the data were further analysed for consistency by counting the number of circular triads within each participant’s dataset and as before, circular triads were computed by the eba package in R with the data split into 20 cycles of each subject’s experimental run (total 600 trials) for each participant. Each of these was subdivided into sequential complete runs of the binary comparison matrix, enabling the calculation of $\zeta$ for each run and allowing the calculation of an average value for this coefficient. In replication of the results from Experiment 2.2, there were some breaks in transitivity across participants, with only one of the respondents demonstrating perfect transitivity (see Table 2.3). When total number of circular triads were analysed in a one-way ANOVA, there was no significant main effect of experimental group on overall number of circular triads (F < 1).

2.9.1 Discussion of Experiment 2.3

While the results of Experiment 2.3 provided no significant main effects to propose that ratio, set size, or group were factors contributing to preferences, there does appear to be a tendency for the golden-sectioned stimuli in all set sizes to receive the lowest scores for Group 4, in which participants were requested to make their preference choices based on “any reason”. The other group categories tended to show the middle ratio as the least preferred in that group. There is some suggestion from these results that the lack of preference for the golden ratio displays arises when the participants are not oriented towards an explicit aesthetic goal. The close match in score function across set sizes would seem to suggest that Group 4 responded in a manner closely linked to the ratio dimensions, but the response criterion implies that this evaluative strategy may have been based on something other than the aesthetic
aspects of the image. One possibility is that the tendency is due to anchoring effects
deriving from the extreme ratios, which biased towards the lower ratios
(Godkewitsch, 1979). Another possibility is that the lower ratios were closer to unity,
so a non-aesthetic evaluative strategy selects these ratios according to their ease of
processing, while a more aesthetically ‘challenging’ stimulus geometry combines with
a pragmatic evaluation in producing negative affect. In summary, the results from
Experiments 2.2 and 2.3 suggest, but very tentatively, the possibility that a lack of
preference for the golden section is due to a response strategy that is most pronounced
when participants are not given aesthetically-coded instructions, and therefore are
freer to assess the image on the basis of fluency evaluations. If aesthetic experience is
characterised by the affordance of cognitive elaboration, then the issuing of response
criteria may inhibit the potential for unconstrained exploration of the image, as
participants may have used some key indexical characteristics of the stimuli to make
their judgements. This idea is similar to Garner’s (1974; Garner & Clement, 1963; see
also Verstegen, 2004) critical realist description of figural structure. Garner suggests
that structure inheres in the arrangement of the stimulus; it is real, but requires
receptivity on the part of the observer. He claims that unless the experimental task has
been ingeniously or fortuitously accurate in capturing a mechanism of interest, an
observer is most likely to detect structure when they are unconstrained by overly-
specific task demands.
Table 2.3
*Coefficients of consistency (ζ) and mean circular triads (CT_M) by participant in Experiment 2.3*

<table>
<thead>
<tr>
<th>Subject</th>
<th>ζ</th>
<th>0.66</th>
<th>0.79</th>
<th>0.81</th>
<th>0.85</th>
<th>0.86</th>
<th>0.81</th>
<th>0.82</th>
<th>0.76</th>
<th>0.74</th>
<th>0.83</th>
<th>0.75</th>
<th>0.70</th>
<th>0.83</th>
<th>0.85</th>
<th>0.87</th>
<th>0.60</th>
<th>0.88</th>
<th>0.78</th>
<th>0.81</th>
<th>0.78</th>
<th>0.53</th>
<th>0.96</th>
<th>0.64</th>
<th>0.72</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT_M</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Figure 2.7 Preference rating results for Experiment 2.3.*
(a) Bradley-Terry-Luce scale scores by set size, ratio and group, and (b) by ratio and group in Experiment 3
2.9.2 Experiment 2.4: Visual Perimetry and Target Localisation in Ratio-Sectioned Mondrian-Type Stimuli.

The results of Experiment 2.1 suggest that a golden section effect is apparent for RTs for the 16-item stimuli. One possible explanation for this is that there is some degree of congruence between these stimuli subsections and the morphology of the visual field – the perimetric hypothesis (Stone & Collins, 1965; Hintz & Nelson, 1970). As the results of Experiments 2.2 and 2.3 provided no clear evidence that golden sectioning would have especial attractiveness as represented in our grid stimuli, the present experiment takes a novel approach to this hypothesis in the use of RT as an indirect measure. Experiment 2.4 was concerned with the idea that if golden sectioning related to the egocentric, horizontal reference frame defined by the visual field, the results of Experiment 2.1 may not generalise to grids presented at other orientations. Aside from studies that have linked golden sectioning with visual field morphology, Experiment 2.3 was also inspired by the more general observation made in Klatzky’s (1998) review that “if the content of an image is congruent with a person’s current field of view (e.g., the perspective to be taken in the image matches the current visual perspective), perceptual processes will support formation of the requisite representation. But if there is a mismatch between the demands of the image and the subject’s current perceptual field, some imaginal process must be performed that transforms the relative positions of person and/or objects”, (1998, p.8).

Experiment 2.4 hypothesized that if participants presented with vertical instead of horizontal grids were subject to a mapping between visual field and grid orientation they would not be slowed in their response to golden-sectioned grids as found in Experiment 2.1.
Method

Participants

Eight naïve participants (4 male, 4 female, mean age 26.125 +/- 4.2 years), all normal or corrected-to-normal vision) participated in the experiment.

Stimuli and Procedure

The methods used in Experiment 2.4 were identical to those used in Experiment 2.1, with the exception that the grids presented now were vertical instead of horizontal.

Results and Analysis

Repeated-measure ANOVAs were carried out on log-transformed mean RT data, with the factors ratio and set size (8, 16, or 32 sub-areas). Trials with response errors (7.9% of all trials) were removed from the data prior to subsequent analyses. As with Experiment 2.1, error RTs tended to be on the whole slower than correct RTs, and an analysis of the probability correct by RT revealed no significant negative correlation. This argues against the correct RT data being contaminated by accuracy-speed trade-offs. An examination of the correct RTs revealed a non-normal distribution with a pronounced positive skew. Shapiro-Wilk tests showed the RT distributions to have pronounced positive skew, and on this basis subsequent analyses were conducted on the anti-logs of the means of the log-transformed RT distributions.

The results of Experiment 2.4 were very closely matched to those of Experiment 1, which was confirmed by a repeated-measures ANOVA that was carried out on the factors ratio and set size. Mauchly’s test for sphericity indicated that the assumptions for sphericity had been violated for all effects and so degrees of freedom were corrected using Greenhouse Geisser estimates for sphericity. There was a significant main effect for ratio ($F_{1.6, 11.4} = 24.06, p < .001, \eta^2_G = 0.19$), and set size,
(F\(_{1.11, 7.8} = 25.00, p < .01, \eta_G^2 = 0.15\)) while the interaction was also significant, F\(_{1.82, 12.73} = 7.39, p < .001, \eta_G^2 = 0.08\).

Figures 2.8 (a) and (b) show the pattern of effects match those of Experiment 2.1, and subsequent analyses were employed identical to those employed to describe the pattern of RT data in Experiment 2.1. As with Experiment 1, RTs decreased in curvilinear relationship with increasing ratio for the 8 item grid. As before, there was also a quadratic trend evident for the 32-item displays for the full complement of ratios, but this was not significant. Again consistent with Experiment 2.1, the 16-item grid was explained by a linear function only following removal of the golden ratio RT, and not for the full range of ratio RTs for that set. Unlike Experiment 2.1, there was also a stronger logarithmic fit of RT over ratio following removal of the golden ratio RTs in the 8-item grid RTs, which was not found for the 32-item grids. This lends some evidence to the (non-discriminate) effect being present in set sizes other than the 16-item one, but if so, it is far less pronounced.
Table 2.4
Linear and curvilinear regression results for log-transformed means in Experiment 2.3 for the full range of ratios and for the range excluding the golden ratio

<table>
<thead>
<tr>
<th>Set Size</th>
<th>Including Ratio 4 (1:0.618)</th>
<th>Excluding Ratio 4 (1:0.618)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Linear R²</td>
<td>Logarithmic R²</td>
</tr>
<tr>
<td>8</td>
<td>.707 (F₁,3 = 7.22)</td>
<td>.776 (F₁,3 = 10.4)*</td>
</tr>
<tr>
<td>16</td>
<td>.637 (F₁,3 = 5.27)</td>
<td>.667 (F₁,3 = 6)</td>
</tr>
<tr>
<td>32</td>
<td>.550 (F₁,3 = 3.66)</td>
<td>.722 (F₁,3 = 7.79)</td>
</tr>
</tbody>
</table>

Note *p < .05 **p < .01 ***p < .001
Error Analysis

As in Experiment 2.1, errors were analysed in a repeated measures ANOVA of arcsine transformed inaccuracy proportions with the same main terms as the RT analysis. There were significant main effects for both ratio and set size ($F_{4, 28} = 40.91$, $p < .001$, $\eta^2 = 0.37$; $F_{2, 14} = 34.36$, $p < .001$, $\eta^2 = 0.31$) and a significant ratio x set size interaction, $F_{8, 56} = 7.9$, $p < .001$, $\eta^2 = 0.21$. As was the case in Experiment 2.1, simple pairwise comparisons showed that the 16 item grids produced substantially more errors than the other grid conditions ($p < .05$ and $p < .001$ for comparison with the 8 and 32 item grids respectively), while the 8 and 32 item grids did not have significantly different patterns of error production. In all grid conditions, errors decreased with increasing ratio. Error production was not increased for the golden ratio grids arguing against a significant influence of target section discriminability resulting in the slowed RTs to golden section grids.

Figure 2.9 Mean error percentages for each ratio for Experiment 2.4.

Error bars represent standard errors.
2.9.3 Discussion of Experiment 2.4

The similarity between the results of Experiments 2.1 and 2.3 suggest that grid orientation is of little influence on section discrimination, although this conclusion requires qualification because of the substantial increase in RTs alongside a more substantial reduction in errors in Experiment 2.3 relative to those of Experiment 2.1. The vertical displays appear to be more difficult to process, so cause a general increase in reaction time, possibly resulting in fewer errors due to extended processing. These results do not bear strongly on the golden section RTs as the pattern of errors does not change substantially for those displays in this experiment. Instead, the results of Experiment 2.3 in combination with Experiment 2.1 do not support the proposition that the golden section effect depends on the similarity of golden sectioning to the morphology of the visual field.

2.10 General Discussion

The results presented in this chapter showed that participants had no general tendency to like or dislike golden sectioning as represented by our stimuli. Instead, in Experiments 2.2 and 2.3, people tended to prefer stimuli at the extremes of the range. In Experiments 2.1 and 2.4, in the middle set, there was some slowing in RTs for golden-sectioned grids which exhibited a discontinuity in the linear function that otherwise described RT by ratio (a general relationship whereby RTs were faster for higher ratios). The general tendency was for responses to become faster when the target is in a smaller areal ratio to its neighbouring sub-area The pattern of responding ruled out the interpretation that performance was due to visual search of the grids, with the middle set having the slowest RTs of the three sets, and the results of Experiment 4 argue against the possibility that the effect is due to visual field morphology. The effect appears to dominate at the lower ratios, while the 32-item
grid exhibits a non-linear functional relationship which is unaffected by golden sectioning. As such, it appears likely that the results derive from a relationship between golden sectioning and RT, but it is complex and possibly relates to discriminability at the target subsection.

There was limited evidence for serial search in the smaller sets (8 and 16-item) but an apparent golden section effect, while in the largest set (32-item) there was no effect of golden sectioning, but some evidence for serial search in the curvilinear relationship between ratio and RT. This suggests that the location of the target is determined quickly in the smaller sets, but that responding is delayed for golden section stimuli. Errors increased for 16-item grids, but this was an aspect of that grid size rather than golden sectioning specifically.

The lack of a pronounced tendency for golden-sectioned stimuli to affect preferences in the paired comparison experiments presents an interesting problem, as the putative effect is traditionally couched in terms of attractiveness. One possible explanation for this is found in Experiment 2.3, where there was some evidence to suggest that the over-constraint of task-orientation may have precluded golden-sectioned stimuli from eliciting conscious aesthetic experience. As such, the effects on RT may be largely compatible with the perceptual fluency approach to aesthetic preference, and the nadir in preference ratings when participants are issued with no specific response criterion for the pairwise comparisons was for the golden-sectioned stimuli. This matches with the idea that less fluent stimulus processing, reflected by slowed RTs, would lead to a lack of preference. This idea is controversial, however, as it presupposes a unidirectional correlation between fluency and aesthetic pleasingness that is possibly not applicable to much of aesthetic experience (Carroll, 1990; Graham, 2005). Furthermore, the pattern of preferences overall do not match
the RTs to a close enough degree to offer this as unequivocal support for the perceptual fluency approach.

In conclusion, these results offer some support for the golden section hypothesis, but further investigation of the relationship between the golden ratio and early visual processing needs to be conducted before a substantial contribution to this area is made. The following chapter presents the results of three experiments which aimed to explore this phenomenon more fully.
Chapter 3: The Golden Section and Perceptual Processing

3.1 Introduction

In the previous chapter, Experiments 2.1 and 2.4 showed that the golden-sectioned stimulus response times (RTs) in the middle set size produced a deviation from an otherwise linear negative function relating RT to ratio, specifically exhibiting slowed RTs for these stimuli. In Experiments 2.2 and 2.3, no corollary effect was found in reports of aesthetic preference (apart from a non-statistically-significant tendency for group 4 to dislike golden sectioning). Due to the lack of a linear model explaining longer RTs as a function of increasing set size; it is not possible to attribute the effect on performance to search mechanisms. On the basis of Experiment 2.4, there is no evidence that the effect results from the similarity between the golden-sectioned grids and the general morphology of the visual field (the perimetric hypothesis). The present chapter details three experiments with the aim of isolating the nature and locus of the responsible mechanisms. Specifically, and on the basis of the results detailed in the previous chapter, we tested the general hypothesis that the golden section effect has its influence at the level of perceptual processing via the dynamics of neural coding. This is based on the theoretical proposal that natural image statistics constrain the functions of the visual system and are therefore capable of relating processing to stimulus characteristics that might be relevant to areal contrast coding.

3.2 Experiment 3.1: Target Localisation in Equiluminant Mondrians

It may be possible to isolate the discrimination process influenced in the previous experiments by removing luminance contrast as a source of information for target localisation. Aside from expediting localisation, high luminance contrast is shown to relate to visual discomfort in the same way as the power spectra of Fourier
analyses; lower luminance contrast is unusual in nature, and correlates with higher discomfort ratings (Juricevic, Land, Wilkins, & Webster, 2010). Juricevic et al. (2010) suggest that luminance has a complex relationship with other stimulus dimensions, so the stimuli used in the previous experiments may have affected the results by having uncontrolled luminance levels. Equiluminance (or isoluminance) is shown to exhibit a motive quality through the differential activation of chromatic or luminance coding, proposed to be instantiated in visual pathways implicated in spatial or motion processing respectively (Livingstone, 2002). Abrupt luminance transitions affect target pop-out far more effectively than colour transitions (Theeuwes, 1995). As such, there is the possibility that luminance contrast interacted with the effect, which may explain why the set size with the slowest RTs was the one in which the effect was demonstrated. Luminance changes therefore expedite responding by inducing “pop out” and enabling the localisation of the target. By using isoluminant stimuli in the present experiment, we examined the possibility that the previous effect was due to luminance contrast between the target section and its surrounding sections. More specifically, we investigated whether a set size effect would be produced if luminance contrast was eliminated as an influence. We predicted longer RTs, but hypothesised that if the results were similar to previous, then we could exclude luminance contrast as a potential confound in the previous results.

Method

Participants

Nine naïve participants (4 male, 5 female, mean age 23.10 +/- 4.1 years), all normal or corrected-to-normal vision) participated in the experiment.
**Stimuli and Procedure**

The methods used in Experiment 3.1 were identical to those used in Experiment 2.1, with the exception that the grids presented now were equiluminant\(^9\). The task involved determining whether the target section was “light green” or “dark green”, defined by RGB values.

**Results and Analysis**

Repeated-measure ANOVAs were carried out on log-transformed mean data with the factors Ratio and Set Size (8, 16, or 32). Each of these will be described in turn. Trials with response errors (11.17% of all trials) were removed from the data prior to subsequent analyses. As with Experiments 2.1 and 2.4, error RTs tended to be on the whole slower than correct RTs, and an analysis of the probability correct by RT revealed no significant negative correlation, so speed-accuracy tradeoffs are unlikely. Shapiro-Wilk tests showed the RT distributions for each condition, and on aggregate, to be significantly non-normal (all \(p’s < .001\)). As in previous experiments, the antilogs of the mean log-transformed RTs were submitted for analysis.

As can be seen from Figure 3.1, the results of Experiment 3.1 were very closely matched to those of Experiment 2.1 and 2.4, and this was confirmed by repeated-measures ANOVA that was carried out on the factors ratio and set size. Mauchly’s test for sphericity indicated that the assumptions for sphericity had been violated for the effect of ratio and the interaction of ratio and set size, so degrees of freedom were corrected using Greenhouse-Geisser estimates for sphericity. There was a significant main effect for ratio (\(F_{2.04, 16.32} = 19.89, p < .001\)) and Set Size (\(F_{2, 16} = \)

---

\(^9\) True equiluminance is difficult to achieve in practice, due to personal variability in points of equality of luminance, as well as heterogeneity across the visual field (Livingstone, 2002). The stimuli used in this experiment are referred to as equiluminant in order to define them in relation to other experiments, but in reality are simply closer to equiluminance than the previous stimuli had been. This is taken to be a sufficient manipulation of the variable of interest (discriminability) to be adequate for our purposes. To foreshadow subsequent results, the RTs for this experiment were slowed in comparison to Experiment 1, so the functional definition is taken to be a valid one.
16.99, p < .001) while the interaction was also significant (F_{3.7, 29.62} = 7.54, p < .001).

Overall, RTs from the middle set size were slower than the other set conditions, and again the ratio 4 (golden) RTs appeared to cause a discontinuity in a linear or curvilinear model fit (see Figure 3.1 [a] and [b]). Unlike Experiments 2.1 and 2.4, the 16-item grid was not explained by a significant linear function following removal of the golden ratio RT, however excising this ratio did allow a logarithmic fit, which was not possible for the full range of ratio RTs (see Table 3.1).
Table 3.1
Linear and curvilinear regression results for log transformed mean RTs in Experiment 3.1 for the full range of ratios and for the range excluding the golden ratio

<table>
<thead>
<tr>
<th>Set Size</th>
<th>Linear R^2</th>
<th>Logarithmic R^2</th>
<th>Quadratic R^2</th>
<th>Linear R^2</th>
<th>Logarithmic R^2</th>
<th>Quadratic R^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>.475 (F_{1,3} = 2.72)</td>
<td>.601 (F_{1,3} = 4.52)</td>
<td>.596 (F_{2,2} = 1.47)</td>
<td>.698 (F_{1,2} = 1.47)</td>
<td>.892 (F_{1,2} = 16.56)</td>
<td>.993 (F_{2,1} = 66.8)</td>
</tr>
<tr>
<td>16</td>
<td>.641 (F_{1,3} = 5.36)</td>
<td>.672 (F_{1,3} = 6.14)</td>
<td>.656 (F_{2,2} = 1.91)</td>
<td>.873 (F_{1,2} = 13.75)</td>
<td>.939 (F_{1,2} = 30.69)*</td>
<td>.962 (F_{2,1} = 12.75)</td>
</tr>
<tr>
<td>32</td>
<td>.348 (F_{1,3} = 1.61)</td>
<td>.522 (F_{1,3} = 3.27)</td>
<td>.902 (F_{2,2} = 9.17)</td>
<td>.255 (F_{1,2} = 6.86)</td>
<td>.446 (F_{1,2} = 1.61)</td>
<td>.897 (F_{2,1} = 4.37)</td>
</tr>
</tbody>
</table>

Note *p < .05 **p < .01 ***p < .001

Figure 3.1 Exponents of mean log-transformed RT distributions by set size and ratio for Experiment 3.1

(a) Includes the three set sizes and (b) is collapsed across set sizes
Error Analysis

As in Experiments 2.1 and 2.4, errors were analysed in a repeated-measures ANOVA of arcsine-transformed inaccuracy proportions with the same main terms as the RT analysis. Main effects for both ratio and set size (F_{2,35,16.48} = 12.26, p < .001; F_{1.82,12.71} = 4.90, p < .05) and a significant ratio by set size interaction (F_{4.11,28.74} = 2.95, p < .05) were observed (all effects are Greenhouse-Geisser adjusted). Simple effects analyses showed that there was no significant linear function describing error proportion by ratio for set size 32, but both the 8-item and 16-item displays demonstrated a linear decrease in errors as ratio increased (R^2 = 0.16, F_{1,38} = 7.05, p < .05; R^2 = 0.30, F_{1,38} = 16.58, p < .001). Therefore the results do not suggest an increase in errors for the golden-sectioned stimuli in the middle set that matches the RT data. As before, this argues against a significant influence of target section discriminability as resulting in the slowed RTs to golden section grids (see Figure 3.2).

![Figure 3.2](image)

*Figure 3.2* Mean error percentages for each ratio for Experiment 3.1.

Error bars represent standard errors
3.2.1 Discussion of Experiment 3.1

As evidence suggested that the effect of golden sectioning is at the level of local target discrimination and there may have been a complex interaction of luminance with the previous effect, Experiment 3.1 aimed to assess the potential contribution of luminance contrast to target localisation. Response times were slower than in Experiments 2.1 and 2.4, which was expected based on the relative difficulty in differentiating equiluminant targets. Equiluminant displays are known to be differentially represented by cortical pathways exiting the occipital cortex to the temporal and parietal lobes (Ungerleider & Mishkin, 1982). Specifically, the former (ventral) stream is sensitive to high spatial frequency and colour contrast, enabling it to establish target identification, whereas the latter (dorsal) stream depends on coarse frequency scales and luminance contrast, making target localisation more difficult for images in which the differentiation between segments is defined by colour and not luminance (Goodale & Milner, 1992; Milner & Goodale, 2008, Zeki, 1973; 1978; 1983; it should be noted however that the amount of co-operation between dorsal and ventral areas in processing of colour is not negligible, see e.g. Claeys et al., 2004). Equiluminance has the effect of making the target spatially ambiguous or appear to have a motive quality (e.g. Livingstone, 2002). Our aim was to determine whether the previous effect was due to an interaction between this property and the golden-sectioned stimuli, but the results show that the effect appears to be due to something other than luminance contrast.

Unlike previous experiments, the removal of the golden-sectioned log-transformed mean RTs did not allow for the statistically significant fitting of a linear function, but it did produce a logarithmic fit. This non-linearity may be due to the introduction of a luminance profile that is unusual in nature, and therefore encourages
an atypical perceptual response (e.g. Juricevic et al., 2010). On the basis of these results, it appears that some generalised aspect of perceptual responding to golden-sectioned ratio grids - in the 16-item displays especially – is affected, but is not entirely the result of luminance contrast sensitivity.

3.3 Natural Image Statistics as Constraints on Visual Processing

3.3.1 Effects of Fractal Dimension

From the results of the RT data in Experiments 2.1, 2.4, and 3.1, it appears that the observed effect of the golden section grid is due to processing of the target location, rather than global search mechanisms. The specificity of the effect to the 16-item grids, however, argues for some role of the whole composition rather than golden sectioning specifically. In all of the stimuli, the grids involved iterative sectioning of the entire visual object. For the golden ratio each iterative subdivision of its elements produced an additional self-similarity – the larger of the two sections bears the same relation to the smaller section as the previous iteration’s smallest section bears to the larger of the current one; in other words they are self-similar at all scales, or fractal. This produces a level of continuity in the recursive algorithm that the other ratio subdivisions do not have, and may exhibit less visual complexity as a result. The results of the sectioning procedure adopted here is a Mondrian-style appearance, which is visually simple in comparison with the more complex appearance of natural fractals or some artistic works, e.g. Jackson Pollock’s paintings (Taylor, Spehar, Clifford, & Newell, 2008). Fractal dimension covers a range, with values close to 1 correspond with the sparseness of Mondrian-type images, while more complex fractals are closer to 2 (see Taylor, 2004; also Taylor et al., 2008 for a list of natural patterns and their fractal dimensions). While there is no easily discernible value that correlates with aesthetic preference, some values associated
with preference have been found at 1.3 (Aks & Sprott, 1996; Hagerhall et al., 2004), 1.3-1.5 (Taylor, 2001), and 1.8 (Pickover, 1995). The power spectra of natural scenes are similar to those shown in images that generate high preference ratings across cultures (Melmer, Amirshahi, Koch, Denzler, & Redies, 2013). Aside from visual art (e.g. Redies, Hasenstein, & Denzler, 2007; Taylor, Micholich, & Jonas, 1999a, 1999b), fractality is also shown to be present in some types of music (Beauvois, 2007) and architecture (Joye, 2007). This may imply that the normative aesthetic value assigned to the golden ratio in these experiments may be due to its fractal dimension. The RT results in the previous experiments may imply that fractal dimension affects performance, and the restriction of the effect to the golden-sectioned stimuli in the intermediate set size can be predicted on the basis of a difference in its fractal dimension.

In investigation of this, RT data were compared to fractal dimension (D) for all ratios. Fractal dimension is a means of characterising patterns by relating their complexity to a ratio of change in detail to change in scale. For a pattern to be considered fractal, its fractal dimension must exceed its topological dimension. This goes beyond regular self-similarity when it is not simply a similarity across spatial scales, but a detailed self-repeating generative pattern (e.g. a Mandelbrot set, Sierpinski triangle, or Koch snowflake). Euclidian geometry is insufficient to characterise the complexity of such sets, so alternative approaches have been developed for fractal analysis. Figure 3.4 illustrates the Minkowski-Bouligand dimension (or box-counting dimension, D_{BC}) of the Ratio conditions. This procedure determines the fractal dimension of a set by a count of the number (N) of boxes of size R that are required to cover it. For images, boxes only cover foreground pixels. If an object is one-dimensional, for example a line, then N = 1/R, because as the squares
decrease in size, more are necessary to cover the object. In a 2D object, this becomes 
\((1/R)^2\) and for three dimensional objects becomes \((1/R)^3\), and so on. \(N\) is 
approximated by the power law:

\[
N = k \left(\frac{1}{R}\right)^d
\]

As the number of boxes increase, the number of dimensions decrease (in Figure 3.3 
this is in the range 2 to 1, where 1 might characterise a line, and 2 a filled square). In 
the log-log approach, the pattern is sampled via arbitrary scaling and the slope of the 
logarithmic regression line relating \(N\) to scale or size is

\[
D_{BC} = \frac{\ln(N)}{\ln(1/R)}
\]

Where scale is approximated by the calibre or size of the boxes, \(R\) (Karpieren, 2004). 
As grid calibre decreases, the box count increases, and the scaling rule is 
approximated by the logarithmic regression line.

*Figure 3.3* Box counting dimension for ratio-sectioned Mondrian stimuli.
3.3.1.1 Fractal Analysis of Experimental Stimuli

Fractal Analyses of the 16-item stimuli were conducted using the boxcount package (Moisy, 2006) in Matlab. As the number of boxes increases, the number of dimensions decreases from 2 to 1. Logistic regression analysis of sample points in this range against RT showed a correlation, but a non-significant one (the best correlation was for fractal dimension value of 1; McFadden’s pseudo $R^2 = .418, \chi^2_4 = 6.73, p = \text{ns}$). The ability of this function to explain the results was only slightly affected when golden-sectioned stimuli RTs were excluded ($R^2 = .406$), which suggests that there may be some general relationship between fractal dimension and RTs, but it does not explain the results obtained in previous experiments. A point to note with regard to explaining any potential relevance to aesthetic concerns is that while box-counting is mathematically equivalent to spatial frequency power spectra (Knill, Field, & Kersten, 1990), the former is proposed to be less relevant to aesthetics as fractal patterns are scale invariant, but also self-similar (Graham & Redies, 2010). While box-counting data are shown to correlate with some abstract paintings (e.g. Taylor et al. 2008) fractal dimension may have limited applicability as a general aesthetic principle, as self-similarity is not common in art or natural scenes. As such, the spatial frequency power spectra may be more relevant to explaining the pattern of RT data.

3.3.2 Spatial Frequency Channels in Visual Perception and Aesthetics

According to the spatial frequency theory of vision, perceptual processing involves a large number of overlapping psychophysical channels with different frequency and orientation sensitivities (Palmer, 1999). Best considered as an algorithmic explanation of processing (e.g. Marr, 1982), each channel responds to a limited range of dimension values. Stimuli used to investigate this theory are usually
luminance gratings which modulate sinusoidally, and contrast sensitivity can be determined by manipulating the frequency of these gratings.

All natural scenes contain information at multiple spatial scales (coarse and fine), and these are simultaneously coded by the visual system so that we can extract the scale information that is relevant to us at the time. This idea was first proposed by Campbell and Robson (1968), who derived the contrast sensitivity functions for a range of frequency values of sine-wave gratings. This involves determining the lowest contrast for which observers can detect the difference between a sine-wave grating and a grey field (i.e. their ability to discriminate differences in amplitude for the waveforms), usually performed by using the method of adjustment. Sensitivity is the inverse of the threshold function, and follows an asymptotic inverted-U shape, where intermediate frequencies are most easily detected. For humans, this is in the range of 3-4 cycles per degree of visual angle (where 1 degree = ~⅓ mm; Robson & Campbell, 1968). Psychophysical evidence for spatial frequency analysis in the visual system has been almost entirely confirmatory (see e.g. Henning, Hertz, & Broadbent, 1975; de Valois & de Valois, 1980).

Visual cells can be characterised as filters, or stimulus-response mechanisms that favour certain stimuli (Shapley & Lennie, 1985). A low-pass filter responds to low-frequency information, so effectively removes high-frequency information. The neural implementation of the multichannel theory is that striate cortical cells may be performing something like a local (because of the limited extension of receptive fields) spatial frequency analysis (Palmer, 1999). This is referred to as local because of the limited spatial extension of individual receptive fields, and can be assessed empirically by using sinusoidal gratings which fade out according to the visual angle responsivity of a receptive field. These stimuli are called Gabor functions or wavelets,
and are constructed by applying a Gaussian envelope to standard sine-wave gratings. Receptive fields of V1 cells follow the same pattern as the profiles of Gabor functions (De Valois & De Valois, 1980). De Valois, Albrecht, & Thorell (1982) showed that simple and complex cells were sharply tuned to small ranges of spatial frequency. The occipitotemporal and occipitoparietal pathways, although co-operative and anatomically interlocking, carry information from retinal ganglion cells with different size receptive fields (due to different levels of dendritic arborisation), so enable each pathway with different gross spatial sensitivities (Blakemore & Campbell, 1969; Campbell & Robson, 1968). However, Shapley and Lennie (1980) note that receptive field sizes are actually heterogenous across the visual field, and a possible revision of the idea that they provide the neural system implementation of spatial frequency analysis would be to propose a number of “patches” of Fourier analysis that are an intermediary between strict Fourier analysis and local feature detection.

Recent work has focused on the possibility that aesthetic experience may be related to certain specific characteristics in the statistics of Fourier spectra resulting from analysis of entire images, and there appears to be a similar power spectrum in visually pleasing art images and natural scenes (Graham & Field, 2007; Graham & Redies, 2007; Redies, et al., 2007) as well as musical compositions (Voss & Clarke, 1975). As frequency increases, the radially-averaged power decreases consistent with a power function of around -2 in log-log plotting. Redies (2007) suggests that artists exploit the adaptation of the visual system to natural scene structure in order to produce an efficient and sparse neural code for artworks. This relates aesthetic experience to perceptual fluency.
3.3.2.1 Spatial Frequency Analysis of Experimental Stimuli

All stimuli were processed by a bank of log-Gabor filters rather than Gabor filters to approximate the 1/f amplitude spectra present in natural images (Field, 1987). This also provides a model for visual system processing, as neural responses are symmetric on the log frequency scale. Furthermore, the DC (mean greyscale value) is zero for log-Gabor filters, allowing a more complete spectrum of the image frequencies to be represented than using the mean of the uncompressed scale (Boukerroui, Noble, & Brady, 2004). Filter banks were averaged from 8 points of origin, and rank-order correlation showed that spectra did not differ across each sweep. As can be seen from Figure 3.4, the golden-sectioned 16-item stimuli were unique in the range of stimuli used in these experiments in that they exhibit an absence of information in the 4-7 c/degree range. A possible reason for this may have been the number of light or dark sections of the grid that border the target area and are of the opposite polarity to the target, however there was no evidence that this factor systematically influenced RTs ($R^2 = .04$, $p = ns$). As such, we focused on the possibility that RTs were affected by spatial frequency processing of the image spectra.
Figure 3.4 Power spectra of experimental stimuli resulting from spatial frequency analysis.

Heavy lines represent spatial frequency information for golden-sectioned stimuli.

3.4 Experiment 3.2: The Influence of Retinal Spatial Frequency and Task Specification on Target Localisation

The pattern of results has so far demonstrated that the middle set size exhibits a golden section effect whereby excision of this ratio allows a significant linear or curvilinear function to be fit, while retention of this ratio does not. Paired comparison results indicate no special status for the golden sectioned displays at the level of conscious aesthetic preference. The performance results may be due to the power spectrum of the image’s spatial frequency, as the golden-sectioned stimuli in the middle set exhibit a lack of frequencies in the 4-7 c/degree range. In order to determine if this was the possible locus of the effect, we manipulated the distance between the observer and the stimulus display over the course of two experimental sessions. The aim of this was to establish whether spatial frequency coding or some other process involved in discrimination was being affected. If the relationship
between retinal spatial frequency and object spatial frequency is manipulated in this manner, there is usually very little difference in the contrast sensitivity function as a result (Campbell & Robson, 1968; Parish & Sperling, 1991), implying that the spatial frequency of the object is the critical factor. If the previously observed effect is due to the unique spectral profile of the golden-sectioned stimuli, we would predict a slowing in RT for the greater distance from the display, but this would be a non-specific finding relating to acuity. Replication of the previous RT results would support the role of spatial frequency.

Experiment 3.2 also addressed the possibility that the effect in the previous experiment was due to decisional rather than discrimination processes. As such, rather than having subjects locate the target and then make a decision about its colour, we asked them to determine if one pre-specified target shade was the smallest subarea in the display.

**Method**

**Participants**

Twelve participants (7 female, 5 male, mean age of 24.0 years +/- 5 years, all with normal or corrected to normal vision) took part.

**Apparatus and Stimuli**

As the middle set size has been consistently more affected, the 16-item stimuli were used to investigate the effects of decision-making and spatial frequency on performance. The experimental apparatus and stimuli were otherwise identical to experiments 2.1, 2.4, and 3.1.

**Design and Procedure**

Participants completed the same task as before, but instead of searching for the smallest target and choosing which colour it corresponded to, they were given the task
of searching for a particular target shade, and responding appropriately. In other words, the task was changed from having a “Light/Dark” response to a “Yes/No” response. Participants completed 800 trials over 20 blocks where the target was one of the shades of grey from Experiments 2.1 and 2.3. The experiment consisted of two sessions, in counterbalanced order – one in which the display was 120cm from the participant, and another where it was 60cm. For analysis, the experimental condition (120cm) was coded as Session 1, and the control condition (60cm) was Session 2.

Results and Analysis

Repeated-measure ANOVAs were carried out on log-transformed mean data, with the factors ratio and session (1 and 2). Trials with response errors (13.5% of all trials) were removed from the data prior to subsequent analyses. As with Experiment 3.1, error RTs tended to be on the whole slower than correct RTs, and an analysis of the probability correct by RT revealed no significant correlation. This argues against the correct RT data being contaminated by accuracy-speed trade-offs. An examination of the correct RTs revealed a non-normal distribution with a pronounced positive skew. Shapiro-Wilk tests showed the RT distributions for each condition and session to be significantly different to a normal distribution, so as in previous experiments, the anti-logs of the mean log RTs were submitted for analysis.

As demonstrated in Figure 3.5, the results of Experiment 3.2 were very closely matched to those of Experiments 2.1, 2.4, and 3.1, which was confirmed by a repeated-measures ANOVA that was carried out on the factors ratio and session. Mauchly’s test for sphericity indicated that the assumptions for sphericity had been violated for the main effect of ratio and so degrees of freedom were corrected using Greenhouse-Geisser estimates for sphericity. There was a significant main effect for ratio ($F_{0.49, 4.91} = 47.35, p < .001$), and session ($F_{1, 10} = 15.10, p < .001$), but their
interaction was not significant, $F_{4, 40} = 2.49, p = ns$. The main effect of session indicated a significant RT decrease between session 1 and session 2, which might indicate some non-specific practice effects. Figures 3.5 (a) and (b) show the pattern of effects are almost identical for the two sessions, and broadly match those of the middle set size in Experiments 2.1, 2.4, and 3.1 – RTs decreased non-linearly as a function of increasing ratio, with an apparent discontinuity in a linear trend at the Ratio 4 RTs. Subsequent analyses were used similar to those employed to describe the pattern of RT data in previous experiments, but with session as a factor instead of set size.

![Figure 3.5 Exponents of mean log-transformed RTs by set size and ratio for Experiment 3.1](image)

(a) Mean of anti-logs of log-transformed RT distributions by ratio in experiment 3.2 for each observer/display distance and (b) collapsed across conditions. Error bars represent standard errors.

As before, we undertook the strategy of linear and curvilinear model fitting with or without the golden sectioned ratio RTs. Unlike previous experiments, the relationship between ratio and RT was explained by a linear function prior to removal of the golden ratio RT ($R^2 = .807, F_{1,3} = 12.53, p < .05$ and $R^2 = .8, F_{1,3} = 16.98, p = .026$ for Session 1 and 2 respectively), but as predicted from previous results, this result was emphasised following removal ($R^2 = .972, F_{1,3} = 68.79, p = .014$, and $R^2 =$
.974 F\textsubscript{1,3} =112.44, p = .009). There was also a much stronger logarithmic fit of RT over ratio following removal of the golden ratio RTs for both sessions (see Table 3.2).

\textit{Error Analysis}

As in previous experiments, errors were analysed in a repeated measures ANOVA of arcsine transformed inaccuracy proportions with the same main terms as the RT analysis. A main effects for ratio was observed (F\textsubscript{4,40} = 42.6, p < .001; Greenhouse-Geisser corrected), but not for session (F\textsubscript{1,10} =.486, p <.05) or the interaction of ratio and set size, F\textsubscript{4, 40} = .981, p = ns. As in previous experiments, errors decreased with increasing ratio, and error production was not increased significantly between sessions. In session one, errors in the golden ratio stimulus category were significantly reduced relative to the Ratio 1 and Ratio 2 conditions (both p’s <.05), reduced but non-significantly from Ratio3, and increased slightly but non-significantly compared to ratio 5. Again, this provides evidence against an explanation of the golden ratio effect in terms of difficulties with target discrimination. For session two, the decrease in errors in the golden section condition was significant for comparison with Ratios 1, 2, and 3 (all p’s < .05), but there was a significant increase in errors for that condition relative to Ratio 5 (p <.05). Overall, the decrease in errors across ratios does not provide evidence for an increase in discrimination difficulty for golden-sectioned stimuli, but suggests a general difficulty with discriminating targets in the ratios close to unity.
Figure 3.6 Mean error percentages for each ratio and distance for Experiment 3.2.

Error bars represent standard errors. Session 1 = 120cm, Session 2 = 60cm

3.4.1 Discussion of Experiment 3.2

Experiment 3.2 provided some support for the hypothesis that the previous results were due to spatial frequency of the stimuli rather than more general discrimination processes. The functional relationship between RT and ratio is similar regardless of viewing distance. In this experiment, however, it is notable that there is a significant linear relationship between RT and ratio for the full range of stimuli including golden-sectioned displays. Nonetheless, model fits were noticeably improved by excision of the golden-sectioned stimuli, and the results may be affected by the shortening of RTs in this experiment. Previously, the middle set exhibited the effect and was the set with the slowest RTs, while the current experiment had a restricted range of stimuli and faster RTs, so this may have inhibited the effect somewhat.

The results also demonstrate that the change in task did not greatly affect the pattern of results. The RT data are still best described by a linear function following
removal of golden-sectioned RTs, implying that the affected mechanisms are unlikely
to be related to decision-making.
Table 3.2
Linear and curvilinear regression results for log-transformed mean RTs in Experiment 3.2 for both sessions.
Session 1 = 120cm distance, Session 2 = 60cm distance

<table>
<thead>
<tr>
<th>Session</th>
<th>Including Ratio 4 (1:0.618)</th>
<th>Excluding Ratio 4 (1:0.618)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Linear $R^2$</td>
<td>Logarithmic $R^2$</td>
</tr>
<tr>
<td>1</td>
<td>.807</td>
<td>.806</td>
</tr>
<tr>
<td></td>
<td>($F_{1,3} = 12.53$)*</td>
<td>($F_{1,3} = 12.49$)*</td>
</tr>
<tr>
<td>2</td>
<td>.85</td>
<td>.814</td>
</tr>
<tr>
<td></td>
<td>($F_{1,3} = 16.98$)*</td>
<td>($F_{1,3} = 13.11$)*</td>
</tr>
</tbody>
</table>

Note: *$p < .05$ **$p < .01$ ***$p < .001$
3.5 Experiment 3.3: Golden Sectioning and Spatial Frequency Coding

The results of the spatial frequency analysis of the 16-item grids shows that the golden-sectioned stimuli exhibit a notable lack of spatial frequencies in the 4-7 c/degree range, while other ratio-sectioned grids have spatial profiles including amplitudes in that range. Experiment 3.3 aimed to eliminate this difference by introducing noise to all ratios uniformly, thereby increasing amplitude to provide a similar spectral profile (with no frequency lacunae) across stimuli. This involves random black or white pixels across 20% of the image, raising any zero values in the amplitude spectrum above zero. The prediction being made in this case is that the effect whereby the golden section causes a discontinuity in the RT/ratio function will be removed. Specifically, it is predicted that the noise manipulation will not exhibit a linear relationship between ratio and RT following removal of the Ratio 4 RTs in the 16-item set.

Method

Participants

Ten participants (4 female, 6 male, mean age of 22.63 years +/- 2.5 years, all with normal or corrected to normal vision) took part.

Apparatus and Stimuli

Experiment 3.3 restricted stimuli to the same set size as Experiment 3.2, but altered the appearance of the previous stimuli with a layer of noise (stimuli with random black or white pixel noise mask) created using random white or black pixels across 20% of the display. This was implemented by using the hurl function in gIMP (v.2.0). In all other respects the apparatus and stimuli were unchanged from the previous experiment (see Figure 3.7 for an example of noise-masked stimuli).
Figure 3.7 Example of noise-masked stimuli used in Experiment 3.3

**Design and Procedure**

The response being elicited was the same as Experiments 2.1, 2.4, and 3.1; i.e. the participant was required to decide if the target was light or dark, and the stimulus displays were greyscale. Participants completed 400 trials over 20 blocks.

**Results and Analysis**

Trials with response errors (228 trials across participants; 2.85% of all trials) were removed from the data prior to subsequent analyses, as well as those responses exhibiting unusually short (<200ms) latencies. As with Experiment 3.1, error RTs tended to be on the whole slower than correct RTs, and an analysis of the probability correct by RT revealed no significant correlation. This argues against the correct RT data being contaminated by accuracy-speed trade-offs. An examination of the correct RTs revealed a non-normal distribution with a pronounced positive skew. Shapiro-Wilk tests showed the RT distributions for each condition and session to be significantly different to a normal distribution, so as in previous experiments, the anti-logs of the mean log-transformed RTs were submitted for analysis.
As before, linear and curvilinear models of exponents of the log-transformed RTs against ratio were applied to the data for the full range of ratios, and the range excluding the golden sectioned stimuli. Unlike the experiments from the previous chapter using similar (but noiseless) stimuli, the results of Experiment 3.3 indicated that a linear model was significant for the full range of ratios (see Table 3.3). As can be seen from Figure 3.8, the function was likely to be curvilinear in form, and there was a slight improvement in model fit for the polynomial regression coefficients. However, while curvilinear fits improved in the range excluding the golden ratio RTs, the significant linear function was nearly identical in each situation. This implies that the previously observed discontinuity in the RT/Ratio function for golden sectioning was not produced in the present experiment.

![Figure 3.8](image.png)

*Figure 3.8 Exponents of mean log-transformed RTs by set size and ratio for Experiment 3.2*

As expected on the basis of Figure 3.9, there was a significant linear trend relating ratio to RT regardless of whether or not the golden ratio stimulus RTs were included (see Table 3.3). It should be noted however that removing the golden section RTs nonetheless saw an improvement in the ability of the linear model to describe the data.
**Error Analysis**

As in previous experiments, arscine-transformed error proportions were analysed against ratio, and a significant linear decrease in error production was observed for increasing ratio ($R^2 = 0.46, F_{1, 38} = 40.35, p < .001$). This follows the same pattern as previous analyses and matches the RT data, indicating that the same general mechanisms are likely involved in each (i.e. that performance is easier as ratio sections are more uneven; see Figure 3.9).

![Figure 3.9](image)

**Figure 3.9** Mean error percentages for each ratio for Experiment 3.3.

Error bars represent standard errors

**3.5.1 Discussion of Experiment 3.3**

The results indicated that while the removal of the golden-sectioned stimuli improved model functions for the curvilinear models, there was a significant and almost identical linear function relating ratio to RT for the full range of ratios and the restricted range. Along with the error data, this implies a simple functional relationship likely to be due to increased difficulty in differentiating a target area from its neighbour when they are similar in size. The use of noise in this experiment supports the hypothesis that the RTs would exhibit a diminution or removal of the
golden section effect from previous experiments. Whether this is due to the spatial frequency profile of the golden-sectioned stimuli inhibiting efficient processing at the level of neural dynamics remains an open question, but the results presented here do suggest that this is a possibility worthy of further investigation.
Table 3.3
Linear and curvilinear regression results for log-transformed RTs in Experiment 3.3 for the full range of ratios and for the range excluding the golden ratio

<table>
<thead>
<tr>
<th></th>
<th>Including Ratio 4 (1:0.618)</th>
<th>Excluding Ratio 4 (1:0.618)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Linear R²</td>
<td>Logarithmic R²</td>
</tr>
<tr>
<td></td>
<td>.900</td>
<td>.961</td>
</tr>
<tr>
<td></td>
<td>(F₁,₃ = 26.97)*</td>
<td>(F₁,₃ = 74.25)**</td>
</tr>
</tbody>
</table>

Note *p < .05 **p < .01 ***p < .001
3.6 General Discussion

The present chapter investigated the potential mechanisms that may underlie the previous finding that RTs tend to be substantially slowed given 16-item grids, relative to an otherwise linear decrease in RT with ratio. The pattern of RTs did not correspond with that expected of visual search except in the 32-item grids, indicating that it was unlikely that the slowed processing in the less dense grids was related to the mechanisms responsible for serial-attentional deployment, and instead related to local processing at the target section location. The results of this chapter presented evidence that the effect may be due to the encoding of stimulus features according to neural dynamics.

Based on Experiments 3.1 and 3.2, the results appeared unlikely to be confounded by luminance contrast or the influence of decision-making mechanisms. We assessed the possibility that the self-similarity (fractality) of the golden-sectioned stimuli was the factor determining the effect. The decrease in RT with increasing ratio appeared to correlate with fractal dimension, but this tendency was not significant and was not improved when golden sectioned stimuli were removed. This suggests that there is some general perceptual sensitivity to fractal dimension, but that the golden-sectioned grids did not represent a special case characterised by differences in complexity to the other grids. The next possibility we investigated was the spatial frequency of the stimuli. Spatial frequency analysis of the 16-item stimuli showed that the golden-sectioned grids had a frequency lacuna around 4-7 c/degree that was not represented in the other stimuli. Noise was added uniformly across stimuli in order to give them a similar frequency spectrum and engender similar RT functions, and this manipulation succeeded in removing the previously observed effect.
The results of the previous chapter argued against the status of the golden section as an aesthetic principle. However, the present chapter offers some evidence that discriminative judgements are affected by the spatial structure of golden-sectioned stimuli. The lack of preference for golden-sectioned stimuli would support the idea that processing fluency affects aesthetic experience (Reber, Schwartz & Winkielman, 2004). However, the direction of this effect is in the opposite direction than expected for a proposed normative aesthetic role for the ratio; rather than being most preferred, it was closer to being least preferred (see previous chapter). The experiments detailed here appear to argue against the proposal that low fluency is inherently negative (or vice versa). Our golden-sectioned stimuli are shown to exhibit a spatial-frequency profile which affects perceptual performance by slowing processing, and aesthetic interest would also appear to be aroused by a wider variety of potential stimuli than the processing fluency can account for – visually or cognitively dissonant images are often enjoyed (e.g. Arnheim, 1950, Carroll, 1990, Solso, 1996) and simple images may appear boring (Palmer & Griscom, 2012). Carroll (1990) emphasised “the way in which artistic presentation of normally aversive events or objects can give rise to pleasure or compel our interests” as an important question for aesthetics. Outside of the lack of preference effects, the golden section is shown here to affect performance negatively through low fluency.

Our findings are also unexpected given research showing that ratings of discomfort correlate closely with information in the Fourier spectrum range of 3 c/degree ± 1 octave (Fernandez & Wilkins, 2008) which is the peak of the human contrast sensitivity function (CSF; Campbell & Robson, 1968). This latter finding is consistent with the proposal that human perceptual processing has adapted to process natural scenes efficiently, and that unnatural scene statistics may be experienced
negatively. The inverted-U profile of the human CSF is therefore a result of high and low spatial frequency information being more common than intermediate frequencies. A maximally-adaptive visual system is one that emphasises sensitivity for more uncommon spatial frequencies in the visible range. When there is an excess of energy in the maximally sensitive range, the subjective experience may be stressful or aversive (Fernandez & Wilkins, 2008). The present research shows the opposite effect, however; a lack of information in the peak of the CSF does not cause greater preference ratings. Instead, it seems that the spatial frequency lacuna observed in the middle of the range for normal visual perception is deleterious to performance. Again, this aligns with the perceptual fluency account by proposing that natural scene statistics that are unusual in nature are processed less fluently, but the historical status of the golden section appears to argue against the idea that less fluent processing precludes aesthetic effect. Instead, it is possible that unusual spatial frequency power spectra may prolong processing, and that this may lead to extended elaboration of the visual object. This proposal is discussed in more detail in relation to visual structure in the next chapter.

A possible mechanism for the impeded processing for golden sectioned stimuli comes from Pletzer, et al. (2010), who suggested that inter-phasic synchronisation in neural assemblies may be affected by the golden ratio. The present results do not conclusively support this proposal, but do offer some supporting evidence that spatial frequency coding affects performance. If pattern grouping is performed by the synchronisation of phase and/or frequency into a common neural code, then subarea comparison is achieved by the differences for these parameters between cell assemblies. If coding is inefficient or absent it may lead to difficulty in resolving a comparison due to the lack of a common integrative framework forsubsectioned
areas. This possibility requires further supporting evidence, however, and would benefit from further research. Nonetheless, the research presented here is based on the tentative proposal that golden sectioning can influence processing dynamics via spatial structure, but this must be subject to further research before a satisfactory reconciliation with these other findings can be drawn. The findings reported in these chapters may provide some new direction for such research.
Chapter 4: Visual Balance and Search Performance

4.1 Introduction

Visual balance is a fundamental compositional principle in art and design, but its relationship with physical balance or symmetry is not straightforward. While symmetrical images necessarily appear balanced, asymmetrical ones can also appear to exhibit balance (Samuel & Kerzel, 2013; Wilson & Chatterjee, 2005). Symmetry refers to the extent to which a pattern has elements that map onto other elements (Tyler, 1995). Mach (1886/1959) describes three types: translational (involving patterned repetition), reflectional (mirror symmetric), and rotational (rotated about an axis). Gestalt grouping principles such as good continuation, closure, proximity, and similarity are defined as such by being effective at improving processing efficiency of visual patterns (e.g. Machilsen, Pauwels, & Wagemans, 2009) and symmetry has been proposed to be in the same category (Biederman, 1987; Wagemans, 1995; Wolfe & Horowitz, 2004). Symmetry has ecological relevance due to the likelihood that a symmetrical organism is facing directly at or away from the observer, and is shown to be easily and rapidly identifiable (Machilsen et al., 2009; McManus, 2005; Roggeveen, Kingstone, & Enns, 2003, 2004). While this is especially true for reflectional symmetry (Kahn & Foster, 1986), Garner (1974) showed that observers are also sensitive to rotational symmetry when making classification judgements, and that these correlate with goodness ratings. All three types of symmetry involve the implicit coding of an object-centred frame of reference in order to map co-ordinates to one another, or be robust to such transformations (Gershoni & Hochstein, 2011; Palmer, 1999).

Less amenable to definition is phenomenal balance - also called pictorial (Gershoni & Hochstein, 2011; Niekamp, 1981) or compositional balance (Nodine,
Locher, & Krupinksi, 1993). This may occur in the absence of physical balance (see Figure 4.1 [a] – [e]), a fact that has long been exploited by artists and designers (Locher & Nagy, 1996; McManus, Edmondson, & Rodger, 1985; McManus, Stöver, & Kim, 2011). Comparable with the rapidity of symmetry detection (e.g. Rogeveen, et al. 2003), balance is quickly and consistently noted by observers (occurring ~100ms after first fixation; Locher & Nagy, 1996; Ognjenovic, 1991). This suggests that it may be an intrinsic aspect of normal stimulus organisation. However, while symmetry appears to be important in art, aesthetic experience seems to depend on phenomenal rather than physical balance. For example, preference for symmetrical images appears to diminish with age (Campbell, 1941; Lund & Anastasi, 1928) or art training (Samuel & Kerzel, 2013). Famous paintings are often those that exhibit phenomenal balance rather than the symmetrical arrangement of stimulus features (Gershoni & Hochstein, 2011; Puffer, 1906). Locher, Gray, and Nodine (1996) define phenomenal balance as the grouping or organisation of perceptual forces so that they compensate each other. The origin of these forces is a topic of debate, however. For example, the centre of mass of a frame can be conceptualised as a fulcrum across which a lever-type action lends phenomenal balance to heavier objects close to the fulcrum against lighter objects further from the fulcrum (McManus, Stover, & Kim, 2011; Samuel & Kerzel, 2013). In this case visual weight is determined by size, colour, or luminance contrast of objects and how they balance across the centre of mass of the frame (Gershoni & Hochstein, 2011).
Figure 4.1 Arnheim’s (1954) structural skeleton and some examples.

(a) Based on Arnheim’s illustrative depiction of a frame’s internal perceptual dynamics. Dashed lines represent areas of equilibrium or minimal energy expenditure for the creation and maintenance of a perceptual representation. These are arranged around the centre, the corners, and principal and diagonal axes. (b) and (c) Left and right sides of the frame contribute differently to the appearance of motion and preference ratings (Levy, 1976), and this is accentuated by the facing direction of the figure (an inward bias; Palmer, Gardner, & Wickens, 2008). Both of these factors combine in producing a more pronounced implied motion in (b) in comparison with (c). Structural instability produced by axial misalignment may similarly affect implied motion by appearing unresolved or unsettled (Arnheim, 1954). (d) An example of an imbalanced display. According to Arnheim (1954), the disc appears unsettled because it lies close to, but not aligned with, the axial structure of the frame. A minimal perceptual organisation is almost provided by the object’s placement, resulting in a motive phenomenal quality as the representation tends towards equilibrium. (e) An example of a balanced image in the absence of symmetrical arrangement. Images 4.1 (a), (b), and (e) are copyright Per Åström, used by permission.
4.2 Physical and Phenomenal Balance

A major difficulty in identifying phenomenal and physical balance is that phenomenal space itself is not distributed homogenously across the visual field. Instead it is anisotropic; certain positions within the scene are inherently more balanced than others. Left and right sides of a symmetry axis contribute differently to the expression of motion or action, with objects on the left facing to the right being those most suggestive of motion (Gaffron, 1950; Palmer, Gardner, & Wickens, 2008; Wilson & Chatterjee, 2005; see Figure 4.1 [b] and [c]). There may also be an experiential effect of gravity on up-down anisotropy (Arnheim, 1954; Niekamp, 1981). The result of interactions with the interior forces of the visual scene is that the perceived mass and volume of objects is dependent on their contextual arrangement – they are holistic or configural (Pomerantz & Portillo, 2011). The manner in which the position of an object in relation to the frame and/or to other objects within it may be important for visual saliency, with items being easily grouped if they organise along the skeletal structure of the frame.

4.3 Balance, Emergence, and Gestalt

Such an account proposes that computations occurring early in the microgenesis of visual perception are sufficient to code balance as an emergent feature of stimulus organisation. This idea has obvious relevance to the holistic units of Gestalt grouping principles, especially the expressiveness that can arise from visual organisation, and it is closely tied to aesthetic experience. The Gestalt approach to aesthetics was initiated by Koffka (1935), who discussed the physiognomic or tertiary qualities of perceptual organisation in terms of the expressive qualia that emerge from the configural arrangement of stimulus features. This theory was extended by Arnheim’s (1954) discussion of the internal perceptual forces of a frame (see Figure
According to Arnheim, the structural axes of the frame (and especially their point of intersection at the centre) represent locations of equilibrium of organisational tendencies, so that an object placed just outside of one of these locations will appear unsettled or imbalanced (Figure 4.1 [d]). One of the most valuable aspects of Arnheim's model is that it allows for a systematic link between physical and phenomenal balance, but without necessarily relying on the symmetry of the configuration. The proposal has some support from studies on preferences for locating a stimulus within a frame (Niekamp, 1981; Palmer & Guidi, 2011). Structural misalignment is also shown to cause preferential orientation in 5-day old domestic chicks, implying the coding of balance in phylogenetically ancient visual mechanisms (Elliott, Rosa-Salva, Mulcahy, & Regolin, 2012).

Emergent features (EFs) in perception are defined as such by being configural properties that are nonetheless consciously recognised before their constituent elements (Pomerantz, Sager, & Stoever, 1977; Pomerantz & Portillo, 2011). A canonical example is the Kanisza (1979) triangle which is instantly observable, whereas attention to the Pac-men that suggest the shape is more difficult. Configural superiority is the defining feature of EFs, and is experimentally observable in improved discrimination speed and accuracy for configurations versus their components. In relation to attentional processing, this conforms to the idea that early mechanisms guide attention to areas of salience across the visual field by coding the relational features of objects (Duncan & Humphreys, 1989; Itti & Koch, 2001; Koch & Ullman, 1985). In visual search experiments, Duncan and Humphreys (1989) found that grouping and segmentation rely upon the preattentive coding of relations between elements of the visual display. Conci, Müller, & Elliott (2007) have also shown that amodal completion facilitates efficient localisation, with global surface being more
influential than the bounding contours, implying that an emergent configural attributes can guide attention. Roggeveen, Kingstone, and Enns (2003, 2004) found evidence for grouping based on inter-item symmetry prior to attention, and this is due to target-distractor relations rather than distractor-distractor symmetry. If phenomenal balance is accorded the status of a Gestalt or emergent feature, it would occur at early enough visual processing to exhibit more efficient search than an unbalanced alternative.

4.4 Balance and Perceptual Processing

The ease and immediacy with which we can decide if we like or dislike a visual composition suggests that aesthetic experience is an intrinsic aspect of normal visual perception (e.g. Chen & Scholl, 2014; Palmer, Schloss, & Sammartino, 2013). People can quickly and consistently establish preferences between similar alternative images (David, 1969; Titchener, 1901), even when the representational content is low (e.g. Mather, 2012). Basic perceptual processing, such as that of avian species, has been shown to facilitate clear and consistent visual preferences that are likely to be a result of the way in which attention is influenced by visual structure (Clara, Regolin, Zanforlin, & Vallortigara G, 2006; Clara, Regolin, & Vallortigara, 2007; Elliott et al., 2012; Endler, 2012; Endler, Endler, & Doerr, 2010). These results confirm that phylogenetically ancient, and most likely early sensory brain mechanisms mediate grouping. Early visual organisation processes may exert an influence on the way that emergent qualities guide attention, and imbalance may furthermore provide some expressiveness that might translate into aesthetic preference. One of the earliest findings from experimental aesthetics is that people prefer unequal line or area subdivisions (Fechner, 1876, 1997; Pierce, 1894, 1896; Puffer, 1906; Valentine, 1913), even when they try to determine the most physically balanced bisection
(Langfield, 1920; Puffer, 1906). Given the breadth of the stimuli and methods used to obtain these results, this does not appear to be a problem of visual acuity but instead a tendency for asymmetrical stimuli to appear balanced. Research on Arnheim’s model of visual balance within experimental aesthetics has produced equivocal results (McManus, Stöver, & Kim, 2011; Samuel & Kerzel, 2013; Wilson & Chatterjee, 2005). Niekamp (1981) suggests that inconsistent findings may be a result of indexing balance as an additive property of basic dimensions such as size or colour, without also accounting for the relative positioning of test stimuli or the spatial anisotropies of the frame.

The present study aims to investigate the influence of phenomenal balance at the level of visual perceptual organisation, by operationalising ‘balance’ as axial alignment according to the skeletal structure of a frame (Arnheim, 1954). A visual search paradigm was used in order to assess the results in terms of search efficiency, with the aim of attributing balance coding to a pre- or post-attentive process. Specifically, we predicted that a balanced target would be localised faster than an unbalanced alternative by affecting attentional capture by visual saliency (for the target defined by size). This was based on the hypothesis that target alignment with the skeletal structure of a frame has an influence on the relative processing efficiency of a balanced versus imbalanced target. Specifically, we tested whether a configural superiority effect exists for balance at the level of a basic feature. Experiment 4.1 assessed search for a balanced or imbalanced target amongst axially-aligned distracter items. The results suggested that an effect of balance was due to two perceptual orientations; a local one in which structure is defined by the distracter set (and is most apparent in the largest set), and a global one in which balance depends on the skeletal structure of the frame (shown in the middle set size). In order to accept this
interpretation, a second experiment attempted to dissociate these processing phenomena by using an axially-misaligned distracter set. In agreement with expectations based on findings from Experiment 4.1, the second experiment showed that visual balance influenced performance. Consistent with predictions regarding organisational grouping of balanced items, the on-axis targets were associated with faster RTs, regardless of whether balance was defined by the frame or by alignment with distractor items.

4.5 Experiment 4.1: Visual search in an Axially-aligned Distracter Array

Method

Participants

Twenty-one participants (11 female, 10 male; mean age 25) took part in Experiment 4.1. Participants are shown to be able to detect imbalance (or asymmetry) easily and at very short stimulus presentations (100ms; Locher & Nagy, 1996; Locher & Wagemans, 1993), so we predicted a medium to large effect (based on Cohen’s 1988 guidelines) and the sample size reported here was sufficient to detect this effect at 80% power and a 0.05 significance level (Champely, 2012). All subjects had normal or corrected-to-normal vision. Approval of research ethics was given by the institutional Research Ethics Committee at NUI Galway.

Apparatus and Stimuli

Stimuli were generated using E-prime 2.0 Professional (Psychology Software tools Inc.) on a Pentium 4 PC running Windows XP. Stimulus control and presentation were programmed using native E-prime macro scripts with system integrity checked by means of an E-prime refresh clock test which gave a diagnostic classification of ‘good’ (measurement error +/- 1 ms). Stimuli were presented on a 19” Magic Displays monitor, model CPD-4402 Trinitron with resolution 1024x768
and refresh rate 75 Hz. Responses were recorded using a two-key Ergodex® DX1™ Input System.

The stimuli used in the experiment each consisted of a centrally-positioned frame whose dimensions were 444 x 444px (see Figure 4.2). Within the frame, a number of filled circles were arranged according to alignment with the main axes of symmetry of the frame (horizontal, vertical, and major and minor diagonal axes). In half of the trials, all of the circles were the same radius (7 pixels [px]), and in the other half one of the items – the target - was larger (r = 10px), subtending about 1° of visual angle. A larger (rather than smaller) target was chosen in order to maximise the amount of axial imbalance of the target. In trials where a target was present, half had the target item aligned with the structural axes, and half were misaligned by 50% of the total diameter of the circle. These target item dimensions and position were optimised in pilot experiments. Distractor arrays consisted of 4, 8, or 16 item displays. Stimuli were dark grey (RGB: 75, 75, 75) on a light grey background (RGB: 192, 192, 192).

![Figure 4.2 An example of stimuli used in Experiment 4.1.](image)

The participant was required to determine if a target (defined by being slightly larger in radius to distracter items – illustrated here above the main horizontal meridian, and to the left of the main vertical axis) is present or absent. This target could be aligned with an axis of symmetry of the frame (“on-axis” as here) or misaligned with the structural axes (“off-axis”). The dashed lines of symmetry depict the frame’s structural axes and the dotted lines refer to a uniformly shifted axial structure defined by the distractor items. Neither were revealed to participants and are shown here for illustration purposes.
**Design and Procedure**

Experiments 4.1 and 4.2 used a within-subjects design and a 2AFC speeded response task in which participants were asked to respond as rapidly and accurately as possible to the presence or absence of the target. Each trial started with the presentation of a central fixation cross for 500 milliseconds (ms). The fixation cross was then immediately replaced by the search array, which remained onscreen until a response was recorded. In case of an erroneous response or a time-out (i.e. after a period of 2,500 ms without response), feedback was given by a computer generated tone and an alert was presented for 500 ms at the centre of the screen. Each trial was separated from the next by a variable interval between 500 – 1,000 ms. Following a 40-trial practice session, participants completed 384 trials in one 12-block session (with 32 trials per block). There were an equal number of target-present and target-absent trials (192) and an equal number of off- and on-axis target present trials (96). Half of target-present trials were presented in the central 50% of the frame, and the remaining half in the periphery. Presentation was fully randomised across all blocks and separately for each participant. The experiment was carried out in a sound proof booth with low ambient lighting with a chin rest used to ensure distance between participants and monitor was kept constant at 55 cm.

**Analysis and Results**

Trials with response errors (6% of total) were removed from the data prior to subsequent analyses. Accuracy was not found to be a predictor of RT, $b = -1496.8$, $t(7) = -1.02$, ns, and did not explain a significant proportion of variance in RT, $R^2 = .13$, $F(1,7) = 1.032$, ns. Since there appeared to be no tendency for increased errors for faster RT’s, we concluded the data were unlikely to be contaminated by speed-accuracy trade-offs. Shapiro-Wilks tests of normality indicated that the RT
distributions were significantly non-normal, exhibiting pronounced positive skew for all conditions, so the data were log-transformed in order to proceed with parametric analysis (Box & Cox, 1964, 1982).

For the log RTs, Mauchly’s test of sphericity was significant for set size and the interaction of set size and target condition (present or absent), so Greenhouse-Geisser estimates were adopted. A repeated-measures ANOVA with the factors target condition (present, absent) and set size (4, 8, 16) revealed a significant main effect for targets ($F_{1, 20} = 14.94, p < .001, \eta^2_G = 0.07$), with a significant increase in RT with increasing set size ($F_{1, 13, 22.65} = 34.22, p < .001, \eta^2_G = 0.05$). The two-way interaction was also significant, $F_{0.57, 11.37} = 16.41, p < .001, \eta^2_G = 0.02$. A comparison of slopes indicates that target detection is conducted at a cost of 4 ms per item (Treisman & Souther, 1985), while participants engage in more exhaustive search (15 ms per item) when there is no target. The approximately 3:1 ratio of target-present:target-absent search is indicative of displays encouraging serial search strategies, while the shallow target search slopes indicates this is modified by a tendency for target search to be guided by parallel target coding.

![Figure 4.3](image)

*Figure 4.3* Exponents of mean log-transformed RTs by set size and target position for Experiment 4.1.
Unlike targets which are detected relatively easily irrespective to display size, the target-absent displays appear to encourage a more effortful, serial search strategy. Consequently, these trials were not considered in further analyses. Search slopes for the target-present trials by condition are both efficient (4.2 ms and 3.9 ms/item for on-axis and off-axis RTs respectively). Based on the assumption that visual balance is a complex feature of stimulus arrangement, the dichotomous variable defining axial alignment (on or off) is expanded to include other factors which are assumed to contribute to balance. In particular, and following the suggestion by Niekamp (1981), target location (especially centrality, anisotropically-weighted positioning, or distance from other items) might provide a more accurate predictive model. Based on the efficiency of target localisation, a specific focus is to investigate whether any differences in performance for axial conditions may be due to intrinsic relational coding.

**Central and Peripheral Target Presentation.** If visual balance is produced prior to attentional integration, then the location of the target should not influence its occurrence (i.e. it should occur in parallel across the visual field). A repeated-measures ANOVA was conducted for the target-present trials only, with the factors target position (on or off-axis), target location (central or peripheral) and set size. The main effect of target location was significant ($F_{1,20} = 92.85, p < .001, \eta_g^2 = 0.14$), and there was a small but significant main effect of target (axial) position ($F_{1,20} = 5.61, p < .05, \eta_g^2 = 0.02$) but the 2-way interaction was non-significant ($F < 1$) as were the other 2-way and 3-way interactions. Paired-sample (Bonferroni-corrected) t-tests were performed in order to compare the RTs for target position (on- or off-axis) for each location and set size. There was a significant difference between RTs for on-axis target stimuli for set size 8 ($M = 604.87, SD = 127.79$) and comparable off-axis
targets ($M = 636.57, SD = 120.72$) only when the target was presented centrally ($p < .05$). No other comparisons were significant. The lack of significant interactions suggests that the statistically significant effect of axial alignment occurs irrespective to the central or peripheral presentation of the target.

![Figure 4.4 Exponents of the means of the log-transformed RTs by set size, position, and location in Experiment 4.1.](image)

**Palmer-Guidi (2011) Goodness-Ratings as Positional Weights.** More detailed investigation of the hypothesis that structural axes affect imbalance at the level of stimulus organisation was undertaken in the context of empirical findings reported by Palmer and Guidi (2011). In their experiments, participants were presented with a circular probe in 77 different spatial bins of a rectangular frame. The task required participants to assign a subjective rating of goodness to each position, leading to a mean rating score for each. The scores tend to be highest at the centre and along principal axes of symmetry and main diagonals, so may provide an operational depiction of the anisotropy of the frame. Assigning an absolute positional weight to the targets in the present experiment has the advantage of introducing empirically-
supported goodness ratings that define structure. This also addresses some other relevant structural effects (i.e. spatial anisotropies) that were measured in the Palmer and Guidi experiments. The effect of an aligned versus a misaligned target should be that locations with higher Palmer-Guidi rating (PGR) values should accentuate the degree of (im)balance according to Arnheim’s theory. Specifically, we believed axial alignment might expedite the processing of aligned targets, but this is not taken to have the same latency advantage at all points along the axes. The effect of spatial anisotropy would therefore be to exhibit the most pronounced visual balance at an axially-aligned and high PGR location. In this regard, PGR should also predict RT, and should interact with the axial position of the target. This is because the organisation of on-axis items into a Gestalt defined by the structural skeleton is assumed to be performed more efficiently when these items are located in areas with higher goodness ratings.

RT data were modelled using a multi-level analysis with participant as the random effect and Palmer-Guidi rating (PGR) and target position as fixed effects for each set size separately for target-present trials only. This type of analysis is promising for experimental aesthetics as it factors inter-individual dependence in the data against the functional relations assumed to characterise aesthetic object-subject interactions (Silvia, 2007). More specifically, seemingly large individual differences in the perception of phenomenal balance may have undermined previous attempts to demonstrate an effective experimental manipulation of the variable (Buswell, 1935; McManus, et al. 2011). However, any functional relationship that characterises the phenomenon may have been overlooked as a result of averaging scores across participants. In a recent example, Silvia (2007) revealed that a multilevel re-analysis of Wilson and Chatterjee’s (2005) data regarding preference for visual balance
produced a more extensive influence of the phenomenon (i.e. across more stimulus form categories) than was previously found. With this in mind, we expected that inter-individual variability will be due to normal individual differences in general perceptual performance as reflected by significantly variable model intercepts. The functional relationship is proposed to be similar between people, reflected by little variability in model slopes. Model coefficients are based on log RTs rather than their anti-logs following transformation, due to the ability to extract parameters from the full complement of data points (Hoffman & Rovine, 2007). Variance inflation factors (VIF) for each set size model with both main predictors included were small (< 2), so multicollinearity was not assumed to be problematic for the data (Field, Miles, & Field, 2012). Parameters were estimated using full maximum likelihood estimation (MLE) instead of its restricted alternative (REML) because it allows comparison of deviance between different model fits (Field, Miles, & Field, 2012).

For all set sizes for the relationship between PGR and RT there was significant variation in the intercept across participants, set size 4: \( SD = 0.199 \) (95% CI = 0.146, 0.272), \( \chi^2(1) = 397.86, p < .001 \); set size 8: \( SD = 0.19 \) (95% CI = 0.14, 0.26), \( \chi^2(1) = 498.78, p < .001 \); set size 16: \( SD = 0.20 \) (95% CI = 0.15, 0.27), \( \chi^2(1) = 466.70, p < .001 \). For all set sizes, there was no significant variation across participants for the slopes of these linear fits (all p’s > .05), so an intercept-only model was adopted in all cases. For the final models including both predictors and their interaction, PGR significantly predicted RT for all set sizes, but target position was not significant for any set. The two-way interaction of these effects was also not significant for any set (see Table 4.1).
On the basis of these results, it appears that the PGR weightings do not provide a more unified demonstration of a structural-axis effect across set sizes.

While it is entirely possible that our participants deviated from Palmer and Guidi’s (2011) subject scores, or that the scores are only relevant to the dimensions of Palmer and Guidi’s rectangular stimuli, it would seem that this formulation of spatial anisotropy may not apply to imbalance as a unitary construct. A caveat when interpreting the results is that no effort was made to systematically exhibit and record every possible spatial bin from the 77 alternatives. In the smallest set, an effect of target position seems to complement the middle set finding in the previous pairwise analyses. However, a lack of interaction effects implies that such results do not derive from PGR placements.

**Target-distracter Inter-Dot Distance by Position.** According to Ziemkiewicz and Kosara (2010), the effects of dynamic tension are most pronounced when the target is to some degree isolated from a cluster of non-targets. The set size condition is operationally similar to a manipulation of this, with the effects of clustering increasing with an increase in the number of items in the display. However,
the target’s embeddedness within (or isolation from) the cluster is convolved in the responses for each set size, so an operationally-defined variable is needed to test this more adequately. One means of accomplishing this is to establish mean inter-dot distance for the target in relation to all other points in the display, and analyse this according to target position. This inter-dot distance (IDD) of the target is calculated as the arithmetic mean of the proximity of the target in relation to all other distracters. Higher figures represent a greater distance between the target and the distracters, and greater isolation overall. While the effect of proximity is probably one in which locating the target closer to other items facilitates performance, this may also accentuate any effect of imbalance. In other words, there may be a more pronounced depiction of balance when the target is closer to a cluster of items, based on the rationale that the alignment of these items (which happens to coincide with the skeletal structure) will define the relevant structure.

RT data were again modelled in a multi-level analysis with participant as the random effect and inter-dot distances (IDD) and target position as fixed effects for each set size separately for target-present trials only. For all set sizes for the relationship between IDD and RT there was significant variation in the intercept across participants, set size 4: $SD = 0.20$ (95% CI = 0.15, 0.27), $\chi^2 (1) = 397.86, p < .001$; set size 8: $SD = 0.19$ (95% CI = 0.14, 0.26), $\chi^2 (1) = 498.78, p < .001$; Set size 16: $SD = 0.20$ (95% CI = 0.14, 0.27), $\chi^2 (1) = 466.70, p < .001$. For all set sizes, there was no significant variation across participants for the slopes of the linear fits (Set size 4: $\chi^2 (2) = 1.31$, ns; Set size 8: $\chi^2 (2) = 1.90$, ns; set size 16: $\chi^2 (2) = 7.20$, ns), so an intercept-only model was adopted in all cases.

For the final models including both predictors and their interaction, IDD significantly predicted RT for all set sizes. Target position was not significant for set
sizes 4 or 8, and the interaction term was also not a significant predictor of RT for these sets. Target position and the interaction of target position with IDD were significant predictors in set size 16 (see Table 4.2 for a summary). The best model for set size 4 according to Akaike’s Information Criterion (AIC; see Table 4.3) was that including IDD only as a predictor. The best model for the 8-item set was that which also included target position but not the interaction term. In this model, target position was shown to be a significant predictor of RT, $\beta = -0.08$, $SE(\beta) = 0.01$, $t(1223) = -5.46$, $p < .001$. In the 16-item set, the final model including both effects and their interaction was best according to AIC.
Table 4.2
*Fixed and random effect estimates for response times as a function of mean inter-dot distance from target to distracters in Experiment 4.1*

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Set Size</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td>8</td>
<td></td>
<td></td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Random Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td></td>
<td>6.22</td>
<td>0.06</td>
<td>97.63</td>
<td>&lt;.001</td>
<td>5.68</td>
<td>0.14</td>
<td>41.51</td>
<td>&lt;.001</td>
<td>5.36</td>
</tr>
<tr>
<td><strong>Fixed Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IDD</td>
<td></td>
<td>0.002</td>
<td>0.001</td>
<td>5.18</td>
<td>&lt;.001</td>
<td>0.01</td>
<td>0.004</td>
<td>6.32</td>
<td>&lt;.001</td>
<td>0.01</td>
</tr>
<tr>
<td>Target</td>
<td></td>
<td>0.04</td>
<td>0.06</td>
<td>0.71</td>
<td>ns</td>
<td>0.04</td>
<td>0.17</td>
<td>0.20</td>
<td>ns</td>
<td>1.08</td>
</tr>
<tr>
<td>IDD x Target</td>
<td></td>
<td>&lt;.001</td>
<td>0.001</td>
<td>-0.77</td>
<td>ns</td>
<td>&lt;.001</td>
<td>0.002</td>
<td>-0.66</td>
<td>ns</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

*Note: B = unstandardised beta coefficients, SE = standard error. IDD = inter-dot distance. Target = axial alignment of the target item (aligned/misaligned), ns = not significant.*

For the largest set, this implies that the effects of IDD change based on the axial alignment of the target – namely that IDD has a more pronounced effect (slower responses for greater distances) in the off-axis condition, but RTs remain roughly the same regardless of IDD for the on-axis condition. This suggests that the interrelationship of local elements within the frame defines structural balance in the 16-item set. Based on the most applicable model for set size 8 according to fit statistics, axial alignment is a predictor of response latency regardless of IDD, with faster RTs for on-axis stimuli.
Table 4.3  
Model fit diagnostics for anti-logs of mean log-transformed response time by inter-dot distance from the target to distractor items for each set size in Experiment 4.1.

<table>
<thead>
<tr>
<th>Set Size</th>
<th>Model 1</th>
<th>Δχ²</th>
<th>AIC</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Model 1</td>
<td>49.16</td>
<td>201.99</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Model 2</td>
<td>0.02</td>
<td>203.98</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Model 3</td>
<td>0.60</td>
<td>205.38</td>
<td>ns</td>
</tr>
<tr>
<td>8</td>
<td>Model 1</td>
<td>58.69</td>
<td>62.45</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Model 2</td>
<td>29.48</td>
<td>34.96</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Model 3</td>
<td>0.44</td>
<td>36.53</td>
<td>ns</td>
</tr>
<tr>
<td>16</td>
<td>Model 1</td>
<td>0.20</td>
<td>245.63</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Model 2</td>
<td>3.04</td>
<td>244.60</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Model 3</td>
<td>4.91</td>
<td>241.68</td>
<td>&lt;.05</td>
</tr>
</tbody>
</table>

Note: AIC = Akaike’s Information Criterion, Lowest values of AIC represent the best model fit. Model 1 includes IDD as a predictor, Model 2 includes IDD and target position (main effects model), and Model 3 includes IDD, target position, and their interaction. Likelihood ratio (χ²) comparison for Model 1 refers to the difference (Δ) in -2LL (log-likelihood) between this model and an intercept-only mode, ns = not significant.

**Error Trials**

Error trials were analysed by investigating the arcsine-transformed error probabilities for the different conditions, based on the possibility that the pattern of results observed in the main RT analysis is due to difficulty with the encoding of targets when those targets are imbalanced. This would predict a similar pattern of results in errors as those shown for RT. A repeated-measures ANOVA with the factors target position and set size showed that there was a significant main effect of target position \( (F_{1.6, 31.91} = 3.70, p < .05, \eta_G^2 = 0.04) \) and set size \( (F_{1.75, 35.05} = 9.51, p < .001, \eta_G^2 = 0.06) \), but their interaction was not significant, \( F_{2.40} = 1.88, p = ns \).

Bonferroni-corrected pairwise t-tests were conducted among each target condition for each set size. In the 4-item set, there were no significant differences for any of the comparisons (all \( p \)'s > .05). In the 8-item set, there was a significant difference between the no-target condition and the off-axis \( (p = .003) \) and on-axis \( (p < .01) \) conditions, with fewest errors in the target-absent condition. However, there was no significant difference between the two target-present conditions. For the 16-item set,
this pattern of results was repeated, with a significant difference between the target-absent trials and the off-axis and on-axis conditions ($p < .05$ and $p < .01$ respectively), but none between the target-present conditions.

4.5.1 Discussion of Experiment 4.1

The results of Experiment 4.1 suggest that the axial alignment of the target may influence its ability to be localised rapidly and efficiently in a visual scene, but that the effect is subtle and may be sensitive to stimulus complexity. Based on comparison of RTs and accuracy profiles, it would appear that targets are easily discriminated regardless of condition. A main effect of axial alignment was produced which was independent of target centrality. Taken together with the relatively flat target search slope, this would tend to suggest that axial alignment is coded by parallel search mechanisms (e.g. is coded in a relational context) and prior to attentional deployment. Alignment was also shown to affect RT in the 16-item set when considered in combination with inter-item structure, while the 8-item set demonstrated an effect of axial alignment that did not interact with inter-item structure. The results support the proposal by Hogeboom and van Leeuwen (1997) that search efficiency can be related to task demands, whereby longer more effortful searches result from the processing of local stimulus structure, and shorter RTs result from global orientation. Imbalance in the 8-item set appears to be characterised by axial misalignment deriving from the skeletal structure of the frame (or at least not necessarily the inter-item structure). In the 16-item set the relative distance from the target to other items produces an interaction with target axial position, consistent with a local structure orientation. The lack of an effect in pairwise comparisons for the latter may be due to search in the 8-item set being around the point of intersection of the interacting organisational tendencies (global and local). It is unexpected that the
four-item set produced no effects in paired comparisons, or either of the multi-level analyses. One possibility is that the task in this case was too easy in either target condition to adequately differentiate performance for either.

4.6 Experiment 4.2: Visual Search in an Axially-misaligned Distracter Array

In Experiment 4.1, an effect of the inter-item distance of the target to distracters interacted with target position in the 16-item set, and there was an effect of target position in the centrally-presented targets in the 8-item set. This suggested that there may have been an influence of the target’s position relative to the global structure in set size 8 (and also a possible contribution of local arrangement), and relative to the local structure (i.e. the configural arrangement of the target relative to other items) in the larger set. A potential way to dissociate these effects is to shift the distracter items away from the structure of the frame – in other words, if the aforementioned interpretation is correct, then a uniformly shifted distracter set which is misaligned according to the global display’s axes will predict its own structure independent of the frame. In this case, we might hope to isolate the structural alignment of the target from the relative influence of the distracter set or the frame. While there appeared to be some mutual contribution of the effects for 8-item sets in Experiment 4.1, in Experiment 4.2 this could not be possible (i.e. the target could not be aligned with both the distracter set and the frame). We predicted that this would produce an effect of balance in all set sizes, but that the locus of the effect would be differentiated by stimulus complexity; the smallest set would be influenced by the skeletal structure, while the larger set would be affected by inter-item structure.
Method

Participants

18 participants (12 female, 6 male; mean age 27) took part in Experiment 2, which was taken to be sufficient to detect a medium to large effect for 80% power at the 0.05 significance level (Champely, 2012). All had normal or corrected-to-normal vision. Approval of research ethics was given by the institutional Research Ethics Committee at NUI Galway.

Apparatus and Stimuli

Stimuli were similar to those used in Experiment 4.1, with the following change: while the target again varied between axial alignment and misalignment, the distracter array was now uniformly misaligned (see Figure 4.5).

![Figure 4.5 An example of stimuli used in Experiment 4.2.](image)

The participant was required to determine if a target (defined by being slightly larger in radius to distracter items – illustrated here above the main horizontal meridian, and to the left of the main vertical axis) is present or absent. This target could be aligned with an axis of symmetry of the frame (“on-axis” as here) or misaligned with the structural axes (“off-axis”). The dashed lines of symmetry depict the frame’s structural axes and the dotted lines refer to a uniformly shifted axial structure defined by the distractor items. Neither were revealed to participants and are shown here for illustration purposes.
Design and Procedure

There were no changes to the experimental procedure other than those used to alter the appearance of the stimuli.

Results and Analysis

Trials with response errors (5.6% of total) were removed from the data prior to subsequent analyses. Error RTs were similar to the correct RTs, and when analysed for proportion correct by RT, there appeared to be no tendency for increased errors for faster RTs, so no speed-accuracy trade-offs were observed ($R^2 = .244$, ns). As before, Shapiro-Wilks tests of normality indicated that the RT distributions were significantly non-normal (all $p$-values $< .001$), so log-transformed RTs were used in the subsequent analyses.

For the log RTs, Mauchly’s test of sphericity was significant for set size and the interaction of set size and target condition (present/absent), so Greenhouse-Geisser estimates were adopted. A repeated-measures ANOVA with the factors target (present, absent) and set size (4, 8, 16) showed that there was a significant main effect of target ($F_{1,17} = 10.03, p < .05, \eta_G^2 = 0.07$), with RTs increasing as set size increased ($F_{1.28,21.8} = 35.45, p < .001, \eta_G^2 = 0.04$). The two-way interaction was also significant, $F_{0.56,9.57} = 10.18, p < .001, \eta_G^2 = 0.02$. As in Experiment 4.1, the comparison of slopes indicates that target coding occurs in parallel, modified by a tendency for the displays to encourage serial search (see Figure 4.6).
Figure 4.6 Exponents of mean log-transformed RTs by set size and target position for Experiment 4.2.

Error bars represent standard errors.

A repeated-measures ANOVA was conducted for the target-present trials only, with the factors target position (on or off-axis), target location (central or peripheral) and set size. As in Experiment 4.1, there was a significant main effect of target location, \( F_{1,17} = 80.98, p < .001, \eta^2_G = 0.13 \), but in addition, a significant interaction between set size and target location, \( F_{0.98,16.69} = 11.8, p < .001, \eta^2_G = 0.01 \). Figure 4.7 suggests that this results from the slope describing RTs to peripherally-located targets, which increases with increasing number of display items, interacting with the flat search function for centrally located targets. This is consistent with the interpretation of Experiment 4.1. The 2-way interactions of set size with target position and target position with target location were non-significant (\( F < 1 \)), although the 3-way interaction between set size, target location, and target position was significant (\( F_{1.9, 32.28} = 6.69, p < .05, \eta^2_G = 0.01 \)).

Based on the significant 3-way interaction, interaction effects and simple effects were examined. First, we assessed the possibility that the set size by target
position interaction would vary by target location (central or peripheral) by conducting 3 x 2 ANOVAs for each of the two levels of target location (with a corrected alpha of 0.025 for family-wise error). When the target was in a peripheral location, there was no significant interaction of set size and target position ($F_{2, 34} = 3.19, p = \text{ns}$), but there was a significant interaction of these variables in the central target location condition, $F_{2, 34} = 11.13, p < .001$. Figure 4.7 shows that there was an increase in RTs across set sizes (i.e. some implication of serial search), and the search slopes appear to exhibit some notable differences in within-set target conditions.

With the data set divided by target position there was a simple main effect of set size for the on-axis ($F_{2,34} = 9.00, p < .001$) and off-axis ($F_{2,34} = 22.45, p < .001$) peripheral targets, and the on-axis ($F_{2,34} = 5.62, p < .025$) but not off-axis targets ($F < 1$) for central target presentation. Along with Figure 4.7, these results suggest that when attention towards the target is facilitated by inter-trial fixation (i.e. the target is central), there is no appreciable increase in RT for additional items in the off-axis condition, i.e. localisation is conducted more or less in parallel, while all other conditions demonstrate an increase in RT across sets. This may imply that the difference in RT for centrally-presented targets is unaffected by an attentional bottleneck, while no such effect is demonstrated for peripheral targets because they are localised serially.
Figure 4.7 Exponents of mean log-transformed RTs by set size, position, and location for Experiment 4.2

Error bars represent standard errors.

**Palmer-Guidi Rating.** As in the previous experiment, for all set sizes for the relationship between PGR and RT there was significant variation in the intercept across participants, set size 4: SD = 0.23 (95% CI = 0.17, 0.32), $\chi^2(1) = 463.27, p < .001$; set size 8: SD = 0.22 (95% CI = 0.16, 0.30), $\chi^2(1) = 508.44, p < .001$; set size 16: SD = 0.21 (95% CI = 0.15, 0.29), $\chi^2(1) = 450.16, p < .001$. For all sets, there was no significant variation across participants for the slopes of these linear fits (all p’s >.05), so an intercept-only model was adopted.

For the final models including both predictors and their interaction, PGR significantly predicted RT for all set sizes. Target position was not significant for sets 4 or 8, but significantly predicted RT in set size 16. The two-way interaction of these effects was significant for sets 4 and 16, but not for the middle set (see Table 4.4). This final model was also the best candidate for the 4- and 16-item sets according to AIC, and the simple PGR/RT model was the best candidate in the 8-item set.
Figure 4.8 shows the relationships between PGR and RT for each set size, and it is apparent that the linear fits are different for each target condition. For set size 4, there was a significant linear function describing PGR by RT for the on-axis ($\beta = -0.07$, SE($\beta$) = 0.013, $t(522) = -5.69, p < .001$) and off-axis conditions ($\beta = -0.04$, SE($\beta$) = 0.015, $t(372) = -2.36, p < .025$), but more variance was explained in the first case ($R^2 = 0.058, F(1,522) = 32.34, p < .001$) than in the second ($R^2 = 0.015, F(1,372) = 5.55, p < .025$). The opposite result was shown in the 16-item set, where there was a significant linear relationship between PGR and RT for the off-axis ($\beta = -0.079$, SE($\beta$) = 0.014, $t(507) = -5.78, p < .001$) but not on-axis condition ($\beta = -0.012$, SE($\beta$) = 0.017, $t(569) = -0.74, ns$). The former also explained a significant proportion of the variance in RT, off-axis: $R^2 = 0.62, F(1,507) = 33.45, p < .001$; on-axis: $R^2 < .001, F < 1$. This is most likely due to the expediency of RT for on-axis targets in each case; In the four item set, the target is on-axis according to the frame, but in the 16-item set, a target that is structurally on-axis is off-axis according to the distracter set (and vice versa).

![Figure 4.8 RT results by PGR scores for Experiment 4.2](image-url)
Table 4.4
Fixed and random effect estimates for response times as a function of Palmer-Guidi ratings in Experiment 4.2

<table>
<thead>
<tr>
<th>Set Size</th>
<th>4</th>
<th>8</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td>B</td>
<td>SE</td>
<td>t</td>
</tr>
<tr>
<td>Random Effects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>6.62</td>
<td>0.07</td>
<td>94.89</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PGR</td>
<td>-0.04</td>
<td>0.01</td>
<td>-3.75</td>
</tr>
<tr>
<td>Target</td>
<td>0.08</td>
<td>0.06</td>
<td>1.34</td>
</tr>
<tr>
<td>PGR x Target</td>
<td>-0.03</td>
<td>0.01</td>
<td>-2.25</td>
</tr>
</tbody>
</table>

Note: B = unstandardised beta coefficients, SE = standard error. PGR = Palmer-Guidi ratings. Target = axial alignment of the target item (aligned/misaligned), ns = not significant.

**Inter-dot Distance.** As in the previous experiment, the average inter-dot distance (IDD) of the target was calculated as the arithmetic mean of the distances between the target and each distracter. RT data were modelled in a multi-level analysis with participant as the random effect and inter-dot distances (IDD) and target position as fixed effects for each set size separately for target-present trials only. For all set sizes for the relationship between IDD and RT there was significant variation in the intercept across participants, set size 4: $SD = 0.23 \ (95\% \ CI = 0.17, 0.32), \chi^2 (1) = 463.27, \ p < .001$; set size 8: $SD = 0.22 \ (95\% \ CI = 0.16, 0.30), \chi^2 (1) = 508.44, \ p < .001$; set size 16: $SD = 0.20 \ (95\% \ CI = 0.15, 0.26), \chi^2 (1) = 450.16, \ p < .001$. For all set sizes, there was no significant variation across participants for the slopes of the linear fits (set size 4: $\chi^2 (2) = 0.003, \ ns$; set size 8: $\chi^2 (2) = .003, \ ns$; Set size 16: $\chi^2 (2) = 1.20, \ ns$), so an intercept-only model was adopted in all cases.

For the final models including both predictors and their interaction, IDD significantly predicted RT for set sizes 8 and 16, but was not significant for the 4-item set. Target position was not significant for set sizes 8 and 16, but was a significant predictor of RT in the 4 item set. The interaction term was not a significant predictor.
of RT for the 4 or 8 item sets, but significantly predicted RT in the 16-item set (see Table 4.5). Likelihood ratios and fit criteria are shown in Table 6, and suggest that for the smallest set, the model including both predictors but not their interaction is the most effective for the data according to AIC. For the middle set, the model including both predictors but not their interaction is the best candidate model, and selecting this model shows that target position is a significant predictor of RT, $\beta = -45.89$, SE($\beta$) = 15.43, $t(548) = -2.97$, $p < .05$. The final model including both predictors and their interaction (as reported) is the best model for the data in the 16-item set.
Table 4.5  
Fixed and random effect estimates for response times as a function of inter-dot distance in Experiment 4.2

<table>
<thead>
<tr>
<th>Parameter</th>
<th>B</th>
<th>SE</th>
<th>t</th>
<th>p</th>
<th>B</th>
<th>SE</th>
<th>t</th>
<th>p</th>
<th>B</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>6.41</td>
<td>0.07</td>
<td>86.33</td>
<td>&lt;.001</td>
<td>6.09</td>
<td>0.16</td>
<td>39.08</td>
<td>&lt;.001</td>
<td>3.62</td>
<td>0.46</td>
<td>7.92</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IDD</td>
<td>0.01</td>
<td>0.01</td>
<td>1.07</td>
<td>ns</td>
<td>0.01</td>
<td>0.02</td>
<td>2.92</td>
<td>&lt;.001</td>
<td>0.02</td>
<td>0.04</td>
<td>6.41</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Target</td>
<td>-0.15</td>
<td>0.07</td>
<td>0.07</td>
<td>&lt;.05</td>
<td>-0.24</td>
<td>0.20</td>
<td>-1.21</td>
<td>ns</td>
<td>2.53</td>
<td>0.49</td>
<td>5.23</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>IDD x Target</td>
<td>0.01</td>
<td>0.01</td>
<td>1.74</td>
<td>ns</td>
<td>0.002</td>
<td>0.02</td>
<td>0.87</td>
<td>ns</td>
<td>-0.02</td>
<td>0.04</td>
<td>-4.91</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Note: B = unstandardised beta coefficients, SE = standard error. IDD = inter-dot distance. Target = axial alignment of the target item (aligned/misaligned), ns = not significant.

Table 4.6  
Model fit diagnostics for anti-logs of mean log-transformed response time by inter-dot distance from the target to distractor items for each set size in Experiment 4.2.

<table>
<thead>
<tr>
<th>Set Size</th>
<th>Model</th>
<th>Δχ^2</th>
<th>AIC</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Model 1</td>
<td>12.93</td>
<td>179.77</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Model 2</td>
<td>5.48</td>
<td>176.29</td>
<td>&lt;.05</td>
</tr>
<tr>
<td></td>
<td>Model 3</td>
<td>3.02</td>
<td>175.27</td>
<td>ns</td>
</tr>
<tr>
<td>8</td>
<td>Model 1</td>
<td>20.02</td>
<td>172.90</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Model 2</td>
<td>15.87</td>
<td>159.03</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Model 3</td>
<td>0.76</td>
<td>160.27</td>
<td>ns</td>
</tr>
<tr>
<td>16</td>
<td>Model 1</td>
<td>3.00</td>
<td>203.99</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Model 2</td>
<td>18.42</td>
<td>187.57</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Model 3</td>
<td>23.94</td>
<td>165.63</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Note: AIC = Akaike’s Information Criterion, Lowest values of AIC represent the best model fit. Model 1 includes IDD as a predictor, Model 2 includes IDD and target position (main effects model), and Model 3 includes IDD, target position, and their interaction. Likelihood ratio (χ^2) comparison for Model 1 refers to the difference (Δ) in -2LL (log-likelihood) between this model and an intercept-only model, ns = not significant.

**Error Analysis**

Error trials were analysed by investigating the arcsine-transformed error probabilities for the different conditions, based on the prediction that the patterns of error would not produce substantive differences between target conditions. In confirmation of this, there was no significant main effect of target position (F_{2,34} =
0.89, ns, $\eta_G^2 = 0.02$), but set size had a significant effect ($F_{1.59, 27} = 3.79, \text{ns}, \eta_G^2 = 0.03$). Their interaction was not significant ($F_{2.34} = 1.94, \text{ns}, \eta_G^2 = 0.03$).

### 4.6.1 Discussion of Experiment 4.2

These results correspond with those of the previous experiment, and demonstrate that when stimuli are more populated, local structure produces an effect of target imbalance that is accentuated by proximity to distracter items. The axial structure of the frame was shown to produce an effect when the target is centrally-presented in the smallest set size (i.e. it is closer to axial convergence) and in the 8-item peripheral target presentation (the target is distant from distractor items which define an alternate structure). A significant interaction between PGR and RT for the 4 and 16-item sets suggests that RTs are faster for on-axis targets (where the axis is defined by the frame and the distracters respectively), and this difference becomes more pronounced in areas of higher PGR.

The results from analysis of the inter-item distance from the target to the distracters are similar to those of Experiment 1: the most notable result is an interaction between IDD and target position in the 16-item set only. The 4 and 8-item sets appear to be affected by target position, but not IDD (set size 4) or the interaction of IDD and target position (both sets). In all target conditions, the pattern of errors was the same as in Experiment 4.1; that is, there was no interaction effect between target position and set size, and no difference in accuracy between target-present conditions. Once again, this demonstrates that the axial manipulation did not appreciably affect the discriminability of the targets. These findings support the proposal that local structure defines imbalance for more populated displays, and that the skeletal structure defines imbalance in smaller sets.
4.7 Experiment 4.3: Paired Comparisons of Stimuli from Experiment 4.1

The previous experiments provide some confirmation of the hypothesis that axial alignment according to the skeletal structure of the frame would influence response time. According to Arnheim (1954), this difference in perceptual organisation confers the structural arrangement of the stimulus with the expressive quality of balance. However, while this might prompt observers to prefer balanced stimuli based on their inherent expressiveness or potential for aesthetic experience, alternatively the effect may not be striking enough to exhibit such phenomena (Palmer et al. 2013). Furthermore, the proposition that balanced stimuli, by producing a more prägnant representation, is therefore preferred (which is similar to the proposal that more fluent processing predicts increased liking; e.g. Reber, Winkielman, & Schwartz, 2004) may not be the outcome. Imbalanced stimuli may be considered to be more aesthetically interesting due to their unsettled appearance, producing the compelling perception of an unfinished action (e.g. Elliott, et al., 2012). As such, a paired comparison procedure was implemented in order to ascertain whether the guidance of attention by visual balance manifested in a consciously aesthetic preference.

Method

Participants

The same participants who took part in experiment one were involved in experiment three.

Design and Procedure

Experiment 4.3 used a paired comparisons procedure to investigate whether a measureable aesthetic preference existed for axial alignment or misalignment. Each item array employed in Experiment 4.1 (an entirely axially-aligned item set or one
item axially-offset) was presented alongside a stimulus array from the opposite axial condition (and presentation orders were fully randomized). In this case however, all items in the array were the same size. For each stimulus category there was a comparison with 3 other stimulus categories comprising the opposite axial condition for each set size (so e.g. the on-axis 4-item display was compared to the off-axis 4, 8, and 16-item displays). Each condition was presented in one of 12 random configurations, with counterbalanced presentation with each of the other possible stimulus categories. Each condition was therefore involved in 72 comparisons and because of repetitions there were 3 such arrangements, giving a total of 216 trials per participant, broken into 9 24-trial blocks. Participants were asked to report which of the two displays they thought was the most aesthetically pleasing (there was no indifference response available). Aesthetic was defined to the subject in simple terms of which of the two images the participant preferred. The stimulus presentation and experimental conditions were identical to those employed in Experiment 4.1.

**Results and Analysis**

Preferences were analysed by investigating the arcsine-transformed preference probabilities for the different conditions (the proportions represented the total number of times a particular axial alignment and set size was chosen from any of the pairs it was presented in). Apparent from Figure 4.9, there was an obvious increase in preferences for more populated sets, and this was shown in a significant main effect of set size, $F_{1.14, 9.15} = 15.3, p < .001, \eta^2_G = 0.59$; Greenhouse-Geisser corrected. There was no significant main effect of axial position ($F < 1$) and no two-way interaction, $F_{16} = 2.41, p = .12, \eta^2_G = 0.036$. Pairwise t-tests were conducted among each target condition and set size. Apart from the comparison between the on-axis 16-item vs off-Axis 8-item ($p = .031$) and the off-Axis 8-item displays vs off-axis 16-item
displays (p = .037), there were no significant differences for preferences for any other comparisons (all p’s > .05).

4.8 Discussion of Experiment 4.3

Experiment 4.3 examined the possibility that the RT performance effects of axial alignment of the target would influence visual preferences. Instead, it appears that participants tended to prefer more populated displays, likely due to the increased possibility of extracting representational content from the random arrangement of dots in those stimulus arrays. Because the main source of interest in this experiment was to test the extent to which inter-set comparisons would differ in preference for axial alignment, this finding does not provide any support for the hypothesis that the perceptual organisation of the displays has affected preferences via implicit axial coding.

![Preference Proportions by SetSize](image)

*Figure 4.9* Arcsine-transformed preference proportions by set size and target position for Experiment 4.3.

Error bars represent standard errors.

4.9 General Discussion

The experiments reported here indicate that visual balance can be operationalised as alignment with axial structure, and that this occurs due to relational
coding processes. Axially-aligned elements in the search display are quickly grouped in a Gestalt so that differentiation of the target on the basis of a feature singleton (size) is performed efficiently. The search slope of the axially-misaligned target also produced a shallow function, implying that the task of differentiating the target was similarly trivial. However, in this case, there appeared to be a difference in the response times between target conditions. This implies that the difference between a balanced and imbalanced target derives from the relational coding of similar stimuli into a textural or figural unit on the basis of axial structure. While performance is faster overall for on-axis stimuli, implying expedited perceptual organisation, in Experiment 4.2 off-axis targets were shown to be localized at around the same RT regardless of the number of items to be searched. Furthermore, as in this experiment the distracter set was also misaligned, this may imply that relational processing occurs faster and target detection is facilitated. The on-axis targets are nonetheless detected faster, suggesting that this process subserves saliency through inter-item dissimilarity (Duncan & Humphreys, 1989). It would appear that all items in the display are grouped by structure prior to differentiating by size, and that this occurs in either response condition, but is more rapid when the items are aligned. In other words, axial alignment facilitates the formation of a relational representation of the image structure.

These results are largely supportive of Arnheim’s prediction regarding the representation of the structural skeleton of a visual display. However, the effects of structural alignment, while producing similar effects at each attentional scale (i.e. axially aligned targets are quicker to locate), differ regarding the local or global stimulus characteristics that define the structure. In other words, when a global orientation is adopted, the skeletal structure defines balance, but when a local
orientation is adopted, the inter-relationships of items define balance. This aligns with some previous research on configural processing during search. Kimchi (1998) found that a configuration of many small elements primed responding at short latencies in a matching task and facilitated efficient visual search, but that a display of fewer larger items required longer exposures to prime responding, and led to inefficient search. This former result implies a configural superiority effect where stimuli are matched based on global organisation and search is performed in parallel, possibly because the stimulus array is grouped as a texture rather than individuated as objects. The latter result seems to suggest a different process of organisation whereby less numerous large objects are easily individuated, but more difficult to group. This is in line with the finding that grouping is prior to segmentation, both in the microgenesis of visual perception, and the ontogenesis of the visual system (Kimchi, Hadad, Behrmann and Palmer, 2005). Furthermore, it is consistent with findings in experimental aesthetics that suggest that as a person becomes more expert or experienced with art, there appears to be an increasing tendency to orient attention to global scene characteristics over local features (Pihko et al., 2011; Vogt, 1999; Vogt & Magnussen, 2007).

Previous attempts to investigate visual balance have relied on intuitive but nebulous operations such as “rightness”, or assumed that balance and symmetry are interchangeable. The evidence reported here offers the possibility that symmetry is not required for a difference in balance to be manifested in perceptual organisation, and suggests that the construction of a visual representation is facilitated by alignment with the axial structure of a scene. This supports the proposal that visual balance is an emergent feature exhibiting a configural superiority effect. In other words, the phenomenon is something other than the sum of its parts, and this “something other” expedites the ability to respond to a balanced scene or object. This potentially offers
an important contribution to understanding of the link between visual scene structure and its phenomenal quality. These results may also underscore the importance of giving consideration to the context of stimulus presentation in visual search experiments. Traditionally, predictions and interpretations from this paradigm are based on the speed of discrimination of a target from non-targets. However, item/background contrast (orientation, luminance, and spatial frequency) has been shown to influence targets and non-targets in different ways (de Vries, Hooge, Wertheim, & Verstraten, 2013), and structural alignment may be considered in a similar way. If an experiment includes randomised target location, and especially if there is a surrounding frame, it may be worth noting that the location of the target as well as its relationship with distracter structure may be contributing to any observed effects.
Chapter 5: Visual Balance as an Influence on Orienting Preferences in Domestic Chicks (*Gallus gallus domesticus*).

5.1 Introduction

The previous chapter provided evidence to suggest that preattentive organisation provides the means by which visual balance can influence performance in visual search. As such, the results offer support for early implicit coding of relational structure in human observers. This may also imply that stimulus organisation occurs in similar manner in visual systems with less complexity than primates, and does not depend on the influence of information processing in higher cortical structure. If this interpretation is correct, then the effect may be due to phylogenetically ancient mechanisms that develop prior to a point of divergence across other visually-guided vertebrates, leading to similar effects on behaviour in species with less complex brain architecture. In this chapter, two experiments are reported in which domestic chicks (*Gallus gallus domesticus*) are shown to exhibit preferential orientation towards displays with a degree of axial misalignment. As in the previous chapter, visual balance is operationalised as axial alignment, and one item in an array of dots is manipulated between conditions (“balanced” and “imbalanced”). Chapter 4 reported that axially-aligned stimuli facilitated faster responding when targets were presented close to fixation, but misalignment may have resulted in a greater reduction in response latency as a function of inter-item proximity for more populated displays. A lack of consistent affects of the same stimuli on aesthetic preference may be due to the complexity of human aesthetic experience; balance may provide a primitive or basic feature for aesthetic experience, but contextual factors must also be present. The ease and consistency with which preference choices are made suggests that all visual experience may be of aesthetic
relevance. However, humans usually require situational factors (such as an aesthetic attitude or a particularly striking image structure) for conscious awareness of these qualitative aspects to be achieved (Palmer, Schloss, & Sammartino, 2013). In line with this reasoning, a possible analogue may be demonstrated in non-humans by preferential orienting behaviour, based on the idea that non-humans will be more reflexively influenced by basic aesthetic features as they require less high-level elaboration for awareness to be affected. The present chapter presents data in support of this prediction.

5.2 Evolution and Divergence of Amniote Vision

Birds, mammals, and reptiles (amniotes) share some close similarities in the function and structure of their visual systems, likely reflecting a common ancestry in the Captorhinidia, the stem reptile for all amniotes (Husband & Shimizu, 2001). This species became extinct 300 million years ago, so shared ancestry is mostly studied by tracing similarities in modern amniotes (Husband & Shimizu, 2001). Similar function and structure does not necessarily imply common divergence from the amniote ancestral line (homology), as there could also be convergent development of similar adaptive strategies given similar environmental and/or biological constraints (homoplasy; Campbell & Hodos, 1970, Rendall & Di Fiore, 2007). For example, the eyes of humans share physical similarity with the unrelated cephalopods (squid, octopus, etc), but this is due to convergence in evolution; amniote eyes are developed from sensitised skin, while cephalopod eyes originate from outgrowths in nervous system tissue (Ings, 2007). However, we make the assumption that the effect of visual balance on perceptual organisation in the previous chapter will be similar in birds due to ancient homologous mechanisms. Birds and humans are known to have common ancestry prior to a point of divergence from another species with similar visual system
structure (i.e. reptiles), but differ from that species in the complexity of their visual processing (Husband & Shimizu, 2001). This make it reasonable to propose that some of the visual cognitive mechanisms in avians and primates are due to homologous evolution, and that microgenetically early perceptual processing may share common characteristics across these species. Microgenesis refers to the evolution of visual perception from the early registration of optical information on the retina to the conscious perception of objects (Smith, 1957; Werner, 1956). Bachmann (2006) explains it as “the short term formation of a psychological process (p.12)”, occurring in the first ~300ms of exposure to stimulus information. This process involves the transformation of stimulus features into a unified relational object code that enables attention and conscious awareness (Crick & Koch, 2003). The previous chapter demonstrated that preattentive processing in humans is influenced by visual balance, so it appears likely that if chicks respond to stimuli in a way that suggests a similar microgenesis to humans, then they should be similarly affected by structural balance.

5.3 Avian Aesthetics

The focus of this chapter is on the effects of visual processing on preference responding in chicks. Some analogies have been made between the orienting behaviour of avians and the aesthetic experiences of humans, with chicks showing preference to respond to the same kinds of images that humans prefer. For instance, domestic chicks tend to prefer images showing human faces which are also rated as more beautiful by humans (Ghirlanda, Jansson, & Enquist, 2002). Nicki and Rogers (1975) found that chicks tended to prefer some level of stimulus complexity up to a point, but then exhibit a gradual decrease in approach behaviour as a linear function of complexity, a finding comparable to that found in humans (but with a lower point of inflection; Berlyne, 1971). Of relevance to the present chapter, visually naive
chicks are shown to demonstrate a spontaneous preference for asymmetric patterns (Clara, Regolin, & Vallortigara 2007). Furthermore, avians are shown to be able to differentiate some features of visual stimuli, which presumably contribute to aesthetic experience. Watanabe (2009) found that pigeons could differentiate paintings based on artist or style, or by being trained to respond to “good” rather than “bad” paintings (as judged by human observers), then generalising this response property to novel stimuli.10

5.4 Comparative Psychology and Visual Experience

Given the difference in relative brain size and architecture, there may seem to be a wide gulf between the possible computational complexity of primate and avian visual perception mechanisms. However, birds are highly visually-guided, and a large proportion of their brains are devoted to visual perception. Avians behave in a way that appears consistent with the proposal that they, like humans, inhabit a perceptual world of objects, so their brains achieve many of the same general competencies in perception (Cook, 2000; Shimizu, 2009; Vallortigara, 2004).

While the previous chapter demonstrated that early visual organisation processes can exhibit an influence of visual balance on perceptual performance, this chapter aims to establish a corollary effect on preference behaviour in visually-guided organisms. As such, some account must be made of the evaluative capabilities of the subjects; chicks must have demonstrable capacity to compare stimulus configurations and choose a preferred one. While the locus of the effect is proposed to be at an early stage of vision (and probably homologous with human or primate effects), this task

10 In Watanabe (2009), paintings were judged as “good” when they had clear or discernible content. It may be unnecessary to describe the pigeons’ ability to assess clarity (interesting in its own right) with terminology as fraught as “good/bad”. In either case, the findings suggest that the evaluations performed by pigeons were consistent with those of human observers, implying similar object-orientated phenomena for each species.
involves some account of goal-direction, working memory, and decision making (Cook, 2000). A good deal of recent research makes the case that avians, like primates, are capable of this kind of behaviour due to comparable neuropsychological mechanisms (Shimizu, 2009).  

5.5 Avian and Human Neurophysiology and Visual Cognition

Birds have been shown to perform many of the visual cognitive functions associated with higher primates. This is in spite of the differences in organisation of bird and mammal forebrains. Most notably, the avian telencephalon is unlaminated (Karten & Shimizu, 1989), which led early researchers to propose that birds had a hypertrophied striatum but no pallium (one of the major components of human cerebrum, including within it sections of cortex and amygdala; Güntürkün, 2005). However, this is not the case, and the avian pallium has since been compared with mammalian neocortex (e.g. Jarvis et al., 2005). While the differences in pallial organisation between species are obvious at a gross level, they exhibit comparable fine anatomy, physiology, and visual cognitive mechanisms. Dopaminergic innervation, a notable feature of the mammalian prefrontal cortex (PFC) and a vital component in the process of sustaining activity levels in working memory, is also found in the pallium (Güntürkün, 2005). Like primates, birds exhibit two overlapping and interconnected modality-specific pathways (Cook, 2000). In birds with eyes placed laterally on the head, the relative importance of each pathway is reversed in comparison with primates (Cook, Cavoto, Katz, & Cavoto, 1997). The major pathway is the tectofugal, which corresponds to the secondary extrastriate pathway in mammals. The secondary pathway is the thalamofugal, which corresponds to the

---

11 Although such aspects of behaviour or responding may be also characterised by phylogenetic convergence on a similar adaptive function (homoplasy; e.g. Rendall & Di Fiore, 2007). Based on the results of the previous chapter (observable effects on performance but not preference in humans), we propose that visual balance influences behaviour at a level homologous to early perception in humans rather than a later homoplastic stage of processing.
primary geniculo-striate pathway in primates (Vallortigara, 2004). Visual performance exhibits gross disturbances if the tectofugal pathway is damaged (Cook, 2000). In both species, information proceeds from primary sensory areas to secondary or association areas as well as frontal areas which transmit afferent signals (Güntürkün, 2005). This makes both the pallium and the PFC areas of convergence of ascending sensory and descending motor signalling. Güntürkün (2005) suggests that there are a large number of similarities in the anatomical networks. The pallium, like the PFC, is at the intersection of sensation and action, has dopaminergic signal modulation, and interconnections with limbic, visceral, and memory-related areas (Güntürkün, 2005). The PFC allows for goal-oriented behaviour – planning and execution of actions, with the ability to assess performance along the way. Working memory performance in pigeons appears to be reduced proportional to the extent of lesioning of the pallium, and sustained firing activity during the physical absence or a related stimulus suggest that the pallium operates like the PFC in subserving goal-directed behaviour (Güntürkün, 2005). What these features argue for is not necessarily a strict analogue in the avian pallium (because the large-scale topography is too different), but a suggestion of commonality between mammals and birds in the visual cognitive operations required for flexible and adaptive survival (Güntürkün, 2005). In other words, the visual cognitive processes implied by these similarities suggest that avians experience the visual world in a similar way to primates, and that responding on the basis of visual preferences may be similar (in kind if not complexity) across species.

5.6 Perceptual Organisation in Avian Early Vision

Both birds and mammals seem to have a perceptual architecture that generalises in a bottom-up fashion from simple features to more complex structure.
Cook and colleagues (Cook, Cavoto, & Cavoto, 1996; Cook, Cavoto, Katz, & Cavoto, 1997) have looked at how pigeons (*Columba livia*) process visual information in order to ascertain the differences and similarities in how birds and mammals see the world. Cook, Cavoto, & Cavoto (1996) used a paradigm based on visual-search experiments in human observers in order to investigate similarities in pre-attentive processing. Pigeons were trained to select a target region that is randomly located within a textural distracter matrix. This region is defined by being different to distractor elements by a single dimension (e.g. a colour) which can be described as feature search, or a combination of dimensions (e.g. a combination of colour and line orientation; conjunction search). The resultant patterns of search efficiency were almost identical to human performance - feature search was not impeded by increasing item number, but conjunction search was. This shows that the way in which attention operates is comparable across species.

Cook, Cavoto, Katz, & Cavoto (1997) used a version of rapid serial visual presentation to look at avian grouping latency, with the prediction that grouping should be as quickly produced in avians as it is in humans. Their experiments involved a target area within a texture stimulus that was defined by colour (e.g. a small sub-matrix of blue squares in a larger matrix of green squares), but the target/distracter colour pair changed every 100ms. Pigeons were easily able to detect the target, which implies coding of the target area occurring at less than 100ms, similar to human organisational latency. The extent to which Gestalt principles of organisation apply to avian visual perception is an open question, and there may be differences. For instance, similar to infant but not adult humans, avians may group features by colour or texture far more easily than shape (Kirsch, Kabanova, & Güntürkün, 2008).
Further evidence for avian perceptual grouping comes from the finding that in birds, a number of visual illusions are shown to be experienced in the same way as in humans. Studies show that domestic chickens (Rosa-Salva, Rugani, Cavazzana, Regolin, & Vallortigara, 2013; Watanabe, Nakamura, & Fujita, 2013), doves (Warden & Baar, 1929), parrots (Pepperberg, Vicinay, & Cavanagh, 2008), and pigeons (Fujita, Blough, & Blough, 1991; 1993) appear to be sensitive to geometrical illusions such as the Ponzo, Müller-Lyer, Wöllner, horizontal-vertical, and Ebbinghaus illusions. Pigeons have also been shown to perceive partially occluded objects (Nagasaka, Hori, & Osada, 2005). These visual phenomena derive from aspects of the whole configuration, so are an aspect of perceptual grouping. The finding that avians experience these illusions in much the same way as humans suggests that their phenomenal experience is at the level of Gestalten, the same as humans. Furthermore, it implies that these phenomena occur early in visual perceptual processing. Taken together, the evidence suggests that the completion of a holistic perceptual representation does not depend on the particular anatomical structure of mammalian visual cortex as it can also be performed by differently structured, avian visual systems.

5.7 Axial Coding in Human and Avian Perceptual Representations

The previous chapter demonstrated that a relational code based on axial structure may be an integral part of preattentive processing. It appears likely that early perceptual organisation in avians is similar to mammalian perceptual processing, so there may be a similar tendency for axial alignment to facilitate the completion of holistic representational units in birds. This may affect visual phenomena by emphasising aspects of structure that may have relevance to object recognition. Theories of human object recognition are often divided into view-centred or object-
centred approaches (e.g. Logothetis & Sheinberg, 1996). View- or template-centred theories are intuitively appealing, as we experience the world according to regularities rather than seeming to intuit object structure with each viewing, and there seems to be some support for this view (Logothetis, Pauls, Bülthoff, & Poggio, 1994; Logothetis, Pauls, & Poggio, 1995; Tarr & Pinker, 1990; Tarr, Williams, Hayward, & Gauthier, 1998). However, when contextual factors such as distance, angle, lighting condition, etc. are taken into account there are an infinite number of potential viewpoints for any object, which might imply an unfeasibly huge number of representations for even a limited set of views, and this would be uneconomical with regard to neural resources. This argument has even more relevance when avian brains are considered. Furthermore, if view-centred approaches apply to avians, then matching rotated or rotating objects should present a problem. However, pigeons can easily match dynamically rotating objects, but have trouble matching static images (Cook & Katz, 1999), and their ability to match rotated still images is better than human performance (Hollard & Delius, 1982).

Object-centred approaches tend to rely on axial transformation of contour and vertex information to establish a representation that is robust to changes in viewing conditions (Blum, 1967; Marr & Nishihara, 1978). In humans, there is some evidence from computational (Cornea, Silver, & Min, 2007), psychophysical (Johansson, 1973; Kovacs & Julesz, 1994; Wang & Burbeck, 1998), and neurophysiological research (Kimia, 2003; Hung, Carlson, & Connor, 2013; Lee, Mumford, Romero, & Lamme, 1998; Lescroart & Biederman, 2013) that object recognition is facilitated by axial transformation. There is also evidence that birds use axial structure to represent spatial maps of the environment, however it is unclear whether this code is medial (Kelly, Chiandetti, & Vallortigara, 2010; Kelly, Durochet, Chiandetti, & Vallortigara,
2011) or principal in nature (Cheng & Gallistel, 2005; Cheng & Newcombe, 2005; Sturz & Bodily, 2011). In humans, both medial (Firestone & Scholl, 2014) and principal (Palmer & Guidi, 2011) axial structure has been shown to have an influence on visual preferences (See Figure 5.1). These results suggest that birds, like humans, rely on a structural relational code for visual experience and orienting behaviour.

Figure 5.1 Principal and medial axes of four-sided polygons.

(a) shows the principal axes and diagonals of a rectangle; Palmer & Guidi (2005) found evidence for visual preferences to be influenced by these axes. (b) shows the medial axial transformation of the same rectangular dimensions, based on Blum’s (1967) grassfire algorithm. This begins with a hypothetical fire lit at all perimeter points, and then progresses inwards. The fire is assumed to quench when all area has been covered, and the resultant quenchpoints represent the medial structure of the object. These axes are most commonly implemented as the midpoints of a trajectory of circles whose edges adjoin opposite contours. Firestone & Scholl (2014) found evidence for visual preferences for this axial structure, however they also found a preference for at least one of the principle axes (the main vertical) so our stimuli largely obviate the need to differentiate between these transformations by using square stimuli (effectively a combination of both transformation schemes) as in (c).

It may be possible that view- and object-centred representations are different aspects of the perceptual process (Hayward, 2012). For instance, viewpoint-dependent representations may be used for higher-order categorisation, then subordinate category assignations may be performed with structural processing (Wolfe, Kluender, & Levi, 2011). It has been suggested that their differential emphasis contributes to aesthetic experience. Cutting (2002) claims that artists use postural motion

---

12 This thesis does not necessarily propose that either a strictly principal or medial account is the correct one. The stimuli used in the present study are based on Arneheim’s structural skeleton for square frames, as well as the stimuli from the previous chapter’s experiments, and as such apply to a combination of principal and medial axial transformation (see Figure 5.1).
information such as contrapposto (arranging a human figure so that it's upper sections are turned in a different direction to the hips and legs) and garment drape to lend dynamic interest to a sculpture. Similarly, the various poses featured in Hokusai’s woodcut sketches (see Figure 5.2) are appreciated for their dynamism, but the postures of the figures in some cases are unusual, and deprived of such signifiers of form as a head (instead we see a hat) or clearly defined arms that might facilitate view-centric recognition. Canonical form and figural symmetry may be more perceptually fluent, but are taken to be uninteresting in such examples. Cutting (2002) suggests that the increased speed of apprehension and classification of canonically viewed objects (see e.g. Palmer, Rosch, & Chase, 1981) reduces the fuller intellectual or emotional impact of increased uncertainty. In other words, perceptual expedience impedes the potential for the visual object to be rich with potential for exploration or elaboration (Bouleau, 1980; Nodine, Locher, & Krupinski, 1993). We proposed that avians are less dependent on elaboration of this kind in order to be influenced by aesthetic primitives such as visual balance, but in terms of our predictions, this proposal aligns with the idea that axial alignment is coded quickly by being easier to organise in early vision, but that the salience of misalignment makes it more amenable to attentional capture by emphasising focus on structure. In other words, the difference between a balanced or imbalanced stimulus influences how much elaboration is facilitated, but this difference is registered early enough to affect avian behaviour.
Arnheim’s (1954) theory predicts that a certain degree of axial misalignment places an object at the intersection of two competing organisational processes – grouping with the axial structure or separate from it - the result of which is that it appears unsettled or indeterminate. The phenomenal effect is that visual imbalance appears to have kinetic energy, suggesting an incomplete action or event. This has obvious relevance to an organism at risk of predation. Our hypothesis can therefore be related to an innate advantage for misalignment to attract focal attention. Attentional deployment becomes a tool used to expend processing resources on the resolution of ambiguities in a visual scene. As such, we predict that chicks will be more interested in misaligned stimuli, reflected by an increased tendency to orient towards them during generalisation.

5.8 Experiment 5.1: Imbalance and Orientation Behaviour in Male Domestic Chicks

Method

Ethics Note

Both experiments reported in this chapter complied with Italian and European Community laws for the ethical treatment of animals and the experimental procedures were conducted under ethics approval at the University of Padua, licensed by
Ministero della Salute, Dipartimento Alimenti, Nutrizione e Sanità Pubblica Veterinaria (permit number 6723).

**Participants**

Participants were male domestic chicks (Gallus gallus domesticus) of the Hybro strain (N = 12), hatched from fertilised eggs obtained from a local hatchery (Agricola Berica, Montegalda [VI], Italy). Following hatching, chicks were reared socially for five days, then individually. Food and water was available *ad libitum* in this period. Prior to training or testing, chicks were food (but not water) deprived for a period of 2 hours.

![Figure 5.3](image)

*Figure 5.3* Experimental stimuli used during (a) shaping, (b) training, and (c) testing phases in Experiment 5.1.

Rows (b) and (c) refer to the BAL (left) and (IMBAL) condition’s correct stimuli (S+). Aligned and misaligned stimuli are identical apart from the position of the second uppermost dot. During the testing or generalisation phase (c), chicks were assessed for their ability to orient towards “spread-apart” versions of the stimuli at the training phase (b).

**Apparatus and Stimuli**

All stimuli were printed on rectangular white card (9 x 6 cm) and consisted of a configuration of grey dots (each with diameter 4mm) within a square black frame (3.2 x 3.2 cm). During the initial shaping phase a stimulus representing a single
central dot was used (Figure 5.3 [a]). During the discrimination training phase stimuli consisted of one aligned and one misaligned (one item offset from axis by 1/8 of its radius) configuration of 3 or 4 adjacent dots (Figure 5.3 [b]). In order to eliminate the possibility of cross-axial alignment, the item closest to the frame was distanced 1/8 of the frame’s length (and height) from the bounding line. The upper left quadrant was selected due to its potential for enhanced relevance and augmentation of the effect of structural misalignment, because this location optimally exploits the effects of visual biases (Arnheim, 1954; Bertamini, Byrne, & Bennett, 2013; Gaffron, 1950; Niekamp, 1981; Palmer, Gardner, & Sammartino, 2008; Wilson & Chatterjee, 2005). The main difference occurring between the two training stimuli was the position of the second dot from the top, which in the misaligned configuration was off-axis (see Figure 5.3 [b]). During the generalization phase, “spread apart” versions of the training stimuli, obtained by increasing the distance between the dots, were used (see Figure 5.3 [c]).

The experimental apparatus (Figure 5.4 [a] – [d]) consisted of a rectangular white-painted cage (33 x 38 x 60 cm) with a slit at the bottom of one of the short walls through which the food-box (6 x 6 x 12 cm) could be introduced. The food-box had a drawer that could be pushed open from outside of the cage by the experimenters in order to allow access to the food. The stimuli were fixed on the top of the food-box (at 45°).
Figure 5.4 Schematic diagram of the experimental apparatus for all experiments.

(a) denotes the moveable partition and (b) refers to the food-box (inset elements (i) and (ii) refer to the stimulus display and food-dispensing drawer respectively). Prior to the beginning of a trial, the chick is located in (c). When the trial begins, the partition is lifted, and the trial arena consists of (c) and (d) combined. During initial training, chicks are shaped to respond to a one-dot stimulus presented on (i), and the chick is then rewarded by gaining access to food contained in (ii) for a few seconds. The experimenter then closes the food box and moves the chick to area (c) behind the partition for the start of the next trial. During discrimination training, this apparatus is changed for one which has two food boxes, and during trials the chick chooses one of these based on the stimulus displayed on (i) for each food dispenser. The stimuli differ in the alignment of one of the dots in a configuration of axially-aligned dots (see Figure 5.3 [b]). Correct responses are those that match the grouping status of the chicks. Subsequent to a correct response (S+), the chick receives a reward in the same way as in initial training. If the response is incorrect the chick is confined in (c) behind the partition for 15 seconds. During the test phase, the chick is rewarded for pecking at either stimulus display (see Figure 5.3 [c]).

Results and Analysis

There was a significant difference between groups in the number of trials that showed generalisation between the training and testing phases ($t_{10} = -3.702$, $p < .005$) with IMBAL chicks ($M = 14.17$, $SD = 2.14$) demonstrating more group-specific responses than BAL chicks ($M = 10.33$, $SD = 1.37$). Furthermore, IMBAL chicks (median = 14.0/20, $t_{5} = 4.78$, $p < .005$) showed performance that was better than at chance (10 out of 20 responses correct), while the BAL chicks did not (median = 10.5/20, $t_{5} = .598$, $p = .576$). There was no significant difference in the number of
trials that groups required to reach the learning criterion ($t_{10} = 1.38, p = .198$), or in the number of correct responses each group performed prior to criterion ($t_{10} = 1.01, p = .336$).

5.8.1 Discussion of Experiment 5.1

The results imply that IMBAL chicks tended to adhere to the response category learned in the training phase, but that BAL chicks showed no systematic tendency to adhere to their response category. The trend in the IMBAL chicks’ performance suggests that all chicks can potentially differentiate the stimuli from each other, and this may reflect a natural innate preference for those configurations. We proposed that the basis of this differentiation is due to visual imbalance. However, the possibility that some other aspect of these configurations provided a basis for preference must be addressed. For instance, the stimuli in this experiment may resemble some ecologically-relevant object shape (such as a mealworm) that might have been more explicitly represented by the axially-misaligned stimulus. This appears unlikely given that the chicks were hatched onsite, so were visually naïve to such objects. Stimuli may furthermore have suggested some other higher-order aspect of the configuration besides balance, i.e. symmetry. To address these possibilities, the same tactic was used to modify stimuli for presentation in a second experiment.
Chicks in the BAL training condition were rewarded for pecking at uniformly axially-aligned stimulus configurations, while IMBAL chicks were rewarded for pecking at the dot configuration with one axially-misaligned element. Correct responses (scores) are stimulus choices during generalisation which are consistent with the training category for each group member. Note: y-axis is truncated.

5.9 Experiment 5.2: Preference for Imbalance Versus Preference for Shape in Domestic Chicks.

Method

In this experiment, we reduced the number of display items in the dot array, while using the same stimulus size as in Experiment 5.1. The effect of this is to emphasise the amount of axial non-coherence of the array overall (⅓ of items are misaligned rather than ¼), as well as to reduce the amount of similarity to ecologically-relevant single objects. This allows the stimuli to depict implicit axial structure rather than being representational.
Participants

Participants were male domestic chicks (Gallus gallus domesticus) of the Hybro strain (N = 8), obtained from the same hatchery and placed in the same socialisation and rearing schedule as in the previous experiment. As before, food and water was available *ad libitum* until 2 hours prior to training or testing, during which food (but not water) was restricted.

![Figure 5.6](image)

*Figure 5.6* Experimental stimuli used during (a) shaping, (b) training, and (c) testing phases in Experiment 5.2.

Rows (b) and (c) refer to the BAL (left) and (IMBAL) condition’s correct stimuli (S+). Aligned and misaligned stimuli are identical apart from the position of the second uppermost dot. During the testing or generalisation phase (c), chicks were assessed for their ability to orient towards “spread-apart” versions of the stimuli at the training phase (b). These stimuli differ from those in Experiment 1 in the absence of the second lowermost dot.

Results and Analysis

Data were log-transformed to correct for non-normality. As in the previous experiment, there was a significant difference between groups in the number of trials that showed generalisation between the training and testing phases ($t_6 = -2.67$, $p < .05$) with IMBAL chicks ($M = 13.00$, $SD = 2.2$) demonstrating more group-specific responses than BAL chicks ($M = 8.5$, $SD = 3.1$). Furthermore, IMBAL chicks (median = 12.5/20, $t_3 = 3.18$, $p = .05$) showed performance that was better than at
chance (10 out of 20 responses correct), while the BAL chicks did not (median = 9.5/20, t3 = -.965, p = .406). There was no significant difference in the number of trials that groups required to reach the learning criterion (t6 = -2.257, p = .101), or in the number of correct responses each group performed prior to criterion (t6 = -1.01, p = .313).

Figure 5.7 Boxplot of scores for each training category in Experiment 5.2.

Chicks in the BAL training condition were rewarded for pecking at uniformly axially-aligned stimulus configurations, while IMBAL chicks were rewarded for pecking at the dot configuration with one axially-misaligned element. Correct responses (scores) are stimulus choices during generalisation which are consistent with the training category for each group member. Note: y-axis is truncated.

5.9.1 Discussion of Experiment 2

The results of Experiment 2 replicate those of the first experiment very closely, with IMBAL chicks tending to adhere to the response category learned in the training phase, and BAL chicks showing a limited tendency to do so. Based on the reduction of stimulus elements, it appears likely that this result is due to coding of axial alignment rather than some preference for the object shape.
5.10 General Discussion

Our results show that imbalanced stimuli are differentiated from balanced stimuli for domestic chicks, and that this difference results in an increased tendency for IMBAL chicks to peck the correct stimulus according to their training category. On this basis, we conclude that chicks in these experiments appear to either more successfully encode the category of imbalance during training, or have an innate orientation towards imbalance during generalisation. There were no significant differences in the number of responses taken to criterion for the two groups in either experiment, so it appears unlikely that the results are due to a difference in acquisition at the training stage. It would seem reasonable to suggest that the orientation towards imbalance occurs during generalisation (and see below for a discussion on how this may relate to non-pragmatic evaluation). In either case, the findings match the prediction that axial alignment affects visual organisation and predisposes orienting behaviour towards imbalanced configurations.

The results indicate that chicks are capable of differentiating stimuli based on alignment, and there appears to be evidence to suggest that chicks prefer the category misalignment. A pilot experiment demonstrated that when frames were rotated by 45° during generalisation, thereby becoming misaligned in all cases, there was no tendency to peck towards either aligned or misaligned stimuli in the test phase. This suggests that the physical features of the stimulus are less important than the axial alignment status defined by their configuration. This preference to orient towards imbalance is consistent with our hypothesis, and on the basis of the previous chapter may be considered to be inter-species. The major structural differences of avian and primate brains argue against this being the result of a higher-level information-processing homology (although such homology has been attributed to the avian
pallium and human PFC), and instead is likely to be a property of brain function relating to early visual cognition and attention.

The tendency to orient towards imbalanced stimuli may have adaptive utility in the Umwelt of the chick. This refers to the “immediate perceptual environment” of an animal, which is a complex relationship between the animal’s physiology, phenomenal experience, and the potential interactions it might have with the environment (von Uexküll, 1934/1957). It seems unlikely that one potential aspect of this, the similarity between the visual stimuli and insect larvae, is operating in the present case, as the chicks were visually naïve and the stimuli in the generalisation phase (and all of Experiment 2) were spread apart so wouldn’t exhibit this similarity. Based on Arnheim’s model, and our predictions derived from it, the data appear consistent with the proposal that imbalance has the appearance of kinetic energy or tension, comparable to an imminent or incomplete action or event. Implied dynamics of this kind have been shown to affect recall for visual scenes, by causing errors in the direction of implied motion during localisation of a remembered item (Freyd, 1987; Ziemkiewicz, & Kosara, 2010). This does not appear to be an effect of mnemonic accuracy, as people tend to be better at remembering tasks which are incomplete rather than complete (Zeigarnik, 1927). Implied motion can also induce aftereffects in the same way as real motion (Pavan, Cuturi, Maniglia, Casco, & Campana, 2011; Winawer, Huk, & Boroditsky, 2008). A potential adaptive advantage to balance coding is that the entire environment could be assessed for potential motion by implied structural characteristics during normal object recognition. This covers potential predators as well as prey. Imbalance may suggest physically compromised prey or, given the fact that other organisms have bilateral symmetry along the axis of elongation (and therefore might imply a predator that is facing directly at or away
from the observer), it may be advantageous to avoid approaching symmetrical structure. Asymmetrical body coloration patterns are shown to correlate with social dominance in European starlings (*Sturnis vulgaris*; Witter & Swaddle, 1994), but a decrease in mating success in zebra finches (Swaddle & Cuthill, 1994). A possible interpretation is that there is no absolute preference for either symmetry or asymmetry and, consistent with our predictions, contextual factors mediate the potential for either of these phenomena to elicit preference behaviours. However, while it should be noted that these points are conjectural, the preference for imbalance may be an aspect of normal visual perception which has arbitrarily been conferred with ecological significance during evolution.

While prioritising attentional deployment to motion, real or implied, has obvious adaptive utility, this may not imply that its origin is straightforwardly pragmatic. A recent controversy in evolutionary aesthetics concerns the neo-Darwinist view that aesthetic value relates to the perception of biomarkers of immunocompetence or mate fitness (Buss, 1989; Cunningham, Roberts, Barbee, Druen, & Wu, 1995), and a classical Darwinist position which claims that aesthetic experience occurs in the absence of direct indicators of mate viability (Endler, 2012; Prum, 2012). Both approaches propose that aesthetic preference can facilitate reproductive success, but they differ in how this is achieved – the former claims that visual preferences are only those which convey fitness, while the latter argues that mates can use a natural preference for some aspects of the visual world in order to attract interest from a potential mate. The classical approach also implies that traits and preferences must therefore be coevolutionary but arbitrary. In other words, traits can become adaptive because of preferences, and preferences will evolve with
developing traits, but the form of the traits themselves is arbitrary (or unrelated to an improvement in fitness signalling).

Our results indicate that preference for visual imbalance is unrelated to a generalised preference for signals of mate viability. The data do not indicate a preference for symmetrical or balanced arrangement, which might be seen to signal immunocompetence (Grammer, Fink, Möller, & Thornhill, 2003; Thornhill & Gangestad, 1993; but see Rhodes, 2006). Also, it is difficult to relate our stimuli to any formal aspects of mate appearance, and while it could be argued that balance is an evaluative quality that is abstracted from visual comparison of candidate mates to all visual structure, it would appear to be an uneconomical strategy to evaluate everything in terms of mate candidacy, instead of limiting to the subset of potential mates. In agreement with the classical Darwinist perspective, aesthetic interest may not derive from mate preference, but may yet be used to attract potential mates. Visually interesting traits or displays can aid reproductive candidacy by drawing the attention of a potential mate. For instance, the elaborate nest building of male bowerbirds (*Ptilonorhynchus nuchalis*) may be part of a strategy to influence the attentional deployment of females and provide an illusion whereby the male appears larger (Endler, 2012; Endler, Endler, & Doerr, 2010). The process utilises innate sensory biases so that bowerbird aesthetic evaluation is best considered as a by-product of natural selection, rather than a direct result of it (Rhodes, 2006). Similarly, while favouring imbalance in the deployment of attention may confer a useful survival tactic for avians, the present research does not offer support for the strong claim that visual organisation evolved because of this utility.

These results indicate that structural alignment relates itself to orienting behaviour by influencing the spontaneous capture of attention by misaligned stimuli.
The behaviour demonstrated here may also indicative of an avian analogue of aesthetic experience in humans. The case for such an interpretation is made on the basis of a number of points. Firstly, the results shown here closely match those of the human performance experiments in the previous chapter, which were themselves in keeping with Arnheim’s predictions based on skeletal structure and visual balance. These results also align with to the finding that axial structure affects visual preferences in humans (Firestone & Scholl, 2014; Palmer & Guidi, 2011). Secondly, the experimental procedures offer a depiction of aesthetic experience that is in agreement with some codifications of this type of experience from philosophical perspectives. As the generalisation phases consist of a procedure whereby the stimulus-response contingency is extinguished via reinforcement for all responses, any preferences showed in this stage are largely dissociated from material needs. The preference for misaligned stimuli exhibited by IMBAL chicks may therefore be comparable to a non-pragmatic evaluative orientation towards the stimuli, which may provide a conceptual link between avian behaviour demonstrated here and human aesthetic experience (Bullough, 1912; Kant, 1790; Kemp, 1999; Palmer, Schloss, & Sammartino, 2013; Stolnitz, 1960; Zangwill, 1992), and may apply to the results shown here. Thirdly, the peak-shift principle, a psychological phenomenon observed in birds and applied to recent theories of aesthetics, may underlie the present data to some extent (Enquist & Arak, 1994; Ramachandran & Hirstein, 1999). The principle states that distorted or super-normal versions of familiar stimuli are preferred to new stimuli or unaltered versions of the familiar stimuli. This is proposed as a means of explaining the preference for novelty in visual art, and is in agreement with the given interpretation of our results, namely that misaligned stimuli are pecked more consistently because they are more interesting aesthetically.
The foregoing presents further evidence for the thesis predictions that visual structure may be conferred with aesthetic value due to the processes of perceptual organisation. While survival and reproductive constraints may have selected for this tendency, it may nonetheless be sensible to discuss aesthetic experience as an aspect of perception which has normative instantiation at a phenomenal level.
Chapter 6: Conclusion

The research presented in this supports the claim that aesthetic experience is fundamentally perceptual (or that perceptual experience is intrinsically aesthetic). This chapter aims to provide some qualifications and implications of this statement. Regarding aesthetic experience, it is necessary to clarify the extent to which perceptual processing influences conscious aesthetic evaluation. This is done by situating the results in terms of previous research in empirical aesthetics which have relied upon preference-based measures, and by offering an explanation for which, seemingly in contradiction to the thesis statement, there were no corollary effects observed in preference measurement given the performance-based procedure data. This chapter also described the relevance of this research to philosophical aesthetics and empirical approaches to visual aesthetics, which deal with perceptual processing. In particular, the data are assessed on the basis of their support for the ideas that visual aesthetics are influenced by perceptual fluency and/or natural scene statistics. Finally, some suggestions for future research are proposed.

6.1 Summary of Findings

Chapter 2 reported the results from four experiments on the effects of the golden section on RT or preference ratings. In Experiment 2.1 and 2.4, the data showed that while there was a general linear decrease in RT as a function of ratio (such that targets of more uneven ratios were easier to detect), the golden sectioned stimuli in the middle set produced an unusual slowing in RT outside of this general function. This effect was not reflected by results from paired comparison experiments using the same stimuli, instead showing some inconsistency in responding, which implies a lack of actual preferences. In Experiment 2.4 the possibility that the effect was due to the morphology of the visual field was examined, and the results were almost identical to
those of Experiment 2.1. Chapter 3 followed up on these results by positing a number of possible mechanisms to explain the results. Luminance contrast, fractal dimension, and spatial frequency were considered and of these, only spatial frequency appeared to affect RT performance in the way observed. These findings are discussed in more detail regarding their relevance to perceptual fluency and natural scene statistics in the next sections.

Chapter 4 reported the results from experiments on visual balance as described by Arnheim (1954) as a factor influencing perceptual organisation. Balance was operationalised by axial alignment, and in Experiment 4.1 off-axis targets tended to be slower to respond to when the targets were centrally presented. The restriction of this effect to the central portion of the screen suggested that some anisotropic effects were influencing responding. However, when this was investigated by using rating scores as positional weights there was no relationship between spatial anisotropy so defined and RT. Another alternative was proposed which suggested that inter-item structure was influencing responding, as this would have been more pronounced in the central area due to greater inter-item proximity at the region of axial convergence. There was shown to be a significant interaction between axial alignment and inter-dot distance for the largest set, which raised the possibility that there were two organisation processes influencing performance; a global one deriving from global structure which operated for the middle set, and a local one based on inter-item structure which affected RT in the largest set. In Experiment 4.2 this possibility was examined by altering the alignment of the distracter set and thereby dissociating axial structure and inter-item structure. In that case, the smallest set demonstrated a linear decrease in RT by location weighting when the target was on-axis and a similar effect was shown in the largest set when the target was off-axis (which was on-axis according to the
structure defined by the distracter set). These results supported the prediction by Arnheim that axial alignment would predict faster responding.

While Chapter 4 produced some results revealing visual balance to have effects on performance, the data from a preference comparison task showed that people simply preferred more populated displays, regardless of alignment of the target. Chapter 5 reported the results from two experiments which sought to assess the extent to which preference could be operationalised by orientation behaviour in non-humans. This allowed us to ascertain the extent to which domestic chicks might prefer to orient on the basis of the visual balance of the target. Our results showed some statistically significant tendencies for chicks to prefer to peck at images exhibiting a misaligned element in otherwise axially-aligned stimulus structure. The major differences in gross neural anatomy in humans and avians suggests that the results cannot be accounted for by specific neural architecture, and in particular the higher-level cognitive capabilities which are often suggested to underlie preferences. Instead, the results appear to suggest that attention is drawn to the holistic image structure which again argues for an influence of the specific perceptual organisation of balanced or imbalanced visual objects.

6.2 Implications of the Experimental Results

While the research presented herein supports the idea that aesthetic principles may affect perceptual performance, the thesis does not claim that a description of how something constitutes a perceptual stimulus is entirely sufficient to characterise its aesthetic relevance (e.g. Seeley, 2009). However, it does advance the idea that intrinsic features of a stimulus can provide aesthetic form if conducive situational factors influence experience. Palmer et al. (2013) suggest that the kinds of factors that can influence the aesthetic appreciation of form include the extremity of the aesthetic
response itself (seeing something very beautiful), the context (e.g. an aesthetic attitude), or the explicit awareness of the potential for such responses (being told to look for beauty in a stimulus in an experiment). The results detailed in this thesis may not have adequately employed these factors to produce results in the preference measurements that matched those of the performance tasks. This may have been the result of the artificiality of the laboratory scenario (the context), or it may be that the specification of the task requirement in the paired comparison task was too difficult to apply consistently (see Chapter 2). Finally, the stimuli themselves may not have been interesting enough to provide the level of aesthetic experience, which would allow judgements to be made in the lab scenario.

This last point may first appear damning to the purpose of this research; if the experimental manipulation of a supposed aesthetic variable did not produce aesthetic experience, then what is its relevance? However, the argument proposed by this treatment is that the critical metric for establishing the validity of visual balance or golden sectioning is not based on introspection or preference ratings. Instead, it is suggested that, following Fechner’s aesthetics-from-below, a basic level of an effect must be demonstrated before it can be determined how this effect interacts with other factors determining higher-level experience. While the principles under investigation in this thesis (golden ratio, visual balance) are supported by the results of performance based experiments, in practice it is difficult to achieve measurement of definite aesthetic preferences deriving from effects at early stages of visual processing. The RT results should therefore be considered in light of the normative status accorded to these principles, as well as the extensive literature on both. McMahon (1999) argues that the universality of perceptual form suggests that aesthetic experience can be predominantly perceptual, or perceptual and cognitive, but not predominantly
cognitive. The finding that the phenomena under investigation were demonstrated to have an influence on perceptual processing while not producing preference effects is thus largely supportive of the idea that all perception is potentially aesthetic.

As described in the introductory chapter, the way in which stimulus structure can influence aesthetic experience has been explained by the perceptual fluency approach and the related theories that refer to natural scene statistics. The present research is concerned with the instances where perceptual representations are affected by stimulus structure based on what might be considered general principles of aesthetics. As such, these principles were discussed in terms of their ease of processing and relationship with natural image statistics and the possible constraints on the visual system caused by ecological adaptation. Results do not provide clear support for the perceptual fluency account, but demonstrate some convincing links between adaptive constraints and the statistical properties of the natural environment.

6.3 Perceptual Fluency

The experiments reported here do not establish a clear link between positive aesthetic evaluation and perceptual processing as there were no significant trends in paired comparisons of the experimental stimuli. However, some aspects of the findings argue against a unidirectional positive influence of perceptual fluency on aesthetic experience. Across several experiments in Chapters 2 and 3, golden-sectioned stimuli were shown to result in slowed responding for the middle set size (16-area) in comparison with the general RT by ratio function, which itself was an otherwise linear relationship. If perceptual fluency theory applied in this instance, then it should be the case that a similar pattern of preferences would be produced, but the data did not support such a prediction. In Chapter 2 there were no systematic tendencies to favour one or another of the ratio-sectioned stimuli (and the fastest RTs did not
produce the most preferred stimulus category). It may have been the case that participant responses were poorly reflective of actual preferences, so the possibility was considered that the task orientation of the observers was being affected by the response criterion. In Chapter 3 this possibility was investigated by manipulating the types of instructions that participants received. In three of the four instruction groups the results were similar to the previous chapter in not demonstrating a pronounced preference for any of the stimuli. Interestingly, when participants were less constrained by a response criterion (they were asked to select a preferred stimulus for “any reason” they liked) there was a noticeable trend for the golden-sectioned stimuli in the middle group to be least preferred. This cannot be taken as strong support for perceptual fluency theory for three reasons: firstly, the ratings nadir was not significantly different to ratings for neighbouring stimuli. Secondly, the RTs obtained by the stimuli sectioned with a ratio close to unity were the slowest, but not the least preferred in this group and stimulus range, as would be expected by even the most conservative reading of perceptual fluency theory. Finally, the applicability of these results to perceptual fluency theory may be challenged by the normative status of the golden section in art and design. If the experiments reported here failed to find clear preference tendencies, we might turn to general compositional principles to compare against our RT data. In this case (and as reported in Chapter 2) the normative status of the golden section may be disputed, but it is seldom claimed that golden-sectioned stimuli have the least aesthetically pleasing dimensions or proportions. It therefore becomes difficult to interpret the RT data reported here according to perceptual fluency theory.

As described in the previous section, RT results in Experiments 4.2 and 4.3 seem to support Arnheim’s claim that axial alignment facilitates perceptual
organisation and can therefore provide a qualitative aspect to a scene (i.e. visual
balance). As in Experiments 3.2 and 3.3, there were no corollary findings in the
preference task using the same stimuli as Experiment 4.1, which again fails to offer
support for perceptual fluency theory, which in this case would predict that axially-
aligned stimuli would be preferred. Furthermore, axial misalignment appeared to
present avian subjects with an orientation preference, which would by unexpected if
the perceptual fluency account was to reconcile these results with human RTs. Taken
together, these findings do not offer support for perceptual fluency as a source of
positive evaluation, and this is not improved by shifting the goalposts and referring to
conceptual fluency. The idea behind conceptual fluency is that the fluent construction
of a conceptual representation may facilitate positive evaluations. This would predict
that balanced displays, being faster to process, are preferred, and this was not the case.
Furthermore, conceptual fluency is difficult to apply to non-representational stimuli
(e.g. Lanska, Olds, & Westerman, 2014). Both perceptual and conceptual fluency are
also subject to more general criticisms, such as the finding that people do not exhibit a
linear relationship between preference and visual complexity (Berlyne, 1971) and
familiarity with objects or scenes does not necessarily cause an increase in preference
(the “mere exposure effect”; Meskin, Phelan, Moore, & Kieran, 2013).

6.4 Natural Scene Statistics

The experiments regarding the golden section detailed in Chapters 2 and 3
offer some support for the claim that natural image statistics affect the processing of
golden sectioned stimuli relative to other ratio-sectioned displays. This was not found
to be due to visual field morphology or luminance contrast sensitivity, so the
statistical properties of the stimuli were analysed. Fractal dimension is one property
whereby iterative subsectioning based on the golden ratio would be differentiated
from other ratios (being more self-similar than others). However, while respondents did appear to exhibit some relationship between RT and fractal dimension, the golden-sectioned stimuli were not differentiated from this general function. Next the stimuli were analysed for spatial frequency via log Gabor filtering. This showed that golden-sectioned stimuli have a lacuna in the power spectrum at the 4-7c/degree range. RT results were shown to be unaffected by observer distance, so the power spectra of the images were made more similar (i.e. the lacuna in the golden-sectioned stimuli was removed) by introducing random noise to all displays. This succeeded in removing the previously observed discontinuity, producing a simple linear RT by ratio function. As the human contrast sensitivity function has its maximal sensitivity at 3-8c/degree range, the results imply that impeded processing efficiency may be due to a spatial frequency profile which is unusual in nature. A possible mechanism for the effect may also be due to the dynamic encoding of cell assemblies related to areal contrast. If subareas are being compared, a common framework is required, but Pletzer et al. (2011) have shown that inter-phasic synchronisation is most impaired when neural firing patterns are in golden-ratio relationships with each other. As such, the comparison of the target section (or indeed any of the interior sections) is more difficult when their relationship is characterised by this ratio, via the populations of neurons responsible for representing the spatial dimensions of the stimuli.

The experiments detailed in Chapters 4 and 5 demonstrated that visual balance, as implemented by alignment with axial structure, can influence performance in humans and orientation preference in avians. These facets are assumed to be different aspects of the same phenomenon, but are difficult to reconcile according to perceptual fluency theory because humans demonstrate slowed responding for imbalance, but chicks tend to prefer imbalanced stimuli. Interpreting the results
according to perceptual organisation allows the possibility that preferences are affected by adaptive constraints on visual processing and attentional deployment. In this view, delayed responding may lead the observer to expend more attentional resources on a stimulus in order to resolve potential ambiguities in the perceptual representation. Speculatively, this might lead to positive or negative affect based on contextual or situational relevance; expedited processing may be valuable when the ambiguity does not have ecological import, but there may be intrinsic affective reward if the delayed responding assists in deploying attention to something with ecological significance. The relationship between this and aesthetic experience remains somewhat conjectural, but one suggestion operates like an extension of Kantian aesthetics described in the introduction, whereby emphasis on this process may itself be a source of aesthetic pleasure. As described by McMahon (1999):

Certain relations in the object, in the course of being perceived, challenge or stretch the relevant perceptual principles in an unprecedented or non-typical way. On the other hand, the relation of the elements within some objects…might epitomise the perceptual principles. Perhaps when the principles are invoked in any way which is likely to draw our attention from straight-forward object recognition to the process of perception as a solution to a problem, then we are experiencing beauty (p. 16).

The idea that natural scene statistics may constrain perceptual processing, and that normal perception can therefore influence aesthetic experience, is in line with the claims made in this thesis, and the evidence presented in these chapters offers support for this proposal.

6.5 Future Research and Concluding Remarks

The preceding chapters have demonstrated that low-level perceptual effects may be ascribed to aesthetic principles, lending them the status of basic features that
determine a qualitative aspect to visual images, which might facilitate aesthetic experience. The results were not entirely compatible with fluency theory, and due to the absence of representational content for our stimuli are not convincingly supportive of conceptual fluency either. Instead, the data seem to lend themselves to the claim that natural scene statistics may affect performance, and that adaptive constraints on visual processing have determined that some image structures are naturally more compelling or open to elaboration. This is proposed to be a more satisfactory description of aesthetic interest than fluency theory, as it does not claim that difficult processing is necessarily poorly favoured.

Sammartino (2011; Sammartino & Palmer, 2012) have argued for the concept of “representational fit”, whereby the appropriateness of fluency to the semantic content or contextual setting of the image is a critical factor in whether or not the image is aesthetically interesting. Future research may extend the findings detailed here in order to make a link between these basic visual aesthetic principles and their interaction with meaningful stimuli. Furthermore, the implication of supporting the claim that all perception is potentially aesthetic is not limited to its relevance to how we discuss art. The wider implication is that aesthetic aspects of perception can and should be studied within basic perceptual research, as nothing in experience is without quality.
References


Bertamini, M., Byrne, C., & Bennett, K. M. (2013). Attractiveness is influenced by the relationship between postures of the viewer and the viewed person. *i-Perception, 4*(3), 170-179.


Cunningham, M.R., Roberts, A.R., Barbee, A.P., Druen, P.B., & Wu, C-H. (1995). “Their ideas of beauty are, on the whole, the same as ours”: Consistency and
variability in the cross-cultural perception of female physical attractiveness.


Kant, I. (1790). *The critique of judgement, part I: Critique of aesthetic judgement*.


Konečni, V. J. (2004). The “golden section” as aesthetic idea and empirical fact.

*Visual Arts Research, 2*(59), 75-86.


dissertation]. University of California, Berkeley. Retrieved from:


Appendix 1:

Paired Comparison Scaling
Introduction

The paired-comparison experiment represents an attempt to derive an interval scale for a psychological magnitude with no clear or obvious physical analogue. Some common applications are in taste testing, colour comparisons, and personnel ratings (David, 1969). The procedure involves participants making preference judgements across multiple paired-object displays. A set of stimuli is judged, usually involving the presentation of all possible pairs of stimuli types, and a respondent chooses one of each pair according to some choice criterion. If there are more than two stimuli, a matrix of all possible comparisons is used. The assumption is usually made that each comparison is independent – what Luce (1959) calls the “choice axiom” (or independence-from-irrelevant-alternatives assumption; IIA), which states that the probability of selecting one item over another from a set of items is not affected by the presence or absence of other items in the set (Luce, 1959). If A is preferred to B in the set \{A, B\}, and item X is added to the set - \{A, B, X\} – it is assumed that B is not now preferred to A (Krabbe, 2008). Some theoretical assumptions are made in order to apply a mathematical model, thereby allowing the construction of a scale of dimension scores.

This procedure is preferable to asking participants to assign the values by some subjective criterion, as in such a case the scale ratings would have poor validity (and it is likely that reliability would be sub-optimal also; Tsukida & Gupta, 2011). It is also difficult to rate prominent outliers by simply asking them to rate or scale items in isolation (Brown & Peterson, 2009). Titchener (1901) suggested that while we might not find it possible to consistently assign an absolute measurement to the amount of pleasantness or unpleasantness in an object, we can easily say which of two objects is the more pleasant. Nunally (1976) claims that this is due to the infrequency
of absolute judgements in everyday life. People are quick and consistent in applying relative preference, but can be inconsistent when it comes to assigning absolute magnitude (Brown & Peterson, 2009). The method of paired comparisons takes this asymmetry into account.

This method was first proposed by Fechner (see David, 1969), but given its most influential and detailed conceptual foundations by Thurstone (1927). In line with classical psychophysics, Thurstone (1927) makes the assumption that the proportion of “Stimulus A is greater than Stimulus B” judgements is based on differences in their respective sensations. If A is preferred to B 50% of the time, we can say that their psychological magnitudes are equal. As A>B (or vice versa) tends towards 100%, we can assume a larger difference in magnitude. Due to natural fluctuations in the nature of perceptual processing, the magnitude will not be the same every time an object pair is encountered (it may be assumed that the stochastic nature of the magnitude fluctuations also applies to the stimuli encountered singly). Thurstone argued that a spread of “discriminal processes” occurs, and that the distribution would be normal with the mean as the scale value for the stimulus, and the “discriminal dispersion”, or variance, being the other defining parameter. The assumption of normality makes computation tractable, and it is the same rationale that underlies the Gaussian assumption in Signal Detection Theory (Green & Swets, 1988). Thurstone’s model rests on the assumption that the quality of an option is a Gaussian random variable (see Figure A1.1).

When a respondent observes a stimulus, that stimulus has an assumed fixed value on a psychological continuum. When they respond to it, they are randomly sampling a value from the discriminal dispersion of the stimulus. When two objects are compared, there are two sampled values from two discriminal dispersions. A
choice between them is based on the difference in those sampled values. The more those dispersions overlap, the more the observed difference response is free to flip randomly between positive and negative. In any given trial, the item chosen is the one with the largest perceived value. The stochastic preference event is formalised by a value \( U \) which can be decomposed into deterministic and random (error) components:

\[
U_{ip} = V_{ip} + \varepsilon_{ip}
\]

Where \( V \) is the expected value, and \( \varepsilon \) is the stochastic element, or error term. Two separate items will have the value functions:

\[
U_{ip} = V_{ip} + \varepsilon_{ip} \\
U_{jp} = V_{jp} + \varepsilon_{jp}
\]

If the error distributions of these two values overlap, there may be some inconsistency in the momentary expression of the expected values \( V \) by their \( U \) values. The probability that item \( j \) will be preferred is:

\[
P(U_{jp} > U_{ip}) = P(V_{jp} + \varepsilon_{jp} > V_{ip} + \varepsilon_{ip})
\]

which becomes more likely as the difference between the expected values of \( j \) and \( i \) become larger, or as the error distributions overlap less. Because we are assuming that the values of the preference event are randomly drawn from the independent and normal distributions, the difference relations will also be normally distributed (Brown and Peterson, 2009). Thurstone (1927) suggested that error could come from the respondent and/or the researcher, but when the model is implemented error is usually attributed to the perceptual processing of the respondent. The law of comparative judgement is the mathematical model for a paired-comparison experiment. Before
describing the experimental implementation of the model in more detail, it is necessary to describe this principle.

**The Law of Comparative Judgement.**

The Law of Comparative Judgement (LCJ) is a probabilistic choice model for estimating scale values on the basis of modal values relating to binary choices between stimuli (Krabbe, 2008; McIver & Carmines, 1981). It is based on the following parameters (using the notation of Gescheider, 1997):

\[ \psi = \text{Discriminal process or sensory magnitude.} \]

\[ S_A, S_B = \text{Stimuli A and B, which result in:} \]

\[ \psi_A, \psi_B = \text{Mean sensory magnitudes of } S_A \text{ and } S_B. \]

The momentary difference between these varies from trial to trial.

\[ \sigma_A, \sigma_B = \text{Standard deviations of } S_A \text{ and } S_B. \]

The mean of \( \psi_B - \psi_A \) is \( \psi_B - \psi_A \), because the mean of the difference between the distributions is the difference of their means, as it represents random sampling from the two normally distributed functions (Brown & Peterson, 2009; Tsukida & Gupta, 2011).

The standard deviation, \( \sigma \), of the difference between \( \psi_A \) and \( \psi_B \) is

\[ \sigma_{\psi_A - \psi_B} = \sqrt{\sigma_{\psi_A}^2 + \sigma_{\psi_B}^2 - 2r_{\psi_A,\psi_B} \sigma_{\psi_A} \sigma_{\psi_B}} \quad [1] \]

Where:

\[ \sigma_{\psi_A}, \sigma_{\psi_B} = \text{Standard deviations of } \psi_A \text{ and } \psi_B \]

\[ r_{\psi_A,\psi_B} = \text{Correlation between momentary pairs of } \psi_A \text{ and } \psi_B \text{ values.} \]
Figure A1.1 probability distributions of $\psi_A$ and $\psi_B$

![Figure A1.1](image1.png)

$\psi_A$ and $\psi_B$

Figure A1.2 probability distribution of the difference between $\psi_A$ and $\psi_B$ as estimated by $\mu_A$ and $\mu_B$.

In Fig A1.2, the shaded area to the right of the zero point is the proportion of times that $\psi_B > \psi_A$, and that point can be converted to a z-score by the standard formula:

$$z = \frac{X - \mu}{\sigma}$$

Where $X$ is a particular score, $\mu$ is the set of scores, and $\sigma$ is the standard deviation of the set. Because we are getting the z-score of the zero point (i.e. the probability that $\psi_B > \psi_A = 0$), the equation becomes

$$z = 0 - \frac{(\psi_B - \psi_A)}{\sigma_{\psi_A - \psi_B}}$$  \[2\]

The zero point on the $\psi_B - \psi_A$ distribution corresponds to a z-score of $(-1 \frac{(\psi_B - \psi_A)}{\sigma_{\psi_A - \psi_B}})$
If we already possessed the scale values, we could therefore derive the probability of $S_A$ being judged greater than $S_B$. In actual paired-comparison experiments, those values are unknown, so instead get the empirical proportion of $S_A > S_B$, and derive the scale values from it. To achieve this, the above procedure is essentially reversed. First convert the proportion of $S_A > S_B$ to a $z$-score, then multiply it by $\sigma_{\psi_A} - \sigma_{\psi_B}$, which gives the equation:

$$\psi_B - \psi_A = Z_{BA} \sigma_{\psi_A} - \psi_B$$

Or given its fuller expression based on [1],

$$\psi_B - \psi_A = Z_{BA} \sqrt{(\sigma_{\psi_A}^2 + \sigma_{\psi_B}^2 - 2r_{\psi_A\psi_B} \sigma_{\psi_A} \sigma_{\psi_B})} \quad [3]$$

This is the law of comparative judgement.

**Implementation**

The variables under the radical in [3] cannot be determined through experiment, so some simplifying assumptions are made in order to make the derivation more tractable. Thurstone (1927) describes five cases of such assumptions, the most common and computationally straightforward of which is Case V, which assumes that:

- Because the scale value distributions are independent (as described in the preamble above), we can take it that the correlation between observed pairs is zero. The equation then becomes:

$$\psi_B - \psi_A = Z_{BA} \sqrt{\sigma_{\psi_A}^2 + \sigma_{\psi_B}^2} \quad [4]$$

- The discriminant dispersions are equal for each stimulus, so we can arbitrarily set the variance at 1, and the remaining terms under the radical become:

$$\psi_B - \psi_A = Z_{BA} \sqrt{1^2 + 1^2} = z_{BA} \sqrt{2}$$
In order to expand the model to accommodate multiple stimuli, construct an $m \times m$ matrix and using the $i$th row of the $j$th column to represent $D_{i,j}$ (the probability of preferring stimulus $i$ to $j$), average the values (the proportions of $i > j$) for each row to give an scale value for the $i$th stimulus when that stimulus serves as a standard for comparison. Then get $z$-scores for each of these proportions, and multiply them by the combined variance factor in equation [5] above which gives a scale value for the stimulus. In other words, the raw data for the model is a count matrix of the number of times one option was preferred to another in a paired-comparison experiment. Where every possible comparison has been made, the procedure is referred to as balanced.

For $t$ items, there are $t (t - 1) / 2$ pairs. Resulting scale values indicate positions along a dimension, and may be estimated for an individual or aggregated from a set of individuals. Respondents are not offered an indifference option, the rationale for which comes from the theory of stochastic preference, where the probability of true indifference is assumed to be negligible (Brown & Peterson, 2009). Although lacking this option might seem too forceful if we are aiming to assess real preferences, a tendency to indifference can be captured by the data in any case (see below).

**Alternative Approach: Bradley-Terry-Luce Model**

An alternative model for paired comparisons is the Bradley-Terry-Luce model (BTL; Bradley & Terry, 1952; Luce, 1959), which does not make the assumption that item qualities are in a Gaussian distribution. Instead, Gumbel random variables

$$= z_{BA} (\sim 1.41) \quad [5]^{13}$$

---

13 Another approach is to remove the variance coefficient by assuming an identical value of 0.5 for each discriminant dispersion, but this does not have the advantage of expressing the result in terms of the standard deviation of the quality or dimension under investigation (Tsukida & Gupta, 2011).
represent the individual qualities, and while in Thurstone’s model the difference of
two Gaussian random variables is a Gaussian random variable, in the BTL model the
difference in scale values is a logistic random variable (see Tsukida & Gupta, 2011,
for a proof). This allows calculation of the probability of A>B by use of the logistic
cumulative distribution function.

The model for this is:

\[ P(A>B) = \frac{\exp(\mu_A / s)}{\exp(\mu_A / s) + \exp(\mu_B / s)} \]  \[6\]

\[ = \frac{1}{2} + \frac{1}{2} \tanh((\mu_A - \mu_B) / 2s) \]  \[7\]

\( s \) is the scale parameter for the logistic distribution, affecting the variance. When the
scale parameter is set at \( \sqrt{3/\pi} \) the Thurstone and BTL models are roughly equal
(Tsukida & Gupta, 2011).

As in Thurstone’s model, the probability of A>B is estimated by the empirical
count data, so we can use the proportion \( C_{A,B}/C_{A,B} + C_{B,A} \) where \( C_{i,j} \) represents any
count matrix of \( i > j \) choices. Using this equation for A>B and inverting equation [7]
to get the inverse logistic CDF or logit gives:

\[ \mu_A - \mu_B = s(\ln(C_{A,B}/C_{A,B} + C_{B,A}) - \ln(1 - C_{A,B}/C_{A,B} + C_{B,A})) \]  \[8\]

The Bradley-Terry-Luce scale approximations have traditionally been used in
experimental aesthetics due to the comparative simplicity of calculating the logit
rather than the inverse Gaussian CDF. Although this is now trivial given the speed of
computers, due to the fact that the models produce similar results (Tsukida and Gupta,
2011), it is probably advisable to adopt the most common approach within recent
experimental aesthetics (e.g. McManus, 1980; Palmer, Gardner, & Wickens, 2008)\(^{14}\).

Furthermore, Handley (2001) found the Bradley-Terry-Luce model to outperform the
Thurstone model in the scaling of image quality assessments.

\(^{14}\) Some other implementations use the cauchit or probit model instead of the logit (Turner &
Firth, 2010).
Fitting one of the models requires an empirical proportion of preferences, and for score estimation of a two option matrix, Thurstone’s Law gives the maximum likelihood solution. When this extends to multiple options, an optimisation problem must be solved. However, because the calculation of the scale values requires the product of the count matrix and either the inverse Gaussian or the logit, at one or zero this produces $+\infty$ or $-\infty$. When these 0/1 cells occur, the unit normal deviates have to be approximated to avoid the production of infinite standard normal deviations. These approximations are necessarily arbitrary, so the best solution is to adopt maximum likelihood values (Tsukida & Gupta, 2011).

The BTL model assumes that preference scores are the result of a contest between players and the proportion of wins is a function of ability (Turner & Firth, 2010).

**Internal Reliability**

One aspect of paired comparison experiments requires special mention - In a simplified example of three stimuli, each of the three comparisons has two possible outcomes so there are eight possible experimental results, and these data might exhibit what are referred to as circular triads (Kendall and Babington-Smith, 1940). A circular triad is so-called because it implies a contradictory-seeming circularity, e.g. the set of results $\{A > B; B > C; C > A\}$ or $\{A < B; B < C; C < A\}$. This inconsistency on the part of the judge might imply guessing, or the function used to describe the comparison might itself be inconsistent from trial to trial. Either of these could be the empirical demonstration of the null hypothesis that there is no difference between the stimuli. The method of paired comparisons allows these inconsistencies to manifest themselves clearly in the data. The number of circular triads can be estimated from
the scores by Kendell and Babbington-Smith’s (1940) coefficient of internal reliability, \( \zeta \):

\[
\zeta = \frac{1 - C}{C_{\text{MAX}}} \quad [9]
\]

Where \( C \) is the number of circular triads observed, and \( C_{\text{MAX}} \) is the maximum possible number of circular triads. This measure requires the observation of all possible pairs, and it varies between 1 (no circular triads) to 0 (maximum). The number of circular triads for an individual respondent can be directly calculated from the preference scores:

\[
C = \frac{t}{24} (t^2 - 1) - \frac{1}{2} \sum (a_i - b)^2 \quad [10]
\]

Where \( t \) is the number of items in the set, \( a_i \) is the preference score for item \( i \), and \( b \) is the average preference score (so \((t - 1) / 2\)). The maximum possible number of circular triads is:

\[
C_{\text{MAX}} = \frac{t (t^2 - 1)}{24} \quad \text{for } t = \text{odd} \quad [11]
\]

\[
C_{\text{MAX}} = \frac{t (t^2 - 4)}{24} \quad \text{for } t = \text{even} \quad [12]
\]

This allows the internal reliability coefficient to be calculated. David (1969) proposed that the null hypothesis in paired comparison procedures can be given by random or inconsistent judging, and even if a more specific hypothesis is being investigated, the use of equation [9] is an important initial test of paired comparison data.
According to Kendall & Babington-Smith (1940; see also Alway, 1962; Knezek, Wallace, & Dunn-Rankin, 1998; Iida, 2009), circular triads are closely modelled by a chi-square distribution, using the following approximation:

$$
\chi^2 = \frac{8}{k-4} \left[ \frac{1}{3} \binom{k}{4} - d - \frac{1}{2} \right] + v
$$

Where:

$$\binom{k}{n} = \frac{k!}{n!(k-n)!}$$

$$v = \frac{k(k-1)(k-2)}{(k-4)^2}$$

And $k$ is the number of objects, $d$ is the number of observed triads, and $v$ represents the degrees of freedom.

**Appendix References**


Appendix 2:

Bradley-Terry-Luce Scale Scores

for Experiments 3.2 and 3.3
Figure A2.1 Bradley-Terry-Luce scale scores for each participant in Experiment 3.2.
Figure A2.2 Bradley-Terry-Luce scale scores for each participant in Experiment 3.3.
Appendix 3:

Participant Consent Form
Informed Consent

Study Title: Visual perception and aesthetic preference.

Experimenter(s): Paul Mulcahy
Supervisor(s): Dr. Mark A. Elliott

In order to participate in this research study, it is necessary that you give your informed consent. By signing this informed consent statement you are indicating that you understand the nature of the research study and your role in that research and that you agree to participate in the research. Please consider the following points before signing:

- I understand that I am participating in psychological research;

- I understand that my identity will not be linked with my data, and that all information I provide will remain confidential;

- I understand that I will be provided with an explanation of the research in which I participated and be given the name and telephone number of an individual to contact if I have questions about the research.
I understand that participation in research is not required, is voluntary, and that, after any individual research project has begun, I may refuse to participate further without penalty.

By signing this form I am stating that I am over 18 years of age, and that I understand the above information and consent to participate in this study being conducted at NUI Galway.

Signature: ____________________________ Date: ________________

First Name: __________________ Surname: ________________________

Are you enrolled in First Year Psychology? YES NO
(If yes, then research participation credits will be given for participation)

I have explained the above and answered all questions asked by the participant:

Researcher’s Signature: __________________________ Date: _____________
Appendix 4:

Information Sheet for Experiments 2.1 – 4.3
Title of Study: Perception and Visual Aesthetics

You are being invited to take part in this research study. Before you make a decision, it is important for you to understand why the research is being done, and what it will involve. The purpose of this Participant Information Sheet is to tell you about the aims, risks, and benefits of the research study. If you agree to take part, the researchers will request that you sign a consent form. If, at this or any point, there is some aspect of the research participation that you aren’t clear about, the person running the experiment will be happy to explain it to you. Please take as much time as you need to read it. You should only consent to participate in this research study when you feel that you understand what is being asked of you, and you feel that you’ve had enough time to consider fully your decision. Keep in mind that even after consenting to be a part of this research, you are free to withdraw your participation at any time, without risk of penalty.

Thank you for reading this information sheet.

* * *

This research aims to assess the influence of the way that the visual system might influence the kinds of images that people prefer. The experimenters will be using a simple visual task (where your job is to give a response to an image presented on a computer screen) to gauge how certain image properties might influence the ability to respond quickly and accurately in a response task, and how this might relate to preferences.

Do I have to take part?

It is up to you to decide if you want to take part or not. If you do decide to take part, you will be given this information sheet to keep and be asked to sign a consent form.

---

15 Items in square brackets represent deviations from this general template in specific experiments.
Participants from the first year psychology student body at NUI, Galway may redeem course credits for participation, as per School of Psychology guidelines. In all cases, if you decide to take part, you are still free to withdraw at any point, and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect your rights in any way.

**What will happen to me if I take part?**

If you are eligible to take part and fully consent to participation, you will be asked to sign the consent form. Please note that even at this stage you are free to withdraw your consent without penalty. The lab session, wherein the task is completed, will take about an hour.

**Who cannot take part?**

All adult participants can take part. If you wear glasses, please remember to put them on before the experiment begins.

**What must I do to comply with the requirements of the study?**

This particular study requires a reasonably high level of concentration, so during the tasks, please avail of rest periods between trial blocks.

**Who cannot take part?**

There are no exclusion criteria for adult (over 18) participants.

**What does the study consist of?**

[Experiments 2.1, 2.4 – 4.2]:

A task in which you are requested to locate a region of the visual display and make a key-button response to it (the experiment will be explained in more detail prior to beginning).

[Experiments 2.2 – 2.3, 4.3]:

A visual task in which you are requested to choose one of two simultaneously-presented images.

**How long will my part in the study last?**

While the task duration is based on the speed of your responses, so might vary, expect to be in the laboratory for approximately 1 hour (although it will likely be less than that).

**What are the possible benefits in taking part?**

Participation allows you to contribute towards research that will advance our understanding of the link between aesthetic preference and visual perception.

**What are the possible disadvantages in taking part?**

The tests and procedures involved in the study do not involve any risks to your health, and involve simple computer-based tasks.

**What happens at the end of the study?**

The measurements from the various tests will be combined with those from other participants and analysed using data analysis software. There will be no specific ‘results’ from your tests, so you will not be contacted in the future, although we can provide general information on our progress and research. The results of the tests are absolutely confidential, and will only be used to understand more about the area of interest.

**What if I have a complaint during my participation in the study?**

The research team will be available for you to contact if you have any complaints during your participation in the study, at the contact details given below.

**Will my results be confidential?**

All information that is collected about you during the course of the research will be kept strictly confidential, it will only be made available to the researchers involved in
this study and will not be shared with anyone else. Anonymised identity numbers will be used instead of names so there will be no way you may be identified. Any paper records identifying the participant with the study will be stored in a locked filing cabinet and will be destroyed after completion of the study. Future presentations of the findings (in the form of conference presentation, or publication in a scientific journal) will feature the data in aggregate form, or using anonymous individual identifiers (e.g. Participant 1, 2, 3, etc.).

Who do I contact for more information or if I have further concerns?

If you have any further questions or concerns about the study, please contact:

Paul Mulcahy (BA)
School of Psychology
National University of Ireland, Galway
University Road,
Galway.
Telephone: *********
E-mail: p.mulcahy1@nuigalway.ie

If you have any concerns about this study and wish to contact someone independent and in confidence, you may contact: The Chairperson of the NUI Galway Research Ethics Committee, c/o Office of the Vice-President for Research, NUI Galway, ethics@nuigalway.ie

I am interested in being contacted about the possibility of my participation in this study

Name: ____________________________

Mobile phone number: ____________________________

Home phone number: ____________________________

Time I would most like to be contacted: ____________________________

Age: ____________________________