UNIVERSITY OF CALIFORNIA,
IRVINE

Analysis of Preattentive Features Controlling Texture Discrimination

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Psychology

by

Kier Groulx

Dissertation Committee:
Professor Charles Chubb, Chair
Associate Professor Charles E. Wright
Professor Mike D'Zmura

2017

DEDICATION

To Irina, my light through this long journey.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This research was made possible through the support given by my advisors, colleagues, and friends at the University of California, Irvine.

Of course, special mention goes to my advisors, Charlie Chubb and Ted Wright. They have provided countless hours of mentorship, advice, feedback, and general guidance which has proved invaluable throughout this entire process.

Additionally, I would like to thank several members of the Cognitive Sciences department who have invested their time and energy into either contributing experimental feedback or participating in experiments themselves. This research would have not been possible without their contributions.

Experiment 1 of this dissertation was supported by funding from NIH Grant EY07977 provided by co-author Jonathan D. Victor.

CURRICULUM VITAE

Kier Groulx

EDUCATION

Doctor of Philosophy in Psychology — **December 2017**
University of California, Irvine — *Irvine, CA*

Master of Arts in Psychology — **June 2016**
University of California, Irvine — *Irvine, CA*

Bachelor of Science in Computer Science — **June 2014**
University of California, Irvine — *Irvine, CA*

Bachelor of Arts in Psychology — **June 2014**
University of California, Irvine — *Irvine, CA*

RESEARCH EXPERIENCE

Graduate Research Assistant — **2014–2017**
University of California, Irvine — *Irvine, CA*

Undergraduate Research Assistant — **2011–2014**
University of California, Irvine — *Irvine, CA*

WORK EXPERIENCE

Data Scientist — **2017–current**
Blizzard Entertainment — *Irvine, CA*

Teaching Associate — **2009–2010**
University of California, Irvine — *Irvine, CA*

Graduate Teaching Assistant — **2009–2010**
University of California, Irvine — *Irvine, CA*

Senior Lead Facilitator — **2013–2017**
University of California, Irvine — *Irvine, CA*

HONORS AND AWARDS

NSF IGERT Traineeship in Complex Scene Perception	**2014**
Phi Beta Kappa Honors Society	**2014**
Psi Chi International Honors Society	**2014**
UC Irvine Order of Merit	**2014**
UC Irvine Chancellor's Award of Distinction	**2014**
Summer Undergraduate Research Program Fellow	**2013**
Summer Undergraduate Research Fellowships in Information Technology	**2011-12**

PUBLICATIONS

Matchin, W, **K Groulx** & G Hickok, "Audiovisual Speech Integration Does Not Rely on the Motor System: Evidence from Articulatory Suppression, the McGurk Effect, and fMRI", Journal of Cognitive Neuroscience 26(3):606-620, 2014.

Herrera C, P Sun, **K Groulx**, C Wright, C Chubb & G Sperling, "How do the S-, M- and L-cones contribute to motion luminance assessed using minimum motion?", Journal of Vision 13(9):1021-1021, 2013.

CONFERENCE PRESENTATIONS

Groulx, K, T Tran & CE Wright, Poster at the annual meeting of the Psychonomics Society, Vancouver, British Columbia, Canada, November 10, 2017. *Attentional Selectivity Increases for Visual Stimulus Sets that Combine Two Dimensions of Variation.*

Groulx, K & C Chubb, Poster at the annual meeting of the Psychonomics Society, Boston, Massachusetts, November 18, 2016. *Analyzing preattentive visual mechanisms via the superthreshold contrast paradigm.*

Groulx, K, C Chubb, J Victor & M Conte, Poster at the annual meeting of the Vision Sciences Society, St. Pete Beach, Florida, May 17, 2016. *Using the texture-centroid method to analyze the mechanisms sensitive to higher-order image statistics.*

Groulx, K, C Chubb, CE Wright, Talk at the UC Irvine Undergraduate Research Symposium, Irvine, California, May 17, 2014. *Individual Cone Contributions to Motion Luminance.*

Groulx, K & M Warschauer, Talk at the Summer Undergraduate Research Program Symposium, Irvine, California, August 30, 2012. *Cloud-Based Writing in K-12 Schools.*

Groulx, K, AJ Shaka & E Fadeev, Talk at the Summer Undergraduate Research Program Symposium, Irvine, California, August 26, 2011. *A Web Application for Science: An NMR Pulse Sequence Database.*

ABSTRACT OF THE DISSERTATION

Analysis of Preattentive Features Controlling Texture Discrimination

By

Kier Groulx

Doctor of Philosophy in Psychology

University of California, Irvine, 2017

Professor Charles Chubb, Chair

Preattentive processing is an essential step of filtering information from the world into readibly parsable chunks that demand attentional resources. In a visual domain, this skill enables a person to easily devote resources to items that are captured by one or more attention filters. This phenomenon can be historically seen through differences in accuracy and reaction time amongst visual search tasks. However, identifying specifically what filters exist in human vision, how strong these filters are, how many filters there are, and what stimuli maximally activate each of these filters are questions that have largely been unanswered. This dissertation explores several experiments that have developed frameworks designed to begin answering these questions. Chapter 1 uses the centroid paradigm - a psychophysical method aimed at discovering the visual influence of stimulus items through weighted center-of-mass calculations - to explore what features drive discrimination amongst textures with equal energy spectra. We discover that large homogeneous regions, rather than high-order texture statistics, drive texture salience in this domain. Chapter 2 seeks to further understand the mental processes that underlie comparative salience computations for visual texture displays. We do so by proposing a new paradigm that compares two visual textures to a background uniform texture and to each other, which enables us to quantify the specific strength and tuning functions for mechanisms whose existence had been strongly suggested by existing literature. Chapter 3 again uses the centroid method to examine several atomic

micropatterns consisting of various lines and line terminators and asks whether the human visual system contains mechanisms that are selective towards a subset of micropatterns. We find that certain micropatterns exhibit strong salience and "popout" effects, while others are difficult to attend to under similar conditions, suggesting that there exists a small number of mechanisms sensitive to these atomic patterns. Overall, these experiments provide a proof-of-concept as a study of how preattentive processing affects a subset of textures and atomic patterns. The groundwork is then laid to explore many more texture types and feature dimensions (such as color) in the future.

Chapter 1

Introduction

Imagine you are shopping online for various types of food. While you are browsing the images of food presented to you, certain foods may stand out based on what you are looking for. For example, if you are looking for tomatoes, other items that share features with tomatoes may stand out in your vision - the red of the bell peppers, or the close color and circular feature of the oranges. Most items that do not conform closely to your template of tomatoes will be filtered out of your attention. However, it may be difficult to ignore the presence of a brightly flashing advertisement at the top of the page, even though it does not conform to your tomato template. All of the items that you are seeing are being preattentively processed by your visual system so that you can then use feature-based attention to find your tomatoes.

Preattentive processing is useful because it guides your attention and directly contributes to the phenomenon of 'salience'. As you accumulate information about the world through your visual input, objects in your visual field are vying for your attentional resources. How you distribute your attention across these objects determines how salient objects are to you, which can lead to the perception of some objects segmenting (or 'popping out') from the background in a highly noticeable fashion.

However, it is true that even when factoring in attention effects, some objects are naturally more salient than others (such as the advertisement on the website). What may be causing this discrepancy in salience amongst objects? One historically useful method for probing these effects is that of "visual search" (Treisman and Gelade, 1980). In this method, observers are shown a group of items and must state whether any item present in the group contains a specific target feature. For example, observers may be shown a group of letters for a brief duration, then asked whether they observed the letter "O" anywhere in the group that was shown. This task is useful in that it can help us identify broad categories (or "tuning functions") for various visual mechanisms. For instance, we may find support that we have a visual mechanism to detect circular objects if it is easy for us to identify the letter "O" among groups of "T" letters, but difficult for us to identify the letter "O" among groups of "Q" letters. In addition, this task reveals interesting asymmetries as well; it is easy for us to find a "Q" among "O" letters, but difficult for us to find an "O" among "Q" letters. This suggests that the presence of a feature contributes more to our notion of salience than the absence of the same feature.

While visual search offers us a window into understanding the tuning functions and strengths of visual mechanisms, it only affords us a rough understanding of what these values are. The tuning functions we obtain inform us of general categories that we may be sensitive to, but they do not tell us what specific stimuli maximally activate each of these mechanisms. For example, if we have a visual mechanism sensitive to enclosed regions, it is difficult to tell whether this mechanism would be maximally sensitive to "O" objects, "Q" objects, θ objects, elliptical objects, or some other type of enclosed region. Additionally, we may be able to compare mechanism strengths to each other on an ordinate scale, but it is difficult to quantify the strength of mechanisms.

Other methods have emerged that attempt to address this issue. In particular, the "centroid method" proposed by Drew et al. (2010) presents an attention task that captures more in-

formation about the influence of different object types on an observer's judgments. Through this task, observers are presented with a cloud of items and must click on the center of mass of a subset of the presented items; for instance, if shown a number of colored dots, observers may be asked to click on the center-of-mass of only the red dots presented. This computation gives us much more information per trial: not only can we see how well you can compute a center-of-mass for a given feature that we are testing, but we can also see how much other features distract you from computing an ideal centroid. This information affords us an ability to generate a more quantitative analysis of the strength and tuning functions of mechanisms, as we can examine what features are easily confusable and how strong the degree of confusability is.

Chapters 2 and 4 of this dissertation use the centroid method as a useful probe into preattentive visual processes. Chapter 2 uses this method to address a previously unsolvable issue: the issue of what features present in texture displays drive selective attention. We tackle this question by taking two different textures that are empirically distinguishable (they have different fourth-order image statistics), but share many similar traits (such as expected mean luminance and variance). We can then create a continuum of stimulus items by taking discrete steps between these two textures. We can then use the centroid method on this set of items to see what features present in these textures drive our centroid computations. Chapter 4 takes a different approach and applies the centroid method to atomic feature elements, or "micropatterns". Past literature (Julesz, 1981) has posited the existence of "textons" which local image feature ensembles that function as atomic units of preattentive texture perception. We can construct a number of micropatterns that align with these textons and test their relative influence and interference with the centroid method.

Chapter 3 uses a separate paradigm to address the issue of salience within the texture domain. Previous work with texture salience has only looked at the detectability of textures whether you can identify that a texture is present or not. As a result, much of this work

has been limited by noise present in the computations (Poirson and Wandell, 1990). Here we use a two-alternative forced choice paradigm that asks observers to judge which of two textures stands out more at a superthreshold level from a background filled with visual noise. Examining these textures above threshold levels allows us to gain a more complete picture of how salience is determined in the texture domain.

Chapter 2

The features that control discrimination of an isodipole texture pair

2.1 Abstract

Visual features, such as edges and corners, are carried by high-order statistics. Previous analysis of discrimination of "isodipole" textures, which isolate specific high-order statistics, demonstrates visual sensitivity to these statistics in isolation but stops short of analyzing the underlying computations. Here we use a new "texture centroid" paradigm to probe these computations. We focus on two canonical isodipole textures, the "even" and "odd" textures: any 2×2 subblock of "even" ("odd") texture contains an even (odd) number of black (and white) checks. Each stimulus comprised a spatially random array of black-and-white texture-disks (background = mean gray) that varied in their fourth-order statistics. In the Even condition, disks varied along the continuum between random "coinflip" texture and

pure (highly structured) even target texture. In the Odd condition, the disks varied between coinflip and odd target texture. The task was to mouse-click the centroid of the disk array, weighting each disk location by the target structure level of the disk-texture (ranging from 0 for coinflip to 1 for even or odd). A linear model was used to estimate the weight exerted on the observers responses by the different types of 3×3 subblock patterns that occurred in the textures. In each task, a small set of specific subblock patterns dominated performance while other patterns with equivalent target structure were much less influential. These findings show that visual mechanisms sensitive to four-point correlations do not compute "evenness" or "oddness" *per se*, but rather are activated selectively by certain features whose frequency varies across isodipole textures.

2.2 Introduction

Extraction of basic visual features depends not only on differences in luminance at two points (spatial contrast), but also on sensitivity to patterns of luminance at three or more points, also known as "higher-order correlations" (Julesz, 1981; Morrone et al., 1982; Oppenheim and Lim, 1981). Paradigms that utilize isodipole textures constitute a principled approach for probing this sensitivity, since isodipole textures isolate higher-order correlations: by definition, they are matched in terms of their two-point correlations but differ in terms of three- or four-point correlations. The goal of the present study is to determine the nature of the computations that underlie visual performance on tasks involving these textures. We focus on the "even" and "odd" family of isodipole textures (Julesz et al., 1978), as these are a well-studied class of isodipole textures whose structure is defined by a four-point correlation among 2×2 blocks of checks, and whose salient visual features include edges and corners. Examples of both of these types of isodipole textures are shown in Fig. 2.1.

Despite extensive study (Victor and Conte, 1989, 1991; Victor et al., 2005; Victor and Conte,

2012; Victor et al., 2015), the neural computations that underlie visual performance driven by the even vs. odd distinction is as yet unknown. One possibility is that neural circuits extract the four-point correlation that defines these textures. This requires a multiplication of luminance values at four points. However, while primary and secondary visual cortices appear to be the locus of the relevant computations (Victor, 1986; Purpura et al., 1994; Yu et al., 2015), four-point multiplication does not readily map to known properties of cortical neurons. A possible resolution of this discrepancy is that visual performance does not rest on extraction of a four-point correlation *per se*. Other computations could serve as well, provided that they extracted statistical features implied by this correlation such as features that extend over larger regions of space.

To probe the underlying computations at a functional level, we use an extension of the "centroid paradigm" (Drew et al., 2010; Sun et al., 2015), a technique recently developed to investigate the attention filters that observers can achieve for extracting the spatial distribution of various sorts of image statistics. As we show, applying the centroid paradigm to stimuli consisting of patches of even and odd isodipole textures enables a characterization of the filter that the observer uses for selective attention. This characterization will indicate whether the visual system indeed computes the fourth-order correlation that defines the even vs. odd gamut, or rather, that the sensitivity to this distinction is best described in another way.

In a typical application of the centroid paradigm, stimuli are briefly flashed, spatially random scatterings of different sorts of items (e.g., dots of different colors or Gabor patches of different spatial frequencies and/or orientations). The is then tested in different attention conditions, always with the same sorts of displays. In a given attention condition, the observer is instructed to mouse-click the centroid of the items in each display, giving weight to different types of items in accordance with a specified target filter. The data from a given attention condition can be analyzed (using linear regression) to derive the "attention filter" achieved

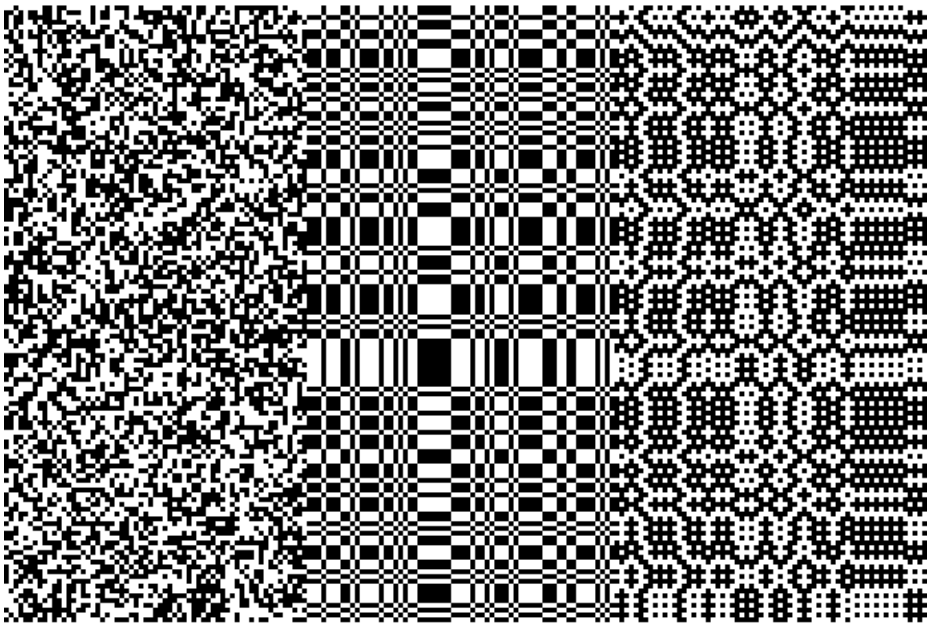

Figure 2.1: An example of isodipole textures. The left panel contains a 'coinflip' texture in which every check is independently determined. The middle panel contains an 'even' texture in which every 2x2 block has an even number of white and black elements, whereas the right panel contains an 'odd' texture in which every 2x2 block has an odd number of each element type (Julesz et al., 1978).

by the observer in that condition, i.e., the function that gives the relative weight exerted on the observer's centroid estimates by all of the different types of items in the stimulus displays.

In the present extension of the centroid paradigm, which we call the "texture centroid paradigm", the stimuli consist of disks that are filled with visual texture, as shown in Fig. 2.2. Each disk in a given display contains a texture whose distribution lies somewhere between a target distribution T and distractor distribution D. The task is to mouse-click the centroid of the display, weighing the location of each texture-disk in proportion to its signal strength, which is given by the distance of its distribution along the continuum from D to T. In order to perform well in a given condition of the texture centroid task, the observer must recruit the neurons in his or her visual system to produce an "attention filter" whose activation by different texture disks matches their signal strengths as closely as possible. The aim of the

current study is to analyze the attention filters deployed by observers in the context of the texture centroid task to differentiate (1) even texture (serving as T) from coinflip texture (serving as D) and (2) odd texture (serving as T) from coinflip texture (serving as D).

The disks used in our displays vary in a texture parameter, called α (Victor et al., 2005; Victor and Conte, 2012). The "even" and "odd" textures are the opposite ends of a gamut parameterized by α, where $\alpha = 1$ corresponds to "even" and $\alpha =$ -1 corresponds to "odd". For any texture disk d in any of the stimuli used in either of these two task conditions, we define

$$\alpha(d) = 2\,(p(d) - 0.5)\,, \tag{2.1}$$

where $p(d)$ is the proportion of abutting 2×2 blocks of checks in d in which the number of blacks (or whites) is even. Note that the extreme values of α, -1 and 1, correspond to probabilities p of 0 and 1, respectively.

In one task condition (the "Even" condition), the target weight of a texture disk d (i.e., the weight the observer strives to give d's location in the centroid computation) is equal to $\alpha(d)$, and in the other condition (the "Odd" condition), d's target weight is $-\alpha(d)$. Thus, in each task condition, the only image features that are explicitly relevant to performance are the 16 different possible patterns of 2×2 blocks of checks that can occur. In the Even condition, the ideal attention filter would give equal weight to all 2×2 block patterns with an even number of black checks and weight 0 to all patterns with an odd number of black checks, and in the Odd condition, the ideal attention filter would give equal weight to all patterns with an odd number of black checks and weight 0 to all patterns with an even number of black checks.

As we shall see, in each condition, the attention filters achieved by our observers deviate systematically from the ideal filter. These deviations provide important insights into the

features in these textures to which human vision is actually sensitive.

2.3 Methods

Five observers (two male) participated in this study at the University of California, Irvine. The observers participated under a UCI-approved IRB protocol with written consent.

Each stimulus consisted of a bounding box subtending 14 $\times$ 14 degrees of visual angle populated with 14 disks each of which subtended 1.06 degrees of visual angle. Each texture disk had an 11-check diameter and comprised 97 checks, each colored either black or white. When we assert that a disk d has $\alpha(d) = s$, this means that in the disk d, the proportion of complete, 2×2 abutting blocks of checks that contained an even number of black checks was as close as possible to $\frac{s+1}{2}$ (the proportion required to make $\alpha(d) = s$).

In the Even condition, two disks d in each display had $\alpha(d) = s$ for $s = 0, \frac{1}{6}, \cdots, 1$, and the observer strove (with trial-by-trial feedback) to mouse-click the centroid of the disks, giving weight to the location of each disk d proportional to $\alpha(d)$ (Eq. 2.1). In the Odd condition, two disks d in each display had $\alpha(d) = s$ for $s = 0, -\frac{1}{6}, \cdots, -1$, and the observer strove to mouse-click the centroid of the disks, giving weight to the location of each disk d proportional to $-\alpha(d)$. A sample of texture disks used in the Even task condition is shown in the top row of Fig. 2.3, and a sample of texture disks used in the Odd condition is shown in the bottom row.

Fig. 2.2 shows a trial in the Even condition. (Not shown is the "bounding box", a thin, black line that circumscribed the square region of the monitor screen in which the stimulus was displayed. The bounding box remained visible throughout the trial.) The observer first viewed (for 1 sec.) a blank (filled with mean gray) field circumscribed by the bounding box. The stimulus was then presented for 300 ms. Then there appeared a blank field with a

cursor in the middle; the observer used the mouse to move the cursor to click on the response location. A feedback display was then presented, which included (1) all the disks present in the stimulus presentation, (2) the observer's response location shown in blue, and (3) the location of the target centroid shown by a green bullseye. The observer could view the feedback for as long as desired. Pressing "Enter" on the keyboard initiated the next trial.

2.3.1 Design

Each observer performed 15 blocks of 100 trials, 1500 trials in all. In blocks 8 through 15, 500 of the last 750 trials used stimuli that were identical to stimuli that had been presented in blocks 1 through 7. This "double-pass" procedure allows us to derive a model-free estimate of the variance of the random noise corrupting responses. Specifically, this estimate is derived by taking the mean squared difference

$$\hat{\sigma}^2 = \frac{1}{4N_{doublepass}} \sum_{k=1}^{N_{doublepass}} \left[(X_{k,1} - X_{k,2})^2 + (Y_{k,1} - Y_{k,2})^2 \right] \tag{2.2}$$

where $N_{doublepass} = 500$, $X_{k,j}$ and $Y_{k,j}$, $k = 1, 2, \cdots, N_{doublepass}$, $j = 1, 2$ are the x- and y-coordinates of the observer's response on the j^{th} presentation of the k^{th} stimulus to be repeated twice. This formula assumes that the horizontal and vertical components of the random noise corrupting responses are independent and have equal variance σ^2. The division by $4N_{doublepass}$ comes from the fact that each term of the sum has expectation equal to $4\sigma^2$.

2.4 Results

Fig. 2.4 plots the weights exerted by disks d with different values of $\alpha(d)$ on the responses of all observers in both the Even and Odd conditions. While the observed weights were approx-

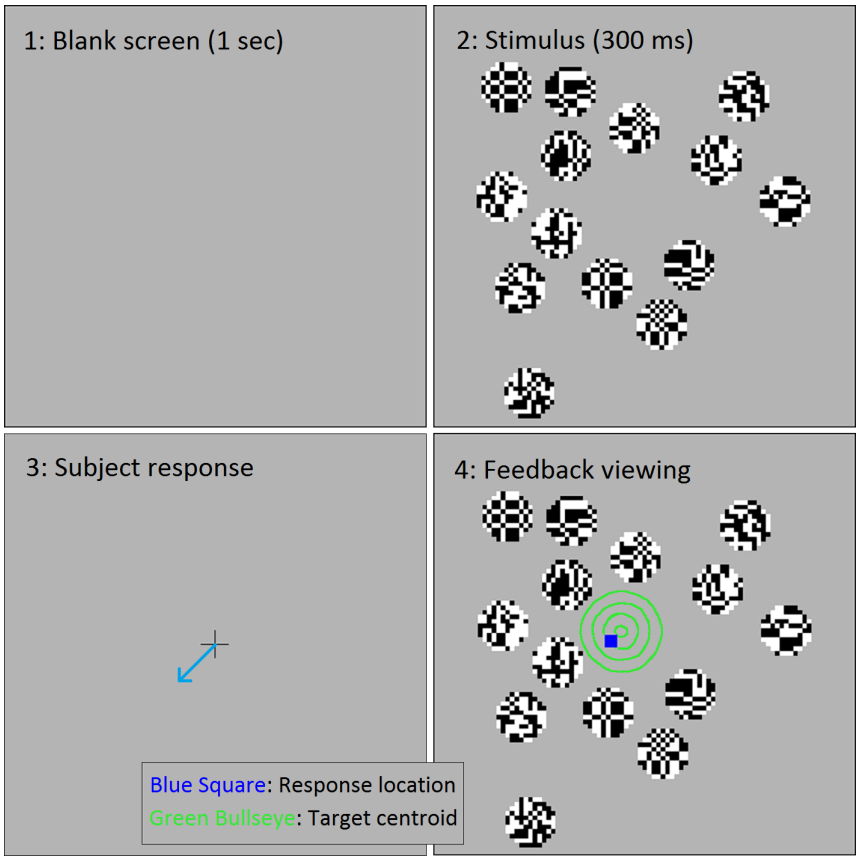

Figure 2.2: The sequence of display images which occur during a trial of the centroid task. After an observer indicates that he or she is ready, the top left panel displays a mean gray field on the screen for one second. Afterwards, the top right stimulus panel displays a number of objects on the screen. The stimulus then disappears and a crosshair guided by the observer's mouse movements appears, allowing him or her to indicate the perceived centroid under the current target condition. Finally, the bottom right panel contains all objects from the original display, the observer's response, and a feedback bullseye centered on the correct response.

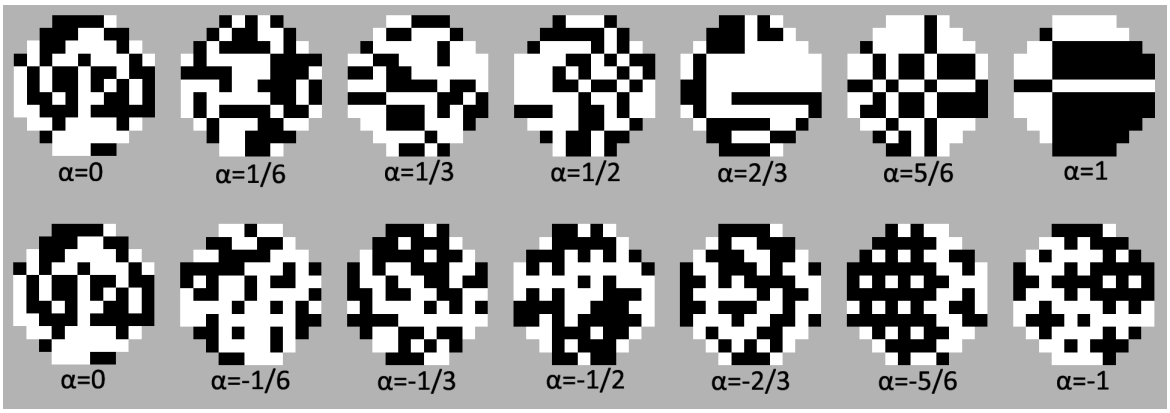

Figure 2.3: *Stimulus texture disks.* The top (bottom) row gives a sample of the seven types of texture disks occurring in a given stimulus (which comprises two disks of each type) in the Even (Odd) condition.

imately proportional to $|\alpha|$ (dashed lines in Fig. 2.4), there were substantial and systematic departures: in all cases, disks with larger values of $|\alpha|$ were given a disproprotionately larger weight. This held both for the even (left column) and odd (right column) textures. To quantify this observation, we compared a linear fit for the weights to a fit that included a quadratic term. F-statistics, shown in Table 2.1, demonstrated that the quadratic term was significant ($p < 0.05$) in four of five observers for the even texture, and one of five observers for the odd texture. Thus, although observers are able to achieve an attention filter selective for even vs. coinflip texture and odd vs. coinflip texture, the filters they are using do not seem to be selective for $\alpha(d)$ *per se*.

The heightened sensitivity of the filters achieved by our observers to disks d with $|\alpha(d)| = 1$ suggests that the filters achieved by our observers are selectively activated by features that emerge with high probability predominantly for texture patches when $|\alpha|$ nears 1. For example, large homogeneous white (or black) patches are relatively common in pure even texture but rare in texture patches d with $\alpha(d) \leq \frac{5}{6}$. It seems likely that such patches may produce strong activation in the attention filters achieved by our observers.

Next, we attempt to determine what these features are. We assume that the centroid esti-

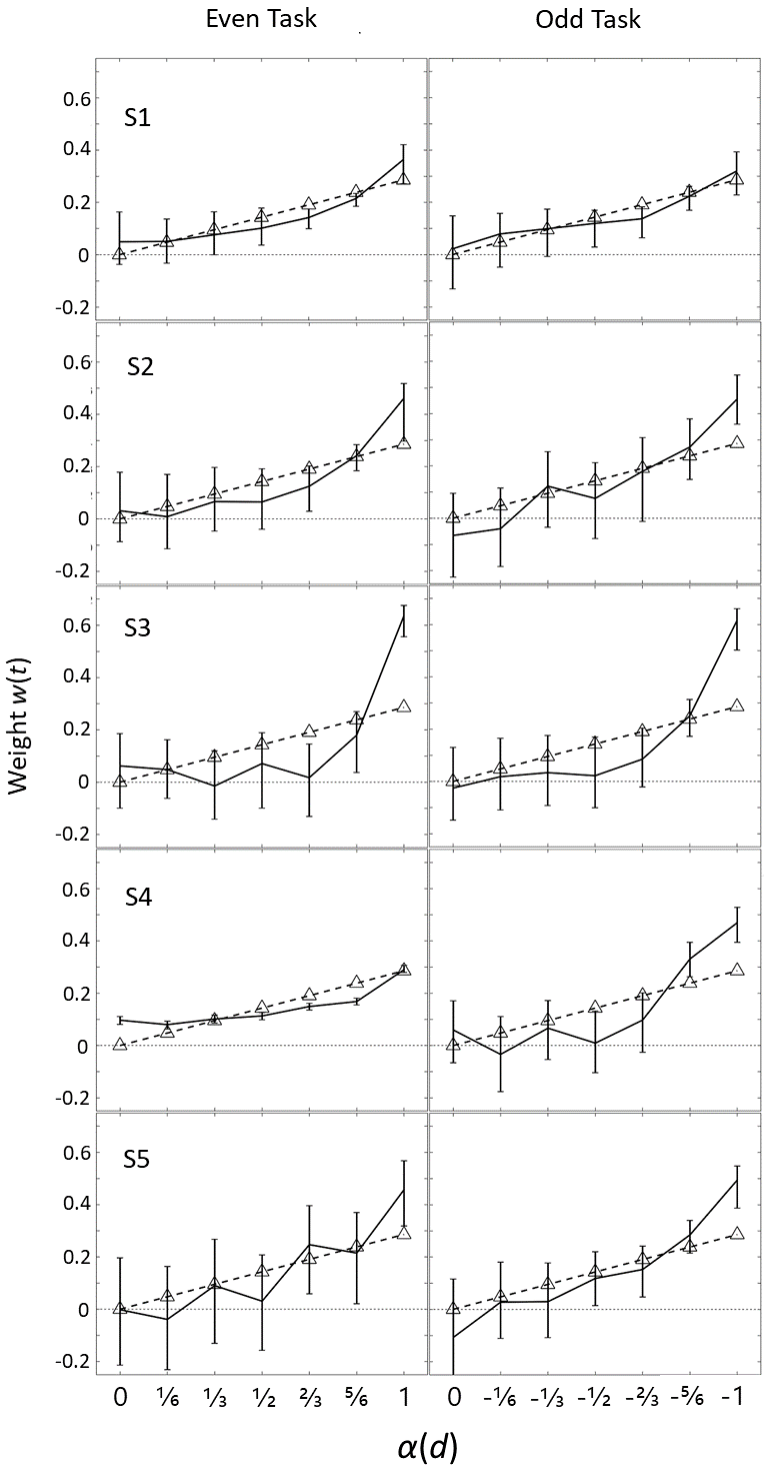

Figure 2.4: *Relative weights of disks d with different values of $\alpha(d)$ for all observers in the Even and Odd conditions.* The left panels show results for the Even task, while the right panels show results for the Odd task. The dashed line in each panel indicates the target weights that observers strove to give disks d with different values of $\alpha(d)$. These seven weights are normalized so that the sum of weights is equal to 1, as are the weights actually exerted on the centroid judgments (black line). Notably, in both tasks, disks d with $|\alpha(d)| \leq \frac{5}{6}$ were consistently underweighted, whereas disks with $|\alpha(d)| = 1$ were overweighted. Error bars give 95% Bayesian credible intervals.

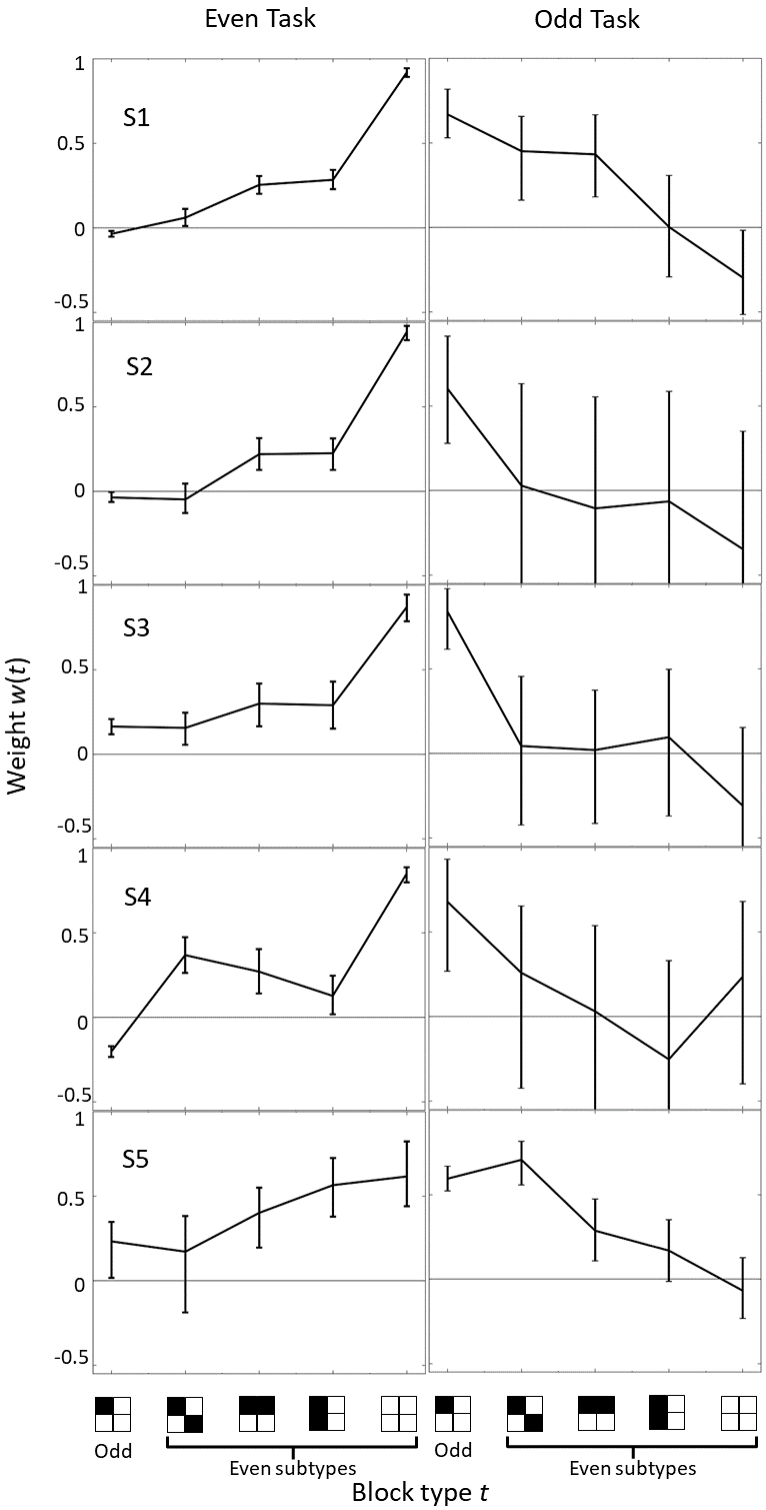

Figure 2.5: *Results from all observers in the* 2×2 *block analysis.* The left panels show performance in the Even task, while the right panels show performance in the Odd task. In the Even task, an ideal filter would give the odd block type a weight of 0, while the even block types would have equivalent weights of 0.25 each. In the Odd task, an ideal filter would give the odd block type a weight of 1 and the even block types a weight of 0. The overrepresentation of the uniformly even block type in the Even task and the variance in weights for the Odd task show that observers are not tuned to α *per se*.

mates of our observers are driven by local image features present in various degrees in different disks. Because even and odd textures have identical statistics for regions of 3 checks or less, the simplest possibility is that the features that support the task are contained within 2×2 blocks. Our first step is to ask: can performance be explained in terms of features confined to 2×2 blocks of checks, but with a weighting that differs from the weighting that would compute α?

The 2×2 *block types.* We assume that the only 2×2 blocks that contribute to the response are those that are entirely contained within a disk, since the blocks with one or more gray checks (from the background) would be equally likely to occur in any disk. There are 16 different 2×2 patterns of this type. We partition these 16 configurations into five equivalence classes (block types) based on the shapes that they contain. The five equivalence classes contain the block patterns shown on the horizontal axis of Fig. 2.5. In addition, each equivalence class is closed under polarity (black/white) inversion, vertical reflection, and horizontal reflection. Specifically, Type 1 includes all eight 2×2 patterns that contain an odd number of black (and white) checks. Type 2 includes the two 2×2 checkerboard patterns in which the top-left and bottom-right checks have polarity opposite the top-right and bottom-left checks. Type 3 includes the two 2×2 patterns in which the top-left and top-right checks have polarity

	Even				Odd			
Subject	$F(1,4)$	p	PV_{Quad}	PV_{Lin}	$F(1,4)$	p	PV_{Quad}	PV_{Lin}
S1	37.6452	0.0036	0.9833	0.8262	4.8453	0.0925	0.9523	0.8946
S2	26.7716	0.0066	0.9696	0.7662	2.2238	0.2102	0.9390	0.9051
S3	13.1455	0.0222	0.8759	0.4682	2.5428	0.1860	0.6575	0.4398
S4	14.9533	0.0180	0.9461	0.7446	21.1995	0.0100	0.9505	0.6879
S5	3.1804	0.1491	0.8741	0.7740	3.9483	0.1179	0.9570	0.9146

Table 2.1: *Results from F tests analyzing observer weight's deviations from linearity.* All observers' weights were analyzed in both the Even and Odd attending conditions to see if a quadratic model would fit significantly better than a linear model. PV_{Quad} represents the proportion of variance explainable by the quadratic model fit and PV_{Lin} represents the proportion of variance explainable by the linear model fit. A quadratic model has a significantly better fit for four out of the five observers in the Even task, but a linear model is sufficient to explain observer behavior for four out of the five observers in the Odd task.

opposite the bottom-left and bottom-right checks. Type 4 includes the two 2×2 patterns in which the top-left and bottom-left checks have polarity opposite the top-right and bottom-right checks. Type 5 includes the two 2×2 patterns in which all checks have the same contrast polarity. All 2×2 block patterns in a given equivalence class are assumed to exert equal weight in determining the responses of the observer.

We proceed to model the results by assuming that

1. observers deploy a "neural filter" to derive a spatial activation map from the stimulus;
2. this neural filter assigns to each location (x,y) that centers a 2×2 block of checks an activation that depends on the type of block;
3. the observer's response is the centroid of the filter activation-map produced by the stimulus, plus noise.

More specifically: let $w : \{1, 2, 3, 4, 5\} \to \mathcal{R}$ give the relative activation produced in the neural filter deployed by the observer by patterns in equivalence classes $1, 2, \cdots, 5$. In the context of the model, w is defined only up to an arbitrary multiplicative constant. We therefore constrain w to sum to 1. For a given 2×2 block b let $\tau(b)$ be the type of b, let $x(b)$ and $y(b)$ be the x- and y-coordinates of the center of b, and let N_x and N_y be independent normally distributed random variables with some standard deviation σ. We assume that the x- and y-coordinates of the observer's response on a given trial are given by

$$R_x = \frac{1}{\text{Total weight}} \sum_{\text{blocks } b} x(b)w(\tau(b)) + N_x \quad \text{and} \quad R_y = \frac{1}{\text{Total weight}} \sum_{\text{blocks } b} y(b)w(\tau(b)) + N_y \tag{2.3}$$

where

$$\text{Total weight} = \sum_{\text{blocks } b} w(\tau(b)). \tag{2.4}$$

The relative weights $w(t)$ exerted by the five block types t are shown in Fig. 2.5 for all observers. Note that if the neural filter used by the observer were actually tuned to Evenness, then $w(t)$ should be equal for all of the four block types that include an even number of black (and white) checks and 0 for the "odd" block type. The results differ dramatically from this pattern. All observers gave higher weight to "solid" blocks than to blocks containing a (horizontal or vertical) bar pattern, and much less weight to 2×2 checkerboard patterns.

Next, we ask whether the unequal (and non-veridical) weightings assigned to each kind of 2×2 block accounts for the observers' behavior. To do this, we compare the estimate of σ (the standard deviation of the random variables N_x and N_y in Equation 3) derived from our model fit with the model-free estimate derived from the double-pass data in Equation 2. These findings are shown in Table 3.1. The column labeled "Even 2×2 $\hat{\sigma}$" gives the estimate of the response-noise standard deviation for all observers derived under the assumption that the data were produced by the model of Eq. 2.3 where the sums range over all 2×2 blocks, and the different types of blocks are those indicated on the horizontal axis of Fig. 2.5. The model-free estimates of response-noise standard deviation derived from the double-pass procedure (Eq. 2.2) are given in the column labeled "Even 2-pass $\hat{\sigma}$." For every observer, the 2-pass estimate of error lies significantly below the 95% credible intervals for error in the model-based estimate of the response. That is, the internal consistency of the observers' responses is greater than predicted by a model based on the features in 2×2 patches. This, in turn, implies that the features controlling the observer's responses are not limited to 2×2 subblocks, or equivalently that the weights from different subblocks combine in a non-additive fashion to produce an observer's response.

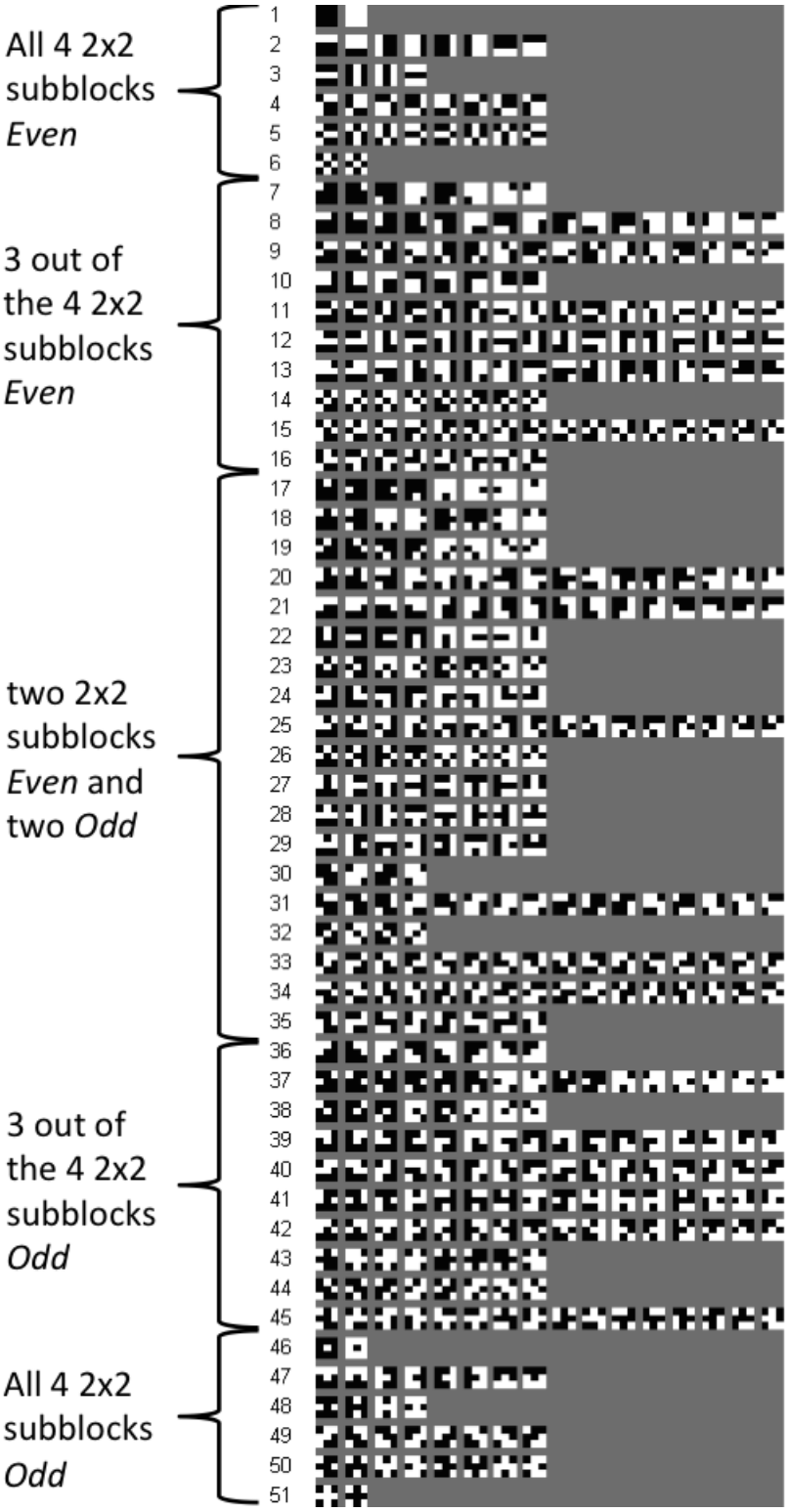

Figure 2.6: All 51 equivalence classes computed from all 512 possible 3x3 subblocks of black and white texels. Each equivalence class was closed with respect to vertical and diagonal flips, polarity inversion, and rotations. Subblocks 1-6 have $\alpha = 1$, subblocks 7-16 have $\alpha = 0.5$, subblocks 17-35 have $\alpha = 0$; subblocks 36-45 display $\alpha =$ -0.5, and subblocks 46-51 have $\alpha =$ -1.

We therefore next ask: can we improve the fit of the model by enlarging our subblocks to 3×3 checks? Fig. 2.6 displays 51 equivalence classes of different 3×3 subblocks. Every equivalence class is closed under contrast inversion, horizontal and vertical reflections, and 90-degree rotations.

We again fit the model of Eq. 2.3 to the data. This time, however, each of the sums ranges over all 3×3 subblocks b, and the type $\tau(b)$ of a given block b is the equivalence class of b, as shown in Fig. 2.6. The results for the Even condition are summarized for observers S1 and S2 in the top two panels of Fig. 2.7. Plotted values reflect the weights exerted by different subblock types on the observers' responses. The filled black data points show weights for which the 95% Bayesian credible intervals do not include 0. Subblock types were sorted on the x-axis in order of descending α; the light gray dots plot 0.4 times the α value for the given subblock type. Note that the largest weight is exerted by homogeneous 3×3 subblocks (i.e., subblocks that either have all white checks or all black checks).

For the Even task, Table 3.1 shows that the model of Eq. 2.3 with 3×3 subblocks accounts for more variance that the model with 2×2 subblocks. This is confirmed by an F test ($F_{9,55} = 5.66, p < 0.001$). Note, however, that although the model based on 3×3 subblocks results in an improved fit, there is still unexplained variance in the observers' behavior (as assessed by the double-pass analysis), indicating that configurations in even larger blocks may play a role.

2.4.1 Specific subblock results

As shown in Fig. 2.7, most of the 51 subblock classes do not significantly contribute to the observers' centroid computations in either task. However, in both the Even and Odd tasks, a small number of subblocks heavily dominated the weight functions of the computed centroid for all observers. In the Even task, 3x3 subblocks whose 9 checks were all one polarity

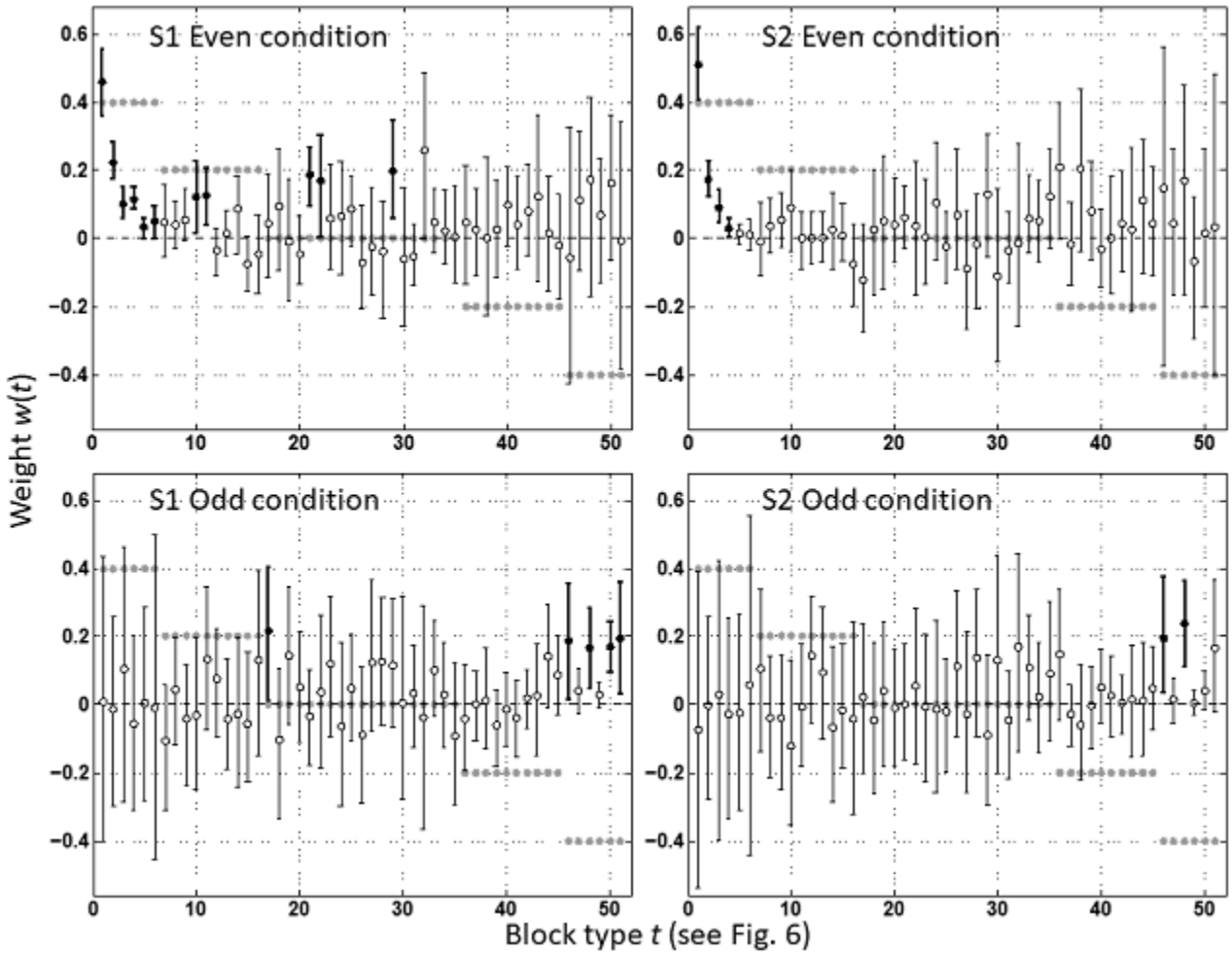

Figure 2.7: *Results of the 3×3 block analysis for observers S1 and S2.* Plotted values reflect the weights exerted by different subblock types on the observers' responses. Error bars give 95% Bayesian credible intervals. Data points whose error bars do not include 0 are filled in black. Points on the horizontal axis correspond to the equivalence classes of 3×3 subblocks shown in Fig. 2.6. In both panels, subblocks were sorted on the x-axis in order of descending α values; to make it easy to see the changes in α, the light gray dots show 0.4α. In the Even task, observers gave the highest weight to 3×3 subblocks whose checks were either all white or all black.

(subblock 1 in Fig. 2.6) exerted the highest weight A_{solid}. Subblocks comprising a single 3-check bar and 6 checks of opposite polarity (subblocks 2 and 3) exerted a weight roughly equal to $A_{solid}/2$. All other subblocks exerted significantly less weight. In the Odd task, a set of subblocks that emphasized anti-homogeneity all significantly contributed to the centroid computation. In particular, subblocks 46, 48, 50, and 51 exerted the strongest weight on observers' responses in the Odd task.

2.5 Discussion

Our analysis shows that although humans are readily able to detect differences in local regions that depend on four-point correlations (as typified by discrimination of "even" and "odd" textures), they do not appear to compute "even-ness" or "odd-ness" *per se*. Instead, human vision appears to be sensitive to certain specific features whose frequency varies across isodipole textures. Moreover, although the statistical structure of the even texture is determined by its fourth-order correlations, observers use a computation that involves more than four checks to compute a region's "evenness". For example, in the even discrimination task, the solid 3×3 subblock (subblock 1) exerted a weight over twice as high as any other

	Even			Odd		
Subject	2×2 $\hat{\sigma}$	3×3 $\hat{\sigma}$	2-pass $\hat{\sigma}$	2×2 $\hat{\sigma}$	3×3 $\hat{\sigma}$	2-pass $\hat{\sigma}$
S1	9.62 ± 0.24	9.31 ± 0.23	8.26	12.84 ± 0.37	12.68 ± 0.32	12.01
S2	16.14 ± 0.40	15.72 ± 0.39	13.79	10.11 ± 0.27	9.99 ± 0.26	9.18
S3	14.90 ± 0.40	14.57 ± 0.32	13.01	18.13 ± 0.44	17.44 ± 0.43	14.74
S4	21.96 ± 0.55	21.39 ± 0.54	20.19	13.44 ± 0.33	12.96 ± 0.32	11.95
S5	9.79 ± 0.24	9.75 ± 0.26	9.35	8.25 ± 0.21	7.98 ± 0.20	7.46

Table 2.2: *Results from the two-pass analysis.* Columns 2 and 5 show the 2×2 subblock model's measure of error with 95% credible intervals, columns 3 and 6 show the 3×3 subblock model's measure of error with 95% credible intervals, and columns 4 and 7 show the model-free estimate of explainable variance taken by comparing the responses of identical trials against each other in the even and odd discrimination tasks, respectively. The 3x3 subblock model captures nearly all of the explainable variance, which can be seen by comparing 3×3 $\hat{\sigma}$ to 2-pass $\hat{\sigma}$ in both tasks.

subblock, while subblocks that differed by only one check (subblocks 7, 17, and 46) exerted negligible weight in the centroid computation. Interestingly, using an entirely different strategy, Victor and Conte previously estimated that the relevant patch size needed to explain human visual sensitivity to α is a 2×4 check region (Victor and Conte, 1989).

The current results suggest that human discrimination of these isodipole textures is primarily driven by variation in local texture homogeneity, rather than variation in α *per se*. In the Even task, the most homogeneous 3×3 blocks exert the greatest weight, and the least homogeneous 3×3 blocks exert the least weight. In the Odd task, subblocks with isolated checks of a different polarity than their surrounding checks exert the strongest weight. While our models have examined the effects of homogeneous 2×2 or 3×3 regions on observers' centroid computations, there may be subblocks of other sizes that may explain more of the unresolved variance that is still present in our model.

Finally, we note that the texture centroid paradigm has enabled us to investigate questions that have not been answerable until now. The texture centroid paradigm can easily be used to explore the features in other types of textures that drive human discrimination, such as other varieties of isodipole texture, colored textures, or natural textures (Portilla and Simoncelli, 2000).

2.6 Acknowledgments

This research was supported by NIH Grant EY07977 to Jonathan D. Victor.

Chapter 3

Superthreshold contrast matching of achromatic texture patterns

3.1 Abstract

Previous work aimed at identifying texture-sensitive visual mechanisms has pointed towards the existence of multiple visual mechanisms responsible for texture salience (Chubb et al., 2004; Silva and Chubb, 2014). However, these studies have not determined (1) what these mechanisms are specifically tuned for (e.g., what stimuli produce maximal activation for each mechanism) and (2) the relative strengths or combination rules of these mechanisms. In addition, previous studies to investigate visual mechanisms at threshold level have generated contours dominated by the effects of internal noise (Hay and Chesters, 1977). We sidestep this problem by measuring superthreshold equisalience contours in a two-alternative forced choice paradigm. Textures were grayscale scrambles whose element histograms differed in mean and variance. On each trial the observer judged (without feedback) which of two texture patches differed with greater salience from uniformly distributed grayscale scramble.

The resulting contours implicate three mechanisms: a light-only mechanism activated only by positive Weber contrasts, an analogous dark-only mechanism, and a gray-tuned mechanism selective for elements of low absolute Weber contrast. These findings are supported by previous research and suggest this paradigm can may be useful for uncovering other visual mechanisms.

3.2 Introduction

When a person visually surveys their environment, they are able to segment and classify unique regions of the visual field. This phenomenon of segregation is driven by the fact that differing patterns of light will trigger differing activation among visual mechanisms that we use to process information about the world. While this visual process is understood when considering object discrimination, what is responsible for discriminating texture, which can be interpreted as a densely clustered ensemble of atomic elements? Studies have sought to probe preattentive mechanisms responsible for this sensitivity by measuring visual salience at threshold levels of detection; however, measures of equisalience at a threshold level offer distorted insight into visual mechanisms because they are (1) governed primarily by internal noise and (2) not reflective of salience judgments computed in everyday non-threshold environments (Georgeson and Sullivan, 1975; Hay and Chesters, 1977; Poirson and Wandell, 1990).

In addition, the question of salience remains relatively unanswered from a modeling perspective. While we know that visual mechanisms governing texture discrimination must differentially activate in order to produce a change in perception across two regions of texture, it is unknown which mechanisms produce the greatest change in salience for visual texture, what the relative strength of each of these mechanisms are, how these mechanisms combine their activation strengths, or even whether the relative strength of these mechanisms

is constant or subject to variance driven by top-down attention effects.

To investigate the unique properties that arise in displays dominated by densely packed visual texture, we focus on the set of textures constructed as 'texture scrambles', which are defined as a random permutation of contiguous squares which conform to a specified luminance histogram H (Chubb et al., 1994, 2004). Texture scrambles with N elements can be formed from the luminance histogram H by taking the probability distribution associated with H, drawing N elements without replacement such that they conform as closely as possible to H (i.e. populating a 800 element patch from a uniform distribution of 8 bins would require 100 elements from each bin of the histogram), and populating the patch randomly with the N elements. H here is limited to a small number of grayscale values, but the selection method allows it to be expanded to include a large range of values, such as HSV color coordinates. We can use these carefully controlled texture scrambles to make inferences as to how human visual texture perception may operate on textures that we encounter in daily life. Visual texture sensitive mechanisms that we possess will likely only be maximally activated by other stimuli, such as natural images; however, because the statistical properties of texture scrambles are easy to manipulate, they provide a useful baseline tool for probing human visual sensitivity.

One paradigm, first developed by Chubb and Silva, uses these texture scrambles to explore how mechanisms operate by identifying which regions of texture naturally segregate from each other. These visual segregation experiments involved placing a target region of scrambled texture T_{Targ} inside a background region of a different scrambled texture T_{Back} (Chubb et al., 2004; Silva and Chubb, 2014). T_{Targ} consisted of a perturbation ρ added to a uniform histogram of grayscale values U, whereas T_{Back} consisted of the same perturbation ρ subtracted from a uniform histogram of grayscale values U. Examples of the textures T_{Targ} used are shown in Figure 3.1. Several perturbations were used to differentiate the two textures from each other, including changes to their expected mean and variance values using the first

through fourth Legendre polynomials. The task presented to observers was to identify which of eight regions T_{Targ} was located in. For each perturbation in which observers were able to perform this task with a high degree of accuracy (that is, significantly above the chance percentage of 12.5% correct), it can be assumed that observers possessed a mechanism (or combination of mechanisms) that was sensitive to that perturbation. While results identified that observers possessed channels sensitive to the average mean luminance value and variance differences across textures, they also revealed that observers possessed a mechanism selectively sensitive to the darkest black located in the stimulus, dubbed the 'blackshot mechanism'. This finding was also of interest due to the asymmetry of the blackshot mechanism; while mechanisms existed to identify deviations both above and below mean luminance, there was no mechanism sensitive to the lightest light used in the stimulus to mirror the blackshot mechanism found.

How can we use texture scrambles to further identify, measure, and understand the visual channels of texture detection? The discrimination of objects in our visual field relies solely on the differential activation of visual channels. If we present stimuli which hold known channels constant (such as holding the average luminance of a spatial region constant), and we are still able to discriminate and segment sections of the presented stimuli, then there must be additional channels working to facilitate this discrimination. Indeed, if two sections of stimuli activated all human visual processing channels in the same fashion, then the two sections should produce the same percept and thus be indistinguishable. However, we can manipulate the image statistics of these texture metamers in an easily discriminable superthreshold environment to identify (1) each visual channel which confers a percept of salience to an observer as well as (2) the strength of each channel by the amount of change of an image statistic needed to produce a noticeable difference in salience.

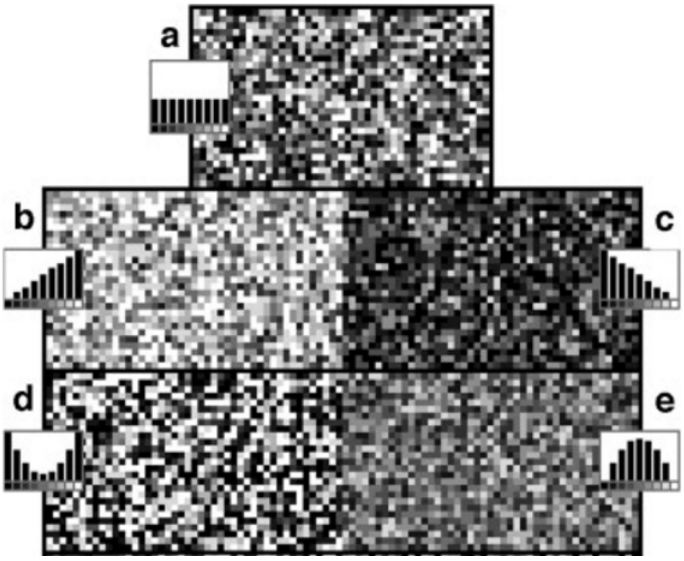

Figure 3.1: Five sample texture patches, each accompanied by its histogram of texture elements. Region a is the background uniform texture U. Regions b and c are formed by combining U with the linear perturbation -λ_1 or λ_1, respectively. Regions d and e are formed by combining U with the quadratic perturbation -λ_2 or λ_2, respectively. Each region here is shown with a maximal perturbation amplitude A of 1 to better identify the perturbation's features; in the experiment, A varied between 0 and 1. From Silva and Chubb 2014.

3.3 Methods

Six observers (four male) participated in this study at the University of California, Irvine. All observers provided signed consent forms under IRB approved protocol 1993-93.

A square display was shown to observers that subtended 15 $\times$ 15 degrees of visual angle and was primarily filled with a uniform grayscale texture scramble U. Two circular texture scramble patches, each subtending 3.5 degrees of visual angle, were presented with centers that laid 3 degrees of visual angle to the left and right of the center of the display. Each of these patches randomly contained the grayscale histogram $U + A_{std}\rho_{std}$ or $U + A_{var}\rho_{var}$, where ρ_{std} and ρ_{var} correspond to perturbations generated by combinations of the Legendre polynomials λ_1 and λ_2 which steer texture histograms away from the uniform texture distribution U. Examples of the pure Legendre perturbations at their maximal values as well

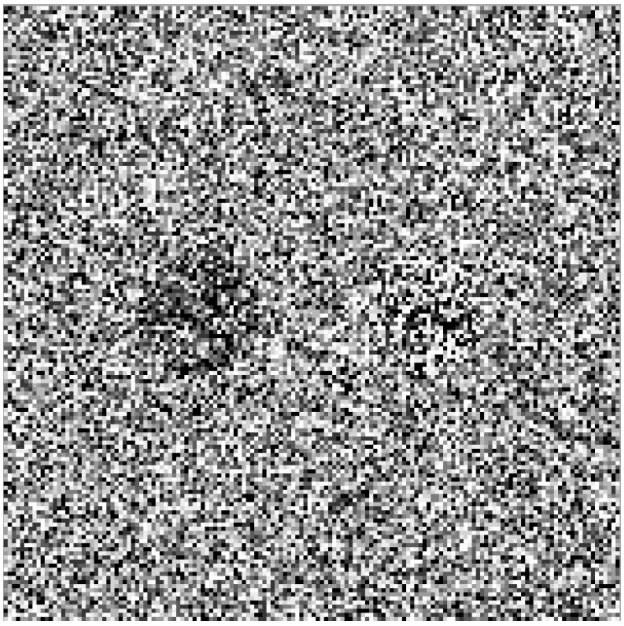

Figure 3.2: An example stimulus image from the superthreshold paradigm. Two patches of texture lie just left and right of the center of the image. The left patch contains texture that has a lower mean luminance than the uniform background, while the right patch contains texture that a larger amount of variance than the uniform background. Observers were tasked to report which patch was more salient with respect to the background. The target texture's perturbation strength was then adjusted via a psychophysical staircase until equisalience was achieved, and this entire process was conducted across 16 different texture perturbations.

as the background texture U are shown in Figure 3.1. $U + A_{std}\rho_{std}$, referred to here as the 'fixed texture', remained constant and salient throughout the trials (that is, A_{std} was fixed at a value of 0.6, which means that the fixed texture was pushed to 60% of its maximal strength, and ρ_{std} was fixed as λ_2, the variance modulating texture, as shown in panel d of Figure 3.1), whereas the perturbation amplitude A_{var} varied across 16 different 1-up/1-down adaptive staircases associated with 16 different values of ρ_{var}. A_{var} could vary from 0, where the perturbation was not represented at all (i.e. the uniform background), to 1, where the perturbation was fully represented, and the staircase was adjusted by 0.01 each step. An example stimulus from this procedure is shown in Figure 3.2. After each trial, the value of A_{var} associated with the perturbation ρ_{var} used in the trial was adjusted by its adaptive staircase to control the strength of the variable texture's histogram with respect to the background. If an observer indicated that the variable texture was perceived as more salient than the fixed texture, A_{var} was decreased to make the variable texture's histogram $(U + A_{var}\rho_{var})$ more similar to the uniform texture, whereas if an observer indicated that the fixed texture was perceived as being more salient, A_{var} was increased to make the variable texture's histogram less similar to the uniform texture. This process was used to estimate the perturbation amplitude A_{var} for which the variable texture appeared equal in salience to the fixed texture; this value of A_{var} is called the "equisalience amplitude" for ρ_{var}.

This procedure is run across 20 blocks, where each block consisted of 10 trials from each of the 16 variable perturbations, totaling to 160 trials per block or 3200 trials overall. Of note is that trials were presented in a mixed format and not sorted into blocks based on running all trials associated with a given perturbation at once; this is so that observers did not acclimate to a given variable texture and so had to constantly reevaluate their salience criteria during the experiment.

3.4 Results

Data collected from six observers are shown in Figure 3.3. Aggregate performance data across all observers is shown as the black contour in the figure. Across every observer's data, each perturbation's adaptive staircase was fit to a psychometric function (the Weibull function) and then plotted as a point on the equisalience contour. Each point on an observer's result contour reflects the average strength of a variable perturbation needed for an observer to reliably differentiate the perturbation from both the fixed perturbation and the background noise; thus, a lower score (i.e. a point closer to the origin) indicates that a stronger sensitivity demonstrated by a given observer to the texture histogram found at that polar angle, as the texture can be presented at a lower strength but still be perceived as equal in salience to other textures that are shown at a higher strength.

Results from this paradigm suggest that all observers' result contours are close in shape to an ellipse, with perturbations dominated by changes in variance across grayscale values being easier to detect than perturbations dominated by changes in mean luminance across grayscale values. However, across observers, details are present that differentiate our result contours from ellipses, which demonstrates that these contours are not primarily governed by noise and that certain visual mechanisms are not balanced in sensitivity across the feature space we have analyzed. The aggregate result contour shape, while resembling an ellipse, is not symmetric with regards to either the x- or y-axis. This demonstrates that, while observers possess mechanisms that are sensitive to both differences in mean luminance and variance of grayscales across texture patches, their sensitivity is not symmetric or balanced. In particular, observers seem to unanimously be more sensitive to perturbations in which the mean luminance is shifted to a darker value rather than a lighter value. An even stronger asymmetry is shown as observers seem to unanimously be more sensitive to a decrease in variance across grayscale values rather than an increase in variance.

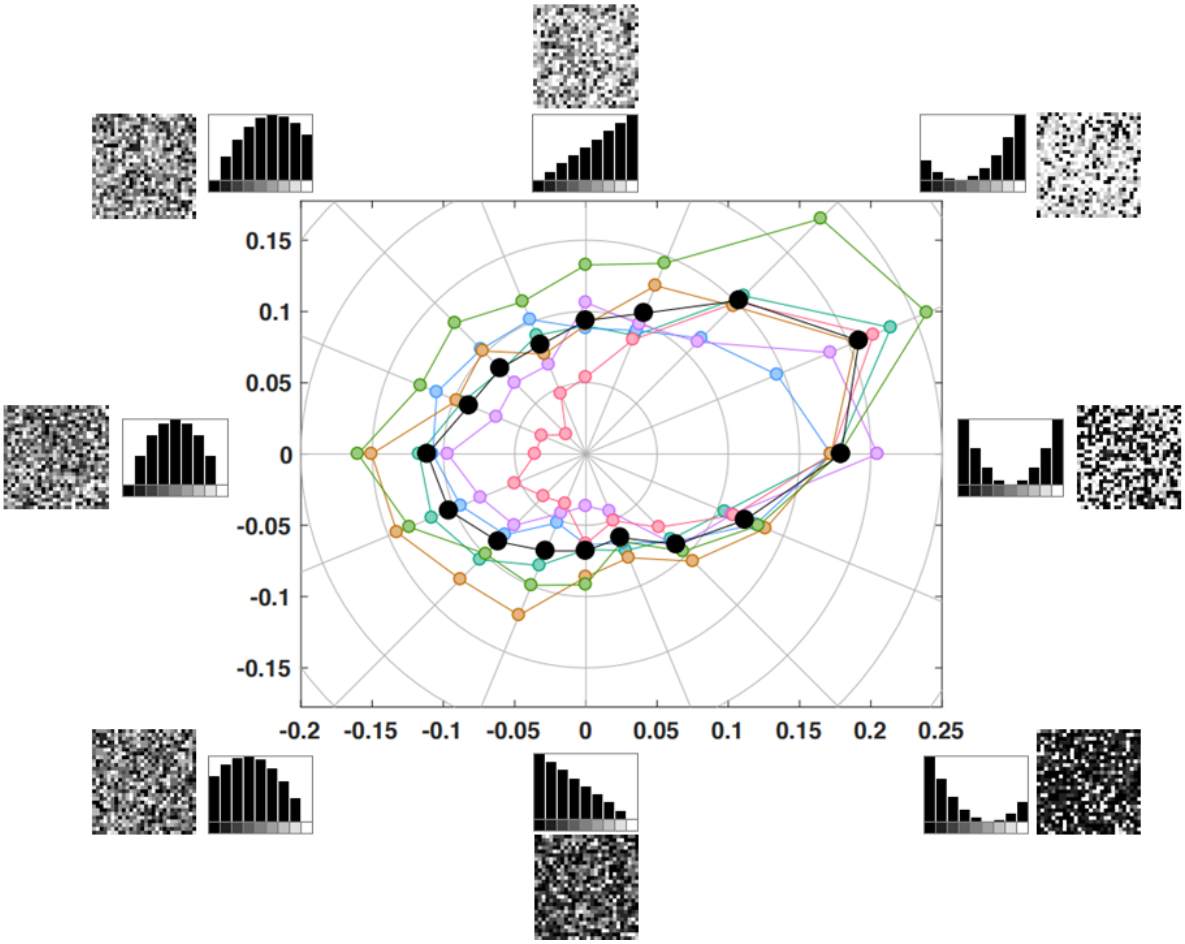

Figure 3.3: Results from the superthreshold paradigm across six observers. The angle of the polar graph corresponds to the perturbation away from a uniform texture, as indicated by the histograms of texel elements on the outside of the graph, while the distance from origin values correspond to each perturbation's strength. Examples of perturbations pushed to their maximal values are shown as inserts next to the histograms. Aggregate sensitivities across all observers are shown as a contour with filled black points.

A model which achieves a good fit for our collected data using a minimal number of parameters follows a Minkowski distance metric. For any perturbation represented in polar space

$$\rho_{A,\theta} = A\left(\cos(\theta)\lambda_2 + \sin(\theta)\lambda_1\right), \tag{3.1}$$

where A represents the weight parameter associated with a possible perturbation, λ_1 and λ_2 represent the two Legendre perturbations modifying our greyscale textures away from a uniform texture, and θ represents a possible perturbation (which here is a combination of λ_1 and λ_2), we can model its salience by the equation:

$$\mathrm{Sal}\left(\rho_{A,\theta}\right) \quad = \quad \left(\sum_{k=1}^{3}\left(AS_k \max\{0, \cos(\theta - \alpha_k)\}^{\beta}\right)^{p}\right)^{\frac{1}{p}}. \tag{3.2}$$

In this equation, k ranges from 1 to the number of mechanisms present (in our present model, there are 3 mechanisms present), S_k represents the weight of the k^{th} mechanism, α_k represents the specific combination of λ_1 and λ_2 corresponding to the k^{th} mechanism, p represents a Minkowski exponent to determine the type of distance metric our mechanisms follow when combined together, and β represents a parameter that applies a nonlinear penalty to textures that do not maximally activate a mechanism. A visualization of this model compared to an aggregate of observer's data can be seen in Figure 3.4. The red curve is defined by all observers' aggregate data, while the blue curve is defined by a best-fitting model assuming three mechanisms are present and combining according to the equation above. All model parameters for both the aggregate data shown above and each observers' individual data are displayed in Table 3.1.

The asymmetry of the equisalience contour generated in Figure 3.4 around the origin suggests that any given mechanism can only confer positive salience responses (i.e. no mechanism can negatively contribute to salience). Our model is sensitive to this feature in that every

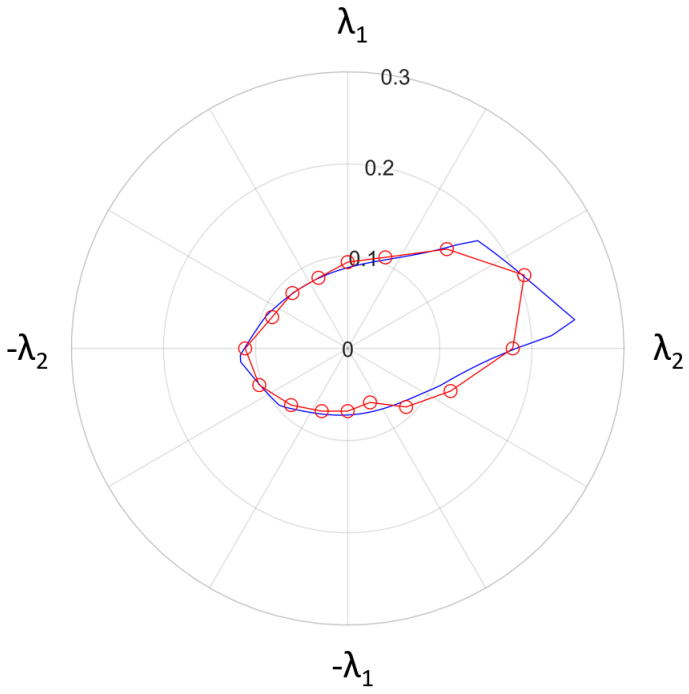

Figure 3.4: Aggregate results from all observers in the superthreshold contrast paradigm, shown in red, as well as a model of best fit from Equation 1, shown in blue. The angle counterclockwise from the right abscissa represents the proportion of both λ_1 and λ_2 present in the perturbation, while the distance from the origin represents the strength of the perturbation amplitude A.

mechanism present is only sensitive to deviations from the uniform histogram in a half-plane of stimulus space centered around the mechanism's strongest tuned area. To formalize this, the k^{th} mechanism in our model is sensitive only to perturbations $\rho_{A,\theta}$ in the half-plane for which $\cos(\theta - \alpha_k) > 0$. In our model, areas where an observer's equisalience contour are close to the center demonstrate areas where sensitivity is high, and therefore these areas correspond to histograms which maximally activate visual texture discrimination mechanisms. We identify only three mechanisms influencing observers' sensitivity in this feature space, as the expanded area present in the top-right quadrant, coupled with the fact that our data is decidedly not symmetric with respect to either the x- or y-axis, suggest that there are fewer than four mechanisms at play in our data.

This model identifies three mechanisms whose sensitivity mechanisms are shown in Fig-

ure 3.5. Starting from the right half of the abscissa of Figure 3.4 and moving counterclockwise, the minima signify a 'light mechanism', a 'grey mechanism', and a 'dark mechanism' that are each tuned toward different sets of grayscale values. We can evaluate the relative strengths of these mechanisms by measuring which minima in an observer's equisalience contour is lowest; alternatively, we can measure relative mechanistic strength by observing which mechanistic strength parameter S_k is strongest from our model fit, shown in Table 3.1. Our data determine the sensitivity functions of our mechanisms only up to an arbitrary additive constant, which does not guarantee that our sensitivity functions have a minima of 0. However, the results that we see are consistent with this convention, and while interpreting our results it makes more sense to view mechanisms as having positive influence contributions than having a mix of positive and negative influences. Of the three mechanisms identified, two mechanisms (the 'light' and 'dark' mechanisms) are maximally activated by a perturbation that selectively targets one half of the range of grayscale values, in a similar fashion to a half-wave rectifier. Our dark-sensitive mechanism was stronger than the other two mechanisms identified; however, this is balanced out by the fact that our grey-sensitive mechanism seemed slightly skewed towards the brighter half of the grayscale spectrume measured.

One could postulate that, since only one 'fixed texture' perturbation is used, the strength of all other texture perturbations are being determined with relation to the one 'fixed texture' perturbation, and that repeated comparisons to a 'fixed texture' may introduce systematic bias to the shape of the equisalience contour. To combat this, we have run a control experiment with one observer in which the 'fixed texture' was modulated through 4 different texture perturbations, which lay on the equisalient values of the positive and negative abscissa and ordinate axes. The results from this control experiment do not seem to deviate in overall shape or amplitude from the results gathered from only one 'fixed texture' perturbation lying on the positive abscissa axis. This serves to emphasize that all points gathered from an observer's salience contour are indeed deemed 'equisalient' and can be used interchangeably with regards to salience from a uniformly distributed texture background.

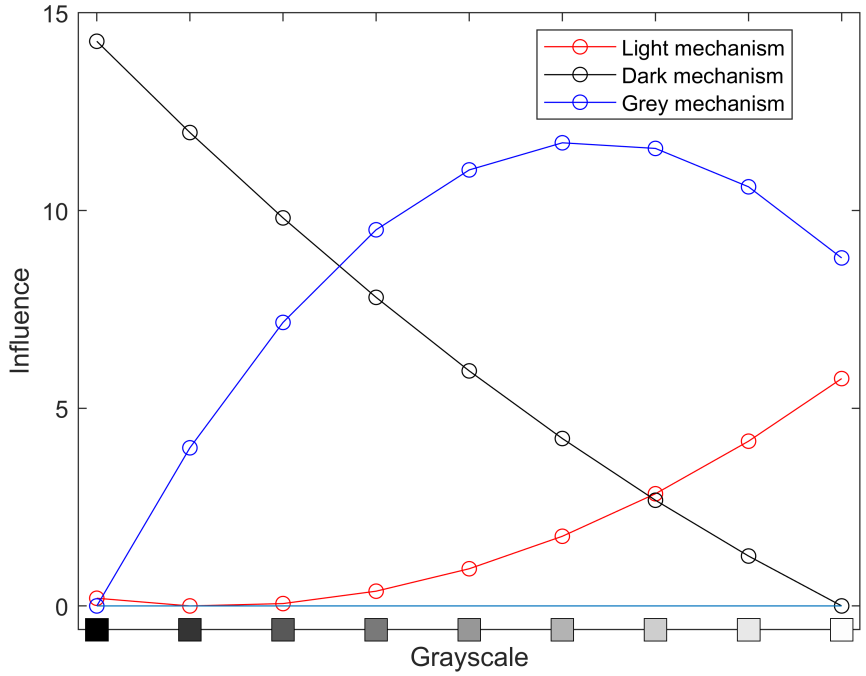

Figure 3.5: Sensitivity functions of the three mechanisms identified by our model in Figure 3.4. Each curve represents the mechanism's strength projected into second-order polynomial space. The light mechanism seems to slightly suppress the 4 darker grayscales while strongly activating for the brighter grayscales shown, while the dark mechanism seems to be symmetrically opposite. The grey mechanism seems to suppress both light and dark values while strongly activating only the contiguous grayscales in the middle of the spectrum. In addition, the dark mechanism identified exhibits a much higher influence than the other two mechanisms identified, suggesting that either (1) the human visual system is more sensitive to dark values than light values or that (2) there are more mechanisms influencing the processing of dark grayscales, such as the blackshot mechanism (Chubb et al., 2004).

3.5 Discussion

Previous work on threshold equisalience judgments has resulted in elliptical contours primarily governed by the noise limiting performance (Poirson and Wandell, 1990). While our current results are elliptical in nature, the asymmetries present across all observers suggest the noise is not limiting performance to the same degree in our results. An especially striking feature of our results is the strong agreement that all observers display in determining sensitivity for the lower right quadrant, which corresponds to perturbations where the mean luminance is shifted to a lower luminance value and the average variance is increased. When this quadrant is compared to the top right quadrant, which corresponds to perturbations where the mean luminance is shifted to a higher luminance value and the average variance is increased, we see that observers both (1) agree on sensitivity values to a much stronger degree and (2) are more sensitive overall to these values. Previous literature has shown that this attention filter sensitive primarily to dark high-variance regions and insensitive to light high-variance regions is present for threshold histogram differences (Chubb and Nam, 2000), and our experiment confirms this result in a superthreshold salience-matching environment. However, unlike the mechanism detected in (Chubb and Nam, 2000), the dark-selective mechanism that we identify expresses influence that is strongly linear across the gamut of

Subject	S_1	S_2	S_3	α_1	α_2	α_3	p	β	Error
S1	16.499	10.0448	6.728	4.523	2.167	0.481	0.996	0.982	0.0000369
S2	13.615	11.731	4.597	4.839	2.218	0.824	1.394	0.663	0.0000347
S3	20.02	27.76	5.744	4.404	2.672	0.675	3.0462	0.347	0.0036
S4	12.118	7.612	4.3	4.875	2.489	0.872	1.037	0.831	0.0012
S5	10.886	9.621	5.677	4.878	2.245	1.054	1.494	0.63	0.0018
S6	10.945	10.012	5.511	4.78	2.267	0.941	1.223	0.801	0.0025
Avg	**13.886**	**11.2**	**5.822**	**4.807**	**2.277**	**1.178**	**1.87**	**0.526**	**0.0000502**

Table 3.1: *Parameters for the best model fit for each observer.* Columns 2-4 show the observers' strength parameters for the light, dark, and grey tuned mechanisms, columns 5-7 show the observers' specific tuning directions (in radians), column 8 shows the model's Minkowski distance parameter, column 9 shows the model's penalty parameter, and column 10 shows the sum of squared deviations from the model and the data for a given observer.

grayscale values we measured.

Our experiment suggests a number of different visual mechanisms that exert influence into the feature space spanned by the Legendre polynomials for our task. While the bottom right quadrant shows the strongest effect, the bottom left and top left quadrants of our task show a fair degree of agreement and high sensitivity values. However, the top right quadrant displays results where observers are both inconsistent with each other and display less sensitivity to detecting perturbations. This pattern of perturbation strength aligns with previous literature of visual mechanisms, in which a high mean luminance, low mean luminance, and low variance mechanism were detected for achromatic texture patterns, but a high variance mechanism was conspicuously absent (Silva and Chubb, 2014). Additionally, the dark-activated mechanism seems to confer about twice as much sensitivity when compared to the other two identified mechanisms. This suggests that either (1) the human visual system is more sensitive to dark grayscale values than light grayscale values, or that (2) the mechanism that we have identified is actually a composite of several mechanisms influencing the processing of dark grayscale values, such as a dark-sensitive value and the blackshot mechanism (Chubb et al., 2004); however, given that the influence of our dark-selective mechanism across grayscales is nearly linear, our results suggest that this mechanism is more consistent with the "down-ramped" mechanism seen in Silva and Chubb (2014) than the sharply tuned "blackshot" mechanism seen in Chubb et al. (2004).

One possible confound for our experiment lies within the subjective nature of an observer's interpretation of 'equisalience'. For instance, while a texture patch of $U + \rho_{var}$ and a patch of $U - \rho_{var}$ can both be discriminable from the background perturbation U, it may be difficult for an observer to identify which patch appears *more* salient with respect to the background, based on how observers (1) may not agree on a weighting scheme or (2) may not ascribe symmetric weights to luminance or variance changes in positive and negative directions. However, the data which we have collected from each observer's equisalience

contour show striking similarities between the values of A_{var} across numerous perturbations. The mechanisms identified in our aggregate data hold individually for all of our observers. From this, we can establish that despite the apparent difficulty of the task, the internal criterion observers seem to be using for salience seems to be quite consistent across observers, even accounting for various individual differences in amplitude of sensitivity.

Of note is that the sensitivity shown in this model is only emblematic of the perturbations being projected into the space defined by λ_1 and λ_2. While these results suggest that we have found a texture that maximally activates a narrowly tuned mechanism (and preliminary testing revealed a lack of sensitivity to Legendre polynomials of degree 3 or higher in this paradigm), there may exist other modifications that observers are sensitive to which cannot be encapsulated by a linear combination of Legendre polynomial modifications alone. To solve this issue, we can run this experiment in an iterative fashion in order to specify exactly what distribution of texture elements maximally activates a mechanism. The strongest activating histogram from one experiment could be set as the perturbation on the abscissa, while a new perturbation function could be introduced on the ordinate. Running this paradigm through several iterations in this fashion will allow our maximally sensitive textures to span several dimensions of sensitivity and will result in a texture which differentially activates a mechanism in the strongest fashion possible. It can then be said that humans have a mechanism which is tuned specifically to that histogram of texture elements and is activated in proportion to the strength of that mechanism's projection on any other textures shown. Constructing an N-dimensional space spanning several different perturbations away from a uniform grayscale texture may also generate a very different contour from the one we present here; however, determining what set of perturbations to use which are orthogonal to one another may prove to be a difficult task.

One last issue to be addressed is the dilemma that this paradigm has only been tested on unidimensional stimuli (that is, the texture elements have only varied on a luminance-isolated

activation continuum). While there are a wide array of unidimensionally varying stimuli that can be tested, this paradigm would gain a great deal of utility if it could be extended to work on multidimensional stimuli. It may very well be possible to test stimuli that vary in two or more dimensions; however, extending this paradigm to function on stimuli that vary in N dimensions would require that researchers construct an equiluminance contour that exists in at least $N + 1$ dimensions, and the number of trials that would be necessary to explore a wide number of perturbations in a 3 dimensional or higher contour space with a high degree of specificity would quickly become overwhelming. While this endeavor could theoretically be taken, it may be a time-consuming process.

Why is this work important? This experimental setup offers a new look into the influence of various mechanisms in a relatively noise-free environment that was not achievable in previous studies. This allows us to both identify preattentive mechanisms and measure how effectively they function at discriminating between various stimuli. In particular, we have been able to specifically identify the tuning functions, relative strengths, and combination rules that govern texture-sensitive mechanisms; these features have not been possible to quantify until now. The scope of mechanisms that can be found is large; previous work in the lab has made advances in identifying mechanisms involved with stimuli differing in size, shape, luminance, and color. Gaining a better understanding of these preattentive mechanisms may allow us to construct a multidimensional space in which each of these feature dimensions resides based upon perceptual discriminability, which would allow us to develop a more complete model of texture discrimination. This may have wide reaching impacts in computer vision, biomedical research, camouflage detection, and further texture perception research.

Chapter 4

Analysis of micropattern features that contribute to salience judgments

4.1 Abstract

Basing his conclusions on experiments using images composed of binary micropatterns made of small line segments, Julesz (1981) introduced the concept of a "texton" to vision research, proposing that these hypothetical local image features were the "basic elements of pre-attentive perception." Julesz's experiments demonstrated the existence of preattentive mechanisms that enabled discrimination between some of his micropattern textures. However, he lacked a method sufficiently powerful to quantitatively characterize these mechanisms. The current project applies the centroid method to this problem. Six micropatterns varying in a range of features are used in stimuli. In different conditions of the centroid task, participants strive to achieve attention filters selective for individual members of this set. Effective attention filters are achieved for three of them: (1) straight line segment, (2) orthogonal crossed line-segments, and (3) orthogonal line segments forming an "L". Un-

der the assumption that an effective attention filter for a given micropattern can only be produced if that micropattern possesses a texton not possessed by the other micropatterns, these findings imply the existence of at least three textons.

4.2 Introduction

How might visual perception lead to the phenomenon of texture segregation? A necessary first stage of visual perception is to sense one or more types of statistical variation in the visual input across space and time. In photopic vision, this sensing begins with the L-, M- and S-cone classes, each of which provides a time-varying spatial map reflecting the distribution of a particular property of the light incident on the retina. However, substantial empirical evidence indicates that (1) human observers do not have direct access to the neural images provided by these photoreceptor classes, whereas (2) human observers do have access to a number of neural images reflecting the spatial distributions of image statistics other than those captured by the photoreceptors – all of which presumably have been derived through various neural transformations from the photoreceptor maps (Robson and Harris, 1980; D'Zmura and Lennie, 1986; D'Zmura, 1991; D'Zmura et al., 1997; Julesz, 1962; Julesz et al., 1978; Julesz, 1981; Kovacs and Julesz, 1993; Webster and Mollon, 1991; Chubb et al., 1994).

How should we think about these higher-order neural images which people can access to control their behavior? Some insight can be garnered by considering some of the basic facts about our visual experience. When we open our eyes, we don't see a single homogeneous field; we see a heterogeneous field composed of numerous spatial regions varying in all sorts of different qualities. It may seem at first that each of these distinct regions is homogeneous, but on closer inspection, these regions may themselves be heterogeneous. Both experiences are important; for example, at a macro level, we may want to easily segment a field of wheat

from a patch of forest, but at a micro level, we may want to observe whether a predator is hiding within a wheat field by identifying any irregularities in how the pattern of wheat in the field is distributed. In doing both tasks, we must understand how our brain parses visual input into distinct regions.

What components of our brain enable this parsing of the visual field? It is known that certain neurons in visual cortex respond only to specific image statistics (Hubel and Wiesel, 1959). Groups of these neurons may be organized into retinotopic maps to form 'field-capture channels' (Silva and Chubb, 2014) which are sensitive broadly across visual space to specific image statistics. Every field-capture channel is continuously updated with the visual input encoded by the photoreceptors in the retina. Each channel then processes the visual image in parallel to produce a transformation, (akin to the 'feature maps' of Treisman and Gelade (1980)) whose pattern of activation reflects the distribution across space of the statistic sensed by that channel. The task of parsing the visual field can then be simplified down to identifying 'boundaries' that mark sharp changes in the activation of one more more field-capture channels.

Julesz and colleagues first attempted to identify new channels that permitted texture segregation by randomly placing several different groups of texture elements in the same image (Julesz et al., 1973). To identify and measure the influence of texture features in visual texture perception, Julesz and colleagues developed a number of "micropatterns" which were composed of several different texture features. These micropatterns varied in the presence and number of line terminators, line intersections, incident angles of intersection, curvature of lines, and number of lines present. Early work by Julesz led to the idea that textures composed of these micropatterns were differentiable based solely on differences between the first-order (probability distribution) and second-order (autocorrelation values) statistics of the stimulus image. This was, however, eventually proven wrong (Julesz et al., 1978); even while the first- and second-order statistics of these micropatterns were constrained, observers

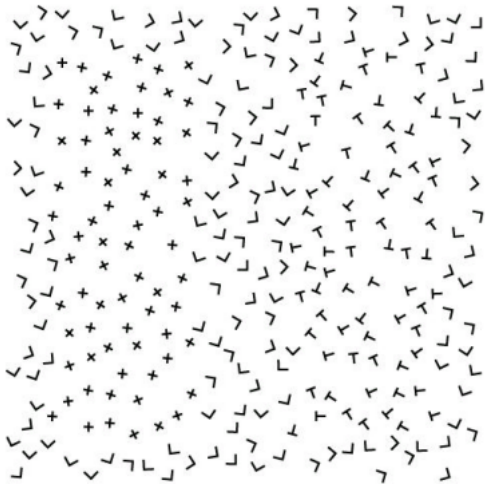

Figure 4.1: *An example of micropattern segregation.* The cross micropatterns on the left half of the figure easily segregate from the L micropatterns that compose the background. However, the T micropatterns on the right half of the image require focal attention in order to segregate easily.

were able to segregate certain target micropatterns from a background of distractor micropatterns. However, segregation was not possible across all sets of micropatterns; for instance, a field of cross micropatterns could be easily segregated from a field of L-shaped micropatterns, but a field of T-shaped micropatterns would not segregate easily from this same field of L-shaped micropatterns without focal attention. This effect can be seen in Figure 4.1. This led Julesz to propose that certain fundamental local image features of micropatterns, or 'textons', were responsible for the perceptual texture segregation observed (Julesz, 1981). Therefore, there must exist mechanisms in the visual cortex which are selectively tuned to certain micropattern features (resulting in effortless segregation) and activate at a weaker level for other features (resulting in effortful segregation).

Although Julesz's work proved useful in identifying many essential features of visual channels, such as the difference between texture elements that could spontaneously segregate and texture elements that required attentional resources to segregate, technological limitations prevented a quantitative analysis of which specific visual mechanisms may be at work and how strong they might be in relation to each other.

The centroid paradigm (Sun et al., 2015), enables us to revisit this question. The centroid paradigm tasks observers with finding the center of mass of an array of randomly placed objects (Drew et al., 2010; Sun et al., 2015). Participants are instructed to observe a display of stimulus objects that vary along at least one feature dimension, such as color, shape, luminance, size, or some other discriminable feature dimension. In "binary" centroid task conditions, all items are either targets or distractors; observers are instructed to mouse-click the centroid of stimulus cloud giving equal weight to all target items while ignoring all 'distractor' items (e.g. attend to red items while ignoring any non-red items). Participants are then shown a brief stimulus display consisting of a mixture of both target and distractor items, which is followed by a post-stimulus mask. After the observer has registered his response with a mouse-click, the original stimulus is presented again, along with feedback showing (i) the stimulus cloud, (ii) the target location, and (iii) the location of the response the observer has just made. The purpose of this feedback is to enable observers to improve their computation over trials. Linear regression is then used to estimate the weights exerted on the observer's responses by the different different types of items in the stimulus displays. This paradigm enables us to take a closer look at the relative influence that each item or combinations thereof can exert on an observer by varying which stimulus items are designated as target or distractor items (e.g. attend to only red items in one case, only blue items in a different case, etc.).

4.3 Experiment 1

4.3.1 Methods

Stimuli consisted of a bounding box subtending 14×14 degrees of visual angle populated with micropatterns which each subtended 0.5 degrees of visual angle. Six micropattern types

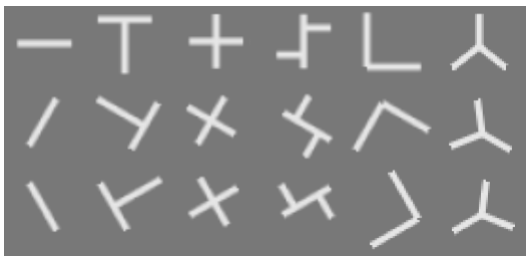

Figure 4.2: *Enlarged template examples of the six micropatterns used in the experiment.* These micropatterns were chosen for this experiment to isolate the texture properties of line intersections and line terminators from other properties, such as texture closure. We refer to these micropatterns, respectively, as a "line," "T," "cross," "off-cross," "L," and "Y." The first row shows the unrotated micropatterns, the second row shows the micropatterns after a 120 degree counterclockwise rotation, and the third row shows the micropatterns after a 240 degree counterclockwise rotation.

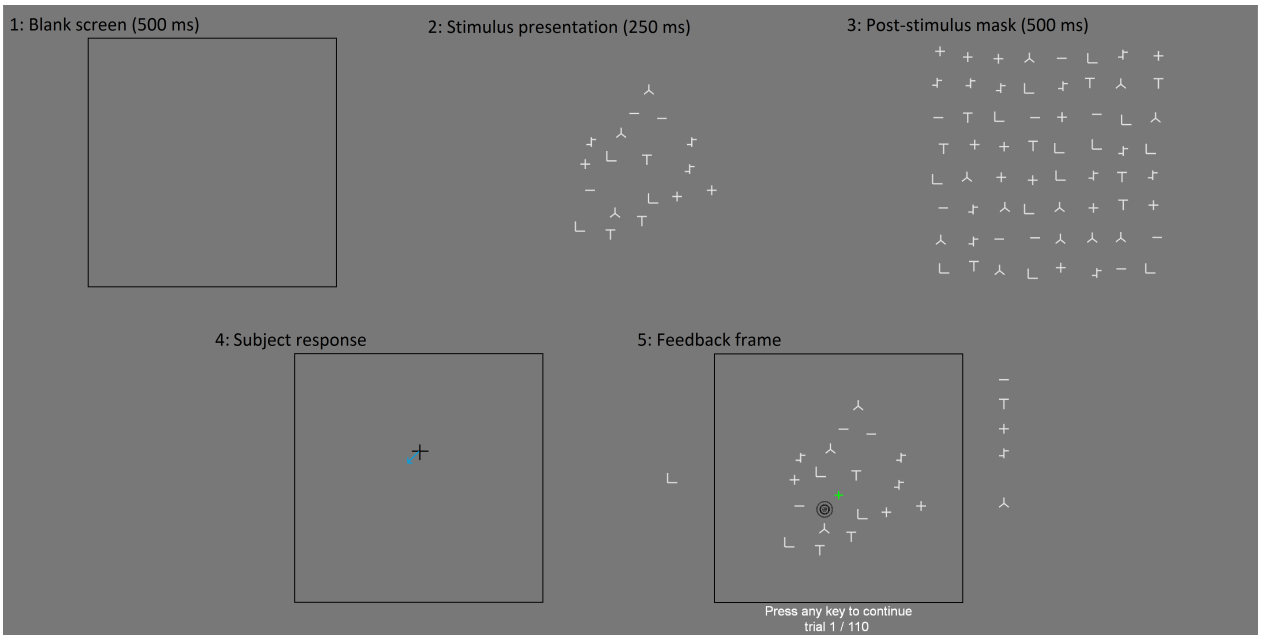

Figure 4.3: *The sequence of display images which occur during a trial of the micropattern centroid task.* After an observer indicates that he/she is ready, the top left panel displays a mean gray field on the screen for 0.5 seconds. Afterwards, the top middle panel displays a number of micropatterns on the screen for 0.25 seconds. The stimulus was then followed by a post-stimulus mask that was displayed for 0.5 seconds, as shown in the top right panel. The stimulus then disappears and a crosshair guided by the observer's mouse movements appears, allowing him or her to indicate the perceived centroid under the current target condition. Finally, the bottom right panel contains all micropatterns from the original display, the observer's response, and a feedback bullseye centered on the target centroid's location.

were used, as detailed in Figure 4.2. On each trial, the stimulus contained three instances of each micropattern type (for a total of 18 micropattern instances per trial). In this first experiment, each instance of a certain type of micropattern was oriented in the same direction across trials and blocks. There were six different, separately blocked conditions. In a given condition, one of the six micropattern types was the target, and the other five types were distractors. The cloud of all stimulus items were generated by distributing all item locations as a bivariate Gaussian density function with its expectation at the center of the presented bounding box and its standard deviation small enough to fit all items within the presented bounding box. Specifically, the dispersion of the stimulus cloud (Sun et al., 2015) was chosen such that 95% of generated trials were valid (here, the dispersion value was 1.81 degrees of visual angle away from the center of the screen), and a gap of 0.2 degrees of visual angle was imposed between all stimulus items to prevent any overlap. Observers were trained to identify each micropattern individually and were instructed to mouse-click the centroid of the three target micropatterns in the stimulus ignoring the 15 distractors.

A single trial consisted of 500 milliseconds of a blank cue frame, 300 milliseconds of a stimulus presentation, a gathering of the observer's response against a blank frame, and a feedback frame presented until the observer proceeded to the next trial. A single stimulus presentation consisted of all 18 micropatterns being painted in white onto a uniform grey background of 52 c/m^2, which was half of the monitor's maximal output. The feedback frame consisted of all the micropatterns present in the stimulus presentation, but with the added features of the location of the observer's response and the location of the target centroid (the centroid of the three target items). A full trial's composite screens are shown in Figure 4.3. Two blocks were run for each attention condition (12 blocks total), and each block contained 110 trials, with 10 trials serving as baseline trials where only target stimuli items were shown in order to capture a lower-bound on response error, for a total of 1320 trials per observer. The order of attention conditions was varied across observers according to a Latin square experimental design.

For this experiment, we selected six different micropatterns to serve as our stimulus object set. These micropatterns, referred to as a "line," "T," "cross," "off-cross," "L," and "Y," are shown in Figure 4.2. These micropatterns are by no means exhaustive and do not intend to reflect the full range of atomic patterns that trigger differential sensitivity in the human visual system. We selected these micropatterns because (1) many of them were previously studied by Julesz and colleagues, (2) they all consist of variations on a small set of atomic lines, line terminators, and line crossings, (3) they mostly reflect various interpretations of simple intersections, and (4) they avoid the principle of closure; that is, there is no space being enclosed by any features of any of these micropatterns. This was done to reduce the number of confounding principles at play with these micropatterns, such that this experiment can serve as a 'proof-of-concept' to examine more varied micropatterns in the future if successful.

4.3.2 Results

An aggregated analysis of five observers' results for the first experiment is shown in Figure 4.4. One can gauge the confusability of a given micropattern, m_2 with another micropattern m_2, by observing the weight exerted by m_2 on the observer's responses when m_1 is the target. Our results agree with the phenomenon shown in Figure 4.1: T's are more confusable with L's than crosses are; also L's are more confusable with T's than crosses are; finally, L's and T's are roughly equally confusable with crosses.

Table 4.1 reports several statistics that reflect the effectiveness of the attention filters achieved in the different conditions. *Efficiency* gives the minimum proportion of items in the stimulus display that would need to be included in the observer's centroid computation in order to achieve responses as systematically controlled as those produced by the observer. *Data-drivenness* reflects the degree to which the observer's responses are determined by the stimulus on the current trial rather than tending toward a fixed default location $(x_{default}, y_{default})$;

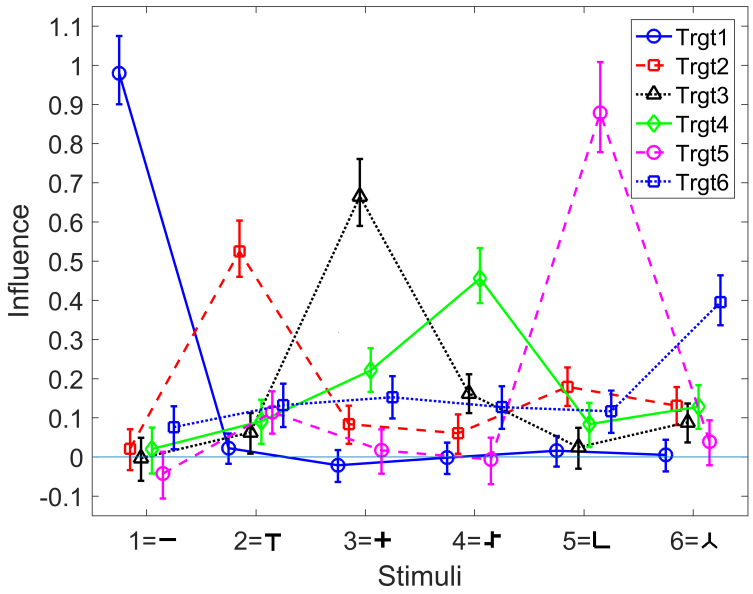

Figure 4.4: *Average attention filters achieved in all conditions in Experiment 1.* Each of the six micropattern types (see Fig. 4.2) served as the target in one condition. The average attention filters achieved in the different conditions are plotted as follows for targets equal to (a) lines: solid blue, (b) T's: dashed red, (c) crosses: dotted black, (d) off-crosses: green, (e) L's: dashed pink, (f) Y's: dotted blue. This task serves as a baseline to demonstrate that there are striking differences in sensitivity as you attend to different micropatterns. There exist various task asymmetries (e.g. the distracting effect of crosses on lines does not necessarily reflect the distracting effect of lines on crosses), which are present in other search literature. There also exist asymmetries in task performance, even though all micropatterns contain only lines, line terminators, or crossings, and all micropatterns lack closure. Error bars give 95% confidence intervals.

Data-drivenness less than 1 indicates a trial-by-trial tendency for the observer's response to be drawn toward a fixed location ($x_{default}, y_{default}$). *Selectivity* is the ratio of the weight given to the micropatterns of the target type in a given condition to the sum of the weights given to distractors. An in-depth explanation of these statistics can be found in Sun et al. (2015).

Every observer's efficiency, data-drivenness, and selectivity in each condition are shown in Table 4.1. From this table, we can get a more nuanced view of what features are being captured in each attention condition. Each observer has a high degree of data-drivenness, which reflects that they are attending to the stimuli present rather than anchoring to a fixed point in space. While efficiency metrics were largely similar between observers, they varied across different micropatterns. In particular, observers were highly efficient in the line, cross, and L target conditions, but largely inefficient in the T, off-cross, and Y target conditions.

	Line			T			Cross		
Observer	Eff	DD	Sel	Eff	DD	Sel	Eff	DD	Sel
S1	0.9	0.85	33.78	0.52	0.91	5.75	0.56	0.8	8.5
S2	0.76	0.79	25.75	0.44	0.69	6.36	0.47	0.75	5.93
S3	0.73	0.89	21.92	0.44	1.1	5.42	0.53	1.05	6.62
S4	0.94	0.65	22.79	0.44	0.64	7.04	0.7	0.59	20.0
S5	0.9	0.93	84.43	0.51	0.88	3.35	0.73	1.06	8.59
Avg	**0.84**	**0.80**	**37.73**	**0.47**	**0.83**	**5.58**	**0.61**	**0.84**	**9.93**
	Off-Cross			L			Y		
Observer	Eff	DD	Sel	Eff	DD	Sel	Eff	DD	Sel
S1	0.44	0.78	4.41	0.76	0.76	23.04	0.49	0.92	4.47
S2	0.41	0.68	2.55	0.56	0.57	9.38	0.25	0.56	1.64
S3	0.5	1.03	6.48	0.62	0.8	12.96	0.34	1.04	3.29
S4	0.18	0.5	4.27	0.82	0.55	17.45	0.23	0.64	4.58
S5	0.53	0.89	3.78	0.75	0.82	16.39	0.39	0.85	3.1
Avg	**0.43**	**0.77**	**4.3**	**0.69**	**0.69**	**15.84**	**0.36**	**0.77**	**3.42**

Table 4.1: Measures of each observer's efficiency, data-drivenness, and selectivity for each target condition in Experiment 1. Note: Averages for efficiency and data-drivenness are arithmetic means; averages for selectivity are geometric means. Each metric was calculated from the formulas shown in Sun, Chubb, Wright, & Sperling, 2015.

This pattern is reflected in the influence functions plotted and reveals striking differences in attentional capture across the six micropatterns shown.

4.4 Experiment 2

4.4.1 Methods

With only three instances of each micropattern type present in Experiment 1, there is a possibility that observers are not truly computing a center-of-mass computation to locate a stimulus cloud's centroid. Perhaps it is possible (at least in some conditions) to explicitly identify the locations of all three target items and then use a triangle-centering strategy to produce the response. We can address this possibility by increasing the number of each type of micropattern displayed from three to five while not increasing the display time of the stimulus screen.

Stimuli again consisted of a bounding box subtending 14×14 degrees of visual angle populated with the same six micropattern types from Experiment 1. Each micropattern was made slightly smaller so as not to overpopulate or enlarge the viewing box and thus only subtended 0.3 degrees of visual angle. On each trial there were five instances of each micropattern type. Each instance of a certain type of micropattern was oriented in the direction of the micropatterns in the top row of Figure 4.2 across trials and blocks. All other factors, including instructions, stimulus generation, stimulus presentation, number of trials, number of blocks, and ordering of attention conditions were otherwise kept the same from Experiment 1 to Experiment 2.

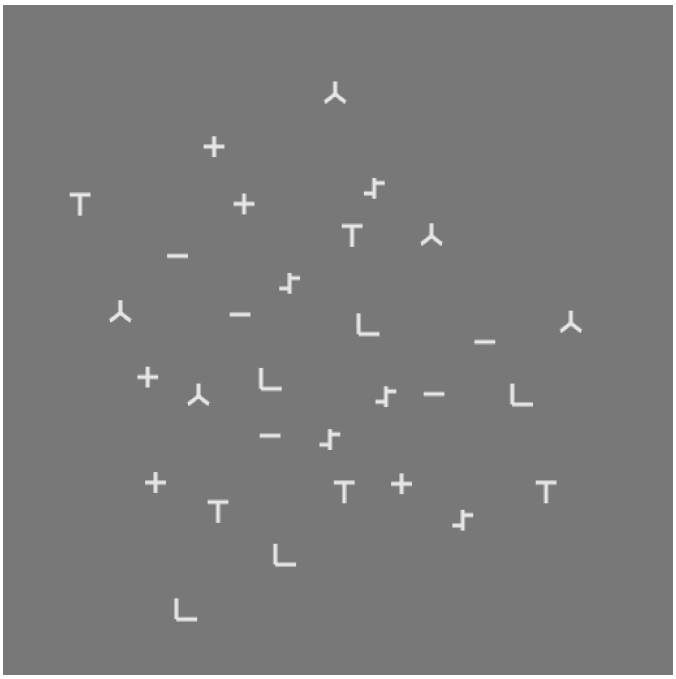

Figure 4.5: *A presentation of stimulus items from Experiment 2.* All six unique micropattern types displayed in the top row of Figure 4.2 are shown five times, consisting of 30 micropatterns appearing per trial. Each micropattern instance was made slightly smaller (sized down from 0.5 to 0.3 degrees of visual angle) to keep the stimulus cloud within the presented bounding box.

4.4.2 Results

An aggregated analysis of five observers' results for the second experiment is shown in Figure 4.6. Every observer's efficiency, data-drivenness, and selectivity in each condition are shown in Table 4.2.

The trend that was shown in Experiment 1 continues through to Experiment 2, albeit with slightly higher levels of influence exerted by distractors. The added noise in the influence computations was due to each centroid being noticeably harder to compute, which could be attributed to the increased number of distractor micropatterns as well as the reduced size of each micropattern instance. Each micropattern type has the same relative influence as in Experiment 1, which confirms that (1) a centroid is being computed rather than a triangulation (or both procedures have sufficiently similar behavioral outputs), and (2) each micropattern has a distinct degree of attentional capture as posited by Experiment 1 with regards to a background composed of various distractor micropatterns that is invariant to size or numerosity.

4.5 Experiment 3

4.5.1 Methods

The micropatterns used in the first two experiments were fixed in orientation. This opens the possibility that the attention filters achieved by our observers depend on the specific orientations of the micropatterns. It is even conceivable that observers are able to achieve neural templates aligned with these fixed-orientation micropatterns. Experiment 3 addresses this issue by seeing whether significantly different weights are conferred to different rotations of micropatterns. In particular, micropattern 4 in its unrotated form is the capital English

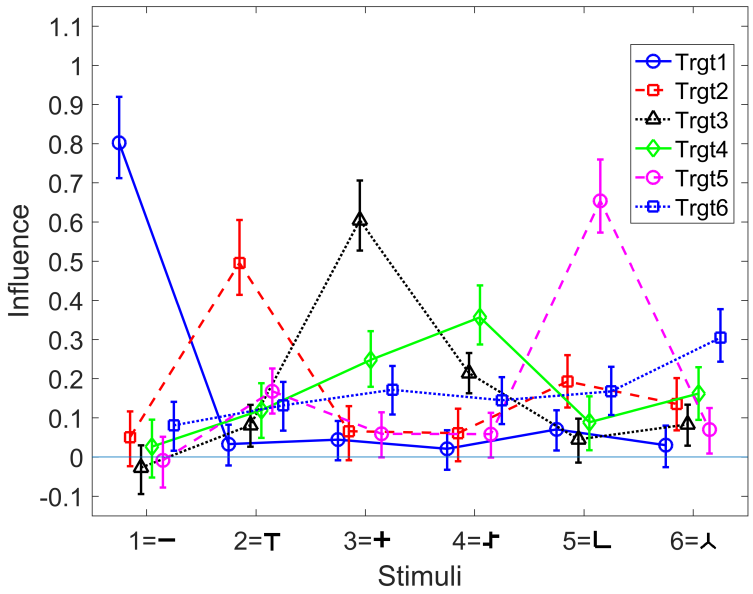

Figure 4.6: *Average attention filters achieved in the six target conditions in Experiment 2.* The six target conditions correspond to the observer being told to attend to each of the six targets shown in Figure 4.2. These results demonstrate that a true centroid computation is being done, and the attention filters are similar to those in Experiment 1. Error bars give 95% confidence intervals.

letter 'T'. It seems likely that this overlearned visual symbol might enable strategies that depend on its upright presentation. When told to attend to this micropattern across all orientations, observers may be vastly more sensitive to the unrotated pattern due to repeated exposure to this letter in everyday life. However, if sensitivity is shown to both unrotated and rotated forms of this micropattern, then there may exist a mechanism in human vision sensitive to the features defined by the relationships between the line segments of which it is composed (Shepard and Metzler, 1988).

Stimuli again consisted of a bounding box subtending 14×14 degrees of visual angle populated with the same six micropattern types from Experiment 1 that each subtend 0.5 degrees of visual angle. Each stimulus (see Figure 4.7) contained one each of all of the 18 micropatterns show in Figure 4.2. All other factors, including instructions, stimulus generation, stimulus presentation, number of trials, number of blocks, and ordering of attention

	Line			T			Cross		
Observer	Eff	DD	Sel	Eff	DD	Sel	Eff	DD	Sel
S1	0.78	1.05	13.93	0.49	0.84	8.68	0.49	0.76	8.92
S2	0.36	0.52	4.93	0.27	0.52	3.95	0.4	0.59	3.4
S3	0.53	0.91	9.05	0.35	0.88	5.46	0.58	1.22	5.3
S4	0.85	0.53	17.56	0.22	0.55	7.44	0.76	0.56	82.62
S5	0.78	1.0	13.87	0.43	0.87	1.95	0.61	1.0	5.4
Avg	**0.66**	**0.80**	**11.87**	**0.35**	**0.73**	**5.5**	**0.57**	**0.83**	**21.13**
	Off-Cross			L			Y		
Observer	Eff	DD	Sel	Eff	DD	Sel	Eff	DD	Sel
S1	0.32	0.92	2.56	0.66	0.86	10.51	0.37	0.9	3.73
S2	0.17	0.36	2.89	0.37	0.58	5.57	0.33	0.58	1.32
S3	0.35	0.87	4.93	0.47	0.9	11.14	0.29	1.11	1.9
S4	0.16	0.66	1.77	0.51	0.58	20.86	0.11	0.54	3.13
S5	0.44	0.95	2.34	0.59	0.97	6.58	0.36	0.85	1.45
Avg	**0.29**	**0.75**	**2.9**	**0.52**	**0.78**	**10.93**	**0.29**	**0.80**	**2.31**

Table 4.2: Measures of each observer's efficiency, data-drivenness, and selectivity in the six different target conditions in Experiment 2. Note: Averages for efficiency and data-drivenness are arithmetic means; averages for selectivity are geometric means. Each metric was calculated from the formulas shown in Sun et al. (2015).

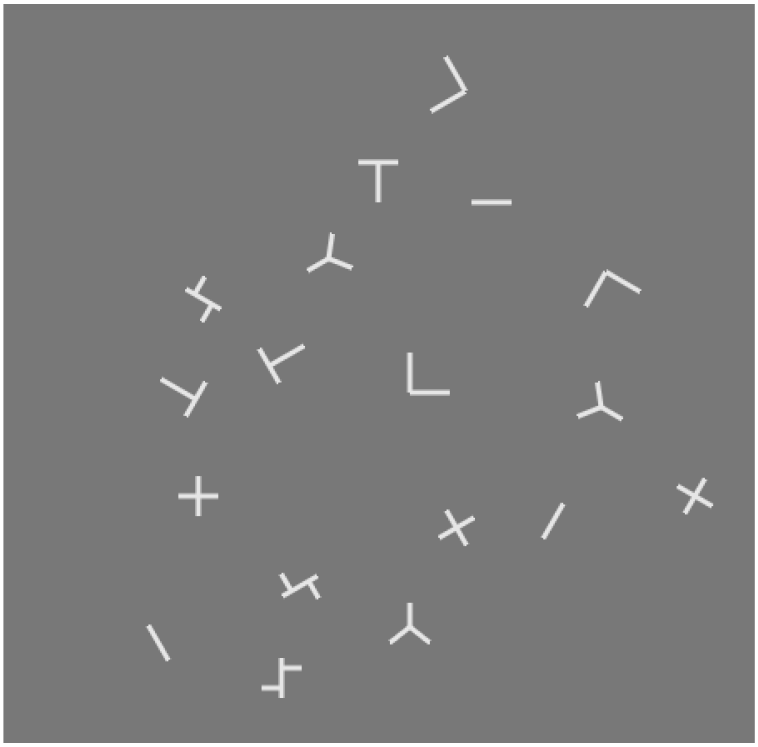

Figure 4.7: *A presentation of stimulus items from Experiment 3.* All 18 items displayed in Figure 4.2 are shown once, consisting of 6 unique micropatterns varying across 3 rotation values (unrotated, 120 degrees of rotation, and 240 degrees of rotation).

conditions were otherwise kept the same from Experiment 1 to Experiment 3.

4.5.2 Results

The average attention filters achieved by five observers in the third experiment are plotted in Figure 4.8. The efficiency, data-drivenness, and selectivity achieved by each observer in each condition are shown in Table 4.3. Overall, sensitivity to micropattern rotation seems to be largely nonexistent, as observers do not show significantly different degrees of influence to instances of micropatterns with different rotations. The largest deviation from equality across different orientations is seen amongst T's, in which the upright 'T' rotation appears to exert stronger influence than either of the other two rotations; however, the 95% confidence intervals of the upright T overlap with those of its two rotations.

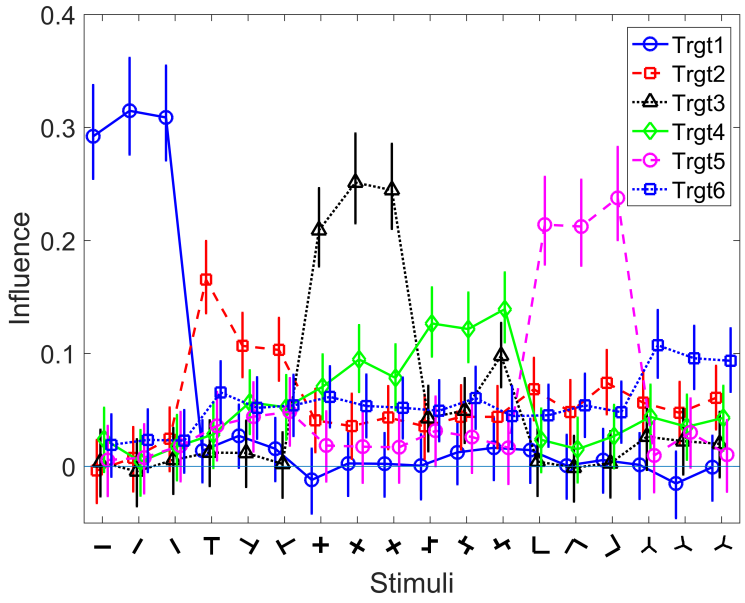

Figure 4.8: *Average attention filters achieved in the third experiment.* The six conditions take the six micropatterns shown in Figure 4.2, each subject to rotations of 0°, 120° and 240°. Error bars give 95% confidence intervals.

4.6 Discussion

Micropattern orientation is irrelevant. The fact that the pattern of results in Experiment 3 echoed the results of Experiments 1 and 2 suggests that even though the orientations of the micropatterns used in Experiments 1 and 2 were fixed, the attention filters deployed by our observers in those first two experiments were actually invariant with respect to orientation. This finding is consonant with the results of Inverso et al. (2016) who tested observers in the centroid task using display that contained equal numbers of four different types of micropatterns: vertical white bars, vertical black bars, horizontal white bars, and horizontal black bars. The four attention conditions of interest were Attend-vertical (all vertical bars were targets, and all horizontal bars were distractors), and analogously, Attend-horizontal, Attend-white and Attend-black. Inverso et al. (2016) found that performance was much better in the Attend-white and Attend-black conditions than in the Attend-vertical and Attend-horizontal conditions. In the Attend-white and Attend-black conditions, for displays

containing either 2, 3 or 4 of each of the four types of bars, observers achieved selectivities around 30 (i.e., they gave 30 times as much weight to targets as they did to distractors) in the Attend-black and Attend-white conditions; by contrast, in the Attend-horizontal and Attend-vertical conditions, observers were largely unable to ignore distractors, giving nearly equal weight to targets and distractors. Thus, even when the orientation difference between micropatterns is made as dramatic as possible, this cue affords very little purchase for feature-based attention. It is hardly surprising, then, that observers made no use of the fixed orientations of the micropatterns in Experiments 1 and 2.

A mystery: What texton distinguishes line micropatterns from the other micropatterns? Strikingly, across all three experiments, the line micropattern afforded the most effective attention filters. This finding suggests that there exists at least one field-capture channel that is activated by lines but not by any of the other micropatterns in our stimuli. Or, to

	Line			T			Cross		
Observer	Eff	DD	Sel	Eff	DD	Sel	Eff	DD	Sel
S1	0.89	0.81	157.67	0.54	1.0	3.16	0.74	0.77	27.74
S2	0.65	0.64	20.35	0.41	0.6	1.95	0.51	0.6	5.85
S3	0.74	0.9	20.35	0.48	1.03	4.46	0.7	1.03	13.7
S4	0.87	0.59	34.32	0.39	0.71	4.01	0.75	0.63	38.64
S5	0.78	0.93	20.51	0.55	0.93	2.04	0.72	1.12	5.39
Avg	**0.78**	**0.77**	**50.64**	**0.47**	**0.85**	**3.12**	**0.68**	**0.83**	**18.26**
	Off-Cross			L			Y		
Observer	Eff	DD	Sel	Eff	DD	Sel	Eff	DD	Sel
S1	0.44	1.06	2.51	0.67	0.91	7.88	0.48	1.02	3.03
S2	0.4	0.55	3.36	0.37	0.56	7.42	0.25	0.4	1.32
S3	0.53	0.94	4.45	0.55	1.03	7.86	0.45	1.12	1.81
S4	0.32	0.72	3.58	0.82	0.61	97.73	0.31	0.77	3.67
S5	0.7	1.19	2.25	0.64	0.98	6.05	0.65	1.17	1.29
Avg	**0.48**	**0.89**	**3.23**	**0.61**	**0.82**	**25.39**	**0.43**	**0.90**	**2.22**

Table 4.3: Measures of each observer's efficiency, data-drivenness, and selectivity while attending to each separate micropattern during Experiment 3. Note: Averages for efficiency and data-drivenness are arithmetic means; averages for selectivity are geometric means. Each metric was calculated from the formulas shown in Sun, Chubb, Wright, & Sperling, 2015.

put this in Juleszian terms: there exists some texton that is present in the line but not in any of our other micropatterns. This raises the question: What could this texton possibly be? Julesz and Bergen (1983) hypothesize that textons include

1. Elongated blobs–e.g., rectangles, ellipses, line segments with specific colors, angular orientations, widths and lengths.
2. Terminators–ends-of-line segments.
3. Crossings of line segments.

Thus, under the original texton theory, the line itself counts as a texton. Note, however, that all three of the T, L and cross micropatterns contain (1) a line segment identical to the line micropattern and (2) as many terminators as the line micropattern (the L) or more terminators than the line micropattern (the T and the cross); in addition, the off-cross contains a line segment of the same length as the line micropattern. It is thus hard to imagine what image statistic underlies the highly effective attention filters our observers achieve in the attend-to-line conditions.

It might be observed that the line micropattern takes up only about half as much area in pixels as any of the other micropatterns. It is possible to design an image transformation to detect this statistical difference; however, by default we expect the output of an image transformation sensitive to micropattern area to increase with micropattern area. A transformation whose response to a given micropattern is an increasing function of the number of pixels it contains would produce (compared to our other micropatterns) minimal activation in response to lines. Such a transformation would be useless for purposes of filtering out the other micropatterns and leaving just the lines. An impressive nonlinearity would need to be applied to the output of this area-driven transformation to leave only the lines and filter out the other micropatterns. It is difficult to imagine that such a neural transformation is built into human vision.

The bottom line about lines: The basic fact suggested by the current experiments is that the line micropattern, unperturbed by connection with other lines, does indeed activate a field-capture channel that the same line fails to activate when it forms a part of a more complicated micropattern. Further experiments will be needed to determine the exact nature of the statistic sensed by this field-capture channel.

Three textons are implicated by the current results. In addition to the Attend-to-lines condition, the Attend-to-crosses and Attend-to-L's conditions also yielded effective attention filters. This implies that there exist distinct field-capture channels that are strongly activated by (1) lines (but not by any of the other micropatterns), (2) crosses (but not by any of the other micropatterns) and (3) L's (but not by any of the other micropatterns).

Concerning crosses, it should be noted that the off-cross micropattern was specifically designed to be as similar as possible to the cross micropattern without including the specific visual property that makes a cross a cross: i.e., the property of including two continuous lines that run across each other. In particular, the off-cross has exactly as many terminators as does the cross, and the off-cross contains two horizontal and two vertical line segments of the same length as the four half-line segments contributing to the cross. Thus, the fact that that the Attend-to-off-crosses condition does not yield effective attention filters underscores the claim that crosses are textons (Julesz and Bergen, 1983).

It is perhaps surprising that the Attend-to-L's condition yields moderately effective attention filters. Julesz and Bergen (1983) found that it takes careful scrutiny to discover a single (randomly oriented) L among (randomly oriented) T's and vice versa. They took this as evidence that T's and L's are preattentively equivalent; i.e., they contain the same textons. There are several points to note in this connection. First, In all three experiments, the average attention filter in the Attend-to-L's condition shows higher activation for T's than it does for any of the other micropatterns. Thus, the current results do indeed suggest that T's are more confusable with L's than are other micropatterns. However, the influence exerted

by L's is nonetheless between 4 and 8 times greater than the influence exerted by T's in the Attend-to-L's condition. We suspect that the critical difference between the Julesz and Bergen (1983) search experiments and the current centroid task experiments is that the T's and L's in the stimuli of Julesz and Bergen (1983) were much more closely packed than the micropatterns in our stimuli. We speculate that the field-capture channel that enables observers to achieve attention filters selective for L's is disrupted if the L's are densely surrounded by T's.

An interesting negative finding. Note that the only micropattern that contained any angles other than right angles is the Y. One might have thought that this feature would enable effective attention filtering. The results show the contrary.

Chapter 5

Conclusion

The methods that we have used in this document have proven effective into answering questions about preattentive visual mechanisms that have, until now, proven unanswerable. The effects of these conclusions will enable deeper discussions in future studies of these mechanisms.

In Chapter 2, we analyzed what features were driving a difference in observers' perception between multiple isodipole textures. We used an application of the centroid method to determine that observers were not sensitive to the higher-order statistics that governed the difference between isodipole textures *per se*. Rather, we determined that observers' sensitivity was governed strongly by sets of local image features, and that certain sets of these features arise much more often in pure instances of certain types of isodipole texture but are not present in other isodipole textures.

Chapter 3 allowed us to further probe how salience was computed from visual texture. We created a method that allows us to examine several different textures and fine tune each different texture so that they were equally salient against a background of visual noise. Examination of these equisalient visual textures allowed us to create a model that identified

and measured the strength of visual mechanisms sensitive to these textures, which largely agreed with findings from past literature.

Finally, Chapter 4 revisited the centroid method with the new stimulus set of micropatterns composed of line ensembles. The centroid method enables us to examine the relative influence of each item under several different attention conditions. We note that of the six items examined, only three of them are easily distinguishable, and there exist micropatterns that are easily confusable despite being visually distinct. These findings agree with past literature, and they also offer insight to justify reproducing this experiment with different sets of micropatterns encapsulating different features in the future.

The work presented here provides insights into how preattentive mechanisms of the visual system function, but more importantly functions as a proof-of-concept for using the two experimental paradigms presented on further texture-related work. We have only used these methods on a narrow domain of stimuli, and many different mechanisms could be easily fit into these frameworks. Areas of initial interest may be the domain of textures that vary in color or micropatterns that contain elements of closure. Hopefully, upon seeing the results that we have demonstrated in this document, the path will be open for further research that will reach towards an even greater understanding of how the human visual system functions.

Bibliography

Charles Chubb and Jong-Ho Nam. Variance of high contrast textures is sensed using negative half-wave rectification. *Vision research*, 40(13):1677–1694, 2000.

Charles Chubb, John Econopouly, and Michael S Landy. Histogram contrast analysis and the visual segregation of iid textures. *JOSA A*, 11(9):2350–2374, 1994.

Charles Chubb, Michael S Landy, and John Econopouly. A visual mechanism tuned to black. *Vision research*, 44(27):3223–3232, 2004.

Stefanie A Drew, Charles F Chubb, and George Sperling. Precise attention filters for weber contrast derived from centroid estimations. *Journal of vision*, 10(10):20–20, 2010.

Michael D'Zmura. Color in visual search. *Vision research*, 31(6):951–966, 1991.

Michael D'Zmura and Peter Lennie. Mechanisms of color constancy. *JOSA A*, 3(10):1662–1672, 1986.

Michael D'Zmura, Philippe Colantoni, Kenneth Knoblauch, and Bernard Laget. Color transparency. *Perception*, 26(4):471–492, 1997.

MA Georgeson and GD Sullivan. Contrast constancy: deblurring in human vision by spatial frequency channels. *The Journal of Physiology*, 252(3):627–656, 1975.

George A Hay and M Susan Chesters. A model of visual threshold detection. *Journal of theoretical biology*, 67(2):221–240, 1977.

David H Hubel and Torsten N Wiesel. Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, 148(3):574–591, 1959.

Matthew Inverso, Peng Sun, Charles Chubb, Charles E Wright, and George Sperling. Evidence against global attention filters selective for absolute bar-orientation in human vision. *Attention, Perception, & Psychophysics*, 78(1):293–308, 2016.

Bela Julesz. Visual pattern discrimination. *Information Theory, IRE Transactions on*, 8(2): 84–92, 1962.

Bela Julesz. Textons, the elements of texture perception, and their interactions. *Nature*, 290 (5802):91–97, 1981.

Bela Julesz and James R Bergen. Human factors and behavioral science: Textons, the fundamental elements in preattentive vision and perception of textures. *The Bell system technical journal*, 62(6):1619–1645, 1983.

Bela Julesz, EN Gilbert, LA Shepp, and HL Frisch. Inability of humans to discriminate between visual textures that agree in second-order statisticsrevisited. *Perception*, 2(4): 391–405, 1973.

Bela Julesz, EN Gilbert, and Jonathan D Victor. Visual discrimination of textures with identical third-order statistics. *Biological Cybernetics*, 31(3):137–140, 1978.

Ilona Kovacs and Bela Julesz. A closed curve is much more than an incomplete one: Effect of closure in figure-ground segmentation. *Proceedings of the National Academy of Sciences*, 90(16):7495–7497, 1993.

M Concetta Morrone, DC Burr, and L Maffei. Functional implications of cross-orientation inhibition of cortical visual cells. i. neurophysiological evidence. *Proceedings of the Royal Society of London B: Biological Sciences*, 216(1204):335–354, 1982.

Alan V Oppenheim and Jae S Lim. The importance of phase in signals. *Proceedings of the IEEE*, 69(5):529–541, 1981.

Allen B Poirson and Brian A Wandell. The ellipsoidal representation of spectral sensitivity. *Vision research*, 30(4):647–652, 1990.

Javier Portilla and Eero P Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *International journal of computer vision*, 40(1):49–70, 2000.

Keith P Purpura, Jonathan D Victor, and Ephraim Katz. Striate cortex extracts higher-order spatial correlations from visual textures. *Proceedings of the National Academy of Sciences*, 91(18):8482–8486, 1994.

JG Robson and CS Harris. Neural images: The physiological basis of spatial vision. *Visual coding and adaptability*, pages 177–214, 1980.

Shenna Shepard and Douglas Metzler. Mental rotation: effects of dimensionality of objects and type of task. *Journal of Experimental Psychology: Human Perception and Performance*, 14(1):3, 1988.

Andrew E Silva and Charles Chubb. The 3-dimensional, 4-channel model of human visual sensitivity to grayscale scrambles. *Vision research*, 101:94–107, 2014.

Peng Sun, Charles Chubb, Charles E Wright, and George Sperling. The centroid paradigm: Quantifying feature-based attention in terms of attention filters. *Attention, Perception, & Psychophysics*, pages 1–42, 2015.

Anne Treisman and Garry Gelade. A feature-integration theory of attention. *Cognitive psychology*, 12(1):97–136, 1980.

Jonathan D Victor. Isolation of components due to intracortical processing in the visual evoked potential. *Proceedings of the National Academy of Sciences*, 83(20):7984–7988, 1986.

Jonathan D Victor and Mary M Conte. Cortical interactions in texture processing: scale and dynamics. *Visual Neuroscience*, 2(3):297–313, 1989.

Jonathan D Victor and Mary M Conte. Spatial organization of nonlinear interactions in form perception. *Vision Research*, 31(9):1457–1488, 1991.

Jonathan D Victor and Mary M Conte. Local image statistics: maximum-entropy constructions and perceptual salience. *JOSA A*, 29(7):1313–1345, 2012.

Jonathan D Victor, Charles Chubb, and Mary M Conte. Interaction of luminance and higher-order statistics in texture discrimination. *Vision research*, 45(3):311–328, 2005.

Jonathan D Victor, Daniel J Thengone, Syed M Rizvi, and Mary M Conte. A perceptual space of local image statistics. *Vision research*, 117:117–135, 2015.

Michael A Webster and JD Mollon. Changes in colour appearance following post-receptoral adaptation. *Nature*, 349(6306):235–238, 1991.

Yunguo Yu, Anita M Schmid, and Jonathan D Victor. Visual processing of informative multipoint correlations arises primarily in v2. *eLife*, 4:e06604, 2015.