UNIVERSITY OF CALIFORNIA,
IRVINE

# Investigating Visual Processing: Salience Maps and Central Foveal Parvo Isoluminance

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Cognitive Sciences

by

Lingyu Gan

Dissertation Committee:
Professor George Sperling, Chair
Professor Zygmunt Pizlo
Professor Charles E. Wright

2024

# DEDICATION

To my mom, my advisor, Professor George Sperling, my best friend Fangfang, my close friend Noa, and the cuttiest dog Sabrina.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to my advisor, Professor George Sperling, for his invaluable mentorship over the past five years. When I first arrived in Irvine, I couldn't have imagined that my PhD journey would be as smooth and enjoyable as it has been. The strong, supportive relationship I have built with George has been a significant part of my positive experience here. I feel incredibly fortunate to have learned from him. Not only did I gain extensive knowledge in psychology and cognitive science, but George also imparted important life lessons, teaching me how to approach challenges with wisdom and balance. For all of this, I am deeply appreciative.

I would like to thank my committee members, Professor Zygmunt Pizlo and Professor Charles Wright, for their invaluable help and advice. Their guidance and support have been crucial to my research and academic development. I am deeply grateful for their contributions.

I would like to thank Dr. Peng Sun, who supervised me at NYU and introduced me to George. The time at NYU was one of the most turbulent periods in my life, and Peng was always there to support me. His assistance was crucial in helping me transition to UCI for my PhD. During my first year at UCI, Peng was always there whenever I needed help. I am deeply grateful for his unwavering support and guidance.

I would also like to thank my family and friends for their constant support and encouragement throughout my PhD journey. Their love, understanding, and belief in me have been a great source of strength. I am truly grateful for their presence in my life and the positive impact they have had on my journey.

Chapter 2 of this dissertation is a reprint of the material as it appears in Proceedings of the National Academy of Sciences (PNAS), used with permission from the National Academy of Sciences. The co-authors listed in this publication are George Sperling and Peng Sun. The co-author George Sperling listed in this publication directed and supervised research which forms the basis for the dissertation.

# VITA

## Lingyu Gan

### EDUCATION

**Doctor of Philosophy in Cognitive Sciences** **2024**
University of California, Irvine *Irvine, CA*

**Master of Arts in Psychology** **2017**
New York University *New York, NY*

**Bachelor of Science in Applied Psychology** **2014**
Sun Yat-Sen University *Guangzhou, Guangdong*

### RESEARCH EXPERIENCE

**Graduate Research Assistant** **2019–2024**
University of California, Irvine *Irvine, California*

**Graduate Research Assistant** **2016–2018**
New York University *New York, NY*

**Undergraduate Research Assistant** **2012–2015**
Sun Yat-Sen University *Guangzhou, Guangdong*

### TEACHING EXPERIENCE

**Teaching Assistant** **2019–2024**
University of California, Irvine *Irvine, California*

## REFEREED JOURNAL PUBLICATIONS

**Deriving the number of salience maps an observer has from the number and quality of concurrent centroid judgments** **2023**
Proceedings of the National Academy of Sciences

**Tapping ahead of time: its association with timing variability** **2020**
Psychological Research

**A mechanism of timing variability underlying the association between the mean and SD of asynchrony** **2019**
Human movement science

**Synchronization to a bouncing ball with a realistic motion trajectory** **2015**
Scientific reports

**Attentional modulations of the early and later stages of the neural processing of visual completion** **2015**
Scientific reports

# ABSTRACT OF THE DISSERTATION

Investigating Visual Processing: Salience Maps and Central Foveal Parvo Isoluminance

By

Lingyu Gan

Doctor of Philosophy in Cognitive Sciences

University of California, Irvine, 2024

Professor George Sperling, Chair

This thesis delves into the mechanisms of visual processing, focusing on salience maps and luminance. The first two chapters examine salience maps. The final chapter explores parvo isoluminance in the central fovea which became an important component in the attempt to measure foveal salience maps.

A salience map, proposed by Koch and Ullman in 1985, is a 2D topographic map that combines inputs from various feature maps to represent the combined salience at each x,y location as a real number. Salience map is mostly used to predict processing priority in the literature. Chapter 1 demonstrates that tasks such as computing distances in the frontal plane, calculating centroids (the center of a group of items), determining the numerosity of a collection of items, and identifying alphabetic letters are substance-indifferent and thus utilize salience maps. Feature-salience alone, without relying on luminance information, is sufficient for these visual judgments.

Chapter 2 investigates the number of salience maps an observer has by assessing their performance in concurrent centroid judgments. In 11 experiments, subjects viewed flashes of item arrays with 28 to 32 items, each with M different features (M = 3 to 8). Subjects then mouse-clicked the centroid of post-cued feature items. Ideal detector response analyses indicated that subjects utilized at least 12-17 stimulus items. Performance in (M-1)-feature

experiments predicting M-feature experiments led to the conclusion that one subject had at least 7 salience maps, while two others had at least 5 each. A computational model identified the primary performance limitations as channel capacity for representing multiple groups and working-memory capacity for maintaining computed centroids.

Luminance describes the visual effectiveness of a light and is typically considered the primary source for spatial visual information. Isoluminance, where different wavelength compositions have equal luminance, offers a way to evaluate the importance of luminance by studying alternative visual processing capabilities when luminance information is unavailable. Whereas isoluminance is almost universally determined under conditions that measure pure magno isoluminance, Chapter 3 used fine-gratings in the central fovea to measure parvo isoluminance. Subjects viewed high spatial-frequency yellow/red and yellow/green gratings, judging their orientation while varying red and green intensity. Results showed higher sensitivity to color than luminance in the central fovea, making foveal parvo isoluminance unmeasurable as color sensitivity overshadowed luminance sensitivity.

# Chapter 1

# Chapter 1: Salience maps for judgments of frontal plane distance, centroids, numerosity, and letter identity inferred from substance-indifferent processing

*Submitted as: Gan, L. and Sperling, G. Salience maps for judgments of frontal plane distance, centroids, numerosity, and letter identity inferred from substance-indifferent processing. [Journal of Vision]. Under revision.*

*Disclaimer:*

1. *Some of the methods in this chapter were adapted from Chapter 2, which has been previously published as: Gan, L., Sun, P., & Sperling, G. (2023). Deriving the number*

*of salience maps an observer has from the number and quality of concurrent centroid judgments. Proceedings of the National Academy of Sciences, 120(21), e2301707120.*

2. *This chapter was previously submitted for preliminary originality checks using Turnitin integrated with the "Test Student" feature in Canvas. As a result, it may show a high similarity index due to the possible inclusion of this work in Turnitin's database. The content is entirely original and authored by me.*

# Abstract

A salience map is a topographic map that has inputs from many different feature maps and summarizes the combined salience of all those inputs at each x,y location as a real number, salience, which is represented in the map. Of the more than one million Google references to salience maps, nearly all use the map for computing the relative priority of components of visual images for subsequent processing. We observe that salience processing is an instance of "substance-indifferent" processing, analogous to household measuring cups, weight scales, and measuring tapes, all of which make single-number substance-indifferent measurements. Like these household devices, the brain also makes substance-indifferent measurements but by a different mechanism: salience maps that collect visual substances for various subsequent measurements. Each salience map can be utilized by many different measurements. The instruction to attend is implemented by increasing the salience of the to-be-attended items so they can be collected in a salience map and then further processed. Here we show that, beyond processing priority, the following measurement tasks are substance indifferent and therefore use salience maps: Computing distance in the frontal plane, computing centroids (center of a group of items), computing the numerosity of a collection of items, and identifying alphabetic letters. We painstakingly demonstrate that defining items exclusively by color or texture not only is sufficient for these tasks, but that light-dark luminance information significantly improves performance only for letter recognition. Obviously, visual features are represented in the brain but feature-salience alone is sufficient for these four judgments.


## 1.1 Introduction

The concept of a saliency map was first proposed by Koch and Ullman [1] Their saliency map was a topographic map that had inputs from many different feature maps and summarized

those inputs at each x,y location as a single real number, now commonly called salience, which was represented in the map. The concept of a salience map was initially elaborated by Itti et al. [2, 3]. Of the more than one million subsequent Google references to salience maps, nearly all use the map for computing the priority of components of visual images for subsequent processing. In contrast, Lu and Sperling [4] proposed that a salience map was used to compute the motion direction of their complex visual stimuli because all simpler computations were excluded. Subsequently, visual centroid (center of mass) judgments [5, 6] and planar distance judgments [7] were reported to utilize salience maps.

Here we aim (1) to clarify the concept of a salience map, (2) to demonstrate that estimates of planar distance, centroids, numerosity, and letter shape utilize salience maps, and (3) to lay to rest the notion that luminance information (light-dark) is required for any of these tasks versus variations only in color or texture. The conclusion is that salience is a critically important brain process for representing information that is independent of the particular features that happen to carry the information to the brain.

### 1.1.1 Substance indifference

Substance-indifferent measurement is a measurement that is indifferent to the substance being measured. In the kitchen, consider these three common substance-indifferent measuring tools: measuring cups, weight scales, and measuring tapes. A measuring cup tells us how much substance it contains, but a measuring cup is indifferent to what the substance is. A cup could contain water or milk, rice or sugar, sand or nails, or any mixture. The defining feature of measuring cups, weight scales, and tape measures is that they don't know what they are measuring, but they provide a real number output that describes the amount.

Because the brain will encounter an indefinitely large number of different visual substances, it needs to be able to make substance-indifferent measurements. However, an important

difference between the brain and the household measuring devices above is that the brain appears to be able to use the same or very similar representation for many different measurements. The analogy would be that we collect the substance to be measured in a container, the salience map, and then empty it onto a cup, weight scale, or a ruler for the computation of volume, weight, or length. This distinction between the representation of salience in a salience map and the subsequent computation of priority is essential in computational models of priority processing [3], has been observed in the brain [8], and it recurs in computational models that propose salience processing for many different tasks [4, 5, 7, 9, 6].

For the measurement tasks considered here, the brain's substance-indifferent measurements require (1) isolating the particular substance to be measured, a process called grouping, (2) representing the group in a salience map so that it can be measured, (3) making the measurement, and (4) associating the measurement with the group identity. These processes are considered in detail elsewhere [5, 6]. Here we are concerned primarily with demonstrating that four additional perceptual tasks, beyond calculating processing priority or motion direction, are solved by substance-indifference processes, and therefore a salience map is an essential intermediate step in each of the four solutions. Luminance processing is sometimes–misguidedly in our opinion–proposed as an alternative to salience processing. Therefore, we take great care to also demonstrate a version of each task that is impervious to luminance processing and requires a higher-level salience process.

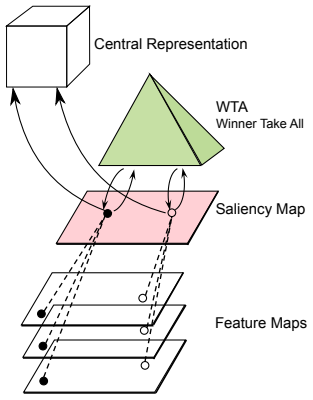



Figure 1.1: (a) The original saliency processing system of Koch & Ullman [1], colors added.

Figure 1.2: Three household substance-indifferent measuring devices: (a) Measuring cup. (b) Weight scale. (c) Measuring tape.

### 1.1.2 Four substance-indifferent judgments

To compute the distance between two target items in the frontal plane, to estimate the center of mass of a set of spatially distributed target items, to estimate the number of target items, or to identify a letter image, the only information required is the locations occupied by the target items. Insofar as these judgments utilize salience maps that record the locations of targets, the accuracy of these judgments should be independent of the particular features used to define targets, provided that the features contain sufficient location information.

## 1.2 General Methods: Substance-mixed paradigm

If frontal plane distance judgments, centroid judgments, and numerosity estimates, utilize saliency maps, then we predict that judgment accuracy can be independent of wide ranges of the substance composition of the targets. To demonstrate substance indifference, we utilize two versions of a substance-mixed paradigm.

In Version I substance mixture paradigm, target items defined by a particular feature are interleaved with distracter items that share nearly all features with the targets. The required judgment has to be made on the basis of a single distinguishing feature. As there is a virtual infinity of potential distinguishing features, the task requires a salience map representation that is independent of the particular critical feature. These tasks also require an "attention filter" to filter out distracter items so as to allow only target items to be represented in a

salience map and thereby to reach subsequent measurement stages [5, 6]. Experiments. 1a, 1b, 2b, and 3b use Version I.

In Version II, there are no distracters, instead different target items are composed of different features so that the only thing target items have in common is being distinguished from the background, i.e., their salience. Subjects have to make a judgment that incorporates all the items on the screen. Expts. 2a and 3a used Version II. The reason for some of these complications is in addition to demonstrating substance indifference, there is at least one condition for each of the tasks that absolutely rules out the possibility that luminance information could be used to solve that task. Details of the apparatus and procedures are given in the Appendices.

## 1.3 Experiment 1a & 1b: Frontal plane distance judgments

### 1.3.1 Procedure and stimuli

Experiment 1a seeks to demonstrate that in judging the distance between two stimuli in the frontal plane, it does not matter whether the stimuli are similar or different, all that matters is knowing their location. Experiment 1b extends these results to stimuli that are invisible to the visual luminance system. In both Expt. 1a and 1b, subjects viewed a computer screen that displayed a fixation cross. A key press was followed in 0.5 sec by a 200 msec exposure of a stimulus containing two target disks followed immediately by a random masking field to terminate visual persistence. Subjects typed their estimate of the separation of the two targets in tenths of inches, and full feedback was provided after each trial. The background was either filled with 142 distracter disks equal in size to target disks but of various shades of

gray to conceal colored targets from a luminance-dependent system, or in control conditions on plain gray backgrounds to maximally expose the two targets.

In Expt. 1a, 15 different substance compositions were tested in a mixed list design, 7 target pairs were identically composed, and 8 pairs were differently composed (shown in the abscissa of Fig. 1.3f).

In Expt. 1b, to demonstrate that luminance is not required for frontal plane distance judgment, targets were either isoluminant to the background or not. More specifically, three colors were used to define targets, red which is approximately isoluminant with the background, green which is approximately isoluminant with the background, and bright white which differs greatly in luminance from the background. These three colors resulted in 6 different pairs of targets.

Figure 1.3i shows two sample stimuli. In 5 out of the 6 conditions, at least one target was isoluminant to the background. If distance judgment requires luminance, it would be impossible for subjects to identify the target defined by isoluminant red or green. The isoluminant colors were calibrated individually for each subject before Expt. 1b. Again, the 142 background disks varied in luminance so that luminance was useless for identifying the nominally isoluminant red or green disks, but luminance was intentionally a vivid cue for identifying the bright white disks.

### 1.3.2 Results

Compared to egocentric distance perception, frontal plane visual distance judgments are rarely studied. Only one paper on frontal plane distance judgments was found in our Google Scholar search. Michael Cook [10] tested the scalability of interval-ordering judgments of distance along a planar horizontal surface and found that the form of the distance scale was

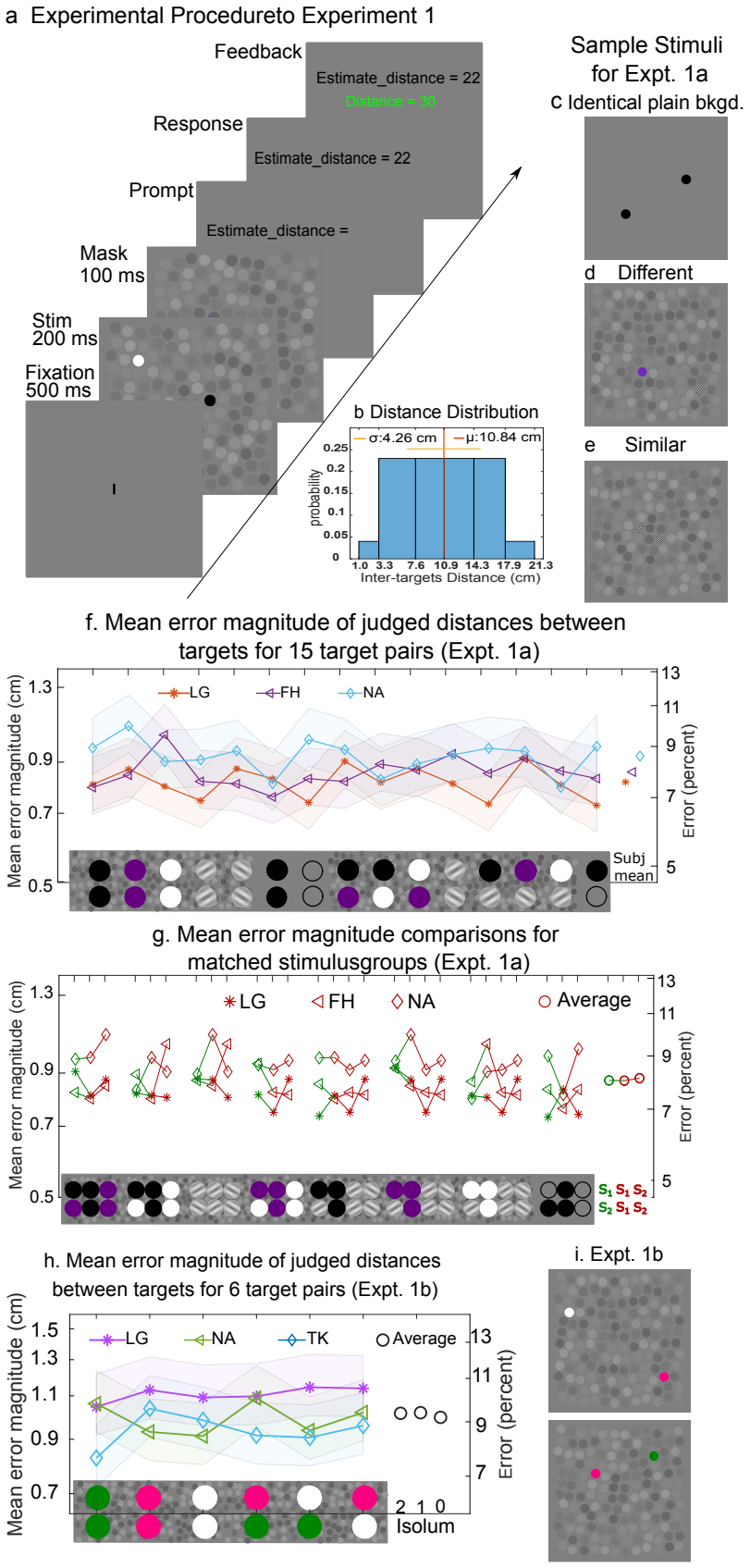



Figure 1.3: Procedure, experimental conditions, sample stimuli, and results for the distance judgments in Expts. 1a and 1b.

(a) Procedure: Each trial began with a 500ms blank field with a fixation bar, a 200ms stimulus display, and a 100ms post-exposure masking field. Subjects were then prompted to enter on the keyboard their estimate of the distance between the two targets. Feedback was provided after each trial.(b) The distribution of the distance between the two targets. (c,d,e) Sample stimuli for the distance judgments of Expt.1 (c) Two black disks. (d) A grating patch and a purple disk. (e) A clockwise-oriented grating patch and a counterclockwise grating patch. (f) Mean error magnitudes of three subjects' distance judgments between 15 types of target pairs (shown at the bottom). The subjects' overall mean error magnitude of 0.87 cm corresponds to an average Weber Fraction of 0.080. The colored area around the data represents a 95% confidence interval. The right-side ordinate is the percentage error of the estimated distance. (g) Groups of 3 and of 4 matched pairs of targets, comparing distances estimated between identical-target pairs (homogeneous, red) with different-target pairs (heterogeneous, green). The last group (extreme right) represents the data averaged over the 8 stimulus groups and the 3 subjects. (h) Three subjects' mean judged distance errors for target pairs shown at the bottom. Targets in red and green were approximately isoluminant to the background. (Distracter disks shown small here actually were same size as stimulus disks.) The open circles above 2,1,0, represent data averaged over the 3 subjects over target pairs that consist of 2, 1, or 0 isoluminant target disks. (i) Two sample stimuli for Expt. 1b.

a power function. A preliminary study [7] that is greatly elaborated here was the first to propose a salience mechanism to account for distance judgments in the frontal plane.

Subjects' performance was measured in terms of the response error – the absolute difference between the true inter-targets distance and subject's judged inter-targets distance. The root mean square error averaged over three subjects and conditions was 1.10 cm, accounting for 93.0% of the variance of the stimuli. The mean error magnitude averaged over subjects and conditions was 0.87 cm. No statistical difference in mean error magnitude was found among the 15 experimental conditions ($F(14, 28) = 0.9824, p = 0.4944$), demonstrating that the accuracy was invariant across conditions. Even in the most difficult trials (condition 11), when the target pair was composed of two differently orientated gratings and the mean luminance of the two target gratings was the same as the mean luminance of the background distracters, judgments were no less accurate than in the easiest trials, where the target pairs were two solid black discs on a blank gray background (condition 6) ($F(1, 2) = 37.76, p = 0.0255$).

For each subject, to further examine whether there was a difference in response accuracy for identically composed disks versus two differently composed disks, 15 experimental conditions were grouped into eight groups (Fig. 1.3g). Five of the eight groups had three pairs of targets, two of which were identical-target pairs; the remaining pair was a mixed-target pair, one target from each of the identical pairs. The remaining three groups each had four pairs of targets. Three of the four pairs were identical target pairs, the fourth pair was a different-target pair whose one target was randomly chosen orientation of grating, and the other target was from the third pair of identical targets. No statistical difference in mean error magnitude was found between identical-target and matched different-target pairs of elements within each group, indicating that there is no advantage in judging identically composed target pairs.

Results for Expt. 1b are shown in Fig. 1.3h. Again, the accuracy was invariant across conditions and no statistical difference in mean error magnitude was found among the 6 experimental conditions ($F(5, 10) = 0.4990, p = 0.7707$). Luminance (versus color or pattern) information was neither necessary nor advantageous, nor was there any advantage for color or pattern similarity for these frontal plane distance judgments: perfect substance indifference.

## 1.4 Experiment 2a & 2b: Centroid judgments

Experiment 2a seeks to demonstrate that judgments of the centroid of a group of items can be substance indifferent, that is, indifferent to the variety of features of which the items are composed. Experiment 2b compares (a) centroid judgments of disks that cannot be distinguished by luminance from numerous surrounding distracter disks with (b) centroid judgments of black and/or white disks that are very obviously distinguished by luminance.

### 1.4.1 Procedure and stimuli

In Expt. 2a, substance-mixed paradigm version II was used. Figure 1.4a depicts the trial procedure. Subjects were presented with a stimulus display containing 16 stimulus items for 300 msec, immediately followed by a 50 msec blank field, and a 100 msec masking field. Subjects were instructed to judge the centroid of all the stimulus items, and feedback was provided after each trial (Fig. 1.4a). Stimulus items were either identically composed or differently composed. In total, 8 different compositions of stimulus items were tested, each in a mixed list (Fig. 1.4b).

In Expt. 2b, to exclude the confounding variable of luminance, substance-mixed paradigm version I was used. Specifically, each stimulus display contained 144 same-size disks, 8 targets, 0 or 8 foils, and the remainder were distracters. Distracters were of the same size as targets and of varying luminance so that colored targets could not be distinguished from distracters on the basis of luminance (Fig. 1.4e-p). Eight combinations of targets and distracters were tested. In four conditions one or two of four colors were used to define targets (Fig. 1.4q): vivid red that is approximately isoluminant to the background, vivid green that is approximately isoluminant to the background, maximally white, and maximally black (Figs. 1.4e,f,h,i,k,l,n, and o). For the other four conditions, each stimulus display contained 128 varied gray-level disks, 8 paired salient distracters, and 8 targets (isoluminant red and isoluminant green was a pair; maximally white and maximally black was a pair; Figs. 1.4j, i,m, and p). These eight experimental conditions were tested in a mixed list. Additionally, two different sizes of disks (large 40x40, small 20x20 pixels) were examined in a blocked design, yielding a 2x8 experimental design.

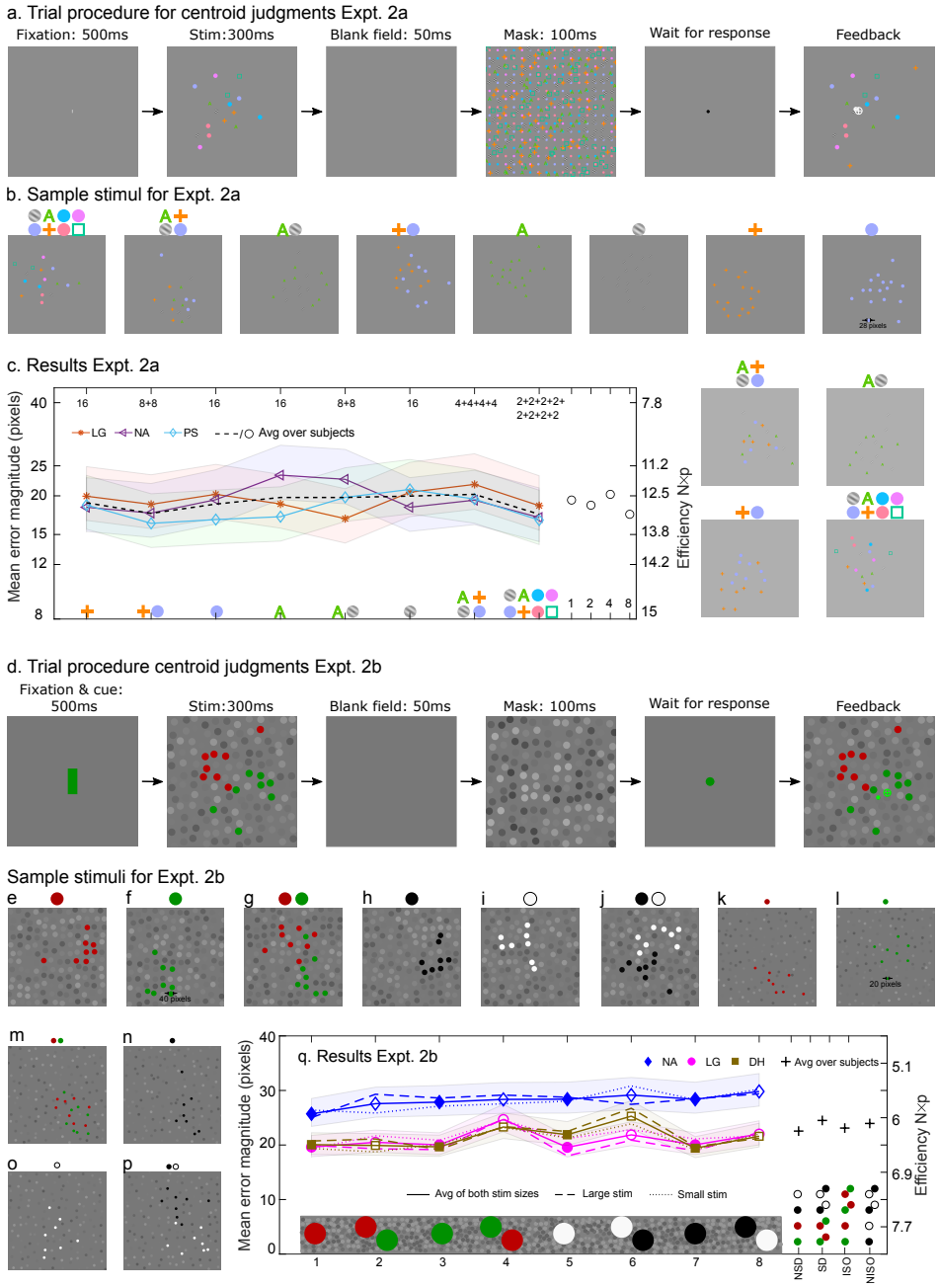



Figure 1.4: Procedure, sample stimulus displays, and results for centroid judgments, Expt. 2a and 2b. (a) Trial procedure for Expt. 2a. Each trial began with a 500ms blank field that contained a fixation point that indicated the to-be-attended color. It followed by a 300ms stimulus, a 50ms blank field, a 100 msec mask, a blank field with a movable cursor that the subject moves to the judged centroid location (centroid of all items), and finally, a feedback display. Feedback showed the stimulus, the centroid of all stimulus items (a large gray plus sign inscribed in a gray open circle), and the subject's response (a smaller gray plus sign inscribed in a gray circle). (b) Sample stimuli for Expt. 2a. To make these sample stimulus items more visible, the gray level of the background of sample stimuli above panel (c) is approximately 77% darker than the actually presented background. The four sample stimuli to the right of panel c show the actual stimulus gray levels.

(c) Results Expt. 2a: Three subjects' mean error magnitudes of judged centroids of all 16 target items for the eight experimental conditions shown below. The numbers at the top of the figure are the number of items per substance in each experimental condition (shown at the bottom). The colored area around the data represents a 95% confidence interval. The right-side ordinate is the corresponding efficiency–the number of stimulus items that an ideal observer has to perfectly process in order to match a subject's performance. (d) Trial procedure for Expt.2b. Everything was the same as Expt. 2a, except that the fixation bar also serves as a pre-cue, indicating which centroid to report. (e-j) Sample stimuli of large disks in Expt. 2b. (i-p) Sample stimuli of small disks in Expt. 2b. (q) Three subjects' mean error magnitude in their centroid judgments for the 8 experimental conditions in Expt. 2b shown at the bottom. Targets in red and green were isoluminant with the background. On the bottom, the left symbol in a pair was the target. Dashed lines represent small stimuli, dotted lines represent large stimuli, and solid lines represent the average of small and large stimuli. The colored area around the data represents a 95% confidence interval of the average of small and large stimuli. "NSD" (No Similar Distracters),"SD" (Similar Distracters), "ISO" (Isoluminant), and "NISO" (Not isoluminant) represent data averaged over 3 subjects and the 4 indicated conditions.

### 1.4.2 Results

Subjects' performance was measured in two ways: *response error*–the difference between the true centroid location of the target feature and the subject's mouse-click response, and *efficiency*–the minimum number of target items that an ideal detector needs to match a subject's mean error.

In Expt. 2a, if centroid judgments were not substance indifferent, then the compositions in which all targets have the same feature would be the easiest trials; whereas composition with 8 different features would be the most difficult. In fact, all three subjects performed both very well and remarkably similarly in all 8 experimental conditions (Fig. 1.4c). The overall mean error magnitude (the distance between the true centroid location of the stimulus items and subject's mouse-click response) averaged over three subjects was 19.06 pixels, which was less than the diameter of a stimulus item (= 28 pixels) in the stimulus array. An ideal detector would require perfect position knowledge of 12.5 of the 16 stimulus items to match the subjects' average accuracy. No statistical difference in mean error magnitude was found

among the eight experimental conditions ($F(7,14) = 0.9879, p = 0.4778$), demonstrating that the accuracy was invariant to the composition of the stimulus item. Also, no statistical differences in mean error magnitude were found between identically composed and matched differently composed stimuli, indicating that there is no statistically significant advantage in judging identically composed stimulus items ($F(2,4) = 3.0908, p = 0.1543$ for plus signs, purple disks and plus signs & purple disks; $F(2,4) = 0.0056, p = 0.9942$ for letter A, Gabor patches and letter A & Gabor patches).

In Expt. 2b, as shown in Fig. 1.4q, the two sizes of stimulus items produced statistically indistinguishable data, so the data of the large and small stimulus items were combined for the following analyses. All three subjects performed very well (Fig. 1.4q). The mean error magnitude averaged over three subjects and all conditions was 23.52 pixels. An ideal detector would require perfect position knowledge of 6.14 of the 8 stimulus items to match this accuracy. The clear result of this experiment is that there is no difference in judging centroids between the four isoluminant and the four maximally obvious luminance conditions ($F(1,2) = 2.0227, p = 0.2909$).

## 1.5 Experiment 3a & 3b Numerosity estimates

There are several approximate regimes of number estimation (e.g., Anobile, Cicchini, and Burr [11]: (1) subitizing, for less than about four items; (2) number estimation, for more than four items but items are distinguishable as unique items; (3) texture mechanisms for items that are densely packed so that items cannot be segmented from each other. Experiment 3a seeks to demonstrate that in the middle range of 9-27, estimating of the number of items in a briefly flashed display, it does not matter whether the items are similar or different; all that matters is knowing their presence. Experiment 3b compares number estimations of stimuli that are invisible to the luminance system with numerosity estimates of highly visible black

and white stimuli.

### 1.5.1 Procedure and stimuli

The number of target items ranged from 9 to 27 for Expt. 3a and from 5 to 13 for Expt. 3b. For both Expts. 3a and 3b, the areas covered by the stimulus items were kept constant for different numbers of items, i.e., the density of items covaried with numerosity. Evidence [12, 13] suggests that perceived numerosity is independent of the density of stimulus items. The procedure and stimuli in Expt. 3a were similar to Expt. 2a, except for the following: (1) the task in Expt. 3a was to estimate the numerosity, and always of ALL stimulus items; (2) the number of items of a particular feature was randomly determined for each trial with the constraint that each feature occurred at least once. Within a session, for each number composition, the total number of items per feature was the same for every feature.

Expt. 3b was identical to Expt. 2b except that (1) Only one item size was tested, the large one; (2) the number of items per feature was randomly drawn from a discrete uniform distribution between 5 and 13. As in Expt. 3a, the task was to estimate the numerosity of the items being cued. Eight experimental conditions were tested in a block design.

### 1.5.2 Results

Response error magnitude, the absolute value of the numerical difference between subjects' estimate and the true numerosity, was used to evaluate subjects' performance. In Expt. 3a, all three subjects performed very well in all 7 experimental conditions with an average error of less than one item (Fig. 1.5c). No statistically significant difference in mean error magnitude was found among the seven experimental conditions ($F(6, 12) = 1.3929, p = 0.2938$). Responses accuracy of two conditions with identically composed targets were not statisti-

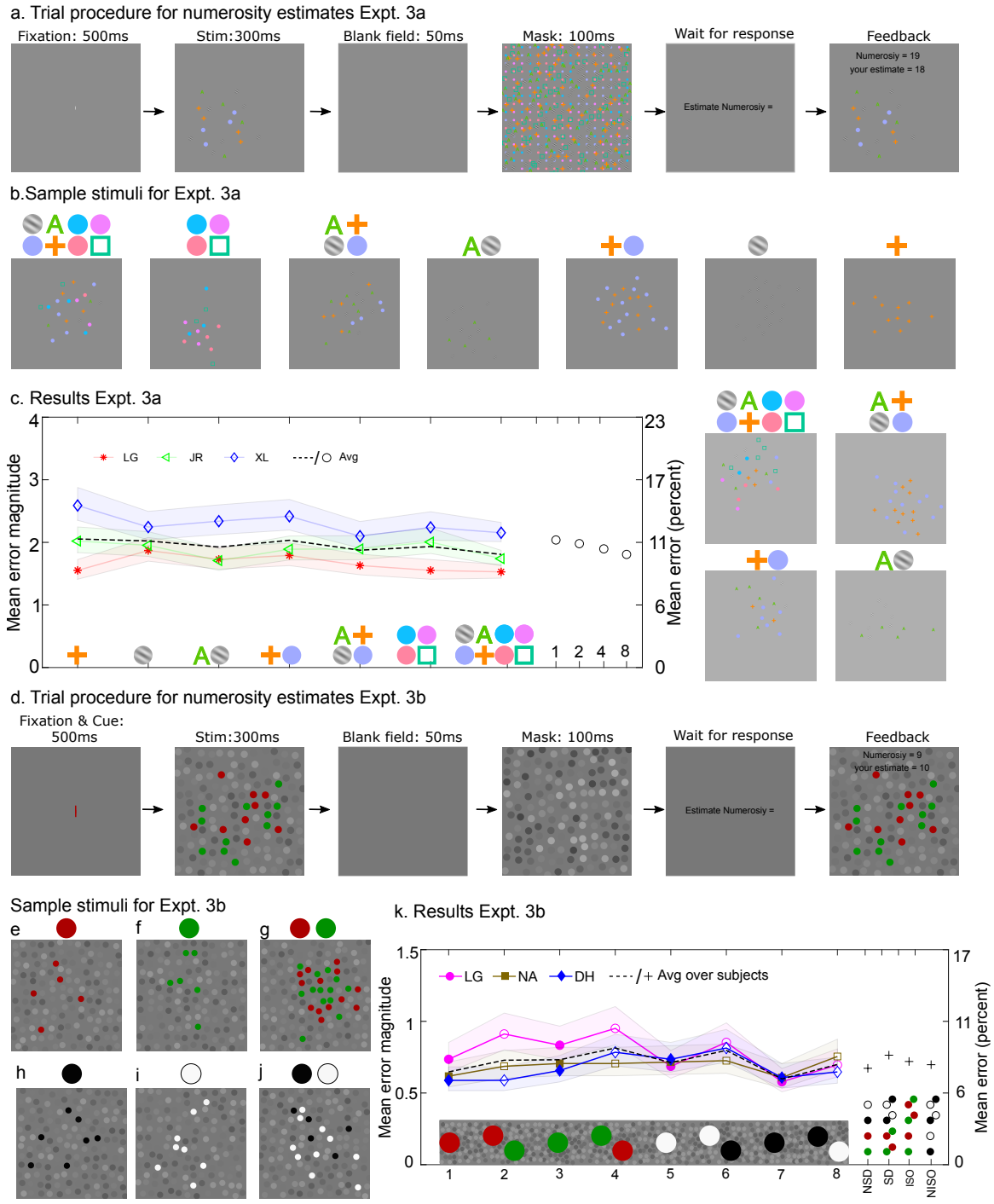



Figure 1.5: Procedure, sample stimulus displays, and experimental results for numerosity estimation, Expt.3a and 3b. (a) Trial Procedure for Expt. 3a. Every trial began with a 500ms blank field with a fixation point, followed by a 300ms stimulus, a 50ms blank field, a 100 msec mask, a field that prompted the subject to type a one or two-digit estimate of the total number of stimulus items, and finally, a feedback display. The feedback display showed the stimulus, the subject's estimate, and the numerosity of the stimulus. (b) Sample stimuli. To make the stimulus items more visible in this reproduction, the gray level of the background of the sample stimuli is darker than the actual background. The four sample stimuli with lighter gray backgrounds show what the stimulus items look like in the experiment.

(c) Results, Expt. 3a. Mean error magnitude of the judged numerosity of stimulus items in seven experimental conditions for three subjects. "1,2,4, and 8" at the far right of the abscissa indicate the experimental conditions in which target items are composed of 1, 2, 4, or 8 different features. The colored area around the data represents a 95% confidence interval. The right-side ordinate is the corresponding error fraction: (mean error magnitude)/(mean number of stimulus items). (d) Trial procedure for Expt. 3b. The task was to estimate the number only of those items that have the same color as the fixation bar. (e-j) Sample stimuli for Expt. 3b. (k) Results: Mean error magnitude of numerosity estimates for three subjects in the 8 experimental conditions shown at the bottom. On the bottom, the left symbol in a pair was the target. "NSD" (No Similar Distracters),"SD" (Similar Distracters), "ISO" (Isoluminant), and "NISO" (Not isoluminant) represent data averaged over 3 subjects and the 4 indicated conditions. The colored area around the data represents a 95% confidence interval. The right-side ordinate is the corresponding error fraction: (mean error magnitude)/(mean number of stimulus items)

cally different from five conditions with differently composed targets ($F(1,2) = 16.3073, p = 0.0562$), although there seems to be a trend for slightly better numerosity judgments with more diverse stimuli. These numerosity estimates were salience-based, not feature-specific. Figure A.1 showed three subjects' pooled numerosity estimates versus the presented number of stimulus items in each experimental condition.

Results Expt. 3b. No statistical difference in mean error magnitude was found among the eight experimental conditions ($F(7,14) = 2.6551$, $p = 0.0575$) that included four isoluminant and four luminance-based numerosity judgments. All three subjects were slightly more accurate in conditions with non-salient distracters than in conditions with salient distracters. This trend was not statistically significant ($F(1,2) = 9.3233, p = 0.0926$ and has no bearing on the issue of luminance-based versus isoluminance-based numerosity judgments. Most relevant: There is no difference in numerosity estimation between the four isoluminant and the four maximally obvious luminance conditions ($F(1,2) = 0.2040, p = 0.6947$).

## 1.6 Letter identification depends on the salience of the substance of which the letters are composed

This section shows examples to demonstrate that letter identification depends on the salience of the substance of which the letters are composed, not necessarily their luminance as is often assumed.[1].

Figure 1.6 shows eight letters each of which is composed of a different substance. The luminance of letters A, B, and L is very different from the background; whereas on a calibrated screen, the luminance of letters C, E, D, H, and K equals the luminance of the background. Isoluminant or not, all eight letters are identifiable.

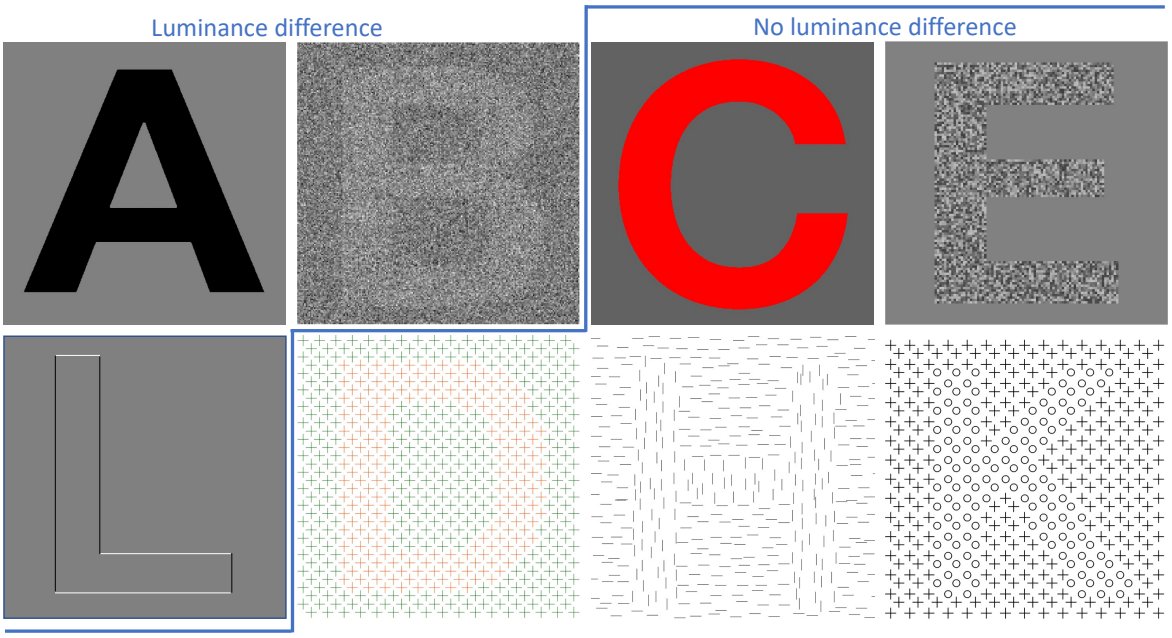
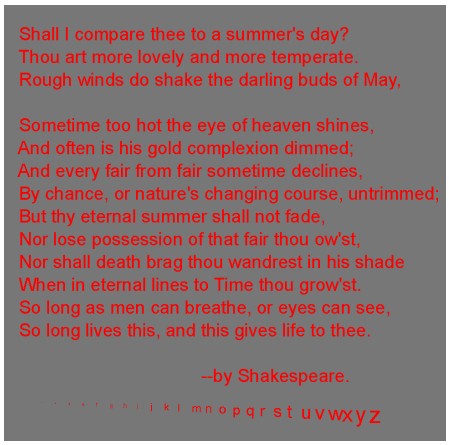



Figure 1.6: Eight letter examples. Letters A, B, and L have luminance cues to shape; on calibrated displays, the other five letters are isoluminant with the background.

[1]Preliminary results were presented at the Vision Sciences Society, 2022. [14]

a. Letters of 14 different colors and of 6 different sizes

b. Shakespeare's Sonnet 18




c. Letters of 25 different contrasts and of 6 different sizes

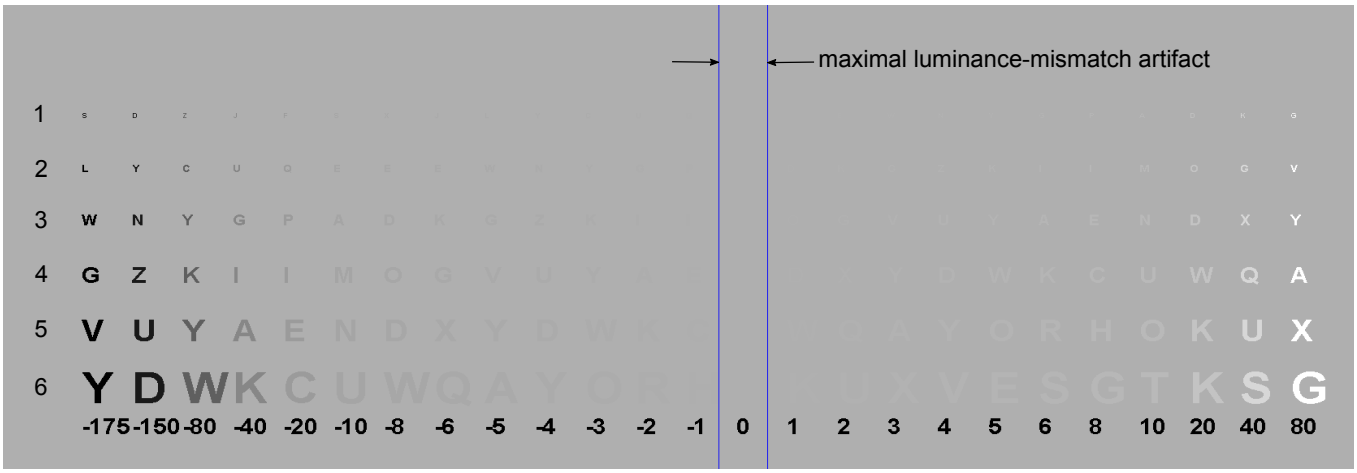



Figure 1.7: (a) Letters of 14 different colors and of 6 different sizes on a gray background. Letters in the same row have the same size, and letters in the same column have the same color. Except for black letters in the first column and white letters in the last column, other letters are iso-luminant with the background on a calibrated display. In supplementary Demonstration 1, the background intensity of Fig. 1.7a can be varied on a display device to find the background intensity value that produces isoluminance (minimum visibility) for any particular letter for that observer. The larger letters always remain visible in this demonstration. (b) A nominally isoluminant text version of Shakespeare's Sonnet 18. (c) Letters of 25 different luminances and 6 different sizes on a background with a gray level of 175 (range 0, 255). Letters in the same row have the same size and letters in the same column have the same luminance. The numbers in the bottom row represent the nominal differences in luminance between letters and the 175 background. The actual contrasts in Fig. 1.7b depend on the Gamma function of the monitor on which the figure is viewed. For typical Gammas of 2.0 to 2.4, the contrast of one unit (bottom row of Fig. 1.7b) is 1.15 to 1.38%.

Figure 1.7a shows letters of 14 different colors and of 6 different sizes on a gray background. Letters in the same column have the same color. Except for the black and white letters (the leftmost and the rightmost columns), on a calibrated display, other letters are more-or-less isoluminant with the background. On a dynamic version of this display, the background intensity is variable so that the viewer can find the intensity that best conceals any particular letter color. The main observation: From viewing monitors or journals at a distance of a meter or so, appropriate adjustment of the background luminance can conceal the smallest letters but larger size letters of all colors always remain visible.

Figure 1.7b shows a nominally isoluminant text version of Shakespeare's Sonnet 18. The journal version may or may not be isoluminant on any particular display. However, in a dynamic version of this sonnet, the background intensity is continuously variable to enable search for a background intensity that maximally conceals the text. In our experience, there is no intensity that fully conceals the text, although some viewers may have to approach closer to perceive a fully readable display. Note that Fig. 1.7b shows that for small low-contrast black-and-white letters this display would be invisible. Isoluminant perception, which depends on photon differences between quite similar long- and medium-wavelength receptors, requires many more photons than luminance perception which depends on the sum, long- plus medium-wavelength receptors, versus zero.

Figure 1.7c shows letters of 25 different luminance levels and of 6 different sizes on a gray background. The luminance table contains 256 gray levels, ranging from 0 (maximally black) to 255 (maximally white). The numbers in the bottom row indicate the nominal differences in luminance between letters and the reference background. The actual contrasts depend on the Gamma functions of the display device. Figure 1.7c demonstrates, as noted by Legge, Rubin, and Luebker [15], that low contrast letters are surprisingly difficult to decipher. That also applies to luminance artifacts that are occasionally alleged to account for the perception of nominally isoluminant chromatic letters, e.g. Knoblauch, Arditi, and Szlyk [16].

While the proposal that salience is an alternative mechanism for letter recognition is new, the fact that luminance is not necessary has a long history. Two examples: Legge, Parish, Luebker, and Wurm [17] found that reading speed increased quite similarly as a function of luminance and of color contrast and concluded, obviously, that luminance is not necessary for letter recognition. Wurm, Legge, Isenberg and Luebker [18] found that both normal and low vision subjects, had faster reaction times to recognize objects when the objects were displayed in color compared to the grayscale version. The point of the Figs. 1.6 and 1.7 demonstrations is not merely that either luminance or color information is sufficient for the perception of letter shape or for reading: It is that stimulus features that enable salience representations of sufficient resolution for distance estimations, centroid computations, and numerosity estimations can also serve letter shape recognition and reading.

## 1.7 Discussion

The main purposes of this paper are to clarify the concept of a salience map, to demonstrate that distance judgment in the frontal plane, centroid judgments, numerosity estimations, and letter recognition utilize salience maps, and to demonstrate that luminance is not necessary for any of these brain computations. Here we further consider the following: Logic requires salience maps. For three tasks there was no loss of accuracy for judgments based on salience alone without luminance information. More generally, luminance is a source of information for salience processes, is a feature of most objects, but luminance is not how the information about the objects–at least for the objects studied here–is represented in the brain.

### 1.7.1 Logic requires salience maps

Though it seems effortless for us to judge the distance between two target items, it requires quite complex and extensive neural circuitry to compute the distance between items that are arbitrarily located in the visual field. It would be overwhelmingly expensive to have a neural circuit for estimating the distance between two black disks, a different neural circuit for estimating the distance between two Gabor patches, or yet another neural circuit for estimating the distance between a Gabor patch and a black disk. The distance computation between two target items has to be made on an internal representation of the target items–a representation that depicts the existence of objects at various locations that is distinct from the representation of the objects. The same logic applies to centroid judgments, numerosity, and letter recognition. For letter recognition, Figs. 1.6 and 1.7 clearly showed that letter recognition depends on the salience of letters. In Fig. 1.6, the letters C, E, D, H, and K, and in Fig. 1.7, the large letters of 12 different colors are isoluminant with the background but nevertheless effortlessly recognized, indicating that luminance is a useful but not a necessary cue for shape recognition. The four tasks described herein (distance, centroid, numerosity, letters) were designed to utilize a flexible spatial representation that is not dependent on any particular feature, neither luminance nor color–just something that distinguishes the targets from the background. Salience that is recorded in salience maps is the name assigned to this property, it is the refined descendent of figure-ground.

### 1.7.2 No loss of accuracy for judgments based on salience.

For Expt.1a, Expt. 2a, and Expt. 3a, the compositions of targets varied within and between different feature domains. As all compositions were randomly mixed in each experiment, subjects did not know what the targets would look like ahead of a trial. If the judgments are made based on features rather than salience, one would expect subjects' performance to vary

with the compositions of targets. However, in all three experiments, subjects' judgments were quite accurate for all experimental conditions and were invariant with the compositions of the targets. When judgments in mixed substance conditions are as accurate as in pure single-substance conditions, it implies that the mixed substance computation on a salience map is the normal computation, there is no better computation. In distance judgments, Expt. 1a, even the easiest possible condition, two black disks on a plain gray background, was no easier than the 8 other more difficult judgments of heterogeneous targets on backgrounds that made single-feature identification of target dots impossible. In centroid judgments, Expt. 2a, and numerosity estimates, Expt. 3a, one-feature conditions were not easier than two-, four-, or eight-feature conditions.

In Expt. 1b, Expt. 2b and Expt, 3b, targets, either isoluminant with the background or not, were presented with luminance distracters. In all three experiments, conditions with targets that were of the same luminance as the background were not harder than conditions with targets that were of vastly different luminances than the background. The results of these three experiments clearly demonstrate (1) that luminance, although very informative, is not required for these three computations. (2) even if judgments of targets having different luminance than the background were made based on a luminance map, these judgments were not better than the judgments necessarily based on a salience map.

Computations based on salience maps can be highly accurate. In judging the locations of centroids in Expt.2a, the mean error magnitude averaged over three subjects was 19.6 pixels (0.56 deg of vis angle). An ideal detector that knows the exact location of each stimulus item and computes a perfect centroid, would have to perfectly process the location 12.6 of the 16 stimulus dots to match the subjects' mean performance. This is in line with the results reported by Sun, Chubb, Wright, and Sperling [19] that following a brief display of 26 dots at least 15 dots were used for the centroid judgments. Subjects report that after a brief display that the location of the centroid appears to be available in memory, it is not

consciously computed from knowledge of the individual dot locations. This is consistent with the observation by Sun et al. [19] that in briefly flashed displays in which subjects utilized 14 of 26 dots for centroid computations, less than two dots were available in a working memory task

### 1.7.3 Letter identification depends on salience, not luminance

In Fig. 1.6a, except for the white letters in the rightmost column, other letters are isoluminant with the background on a calibrated display. The observations that isoluminant letters are more and more recognizable with the increase of the size of the letters and that greenish letters are more recognizable than purple letters demonstrate that letter identification in this display is based on a color difference between letters and background, not on luminance. Twenty isoluminant colors were used to define the letters in this example, but hundreds of other isoluminant colors could have been used to define the letters. It is not likely that for each color, there is a unique internal representation of the letter in that color. For neural efficiency, it's clear that if a letter is identifiable by a virtual infinity of different features, it must have an internal representation that only registers how different that letter is from its surround and is indifferent to the particular features that produce that difference. The defining features are also represented in the brain, elsewhere.

## 1.8 Acknowledgements

Publication costs are supported ONLY by the ARVO's Publication Financial Assistance Program.

# Bibliography


[1] Christof Koch and Shimon Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *Human neurobiology*, 4 4:219–27, 1985.

[2] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 20(11):1254–1259, 1998.

[3] Laurent Itti and Christof Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, 40(10-12):1489–1506, 2000.

[4] Zhong-Lin Lu and George Sperling. Attention-generated apparent motion. *Nature*, 377(6546):237–239, 1995.

[5] Peng Sun, Veronica Chu, and George Sperling. Multiple concurrent centroid judgments imply multiple within-group salience maps. *Attention, Perception, & Psychophysics*, 83(3):934–955, 2021.

[6] Lingyu Gan, Peng Sun, and George Sperling. Deriving the number of salience maps an observer has from the number and quality of concurrent centroid judgments. *Proceedings of the National Academy of Sciences*, 120(21):e2301707120, 2023.

[7] Lingyu Gan, Peng Sun, and George Sperling. Frontal-plane distance judgments between two equal-size items are made on the basis of a salience map. *Journal of Vision*, 21(9):2828–2828, 2021.

[8] Carsten Bogler, Stefan Bode, and John-Dylan Haynes. Decoding successive computational stages of saliency processing. *Current Biology*, 21(19):1667–1671, 2011.

[9] Lingyu Gan and George Sperling. Centroid judgments are substance indifferent and therefore based on a salience map. *Journal of Vision*, 22(14):3832–3832, 2022.

[10] Michael Cook. The judgment of distance on a plane surface. *Perception & Psychophysics*, 23(1):85–90, 1978.

[11] Giovanni Anobile, Guido Marco Cicchini, and David C Burr. Number as a primary perceptual attribute: A review. *Perception*, 45(1-2):5–31, 2016.

[12] Lixia He, Jun Zhang, Tiangang Zhou, and Lin Chen. Connectedness affects dot numerosity judgment: Implications for configural processing. *Psychonomic bulletin & review*, 16(3):509–517, 2009.

[13] John Ross and David C Burr. Vision senses number directly. *Journal of vision*, 10(2):10–10, 2010.

[14] George Sperling and Lingyu Gan. Two-dimensional shape perception is based on a salience map. *Journal of Vision*, 22(3):17–17, 2022.

[15] Gordon E Legge, Gary S Rubin, and Andrew Luebker. Psychophysics of reading—v. the role of contrast in normal vision. *Vision research*, 27(7):1165–1177, 1987.

[16] Kenneth Knoblauch, Aries Arditi, and Janet Szlyk. Effects of chromatic and luminance contrast on reading. *JOSA A*, 8(2):428–439, 1991.

[17] Gordon E Legge, David H Parish, Andrew Luebker, and Lee H Wurm. Psychophysics of reading. xi. comparing color contrast and luminance contrast. *JOSA A*, 7(10):2002–2010, 1990.

[18] Lee H Wurm, Gordon E Legge, Lisa M Isenberg, and Andrew Luebker. Color improves object recognition in normal and low vision. *Journal of Experimental Psychology: Human perception and performance*, 19(4):899, 1993.

[19] Peng Sun, Charles Chubb, Charles E Wright, and George Sperling. High-capacity preconscious processing in concurrent groupings of colored dots. *Proceedings of the National Academy of Sciences*, 115(52):E12153–E12162, 2018.

# Appendix for Chapter 1: Detailed Materials and Methods

For all experiments, all subjects had normal or corrected-to-normal visual acuity. All subjects gave informed consent for participation in the study. All methods were approved by the University of California, Irvine Institutional Review Board.

## A1 Experiment 1a

### Subjects

LG was the senior author. FA and NA were naive about the purpose of the study.

### Apparatus

The experiment was conducted on a Mac Pro 2015 running MATLAB with a Psychtoolbox package. A 60-Hz refresh rate CRT monitor with 1280 × 1024 resolution was used to display the stimuli. Stimuli were viewed from a distance of 58 cm.

## Stimuli

The stimuli were displayed within a 720 by 720 pixels wide square (dva 17.25°) in the center of the screen. In total, 15 conditions, each with a different pair of targets, were tested. For the first 12 conditions, 144 same-size disks were presented in every stimulus display. The diameter of the disks was 50 pixels (visual angle 1.29°). The gray levels of 142 disks were drawn from a uniform distribution of intensities $U(-0.35, +0.35)$, where 0 represents mid-gray, and -1.0 and 1.0 represent the lowest and highest intensities available on the monitor. Two disks were significantly different from the rest of the disks. Their interiors were randomly drawn from one of 15 pairs of disks described in the abscissa of Fig 1.3f. Two disks were targets and the other 142 disks were distracters. For the remaining three conditions, only the two targets (no distracters) were presented. Figures 1.3c, d, and e illustrate stimulus displays. A post-stimulus mask (Fig. 3a) consisting of 144 disks whose gray levels were sampled from a uniform distribution of intensities $U(-0.5, 0.5)$ immediately followed the stimulus exposure.

## Procedure

Figure 1.3a depicts the procedure of each trial. Each trial started with a blank screen with a fixation bar for 500 ms, followed by a stimulus display shown on the screen for 200 ms. A 100ms-duration post-exposure mask field immediately followed the stimulus display to strictly control the duration for which information was visually available. Then subjects were prompted to enter their estimate of the distance between the two target disks on the keyboard. Distance estimates consisted of two integers, the digit was the number inches, the second number was the number of tenths. Subjects were allowed to change their answer before submitting it by pressing "return". Feedback was provided after each trial.

### Number of trials

Each subject participated in three sessions of the experiment, one session per day. Each session was divided into a training section and a formal experiment. In the training section, to facilitate learning, the duration of the feedback was controlled by the subjects. The training section consisted of 20 trials on each of the three training conditions, at the end of which subjects were able to estimate the distance between the targets quite accurately and without demonstrating additional improvement. In the formal experiment, feedback was shown on the screen for 1 second and the inter-trial interval was 1 second. The first two sessions contained 30 trials of each of the 15 conditions and the last session contained 40 trials of each of the 15 conditions. Experimental conditions were mixed in each session. In Expt. 1a, for three subjects, 37, 47, and 38 outlier trials—those beyond 2.5 times the standard deviation of the response error —were discarded from a total of 1,500 trials per subject.

## A2 Experiment 1b

### Subjects

Subjects LG and NA from Expt. 1a plus a new subject TK participated.

### Stimuli, Procedure and Number of trials

The stimulus and the procedure of Expt. 1b was the same Expt. 1a except for the following: (1) The composition of the target pairs. In total, 6 target pairs were tested in Expt. 1b (shown in the abscissa of Fig 1.3h). The red and green were isoluminant to the background

and were tested for each subject before the formal experiments. (2) Each subject did two sessions of the experiment, one session per day. Each session contained 50 trials of each of the six conditions. Experimental conditions were mixed in each session.

# A3 Experiment 2a

## Subjects

Subjects LG and NA from Experiment 1, plus a new subject PS, all were experienced in centroid tasks.

## Apparatus

The experiment was conducted on a Mac Pro 2015 running MATLAB with a Psychtoolbox package. For LG and NA, the stimuli were presented on an ASUS ProArt Display monitor with 1920 x 1200 resolution at a refresh rate of 60 Hz. The monitor screen was 51.8 cm wide x 32.4 cm high. Each pixel was 0.27mm x 0.27 mm. For PS, the stimuli were presented on a SamSung Syncmaster Display monitor with 1680 x 1050 resolution at a refresh rate of 60 Hz. Each pixel was 0.282 mm x 0.282 mm.

## Stimuli and Number of trials

The stimuli were displayed within an 800 by 800 pixel-wide square (Fig. 1.4b) that had a dva of 20.4° for LG and NA, and 22.0° for PS. The stimulus display contained 32 stimulus items. Each stimulus item was inscribed inside invisible circles of 28-pixel diameter, spanning 0.72 dva, that were prohibited from overlapping. The 32 stimulus items had either the same

feature or 2, 4, or 8 different features. Items varied in color, shape, and luminance, which could be less than, equal to, or greater than the background gray level (175). In total, eight different compositions of stimulus items were tested in a mixed-list design, 50 trials per composition. Figure 1.4b shows sample stimuli for each of the 8 compositions.

### Procedure

Figure 4a depicts the procedure of each trial in Expt. 2a. Every trial began with a 500ms blank field with a fixation point, followed by a 300ms stimulus array, a 50ms blank field, a 100 msec mask, a blank field with a movable cursor that the subject moved to the judged centroid location, and finally, a feedback display. Feedback displayed the stimulus, the centroid of the target set as the larger gray plus sign inscribed in a gray open circle, and the subject's response as the smaller gray plus sign inscribed in a gray open circle. In Expt. 2a, For the three subjects, 7, 4, and 3 outlier trials—those beyond 2.5 times the standard deviation of the response error —were discarded from a total of 400 trials per subject. In Expt. 2b, 7, 9, and 10 outlier trials were discarded from each subject's total of 1600 trials.

## A4 Experiment 2b

### Subjects

Subjects LG and NA as in Expt. 2a, plus a new subject DH.

## Apparatus

The experiment was conducted on a Mac Pro 2015 running MATLAB with a Psychtoolbox package. For all subjects, the stimuli were presented on the built-in Retina Display with 1280 x 800 resolution at a refresh rate of 60 Hz. The monitor screen was 30.41 cm wide x 21.24 cm high. Each pixel was 0.22mm x 0.22 mm.

## Stimuli and number of trials

The stimuli were displayed within an 800 by 800 pixel-wide square that spanned 17.86 deg of visual angle (dva) for all subjects. The stimulus display contained 144 same-size stimulus items. Each stimulus item was inscribed inside invisible circles of 28-pixel diameter, spanning 0.72 dva, that were prohibited from overlapping. Eight different combinations of targets and distracters were tested. For four conditions, each stimulus display contained 136 varied gray-level disks and 8 target disks. In each condition, one of the four colors was used to define targets, one most red that is isoluminant to the background, one most green that is isoluminant to the background, the maximally white, and the maximally black (Fig. 1.4e,f,h,i,k,l,n, and o). For the rest four conditions, each stimulus display contained 128 varied gray-level disks, 8 paired salient distracters, and 8 targets (isoluminant red and isoluminant green was a pair; maximally white and maximally black was a pair; Fig 1.4j, i,m, and p). These eight experimental conditions were tested in a mixed list with each condition comprising 100 trials. Additionally, two different sizes of disks were examined in a blocked design, yielding a 2x8 experimental design. The large disks were 40 by 40 pixels and the small disks were 20 by 20 pixels

### Procedure

Figure 1.4d depicts the procedure of each trial in Expt. 2b. Every trial began with a 500ms blank field with a fixation point that had the same feature as the target, followed by a 300ms stimulus array, a 50ms blank field, a 100 msec mask, a blank field with a movable cursor that the subject moved to the judged centroid location, and finally, a feedback display. Feedback displayed the stimulus, the centroid of the target set as the larger gray plus sign inscribed in a gray open circle, and the subject's response as the smaller gray plus sign inscribed in a gray open circle.

## A5 Experiment 3a

### Subjects

LG plus two new subjects JR and XL.

### Apparatus

The experiment was conducted on a Mac Pro 2015 running MATLAB with a Psychtoolbox package. For all subjects, the stimuli were presented on the Acer Predator XB321HK monitor with 3840 x 2160 resolution at a refresh rate of 60 Hz. Each pixel was 0.19mm x 0.19mm.

### Stimuli and number of trials

The stimuli were displayed within an 800 by 800 pixel-wide square that spanned 14.19° dva. The stimulus display contained 9 to 26 stimulus items. Each stimulus item was inscribed

inside invisible circles of 28-pixel diameter, spanning 0.50° dva, that were prohibited from overlapping. The stimulus items had either the same feature or 2, 4, or 8 different features. Items varied in color, shape, and luminance, which could be less than, equal to, or greater than the background gray level (175). In total, seven different compositions of stimulus items were tested in a mixed-list design. Figure 1.5b shows sample stimuli for each of the seven compositions. For a single session, except for the composition of 8 features that had 96 trials, all other compositions had 48 trials. All compositions have the same distribution of the number of stimulus items. Each subject completed four sessions, conducted on separate days.

### Procedure

On each trial, subjects were instructed to estimate the numerosity of all stimulus items. Figure reffig:fig5a depicts the procedure of each trial. Every trial began with a 500ms blank field with a fixation point, followed by a 300ms stimulus array, a 50ms blank field, a 100 msec mask, and a blank field with prompt cueing subjects to enter their estimate of the numerosity of all stimulus items on the keyboard, and finally, a feedback display. Feedback displayed the stimulus, the correct answer, and the subject's response.

## A6 Experiment 3b

### Subjects

Same subjects as in Expt.2b

### Apparatus

Same as Expt.2b

### Stimuli and number of trials

The stimulus in Expt. 3b was identical to Expt. 2b except the following: (1) the number of items per feature was randomly drawn from a discrete uniform distribution between 5 and 13; (2) Only the large size stimulus elements were tested; (3) The eight experimental conditions were tested in a block design, with each condition comprising 102 trials. The order of conditions was randomized for each subject.

### Procedure

The procedure was identical to Expt. 2b except the task was to estimate the numerosity of target stimulus items.

## A7 Supplementary figure

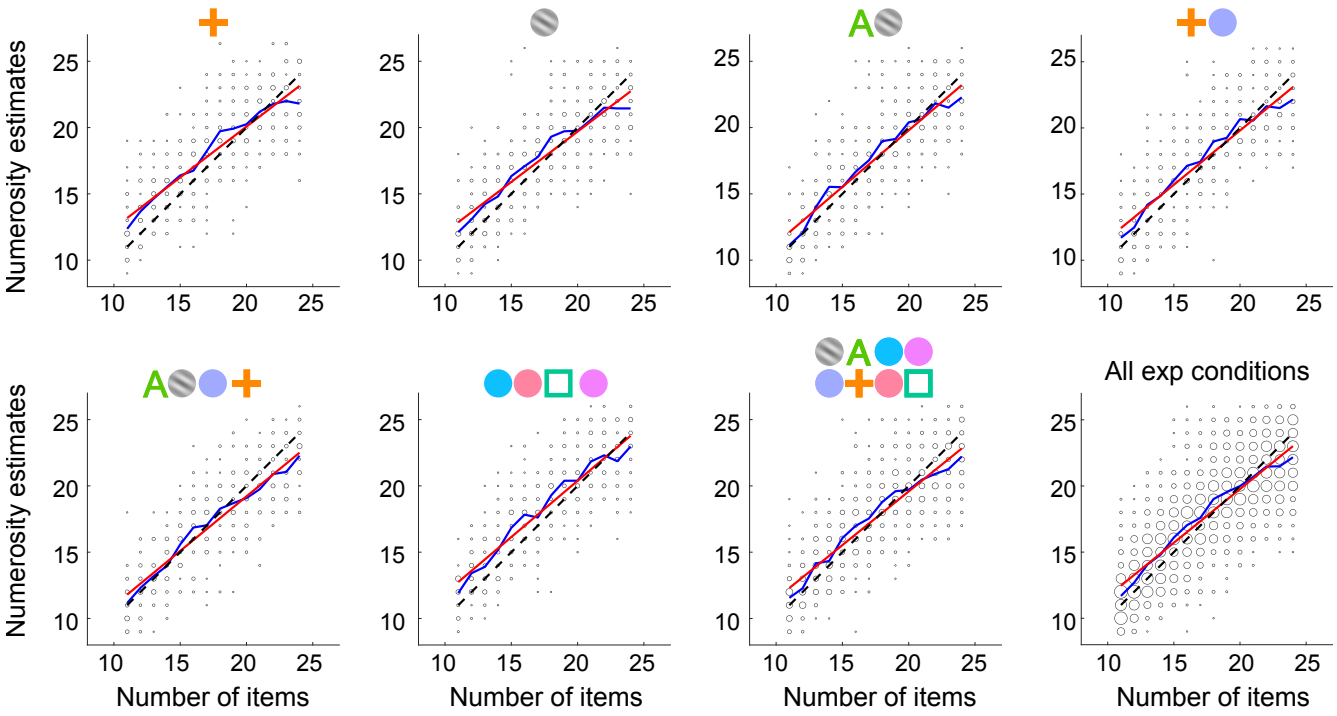



Figure A.1: Expt.3a: Three subjects' pooled numerosity estimates versus the presented number of stimulus items in each experimental condition. The size of open circles is proportional to the frequency of responses. The blue curve is the mean response for each number of items. The dashed black line is the presented number of items. The red line is the best-fitting line to the data.

# Chapter 2

# Chapter 2: Deriving the Number of Salience Maps an Observer has from the Number and Quality of Concurrent Centroid Judgements

*Disclaimer: Chapter 2 is a reprint of the paper published as: Gan, L., Sun, P., & Sperling, G. (2023). Deriving the number of salience maps an observer has from the number and quality of concurrent centroid judgments. Proceedings of the National Academy of Sciences, 120(21), e2301707120.*

# Abstract

Koch and Ullman [1] proposed a 2D topographical salience map that took feature-map outputs as its input and represented the importance "saliency" of the feature inputs at each location as a real number. The computation on the map, "Winner-Take-All," was used to predict action priority. We propose that the same or a similar map is used to compute centroid judgments, the center of a cloud of diverse items. Sun, Chu, and Sperling [2] demonstrated that following a 250 msec exposure of a 24-dot array of 3 intermixed colors, subjects could accurately report the centroid of each dot color, thereby indicating that these subjects had at least three salience maps. Here, we use a post-cue, partial-report paradigm to determine how many more salience maps subjects might have. In 11 experiments, subjects viewed 0.3 sec flashes of 28 to 32 item arrays composed of M, M=3,...,8, different features followed by a cue to mouse-click the centroid of items of just the post-cued feature. Ideal detector response analyses show that subjects utilized at least 12-17 stimulus items. By determining whether a subject's performance in (M-1)-feature experiments could/could-not predict performance in M-feature experiments, we conclude that one subject has at least 7 and the other two have at least 5 salience maps. A computational model shows that the primary performance-limiting factors are channel capacity for representing so many concurrently presented groups of items and working-memory capacity for so many computed centroids.


## 2.1 Introduction

The human visual system can extract summary statistical information from groups of similar items in a brief glance [3, 4]. Such ensemble statistics are interesting because they result from brain mechanisms that can quickly distill a large amount of sensory information for subsequent cognitive processes that have much lower capacity. Prior studies have shown

that humans can form statistical representations, for example, of size [5], spatial location [6], orientation [7], motion [8], brightness [9], hue [10], numerosity [11] and facial expression [12].

Humans are also capable of selectively forming statistics out of spatially intermingled visual stimuli defined by different features. Of particular relevance here are centroids, the mean location of a group of items. Drew et al. [6] report that in a display of interleaved light and dark dots, subjects can selectively compute different centroids for the light and for the dark dots. Sun et al. [13] report that subjects could efficiently compute the centroid of 3 target dots of a pre-cued color even when the target dots were intermixed with 21 dots of 7 other equiluminant distracter colors.

### 2.1.1 Salience maps and centroid judgments

To compute the centroid of a set of items of a particular feature, the target items need to be separated from distracter items and spatial information of the segregated target items is required. A topographical map that inherently contains the required spatial information of the target items would meet the requirement for computing a centroid of the target items. Because the brain is finite, and because there is an infinity of features and feature combinations that could define target items, it's clear that the feature locations must be represented in a topographic map that is indifferent to feature identities. This kind of topographic map was first proposed by Koch and Ullman [1] to represent the overall saliency of items (Fig. 2.1). Their saliency map has inputs from many different feature maps and summarizes the inputs at each x,y location as a single real number, saliency, which is represented in the map. The output of their saliency map was fed into a winner-take-all network that computed the location with the highest saliency. The output of the winner-take-all network could be used to predict priority in visual search and other visual processes. Koch and Ullman [1] is a

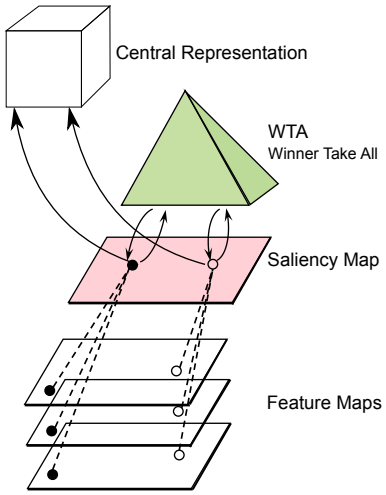



Figure 2.1: Koch and Ullman's [1] original representation of a saliency processing system. (Colors have been added.)

conceptual model. Many subsequent papers clarify, quantify, and further elaborate the original concept, e.g., [14, 15, 16, 17]. A recent Google Scholar search yielded over one million citations to salience maps, most of the relevant citations were for priority.

Sun et al. [2] proposed a similar topographic salience map to compute not a winner-take-all but rather a centroid. Their conception was that the input to a salience map is flexible, rather than hardwired and depends on the task. Sun et al. [2] reported that after a 250 msec view of a display of 24 dots, 8 dots per color, subjects were able to report three centroids, one for each color, with approximately equal accuracy. An ideal detector that knows the exact location of each stimulus item and computes a centroid perfectly would require 13.0 of 24 stimulus items to match the average subject's performance. The number of stimulus items required by such an ideal detector is the lower bound of the number of stimulus items processed by subjects. The aim of the present study is to extend these results to determine whether more than three centroids can be computed concurrently and to determine how many perceptually processed items are incorporated into these centroids. But first, even to draw the conclusion that subjects can concurrently compute three centroids, two possible confounds needed to be excluded.

1) Visual information might be available for much longer than the brief exposure to enable computation of a centroid long after the stimulus exposure was terminated. Indeed, visual stimulus continues to be visible for some time after the offset of the stimulus [18]. However, a post-stimulus mask [19] can strictly control the duration for which the stimulus is visually available. In [2], for 3 of 5 subjects, there was no post-stimulus mask following the stimulus display. Therefore, for these three subjects, visual information is available longer than the duration of the stimulus display; two subjects viewed stimuli followed by masks. No statistical difference was found between the mask data and no-mask data, suggesting that visual information available in any mask-susceptible visual memory after the offset of the stimulus display did not significantly benefit the centroid computation.

2) Suppose subjects computed only a random two of three centroids on each exposure and optimally guessed the third centroid (subsampling). If subjects were to adopt this subsampling strategy, the quality of subject's attention filters–a measure that describes the weights that subjects give to different features–when three centroids are required would be worse than the quality of subject's attention filters when only one centroid is required (subjects know which centroid to report ahead of the stimulus presentation). The subsampling confound is excluded by the observation that the quality of subject's attention filters when subjects reported three centroids was as good as the quality of subject's attention filters when subjects reported only one centroid.

The facts (1) that subjects process at least 13.0 stimulus items (implying more than 4 dots utilized times 3 centroids), (2) that there is no significant impairment in subject's attention filters when reporting three versus one centroid, and (3) that there are no significant differences in error magnitude between the three different concurrent centroid judgments demonstrate that subjects are indeed able to compute three concurrent centroids. Because a salience map is feature-independent, and a single salience map cannot discriminate colors, Sun et al [2] concluded that three salience maps are required to concurrently compute three

centroids.

### 2.1.2 Procedure Outline

The question that we investigate here is: "How many salience maps do subjects have?" Each centroid computation requires its own salience map, so we seek to determine the maximum number of centroids that subjects can compute concurrently. Sun et al. [2] required their subjects to report three centroids following a brief exposure. The problem with asking subjects to report more than three centroids is that the process of reporting the first few centroids strongly interferes with the memory of the still unreported centroids. Therefore, we used a partial-report paradigm[18], a post-cue procedure that requires only one report on each trial (Fig. 2.2). On a post-cued trial, subjects viewed briefly flashed 300 ms arrays containing $M$, $M = 3$, 4, 5, 6, 7, or 8, spatially interleaved features. The number of different dot colors was the main feature variable. When the number of colors was 5 and greater, colors became difficult to discriminate so different black shape items were introduced. $M$ denotes the number of items in a stimulus consisting of different dot colors, or different shapes, or a mixture of colored dots and black shapes. A cue indicating a randomly selected target feature was presented immediately after a 50 ms blank interval and a 100 ms mask that followed the stimulus exposure. Subjects' task was to compute only the centroid of items defined by the cued feature and to mouse-click its location on the same screen that displayed the stimuli.

### 2.1.3 How do post-cued trials in the current centroid paradigm enable us to determine how many salience maps a subject has?

The brief stimulus exposure and the post-cue masking stimulus require all the possibly-to-be-reported centroids to be computed during the brief period in which stimulus information is visually available. On a post-cued trial, subjects don't know which centroid to report until the post-cue appears 150 msec after the onset of the post-stimulus mask. At that point, there is no longer any significant amount of visual information available in any visual memory that is susceptible to post-exposure masking, i.e., by the time the subject receives the post-cue, there is no information available in "visual persistence" [18] or "iconic memory" [20]. Memory that survives a strong post-exposure mask is commonly called "working memory." Sun, Chubb, Wright, and Sperling [21] found that, following a brief display of 26 dots, less than two dots were available in working memory whereas, in the same display, at least 15 dots were incorporated into centroid computations. Therefore, all centroid computations have to be completed on the basis of information that fails to survive the post-exposure mask, and only computed centroids, not individual stimulus items, are retained in working memory. To re-iterate: On a post-cued trial, all the centroids that a subject computes are computed on the basis of visual information that is available only during a 300 stimulus exposure and a 50 msec blank period before the strong post-exposure masking field.

### 2.1.4 Eleven Centroid Experiments (Summary)

The current study used a variant of the centroid paradigm that was originally developed by Drew et al. [6] and was considerably enhanced by Sun et al. [13]. Subjects viewed briefly flashed arrays containing 3, 4, 5, 6, 7, or 8 spatially interleaved features. Figures 2.2 and B.1 show sample experimental stimuli. An initial set of 3 experiments was conducted with 24

items per stimulus. Subjects performed so well in these experiments that, to avoid ceiling effects, the total number of items in each of the main experimental stimuli was increased to $30 \pm 2$, the largest number that allowed us to maintain the 56-pixel center-to-center spacing of items (required for the shape stimuli) within the display boundaries. The subjects' task in all experiments was to move a cursor to the judged centroid location of the cued-feature items (targets) and to mouse-click.

There were four primary trial types (pre-cued versus post-cued trials X multi-item stimuli versus singletons) and several variations.

(1) Pre-cued multi-item trials: The cue indicating the target feature whose centroid was to be reported was presented 500 msec before the stimulus.

(2) Post-cued multi-item trials: The cue was presented immediately after a 50 msec blank interval and a 100 msec mask that followed the stimulus exposure.

(3,4) Pre-cued and post-cued singletons trials: For every type of multi-item trial, there was matched singletons trial. A singletons stimulus consisted of just one item of each feature type. Each singletons item was located at centroid of its feature in the corresponding full stimulus. The task was reporting the location of the cued singleton item. Singleton trials control for difficulties in computing centroids.

(5) Zero-distracter trials: There was no cue, distracters were eliminated and only target dots or target shapes were presented. Zero-distracter trials control for attention-filter imperfections in extracting target items.

(6) Centroid-of-ALL trials: Stimulus items were the same as the pre-cued and post-cued trials, but subjects were instructed to judge the centroid of ALL displayed items. Centroid-of-ALL trials estimate how many items can be processed when the problems associated with separating the items into separate groups are eliminated.

The stimuli of the six kinds of trials and the experimental procedure are depicted in Fig.2.2.

## 2.2 Results

### 2.2.1 Overview

In every experiment, the basic datum from each trial is the centroid response location, represented as response error distance. To gain more insight into the processes by which subjects arrive at their responses, three other statistics are computed as further explained below: A subject's attention filter (the weights of target and distracter items in the filter output) is estimated directly from the locations of a subject's responses. The number of stimulus items that survive to be incorporated into a subject's centroid computation is estimated by an ideal detector that processes only a sufficient fraction of the stimulus items to match the accuracy of subject's centroid judgments. Two ideal detectors are considered: one assumes the subject's estimated attention filter, and the other assumes a perfect attention filter that admits all target items and rejects all others.

### 2.2.2 Mean error magnitude

The most basic measurement of a subject's performance is the response error—the difference between the true centroid location of the target feature and the subject's mouse-click response. Although this difference is a 2D quantity, there were no significant differences between x and y components, therefore they are treated equally. Error magnitude is the Euclidean distance error measured in units of display pixels.

All the data of all the experiments are exhibited in Fig. 2.3a and 2.3b, except that implicit data that are used to compute attention filter weights. Figure 2.3a depicts the response

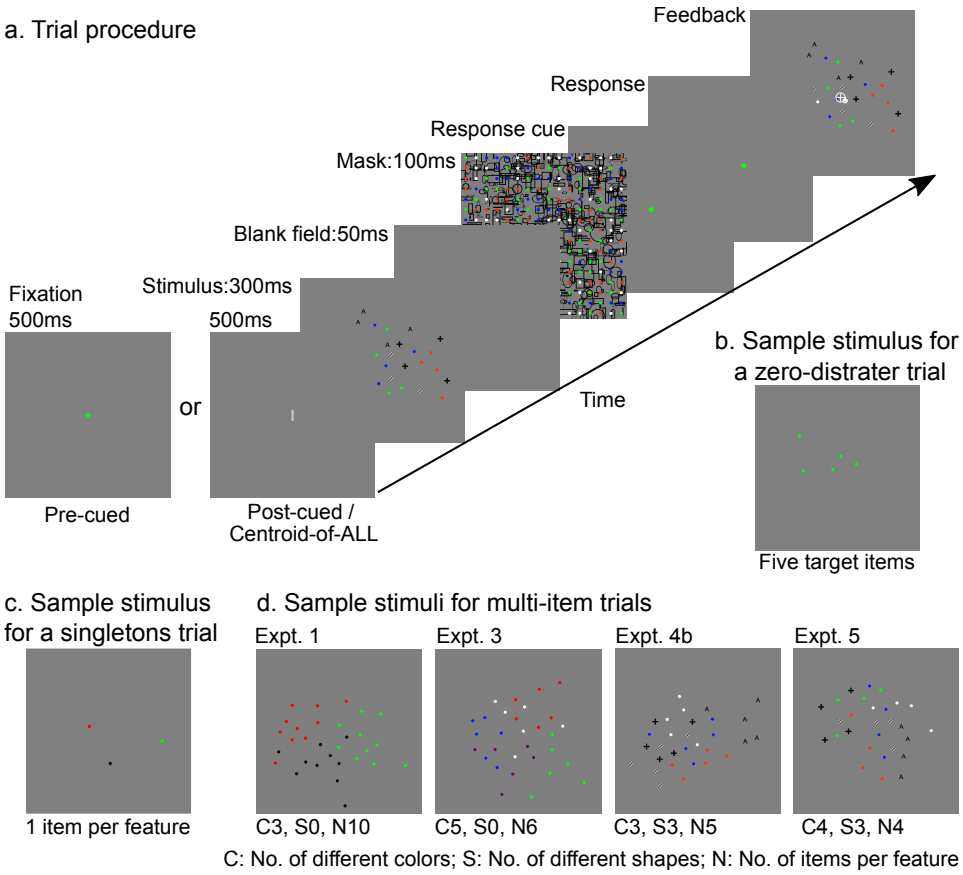



Figure 2.2: Experimental procedures and sample stimuli. (a) Every trial begins with a 500ms blank field with a fixation point, followed by a 300ms stimulus, a 50ms blank field, a 100ms mask, a blank field with an example target stimulus that the subject moves to the judged centroid location, and finally, a feedback display. Feedback shows the stimulus, the centroid of the target feature as the large plus sign inscribed in a light gray open circle, and the subject's response as the smaller plus sign inscribed in a white circle. For pre-cued trials, the fixation point is a target stimulus. For post-cued trials, the fixation point is a white bar, and the cue indicating the target feature is presented 50ms after the post-stimulus mask. Centroid-of-ALL trials have the same procedure as the post-cued trials, except that the task is to judge the centroid of ALL stimulus items. (b) Sample stimulus for a zero-distracter trial. The procedure for zero-distracter trials is the same as for post-cue trials, except that only target items are presented on the screen. (c) Sample stimulus for a singletons trial. For each singletons trial, the stimulus display consists of one item from each feature. The singletons items are located at centroids of the corresponding full stimuli from which they are derived. (d) Sample stimuli as displayed on an 800x800 pixel screen. 'C' in the title under the stimulus display represents the number of different colors of dots in the stimulus display, 'S' is the number of different shapes, and 'N' is the number of items in each feature class. For example, "C3, S3, N5" means that this stimulus contains dots of 3 different Colors, 3 different Shapes, and the Number of items of each feature is 5. There are (3+3)*5=30 items in this display. [1]

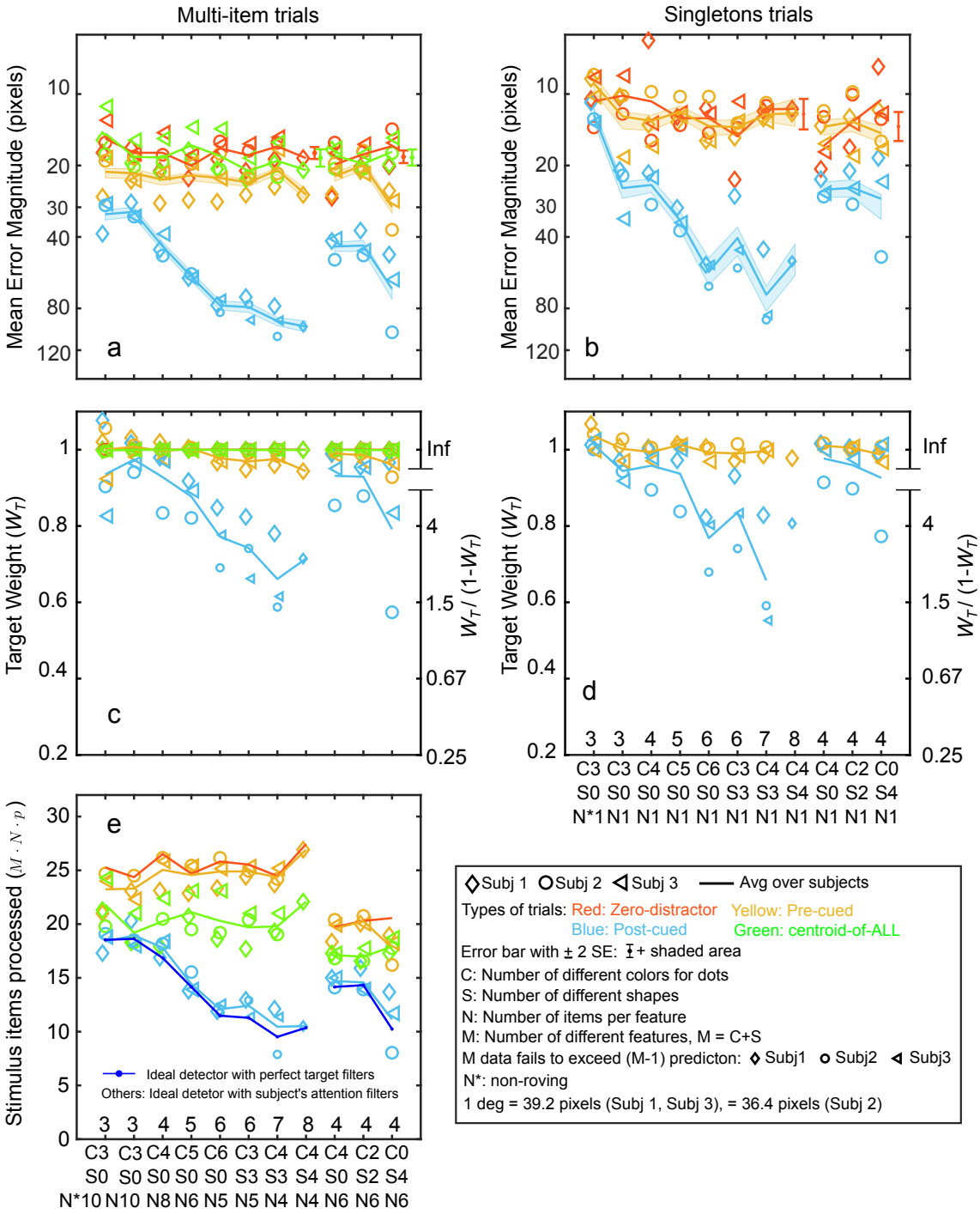



Figure 2.3: Results of 11 multi-centroid judgment experiments for three subjects. In all five panels, the x-axis represents the 11 experiments with their stimulus parameters as coded in Fig. 2.2. The number above each bottom x-axis tick represents $M$, the number of different features the subject must attend. In each of the 11 experiments, the symbols (triangle, diamond, circle) represent a particular subject's data. Small-size data points are from an $M$-feature experiment that failed to exceed the prediction from an available corresponding ($M$-1)-feature experiment. The 8 leftmost data points are from experiments with 30 stimulus items, the 3 rightmost data points are from experiments with 24 stimulus items. Colored lines represent the mean data of three subjects.

(a,b) Response accuracy in terms of mean error magnitude in pixels—the distance between a subject's centroid judgment and the true centroid. The shaded areas around mean post-cued and mean pre-cued data curves represent the confidence intervals ± 2SE (Standard Errors) for each computed mean. Error bars (mean± 2SE) to the right of C4S4N4 apply to the first 8 experiments; error bars on the far right apply to the three rightmost experiments.(c,d) Filter weight for target features (left ordinate). The right ordinate is a ratio, the average filter weight for target features ($W_T$) divided by average filter weight for distracters ($1-W_T$). For centroid-of-ALL trials, there are no distracters, target weight is defined as 1.(e) $M \cdot N \cdot p$, where $M$ is the number of different features, $N$ is the number of items of each feature, and $p$ is the probability of each item surviving an initial decimation process that produces the observed mean error magnitude. $M \cdot N \cdot p$ is the minimum number of stimulus items that an ideal observer with the same attention filter as the subject needs to process perfectly in order to match the subject's performance (upper four curves); the dark blue bottom curves represent the slightly smaller number of items required by an ideal detector with a perfect target filter. [2]

accuracy of multi-item trials in terms of mean response error magnitude averaged over the M different target conditions. The x-axis represents the 11 experiments, and the y-axis represents the response error magnitude in pixels. Response accuracy is best for zero-distracter and centroid-of-ALL trials, slightly worse for pre-cue trials (in 29 of 31 instances), and very significantly worse for post-cued trials. For zero-distracter and centroid-of-ALL trials, subjects' response accuracy was independent of experimental conditions. For pre-cued trials, subjects' response accuracy was independent of experimental conditions with one exception, S2, Expt. C0S4N6, which reflects this particular subject's inability to discriminate 4 shapes. One-way ANOVA tests confirmed that there were no significant differences among experiments for pre-cued ($F(10,20) = 1.385$, $p = 0.256$), zero–distracter ($F(10,20) = 0.721$, $p = 0.696$) and centroid-of-ALL trials ($F(10,20) = 1.125$ , $p = 0.392$). For post-cued trials, subjects' response accuracy dropped significantly as the number of attended groups increased ($F(10,20) = 14.979$, $p = 0.000$).

Figure 2.3b depicts the mean response error magnitude of singletons trials averaged over the $M$ different target conditions. The mean error magnitude of singletons trials is remarkably similar to the mean error magnitude of multi-item trials for pre-cued, post-cued, and zero-

distracter trial types in all 11 experiments. The variability of multi-item data points is smaller than for singletons data because there were 2.5 times more multi-item trials. Response accuracy for zero-distracter and pre-cued trials is independent of experimental conditions, whereas response accuracy for post-cued trials decreases as $M$ increases.

### 2.2.3 Number of surviving target items

When a subject's mean centroid error is 20 pixels, does this represent good or bad performance? We measure the quality of a subject's performance in terms of the minimum number of stimulus items that an ideal detector, which knows the exact location of every item and that computes a perfect centroid, needs in order to match a subject's performance.

Figure 2.4a shows the ideal detector model. The ideal detector has only one source of error, an initial decimation process during which each stimulus item has only a probability p, $0 \leq p \leq 1$, of surviving. The ideal detector considered here has a perfect target filter that passes only the $N$ target items and discards all distracters. The ideal detector knows the exact location of each of the $N \cdot p$ admitted target items and computes a perfect centroid. A Monte Carlo method is used to estimate the mean error magnitude of the ideal detector as a function of p for each of the 11 experiments (curves in Fig. 2.4b). $N \cdot \widehat{p}$ is the minimum number of target items that the ideal detector needs to match its mean error to a subject's mean error.

The curves in Fig. 2.4b depict the error magnitude (abscissa) as a function of the number of surviving target items (ordinate) for the ideal detector for each of the 11 experiments. Because the ideal detector has a perfect filter, its performance depends only on $N$, the number of target items in the stimulus, and on $p$, their survival probability. Five curves are shown for the five values of $N$ plus a sixth curve for the one non-roving condition C3S0N*10. Data averaged over three subjects for the 3 kinds of trials in 11 experiments are shown as points

on the ideal detector curves. All subsequent references to the number of surviving target items are based on these 33 numbers.

### 2.2.4 Target Weight

Response error magnitude informs us about how far a subject's response is away from the correct response, but error magnitude is not informative about the nature of the response computation. In the centroid paradigm, in addition to error magnitude, the particular distribution of errors allows us to estimate a subject's attention filter for each experiment. Attention filters, first used by [6] and refined by [13], describe the weight for each target and each distracter feature that best predicts the judged centroids. Attention filter weights are estimated separately for each experiment, condition, and subject to minimize the sum of the squared distances between responses predicted purely by attention filters and observed responses (Details in Supplementary Materials).

For each subject, in each experiment, the left ordinate in Fig. 2.3c depicts the target weight $W_T$ of multi-item trials averaged over the $M$ features. In zero-distracter trials, only the target feature was displayed, therefore the target weight for zero-distracter trials is always 1. In centroid-of-ALL trials, subjects were instructed to give equal weight to each feature. As no significant weight differences were observed for various different targets in centroid-of-ALL trials, in Fig. 2.3c target weight is plotted as 1. For pre-cued trials, for all three subjects, target weights are nearly 1.0, with one exception, Subj 2, Expt. C0S4N6 where Subject 2 found the 4 shapes difficult to discriminate. Post-cued trials are much more difficult than pre-cue trials, and target weight drops significantly as $M$ increases. The right ordinate in Fig 3c depicts the ratio between the aggregate weights of the target features and the aggregate weight of all the distracter features, $W_T/(1 - W_T)$. The ratio of the average target item filter weight to the average distracter item filter weight is larger by a factor of $M$-1, i.e., it is

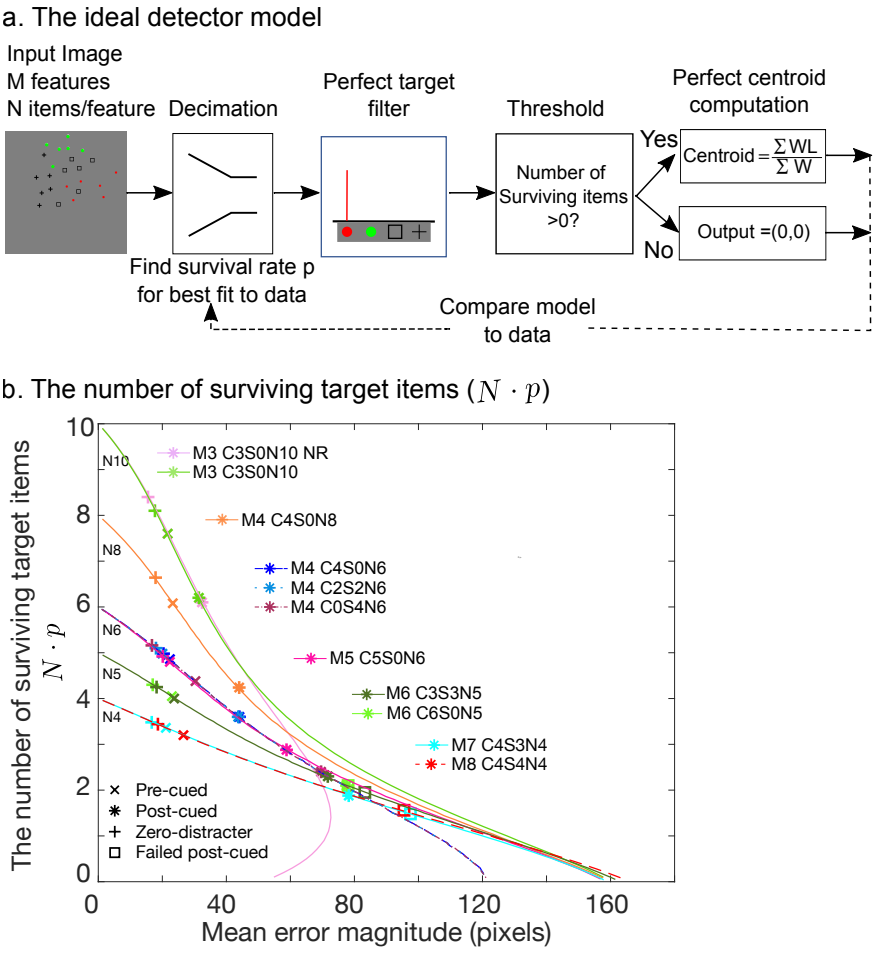



Figure 2.4: The ideal detector model and its expected centroid error as a function of the number of surviving target items. (a) The ideal detector model. The input is a stimulus image. In this example, it is a C2S2N6 stimulus (2 colors, 2 shapes, 6 dots for each feature, 24 dots total). The input image first goes through a decimation process during which each stimulus item has a probability p of surviving. Then a perfect target filter removes all non-target items and passes only surviving target items. If no target item has survived the decimation process, the output is (0,0), the center of the screen. Otherwise, a perfect centroid computation is carried out on the surviving target items. A Monte Carlo method is used to estimate the mean error magnitude of the ideal detector as a function of p (the curves in Fig. 2.4b). The number of surviving target items is the product p*N of the survival rate, p, and the number of target items, $N$ ($N = 6$ in this example). Figure 2.4a is adapted from Fig. 2.5 in [2]. (b) Mean error magnitude (abscissa) as a function of the number of surviving target items (ordinate) for 3 kinds of trials in 11 experiments. The data points show $N \cdot p$, the least number of target items that must be processed by an ideal detector with a perfect target filter to match a subject's performance. The colored lines represent the mean error magnitude of the ideal detector as a function of the number of target items being processed by the ideal detector for each of the 11 experiments using 8 different stimulus configurations. Data points indicated by 'x', '$*$', and '$+$' represent the mean error magnitude averaged over subjects' data from an $M$-feature experiment in which performance was better than could be predicted from the corresponding ($M$-1)-feature experiment. Squares represent post-cue data from an M-feature experiment that failed to exceed the prediction from the corresponding ($M$-1)-feature experiment.

$W_T(M-1)/(1-W_T)$; this ratio is a better indicator of filter selectivity, it is not illustrated here.

Figure 2.3d depicts the target weight in singletons trials' attention filters averaged over $M$ features for each subject in each experiment. Although the target weights of singletons trials are superficially similar to the target weights of multi-item trials, a more detailed analysis (Supplementary Material, Fig. B.4 and B.5) shows that the error distributions differ. Wide multi-centroid error distributions reflect filter error, whereas distributions for singletons are mixtures of narrow and flat distributions indicating memory errors.

### 2.2.5 Number of Stimulus Items Processed

The number of surviving target items is the absolute smallest number of stimulus items that an ideal detector with a perfect attention filter needs to process in order to match a subject's performance (Fig. 2.3e, dark blue curves). While that is a useful characterization of the absolute quality of performance, we here begin to derive a description of how subjects arrive at their performances rather than merely characterizing the quality of their performances. The first step is replacing the perfect attention filter with the attention filter that best characterizes a subject's performance—the subject's *achieved attention filter*. Because a subject's achieved attention filter is rarely perfect, an ideal detector using a subject's achieved attention filter requires more items to match that subject's performance than a detector with a perfect filter. For a stimulus containing $M \cdot N$ stimulus items, the number of stimulus items processed by a detector with an achieved attention filter is expressed as $M \cdot N \cdot \widehat{p}$. Here, $\widehat{p}$ is the probability of survival that equates the mean error magnitude of an ideal detector with a subject's attention filter to that subject's measured error magnitude.

Figure 2.3 depicts the number of stimulus items processed by an ideal detector with a different attention filter for each of 3 subjects, 4 kinds of trials, and 11 experiments, averaged

only over target types. For comparison, the mean number of stimulus items processed $M \cdot N \cdot \widehat{p}$ computed using a perfect filter is shown only for post-cued trials (the dark blue curve). (The other perfect vs actual filter comparisons are shown in Fig. 2.5a.) For all 11 experiments, zero-distracter trials have the highest number of stimulus items processed; pre-cued trials are very similar, and for both kinds of trials there's no significant drop in performance as the number of features to be processed increases. In contrast, post-cued trials have the lowest number of surviving target items, and post-cued performance declines significantly as the number of features increases.

Two things are worthy of notice. First, the mean number of stimulus items processed by an ideal detector with a perfect target filter (the dark blue curve) closely follows the mean number of stimulus items processed by an ideal detector with the subject's attention filter (the light blue curve). The great similarity between these two curves indicates that subjects' attention filters are not a major source of error. Second, the number of stimulus items processed for the pre-cued trial type is only slightly worse than the corresponding zero-distracter trials. This indicates that, in pre-cued trials, when subjects need to form only a single attention filter, this single attention filter virtually eliminates distracters.

### 2.2.6 Number of stimulus items processed corrected for motor error

We are interested here in the costs of computing and remembering multiple centroids. However, even the report of the position of a single stimulus item, the easiest possible condition, has a significant error that we call "motor error". We estimate motor error variance and subtract it from the overall error variance to enable us to better estimate perceptual and memory errors that are of greater interest. Here we report surprising properties of zero-distracter and pre-cued centroid trials that are revealed by subtracting motor error.

We estimate each subject's motor error from zero-distracter singletons trials in which only one stimulus item is shown on the screen and the task is to report the location of this item. The error variance for such trials includes both motor error (positioning the cursor) and memory noise (remembering the target location) and possibly other components. Each subject's motor error for a particular stimulus type is computed as the overall error variance of 50 zero-distracter singletons trials. Motor error variance is then subtracted from the overall error variance in each experiment to produce a set of data that is "corrected for motor error" for that type of stimulus. The number of stimulus items processed by an ideal detector is then re-computed in two slightly different ways based on the data corrected for singletons error: (1) assuming the subject's attention filter and (2) assuming a perfect target filter.

Figures 2.5a and 2.5b depict the ideal detector's estimates of the total number of stimulus items processed, both corrected and uncorrected for motor error, averaged over subjects. These estimates are shown for the 4 types of trials in each of the 11 experiments using the subject's attention filters and also using perfect attention filters.

Figure 2.5b shows that the numbers of stimulus items processed for zero-distracter, pre-cued, and centroid-of-ALL trials after correction for motor error are close to the data points representing the total number of stimulus items. The average difference between the zero-distracter subject's filter curve and the number of stimulus items is 1.37 items, the pre-cued difference is 1.47 items, and the centroid-of-ALL difference is 1.01 items. These small differences indicate an extraordinary processing capacity: For zero-distracter, pre-cued, and centroid-of-ALL multi-item trials, subjects utilized essentially all the $30\pm2$ target items displayed on the screen to compute their centroid response and that the centroid error observed in data that includes motor error is due to the subjects' inability to perfectly output the result of their centroid computations.

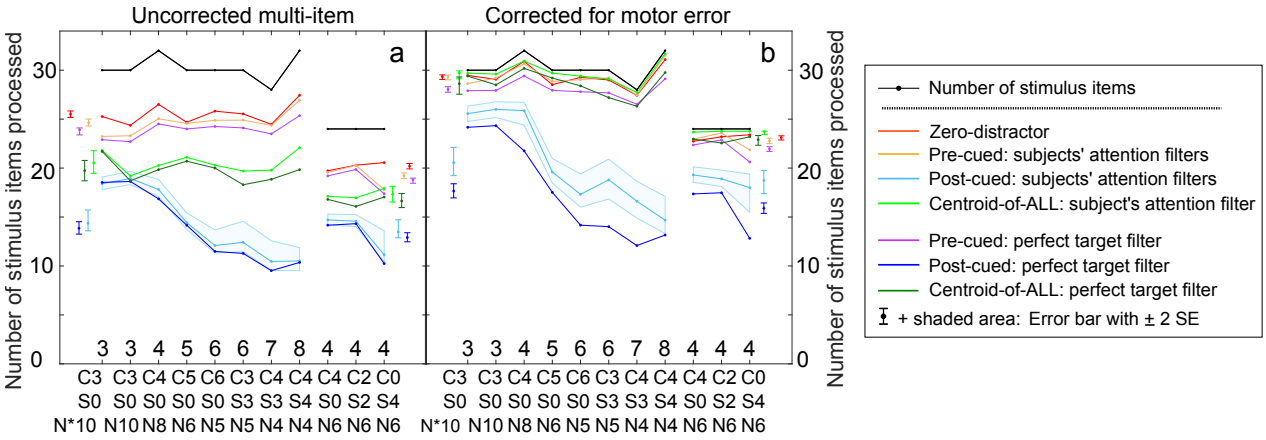



Figure 2.5: Number of stimulus items processed, averages of three subjects, in four types of multi-item trials in each of 11 experiments, uncorrected and corrected for motor error. In (a) and (b), the shaded area around the post-cued curves represents the confidence interval for ± 2 standard errors (SE) for each experiment. Error bars ($mean \pm$ 2SE) at the extreme left of Fig. 2.4 apply to the 8 experiments with 28-32 items; error bars on the far right apply to the 24-item experiments. (a) The mean number of stimulus items processed by an ideal detector with each individual subject's filter and with perfect target filters for pre-cued, post-cued, zero-distracter, and centroid-of-ALL trials in 11 experiments. The black line represents the performance ceiling, the total number of stimulus items in each experiment. (b) The number of stimulus items processed assuming subject's attention filters and also perfect target filters, all corrected for motor error—i.e., the variance of zero-distracter singletons trials (motor error) was subtracted from mean error variance in each experiment and condition.

## 2.3 Discussion

### 2.3.1 How many centroids can observers compute concurrently?

#### Prior studies

[11, 22, 23] attempted to measure the number of features that can be processed concurrently in ensemble perception. The task in all three studies was estimating the number of items composed of a particular cued feature. The criterion for all three studies was the minimum number of item features for which performance was reliably different between the pre-cued and the post-cued conditions. Poltoratski and Xu [22] and Luo and Zhao [23] reported 2 and Halberda et al.[11] reported 3 concurrently processed numerosity estimates.

Figures 2.3 and 2.5 illustrate pre-cued performance that is almost identical to zero-distracter performance. By definition, in zero-distracter trials, there is only one centroid available to be computed, the target centroid. The almost identical performance between pre-cued and zero-distracter trials indicates that, in pre-cued trials, subjects compute only the pre-cued target centroid, i.e., subjects have very good attention filters. Therefore, all that one can learn from the difference between pre-cued and post-cued performance is that more than one feature is processed in the post-cued trials. That post-cued trials have a significantly greater error than pre-cued trials (the measure of capacity in prior studies, e.g. [11, 22, 23]) only tells us that there may be some cost in processing more than one group of items, not that processing a particular number of groups is possible or impossible. To estimate the number of features that can be processed concurrently, we determine the extent to which performance on $M$-1 feature trials can predict performance on $M$ feature trials. When performance on $M$ feature trials exceeds the optimal extrapolation of performance on $M$-1 feature trials it means at least some additional capacity is displayed on $M$ feature trials, versus that the available capacity was merely divided more finely. This capacity analysis is developed in the next section.

**Subsampling analysis: Predicting target weight and errors in M-feature experiments from (M-1) and (M-2)-feature experiments**

How many centroids can a subject compute simultaneously? To estimate the number of centroids that a subject can compute simultaneously, we used the data of ($M$-1)- and ($M$-2)-feature experiments to predict a subject's performance in the M-feature experiment. Specifically, subjects are assumed to have only $M$-1 or $M$-2 attention filters. This results in two kinds of $M$-feature trials: (a) $(M-1)/M$ or $(M-2)/M$ "available-filter" trials in which subjects report a feature using the same filters as in the corresponding $M$-1 and $M$-2 feature trial, and (b) $1-(M-1)/M$ or $1-(M-2)/M$ "unavailable-filter" trials in which subjects

must report a feature they cannot process. Available-filter trials were analyzed as if the stimulus items for which the subjects didn't have a filter were invisible. i.e., these trials are basically similar to trials in experiments with $M$-1 or $M$-2 features, and these stimuli are processed similarly in terms of target weight. On the $1-(M-1)/M$ or $1-(M-2)/M$ "unavailable-filter" trials on which subjects are presented with a feature they cannot process, they click the center of the screen. These strategies are statistically optimal for stimuli that contain items that cannot be processed.

**The three possible outcomes of the subsampling analysis**

When the subject's actual performance with $M$ centroids is better than the predicted performance from $M$-1 or $M$-2 centroids, it means that subjects must have computed $M$ centroids. When the subject's actual performance is equivalent to the predicted performance, it is ambiguous whether (a) subjects could have computed just $M$-1 or $M$-2 centroids and still have produced the observed performance on $M$-feature trials or (b) have computed $M$ centroids but just not better than the $M$-1 prediction. When the subject's actual performance is worse than the predicted performance, it implies that the subject used a sub-optimal centroid-computation strategy.

The review of [2] in the introduction section pointed out that subjects were able to concurrently compute three centroids. Therefore, the stimuli in the present experiments which were designed to find the limits of performance contained 3 to 8 centroids. Figure 2.6 displays the data and the predictions of the subsampling analysis.

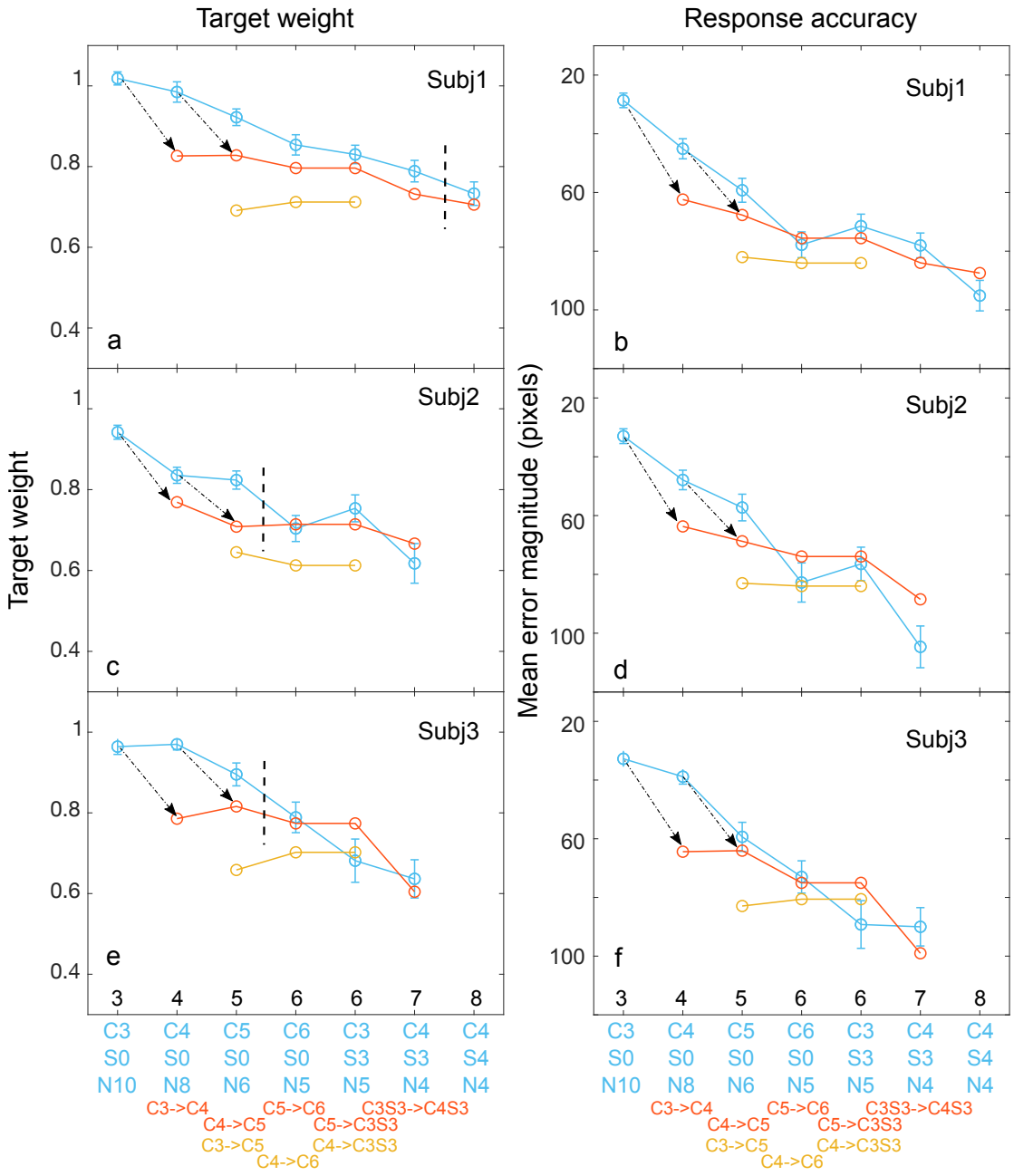



Figure 2.6: A subsampling analysis of multi-item post-cued trials to determine the number of centroids that a subject computes. To demonstrate $M$ computed centroids, subjects' location errors and/or target-filter weights in $M$-feature experiments must exceed predictions from ($M$-1)- or ($M$-2)-feature experiments. For all six panels, the black integers above the horizontal axis represent $M$, the number of different stimulus features. Underneath the figures, $C$, $S$, and $N$ indicate the number of stimulus Colors, Shapes, and target items. In panels a,c, and e, the ordinate is the estimated weight of the target feature; in panels b,d, and f, the ordinate is the observed mean error in judged centroid location. Blue dots indicate a subject's mean data, the error bars indicate ± 2 standard deviations of the mean. Red symbols indicate the result when an ($M$-1)-feature experiment is used to predict performance in an $M$-feature experiment.

To illustrate, the black arrows from blue $M3$ data to red $M4$ predictions and from $M4$ to $M5$ illustrate the source and target of two of the 4 or 5 predictions in each panel. The figure is too crowded to show the other arrows. The yellow symbols indicate predictions of an ($M$-2)-feature experiment. When a prediction of an ($M$-1)- or ($M$-2)- feature experiment equals or exceeds observed performance on an $M$-feature experiment, it means that a subject potentially could have processed only $M$-1 or $M$-2 features and have achieved the observed $M$-feature performance. It does not exclude the possibility that the subject can process $M$ features, it just means that this particular prediction is ambiguous with respect to the number of features actually processed. Experiments on the left of the vertical lines in the first column were the experiments in which subjects' performance in $M$-feature experiments exceeds the predictions from corresponding $M$-1 feature experiments. When a prediction of an ($M$-1)- or ($M$-2)- feature experiment equals or exceeds observed performance on an $M$-feature experiment, it means that a subject potentially could have processed only $M$-1 or $M$-2 features and have achieved the observed $M$-feature performance. It does not exclude the possibility that the subject can process $M$ features, it just means that this particular prediction is ambiguous with respect to the number of features actually processed. Experiments on the left of the vertical lines in the first column were the experiments in which subjects' performance in $M$-feature experiments exceeds the predictions from corresponding $M$-1 feature experiments.

**The results of the subsampling analysis: The number of subject-computed centroids**

For multi-item post-cued trials, Figs 2.6a, c, and e depict observed and predicted target weights; Fig. 2.6b, d, and f depict observed and predicted response errors individually for three subjects when ($M$-1)- and ($M$-2)- feature experiments were used to generate predictions for $M$-feature experiments. Predictions were based on Efron's [24] method of stimulus resampling (200 runs); the error bars were so small they are omitted in the graph. The target weight subsampling predictions are similar to but slightly more discriminating than the response accuracy predictions, so we concentrate on target weights.

For Subject 1 (Fig. 2.6a), the predicted target weights are worse than the observed weight from $M4$ to $M7$ and are equal to $M8$. If either target weight or response accuracy indicates that a subject must have computed $M$ centroids and the other measure is ambiguous (i.e., does not contradict the first measure), that is sufficient to demonstrate the success of an $M$-

centroid computation. (That an $M$-1 prediction equals performance does not mean that $M$-1 was the basis of performance, only that it could have been.) This logic implies that Subject 1 is able to establish at least 7 concurrent attention filters and to compute 7 concurrent centroids because the attention filters are derived from successful centroid performance. For response errors (Fig. 2.6b), Subject 1's $M$-data are better than predicted from $M$-1 data for $M$=4,5; equivalent for $M$=6,7, and worse for $M$=8. Based only on response error, Subject 1 can compute at least 5 concurrent centroids, for $M$=6,7 the response-error data are ambiguous. The failure of Subject 1 to match the ($M$-1)-prediction for $M$=8 means that attempting to compute 8 centroids leads to worse performance than could have been achieved by consciously ignoring one of the eight presented features. Based on attention filter accuracy and not contradicted by the response error analysis, Subject 1 was able to compute 7 but not 8 concurrent centroids.

For Subjects 2 and 3, the target weight (Fig. 2.6c,e) and response error (Fig. 2.6d, f) predictions are more similar to each other than for Subject 1. Following similar reasoning as for Subject 1, we conclude that Subjects 2 and 3 can create 5 concurrent attention filters and compute 5 centroids concurrently following a brief stimulus exposure. As centroids are computed on data that is represented in salience maps, we conclude that our three subjects have at least 7,5,5 salience maps. These numbers of salience maps are significantly greater than the 3 salience maps observed by Sun et al [2]—the maximum possible in their procedure.

### 2.3.2 Process models of pre-and post-cued centroid trials

#### Model of a post-cued centroid trial

Figures 2.7a and 2.7b are elaborations of the model in Figure 2.6 of reference [2], they are based on the data herein. Figure 2.7a illustrates how the retinal image of the stimulus display in a post-cued trial is transformed into a response. Figure 2.7b does the same for pre-cued

trials. The concern here is identifying possible sources of error. Error sources in Fig. 2.7a are (1a) failure to encode stimulus items prior to the point at which the target items have to be segregated (encoding failure), (1b) the number of stimulus items that survive to be grouped; 1a and 1b are not distiguished in the subsequent error analysis and treated together as multi-channel loss. (2) the precision of attention filters (% inclusion), (3) error in the computation of centroids, (4) working memory (the limited capacity to record centroids and the precision of the recording as determined by memory noise), and (5) "motor" error in producing the response. The blue areas in Fig. 2.7a and 2.7b illustrate these five error sources which are estimated in a subsequent section.

**Model of a pre-cued centroid trial**

Pre-cued trials begin with a small fixation stimulus (the pre-cue) that is a sample of the items whose centroid is to be reported. As in post-cued trials, it is followed by the stimulus display, a post-stimulus mask, a reminder screen with a sample of items whose centroid is to be reported, and finally, the response screen. In the pre-cued model, prior availability of the pre-cue is assumed to enable selective filtering that is so effective that only the cued items are encoded and there is no encoding noise (no perceptual loss). As illustrated in Fig. 2.5a and 2.5b, the centroid accuracy of a pre-cued subset is nearly as accurate as the response to zero-distracter trials. The remarkable performance of our practiced subjects indicates that, although pre-cued centroid tasks involve some processes similar to post-cued centroid task processes, pre-cued centroids are completely different because they involve encoding only the pre-cued items.

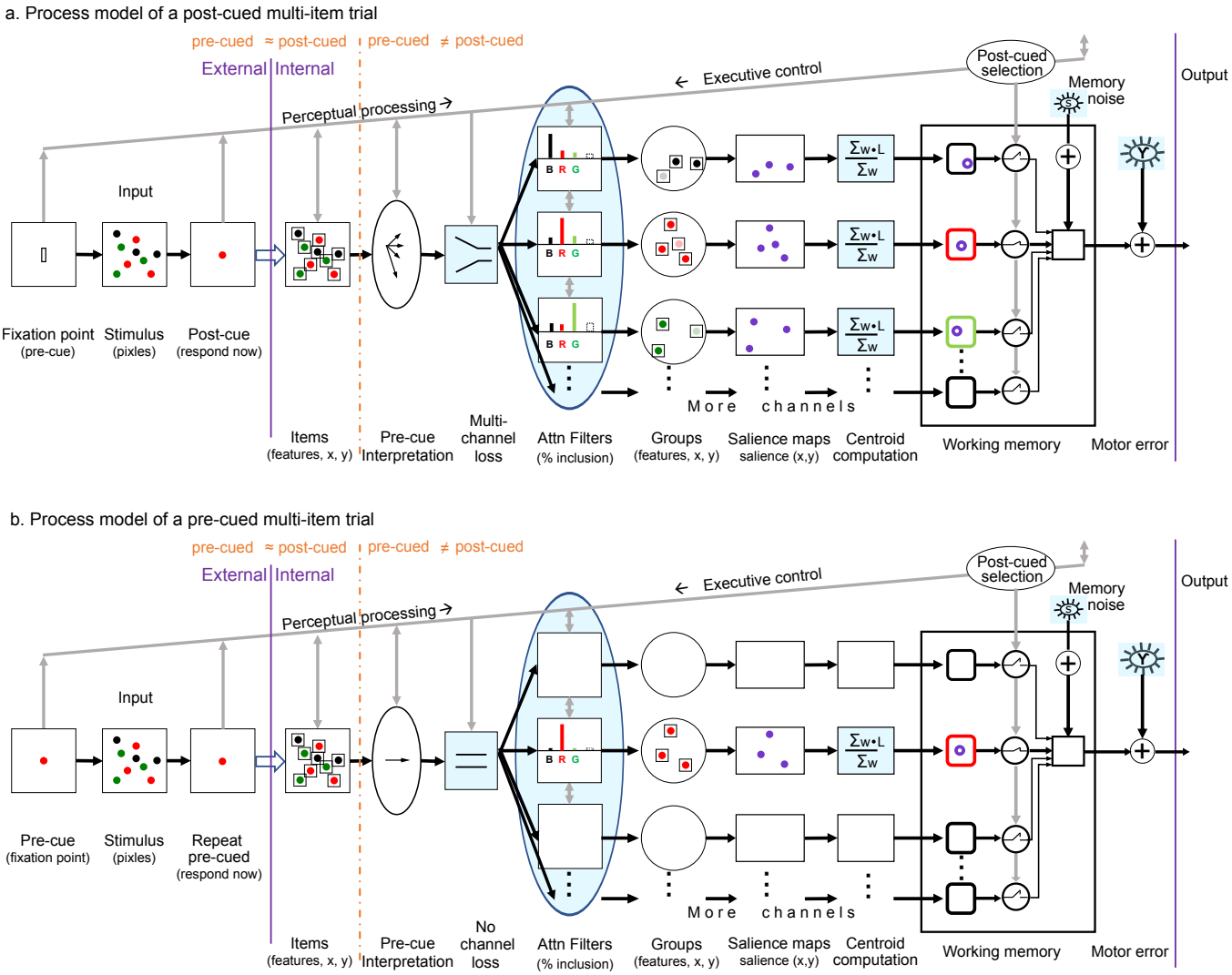



Figure 2.7: Flow chart of the essential components of computational models of post-cued and pre-cued multi-item trials. (a) A computational model of a post-cued multi-item trial in which a stimulus array (9 dots of 3 colors in this example) is transformed into a single response, the centroid of the dots whose color matches the post-cue. The first three components represent the stimulus sequence. The rectangular fixation bar indicates a post-cued trial. The stimulus image is represented as pixels, early visual processing transforms it into an array of items (indicated as small squares) each of which consists of a feature and a location. Interpretation of the pre-cue fixation point indicates that multiple attention filters are needed. Not all stimulus items can be fully processed, some of the multi-channel loss of stimulus information is represented by a decimation process. Every stimulus item has the same probability p, $0 \leq p \leq 1$, of survival. Three attention filters assign a weight to each surviving stimulus item according to its feature (color, in this example). The filter output (weighted items) constitutes a group. The content of each group–the weight and spatial location of each group item–is passed to a salience map. A centroid process computes a centroid location based on the weights and locations of the items in the salience map. In this path, feature information is associated with items only by an executive control process that binds each centroid with its filter label.

The labeled centroids are transferred into working memory and perturbed by memory noise. Response output is triggered by a post-cue that indicates which centroid is to be reported. A mouse-click response is made with additive motor error. Blue color in a component indicates it is a source of information loss in the five-error-source model. (b) Pre-cued multi-centroid model. The components and processes of the pre-cue computational model are the same as the post-cue computational model except for (1) Pre-cued trials begin with a pre-cue indicating which feature-centroid to report, therefore only one attention filter is needed, and it assigns nearly all weight to items whose feature matches the pre-cue. (2) Only one channel is used, and only one centroid is computed. The single centroid stored in working memory is the response. (3) Because the pre-cue defines the feature of a small group of to-be-attended items, there is very little channel loss (only cued items are processed to the level of entering into a group) and little memory loss (only one centroid needs to be recorded).

### 2.3.3 Estimating the sources of error in multi-centroid processing

#### Five error sources defined

Based on the models in Fig. 2.7, the data from multi-centroid and singletons trials enable the derivation of five component errors for each experiment that are assumed to add independently to produce the observed response errors: *1*, loss in each channel that conveys information about a particular class of stimuli from the input all the way to the memory component but not including *2*, attention filter imperfections and *3*, centroid computation errors, each of which is computed separately; *4*, memory error in storing and recalling the computed centroids; and *5*, motor error in producing a cursor response to represent the recalled memory content.

#### Distance and item error estimates

In the prior sections, we considered two ways of measuring error: (a) the pixel distance of the response from the target, and (b) the number of stimulus items an ideal detector must discard (items lost) in order to match a subject's performance level. To produce total pixel error squared for Experiments with $M$ different features, $E_M^2$, the squared component errors

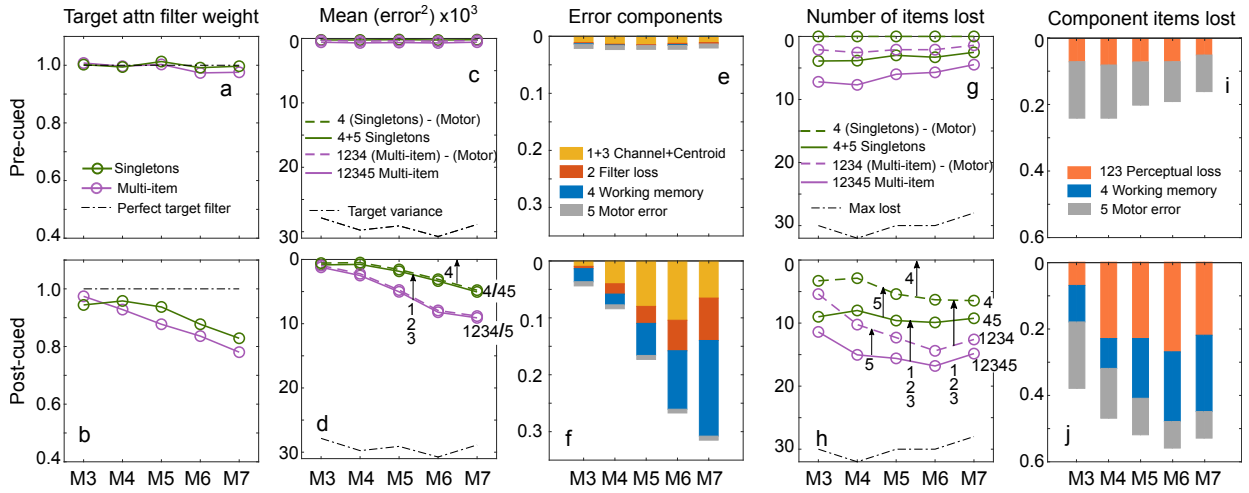



Figure 2.8: Estimates of five additive component error sources in multi-centroid computations: Perceptual errors include (1) channel loss (2) attention-filter error, (3) centroid computation error; post-perceptual errors include (4) storing and recalling the computed centroid (working memory, and (5) motor error in producing a cursor response to represent the memory content. For all ten panels, the x-axis is the number of different stimulus features. The upper row represents pre-cued trials; the bottom row, post-cued trials. Data are only from subjects and multi-item conditions that exceeded subsampling predictions (Fig. 2.6). (a,b) Attention filter weight of targets in multi-item and singletons trials. (c,d) Mean squared error in units of $display\_pixels^2$ for multi-item and singletons trials with and without correcting for motor error (tiny differences are not visible). The numbers under the vertical arrows represent the component(s) measured by the arrows except motor error (5) which is too small to display clearly. (e,f) Error component variances as a fraction of the target location variance. (g,h) The number of stimulus items lost for singletons and multi-item trials, with and without correcting for motor error. (i,j) Component item loss as a fraction of total stimulus items.

of five assumed-to-be-independent stages $j$ add independently. The sequence in which these stages occur is irrelevant.

$$\text{Post-cued error:}\quad E_M^2 = \sum_{j=1,5} E_{j,M}^2 \tag{2.1}$$

Unfortunately, there is no equivalent item theory to derive total item error $I$ from independent component item errors. In an item theory, the items lost at each stage depend on the number and quality of items delivered by the prior stage (interdependence). However, to aid intuition about how well the stimulus information is utilized, we derive a "number of items lost" that would produce the observed "distance error" (and vice versa) after three

processing levels: perceptual, memory, and motor.

To estimate the five component error sources, only the 12/19 conditions in which the analysis of Fig. 2.6 indicates that a subject did not use a subsampling strategy are included. Two different conditions (C6S0N5 and C3S3N5) in which there are 6 different item types are combined. Parameter estimates of mean-square pixel-distance errors and of items-lost errors are made individually for each subject, the averages of these are reported in the text and Fig. 2.8.

*Motor error* $E_5$ is the pixel-distance error observed in trials in which only one item is presented and the subject's task is to mouse-click the location of that item (Fig. 2.3b, Zero-distracter Singleton trials. There are insufficient trials to discriminate motor error in different conditions, so they are all combined in these estimates: $E_5$ =14.8, 14.6 and 13.4 pixels for the three subjects. For reference, the diameter of an individual stimulus dot is 15 pixels (0.39 deg for subjects 1 and 3; 0.43 deg for subject 2). Motor error squared for 10 conditions is shown Figs 2.8e,f; the motor error item loss is in Fig. 2.8i,j. The motor error proportion of the total error variance is much smaller than in the item analysis for two reasons: (1) Squaring distance errors give a much greater relative weight to larger versus smaller errors than does the item analysis. (2) The item analysis asks "if there were no motor error how much better would the total item score have been." The raw item score indicates that at least 12 stimulus items have been processed in every post-cued multi-centroid condition. To further improve this already good accuracy to the level it would be without motor error requires processing more additional items than intuition suggests (Supplementary Material Fig. B.3). This discrepancy between two different measurements (distance, item) illustrates a fundamental measurement principle: The relative importance of various sources of error depends on the particular choices of how error is measured.

*Centroid computation errors* $E^2_{3,M}$ are estimated from trials in which only 4-10 target items and zero distracters are presented (Fig. 2.3a, multi-item zero-distracters trials). The

only other significant source of error in these trials is motor error which is subtracted from the response error to yield the squared centroid computation error $E^2_{3,M}$. Our experienced subjects were so accurate in estimating single centroids that these centroid errors are too small to represent on the scales used for the other errors in Fig. 2.8. Centroid computation errors are probably greater when multiple centroids have to be computed, but we have not found an efficient way to estimate them here.

*Attention filter discrimination* of targets from non-targets is indexed by the fraction of the total filter weight of 1.0 that is assigned to the target filter. The pre-cued attention filter weights in Fig. 2.8a represent virtually perfect target discrimination. For post-cued trials (Fig. 2.8b), filter weights for (multi-item, singletons) decrease from (0.97, 0.94) to (0.78, 0.83) as the number of target types M increases. The distance errors $E^2_{2,M}$ that would be produced by these attention filters are shown Fig. 2.8e,f.

*Memory loss* $E^2_{4,M}$ in Experiment $M$ is the error difference between singletons and motor error averaged over the $M$ target conditions. The only error sources for post-cued singletons are remembering the feature and location of the to-be-cued single item until the post-cue appears, and then moving the cursor to the remembered location of cued item:

$$\begin{aligned} E^2_{4,M} &= \text{Singletons error}^2_M - \text{Motor error}^2 \\ &= (E^2_{4,M} + E^2_5) - E^2_5 \\ I_{4,M} &= (I_{4,M} + I_{5,M}) - I_{5,M} \end{aligned} \tag{2.2}$$

In both the distance error and item error analysis memory loss increases with the number of locations to be remembered.

*Channel loss* $E^2_{1,M}$ is the residual after all the other error losses have been accounted,i.e.,

the total error minus (filter error + centroid error + singletons error):

$$E^2_{1,M} = \sum_{j=1,5} E^2_{j,M} \; - E^2_{2,M} - E^2_{3,M} - (E^2_{4,M} + E^2_5) \tag{2.3}$$

Figure 2.8 shows the result of the above estimates of the channel, filter, memory, and motor components of response error in the multi-centroid post-cued task. The general summary is that channel and working memory are the major sources of error in the multi-centroid task according to both the distance and item analyses. In the distance error analysis, there also is appreciable filter loss and negligible motor error. In the item analysis, filter loss was not isolated, it was incorporated into perceptual loss, whereas motor error was appreciable.

### 2.3.4 Updated representation of the salience system

The process by which items that have a similar color or a similar shape are isolated and separated as a group is not simple, but it is simple to describe and its effectiveness is measured by the selectivity of attention filters. The brain process of computing the centroid of the grouped items has to be independent of the identity of the items because there's an infinity of different possible items. This independence of item content is accomplished by representing the contents of a group of similar items on a salience map that records only the location of items and a real number quantity (salience) at that location. Salience maps were originally proposed to represent the importance (salience) of locations. A winner-take-all computation on the map found the most salient location [1], and that location was then used to guide subsequent processes, such as visual, search, eye movements, and pattern identification (Fig. 2.1 in the Introduction section).

Salience maps have been proposed as the basis of the computations other than processing

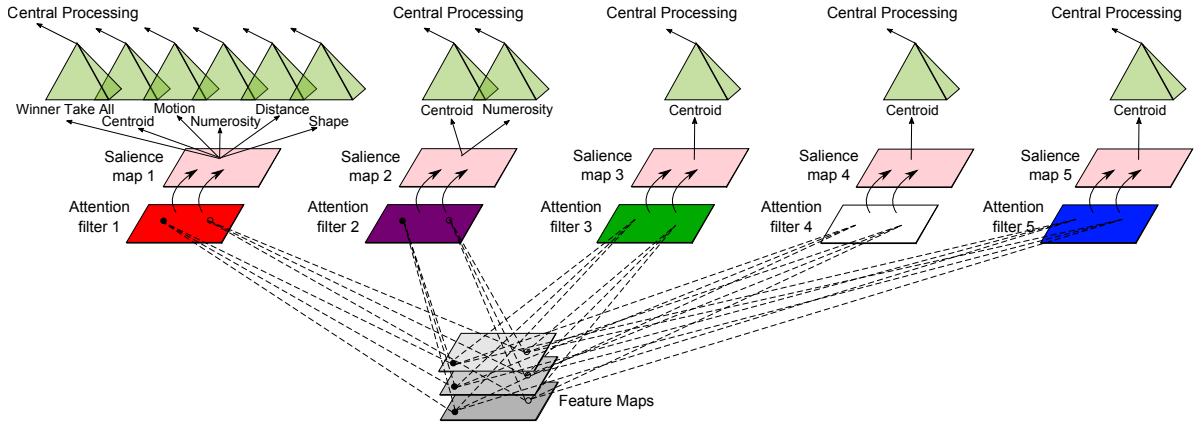



Figure 2.9: Updated representation of the human salience system showing five semi-independent salient maps and the processes that until now have been demonstrated to utilize them. Stimulus information is processed by various feature maps, and the outputs of feature processing are directed to attention filters. Five attention filters are illustrated for the conditions of Expt. 3 in which 6 dots in each of 5 different colors were presented. After the stimulus presentation, subjects were cued to report the centroid of just one set (C5, S0, N6). Filter outputs (the locations and strengths of surviving items) are directed to salience maps. In all the experiments described here, the outputs of the salience maps were directed to centroid computations, but the figure also shows the other computations that until now have been demonstrated to utilize salience maps.

priority: attention-generated motion [25], distance estimation between two points in the frontal plane [26], letter and shape identification [27] and, in the present case, a centroid computation [2]. In contrast to Fig. 2.1, our present understanding of salience maps is represented in Fig. 2.9 which shows the 5 salience maps that all our subjects demonstrated, and the various uses of salience maps that have been demonstrated by this and the studies cited above.

## 2.4 Conclusions

The previous sections demonstrated how subjects perform when they view brief exposures of displays composed of 28 to 32 items composed of 3 to 8 different features (colors or shapes). The task was to judge the centroid of just the items composed of one randomly designated feature. The feature to be reported was made known to the subject either 0.5 sec before

the display (pre-cue) or only after a post-exposure masking field that followed the stimulus exposure (post-cue). The results, centroid estimation distance errors, were interpreted in terms of the number of stimulus items that a statistical ideal observer with perfect knowledge of item identity and location would require to match each subject's performance. In pre-cued trials, as the number of centroids increased from 3 to 8 in displays of 28-32 items, performance remained approximately constant, an ideal observer would have to perfectly process 25 items to match the subjects' performances, and about 30 items if motor error in producing a response is taken into account. On the other hand, on post-cued trials, to match the subjects' performance on displays with three centroids, the ideal observer required perfect knowledge of 17 items; for displays containing 7 centroids, performance declined to 10 items. Performances on $M$-centroid trials that exceeded optimal predicted performance (based on subsampling $M$-1 centroids) indicated that all subjects could compute at least $M$=5 centroids and one subject could compute 7 centroids on post-cued trials. When the critical feature of the items whose centroid was to be reported was shown to the subject before a trial (pre-cued condition) their performance was almost the same as on trials in which distracter items were physically removed and only target items were displayed. This near equivalence demonstrates nearly perfect attention filters on pre-cued trials.

A computational model that characterizes the subjects' centroid performances (Fig. 2.7) consists of perceptual processing (converting pixel images into items composed of features and a location, channeling the items into attention filters, grouping the filter outputs and computing the group centroid), entering the centroids into working memory, and converting the cued memory contents into a motor cursor response. The amount of noise introduced at each of these steps was estimated in terms of response error (pixel distance of response to target) and the number of stimulus items lost at each stage. The two main sources of error were channel loss in processing so many items and memory loss. The distance analysis also found significant attention-filter error, the item analysis found significant motor error, centroid computation errors were minimal. In the context of the models and the

aforementioned error analysis, the accuracy of our subjects in locating the centroids in briefly exposed arrays as accurately as they did requires that the three subjects have processing capabilities equivalent to at least 7,5,5 concurrently available salience maps.

## 2.5 Methods

### 2.5.1 Subjects

The first and second authors (Subj1, Subj2) participated in the experiment. Subj3 was naïve to the purpose of the experiments. The three subjects (two females and one male) ranged from 23 to 40 years of age. All subjects reported having normal or corrected-to-normal vision. Methods were approved by the UC Irvine Institutional Review Board, and all subjects signed informed consent forms. The protocol and signed consent forms were approved by the UCI IRB. Subj1 and Subj2 were experienced in doing centroid tasks. Subj3 had no experience in doing centroid tasks; therefore, prior to starting Expt. 1, Subj3 had four 1-hour centroid training sessions.

### 2.5.2 Apparatus

The experiment was conducted on an iMac intel computer installed with MATLAB 2018b and Psychtoolbox-3 software. For Subj1 and Subj3, the stimuli were presented on an ASUS ProArt Display monitor with 1920 x 1200 resolution at a refresh rate of 60 Hz, mean luminance 62.4 cd/m2. The monitor screen was 51.8 cm wide x 32.4 cm high, pixels were 0.27mm x 0.27 mm. For Subj2, the stimuli were presented on a Samsung Syncmaster 2233BW Display monitor with 1680 x 1050 resolution at a refresh rate of 60 Hz, mean luminance 32.5 cd/m2. The monitor screen was 47.6 cm wide x 29.6 cm high, pixels were 0.282 mm x 0.282

mm.

### 2.5.3 Number of trials

An experiment consisted of four types of blocked sessions: one pre-cued session in which a pre-cue indicated the target feature, one post-cued session in which the post-cue was shown after stimulus and mask, one zero-distracter session in which only the target feature of items was displayed, and one centroid-of-ALL session in which all features were targets. Singletons trials were interleaved in pre-cued, post-cued, and zero-distracter sessions. In both pre-cue and post-cue sessions, each feature was cued in 50 multi-item trials + 20 singletons trials; all features were interleaved in a mixed-list design. In zero-distracter sessions, all features were interleaved, and each feature was tested in 25 multi-item trials + 5 singletons trials. Centroid-of-ALL sessions had 50 trials. The 217 data points of Fig.2.3a and 2.3b summarize 32,210 trials.

### 2.5.4 Stimuli

For pre-cued, post-cued, and centroid-off-ALL trials in each of 8 experiments, the stimulus display contained 28 to 32 interleaved items of $M$ different features. Features were either $M$ different dot colors, $M$ shapes, or a mixture of colored dots and shapes. The colors of dots were red, blue, green, white, purple, and black. The shapes were Gabor patches, plus signs, letter A, and square boxes. All features within an experiment had the same number of items. The number $M$ was 3 for Expt. 1a and 1b, 4 for Expt. 2a, 2b, 2c, and 2d, 5 for Expt.3, 6 for Expt.4a and 4b, 7 for Expt.5, and 8 for Expt.6. (For 3 experiments (Expt 2b, 2c, and 2d) stimuli contained 24 items, composed of either 4 colors or 4 shapes, or a 2+2 mixture.) For pre- and for post-cued trials, on each trial, a random one of the features was designated as the target feature. For zero-distracter trials, the stimulus display contained

only one randomly selected feature from the $M$ features of the corresponding experiment. For centroid-of-ALL trials, all $M$ features were targets. In addition to multi-item stimuli, singletons stimuli consisted of just one item from each feature. The position of each singletons item was the centroid of the feature from the corresponding multi-centroid trial. Figure 2.2 and Fig. B.1 show sample stimuli.

### 2.5.5 Dimensions

Stimuli were displayed for 300 msec within an 800 by 800 pixel-wide square (Fig. 2.2) that spanned 20.4 deg of visual angle (dva) for Subj1, Subj3 and 22.0 dva for Subj2. The diameter of colored dots (filled circles) was 15 pixels, spanning 0.39 dva for Subj1 and Subj3, and 0.42 dva for Subj2. Stimulus items in any shape other than filled circles were inscribed inside invisible circles of 28-pixel diameter, spanning 0.72 dva for Subj1 and Subj3, and 0.78 dva for Subj2 to prohibit items from overlapping. All stimuli were clearly visible on the grey background.

*Centroid standard deviation and item dispersion.* In order to produce useful data, different features have to have sufficiently different centroids. To achieve centroid separation, we first randomly chose centroid locations and then varied item locations around the chosen centroid as follows: Stimulus displays contained $M$ different features with $N$ items per feature for a total of $M*N$ items. Except for Expt. 1a, non-roving C3S0N*10, in each stimulus display, locations of the items were determined by a three-step process: (1) $M$ initial centroid locations were drawn from a bivariate normal distribution centered on the center of the screen and with a standard deviation of 80 pixels. (2) $N$ item locations were drawn from a bivariate normal distribution which centered on each of the $M$ centroids. The sample locations were divided by the standard deviation of the $N$ samples and then multiplied by a fixed constant to ensure that the locations of all generated arrays would

have the same standard deviation (dispersion) of pixels on every trial. (3) The location of an item was re-sampled if its location occurred within 5 pixels of another item. The overall centroid standard deviation that resulted from the initial location of each centroid followed by the repositioning of the centroid due to the sampling and re-sampling of the items within each individual feature was 170.0 ± 10.6 pixels for the 7 experiments containing 28 to 32 stimulus items; 121.2 ± 3.5 pixels for the 3 experiments containing 24 stimulus items. For Expt. 1a, non-roving C3S0N*10 trials, the procedure of generating the locations of items was the same as the procedure described above, except that the 3 initial centroid locations were the center of the screen and the centroid standard deviation was 56.6 ± 2.6 pixels.

## 2.6 Acknowledgement

We thank Ruth Rosenholtz and Misha Pavel for useful suggestions.

# Bibliography


[1] Christof Koch and Shimon Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *Human neurobiology*, 4 4:219–27, 1985.

[2] Peng Sun, Veronica Chu, and George Sperling. Multiple concurrent centroid judgments imply multiple within-group salience maps. *Attention, Perception, & Psychophysics*, 83(3):934–955, 2021.

[3] Dan Ariely. Seeing sets: Representation by statistical properties. *Psychological science*, 12(2):157–162, 2001.

[4] David Whitney and Allison Yamanashi Leib. Ensemble perception. *Annual review of psychology*, 69(1):105–129, 2018.

[5] Sang Chul Chong and Anne Treisman. Representation of statistical properties. *Vision research*, 43(4):393–404, 2003.

[6] Stefanie Drew, Charles Chubb, and George Sperling. Quantifying attention: Attention filtering in centroid estimations. *Journal of Vision*, 9(8):229–229, 2009.

[7] Mouna Attarha and Cathleen M Moore. The capacity limitations of orientation summary statistics. *Attention, Perception, & Psychophysics*, 77(4):1116–1131, 2015.

[8] Julie M Harris, Suzanne P McKee, and Scott NJ Watamaniuk. Visual search for motion-

in-depth: Stereomotion does not'pop out'from disparity noise. *Nature neuroscience*, 1(2):165–168, 1998.

[9] Ben Bauer. Does stevens's power law for brightness extend to perceptual brightness averaging? *The Psychological Record*, 59(2):171–185, 2009.

[10] Nele Demeyere, Anna Rzeskiewicz, Katharine A Humphreys, and Glyn W Humphreys. Automatic statistical processing of visual properties in simultanagnosia. *Neuropsychologia*, 46(11):2861–2864, 2008.

[11] Justin Halberda, Sean F Sires, and Lisa Feigenson. Multiple spatially overlapping sets can be enumerated in parallel. *Psychological science*, 17(7):572–576, 2006.

[12] Jason Haberman and David Whitney. Rapid extraction of mean emotion and gender from sets of faces. *Current biology*, 17(17):R751–R753, 2007.

[13] Peng Sun, Charles Chubb, Charles E Wright, and George Sperling. Human attention filters for single colors. *Proceedings of the National Academy of Sciences*, 113(43):E6712–E6720, 2016.

[14] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 20(11):1254–1259, 1998.

[15] Neil Bruce and John Tsotsos. Saliency based on information maximization. *Advances in neural information processing systems*, 18, 2005.

[16] Laurent Itti and Pierre Baldi. Bayesian surprise attracts human attention. *Advances in neural information processing systems*, 18, 2005.

[17] Hae Jong Seo and Peyman Milanfar. Static and space-time visual saliency detection by self-resemblance. *Journal of vision*, 9(12):15–15, 2009.

[18] George Sperling. The information available in brief visual presentations. *Psychological monographs: General and applied*, 74(11):1, 1960.

[19] George Sperling. A model for visual memory tasks. *Human factors*, 5(1):19–31, 1963.

[20] Ulric Neisser. *Cognitive psychology: Classic edition*. New York: Appleton-Century-Crofts, 1967.

[21] Peng Sun, Charles Chubb, Charles E Wright, and George Sperling. High-capacity preconscious processing in concurrent groupings of colored dots. *Proceedings of the National Academy of Sciences*, 115(52):E12153–E12162, 2018.

[22] Sonia Poltoratski and Yaoda Xu. The association of color memory and the enumeration of multiple spatially overlapping sets. *Journal of vision*, 13(8):6–6, 2013.

[23] Anna Xiao Luo and Jiaying Zhao. Capacity limit of ensemble perception of multiple spatially intermixed sets. *Attention, Perception, & Psychophysics*, 80(8):2033–2047, 2018.

[24] Bradley Efron. Computers and the theory of statistics: thinking the unthinkable. *SIAM review*, 21(4):460–480, 1979.

[25] Zhong-Lin Lu and George Sperling. Attention-generated apparent motion. *Nature*, 377(6546):237–239, 1995.

[26] Lingyu Gan, Peng Sun, and George Sperling. Frontal-plane distance judgments between two equal-size items are made on the basis of a salience map. *Journal of Vision*, 21(9):2828–2828, 2021.

[27] George Sperling and Lingyu Gan. Two-dimensional shape perception is based on a salience map. *Journal of Vision*, 22(3):17, 2022.

# Appendix for Chapter 2

## B1 Estimating a subject's attention filters

The model for estimating a subject's attention filters assumes that the only source of error in centroid computation is an imperfect attention filter. To compute a centroid response based on an attention filter, each item is weighted according to the filter value of its feature. The predicted response is the centroid of the filter-weighted items as described below.

For a trial j that has M different features and N items per each feature, let $W_T(m)$ be the weight assigned to feature $m$ when $T$ is the target feature, let $X_m(j)$ and $Y_m(j)$ be the x- and y-coordinates of the mean location (centroid) of items whose feature is $m$, and $P_{x,T}(j)$ and $P_{y,T}(j)$ be the x- and y-coordinates of a predicted response. Equations B.1 describe the subject's predicted response on trial j:

$$P_{x,T}(j) = \sum_{m=1}^{M} W_T(m) \cdot X_m(j) \tag{B.1a}$$

$$P_{y,T}(j) = \sum_{m=1}^{M} W_T(m) \cdot Y_m(j) \tag{B.1b}$$

$W_T(m)$ is constrained by:

$$\sum_{m=1}^{M} W_T(m) = 1 \tag{B.2}$$

In an experimental condition, for a set of $J$ trials that have the same M features and the same target feature $T$, a subject's attention filter is the $M$ values of $W_T(m)$ that minimizing the sum of the squared distances between the predicted and observed responses. Denote the x- and y-coordinates of a subject's observed response on trial j as $R_{x,T}(j)$ and $R_{y,T}(j)$. The estimated attention filter is the values of $W_T(m)$ as defined in Equations B.1 and B.2 that minimize "Loss" in Equations B.3 and thereby maximize the likelihood of $W_T(m)$.

$$Loss = \sum_{j=1}^{J}[(P_{x,T}(j) - R_{x,T}(j))^2 + (P_{y,T}(j) - R_{y,T}(j))^2] \tag{B.3a}$$

$$= \sum_{j=1}^{J}[(\Delta X_{j,T})^2 + (\Delta Y_{j,T})^2] \tag{B.3b}$$

Minimizing $Loss$ in Equations 3a,b is equivalent to maximizing the likelihood $L$ in Equation B.4:

$$L = \prod_{j=1}^{J} \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}[(\frac{\Delta X_{j,T}}{\sigma})^2 - 2\rho(\frac{\Delta X_{j,T}}{\sigma})(\frac{\Delta Y_{j,T}}{\sigma}) + \frac{\Delta Y_{j,T}}{\sigma})^2]} \tag{B.4}$$

In Equation B.4, $\rho$ is the correlation between $\Delta X_T$ and $\Delta Y_T$. Taking the log of both sides of Equation B.4 yields the log-likelihood ratio $\ln(L)$ which is the normal computational form.

$$\ln(L) = -J[\ln(2\pi) + 2\ln\sigma + 0.5\ln(1-\rho^2)] - \frac{1}{2\sigma^2(1-\rho^2)} \sum_{j=1}^{J}[(\Delta X_{j,T})^2 + (\Delta Y_{j,T})^2 - 2\rho\Delta X_{j,T}\Delta Y_{j,T}] \tag{B.5}$$

Attention filter weights as estimated by Equation B.5 were separately estimated for each experiment, each condition, and each subject to maximize Equation B.5 by using a Markov Chain Monte Carlo (MCMC) algorithm constrained by Equation B.2.

## B2 Supplementary Figures

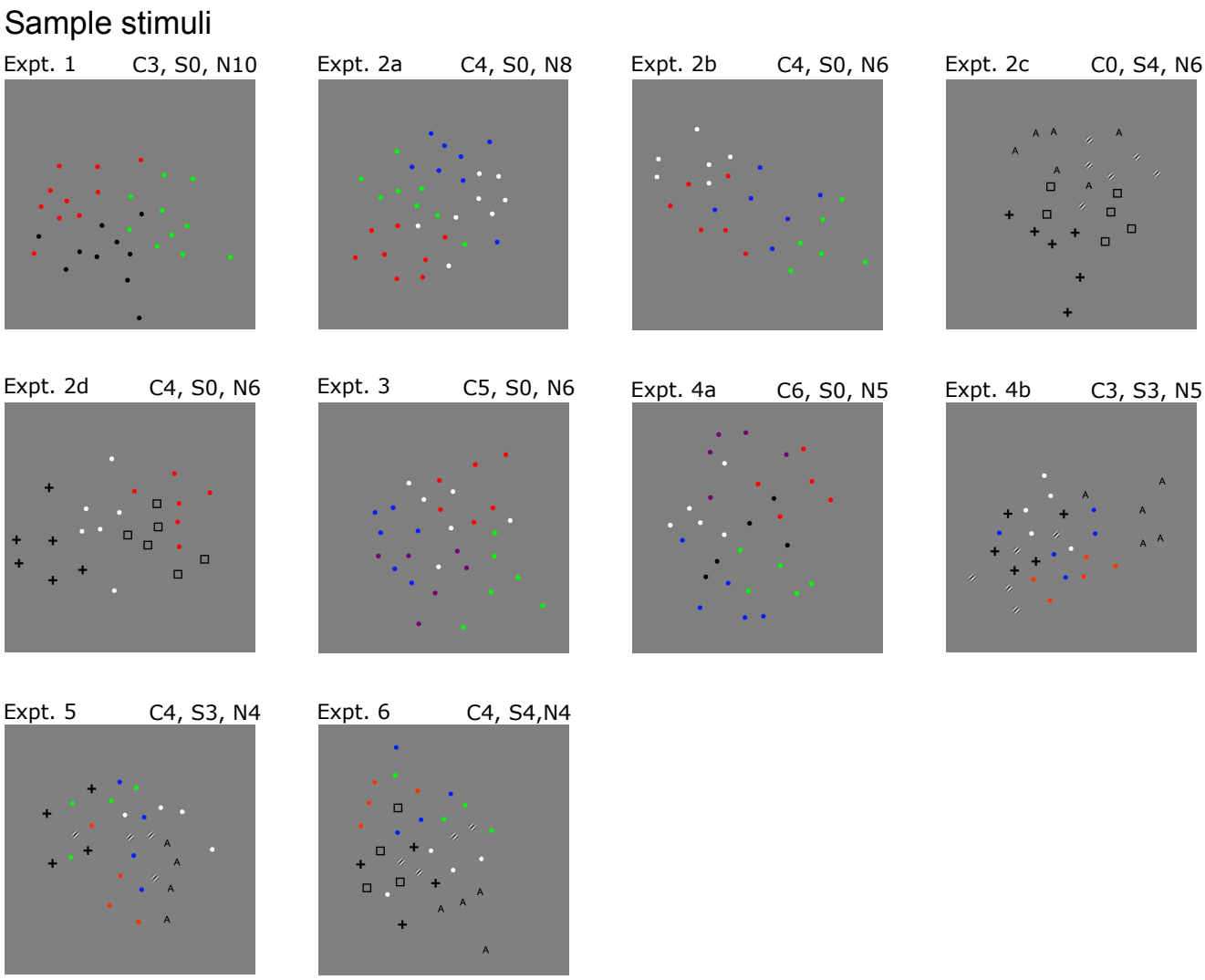



Figure B.1: Sample stimuli for each experiment. 'C' represents the number of different colors of dots in the stimulus display, 'S' represents the number of different shapes, and 'N' represents the number of items in each feature class.

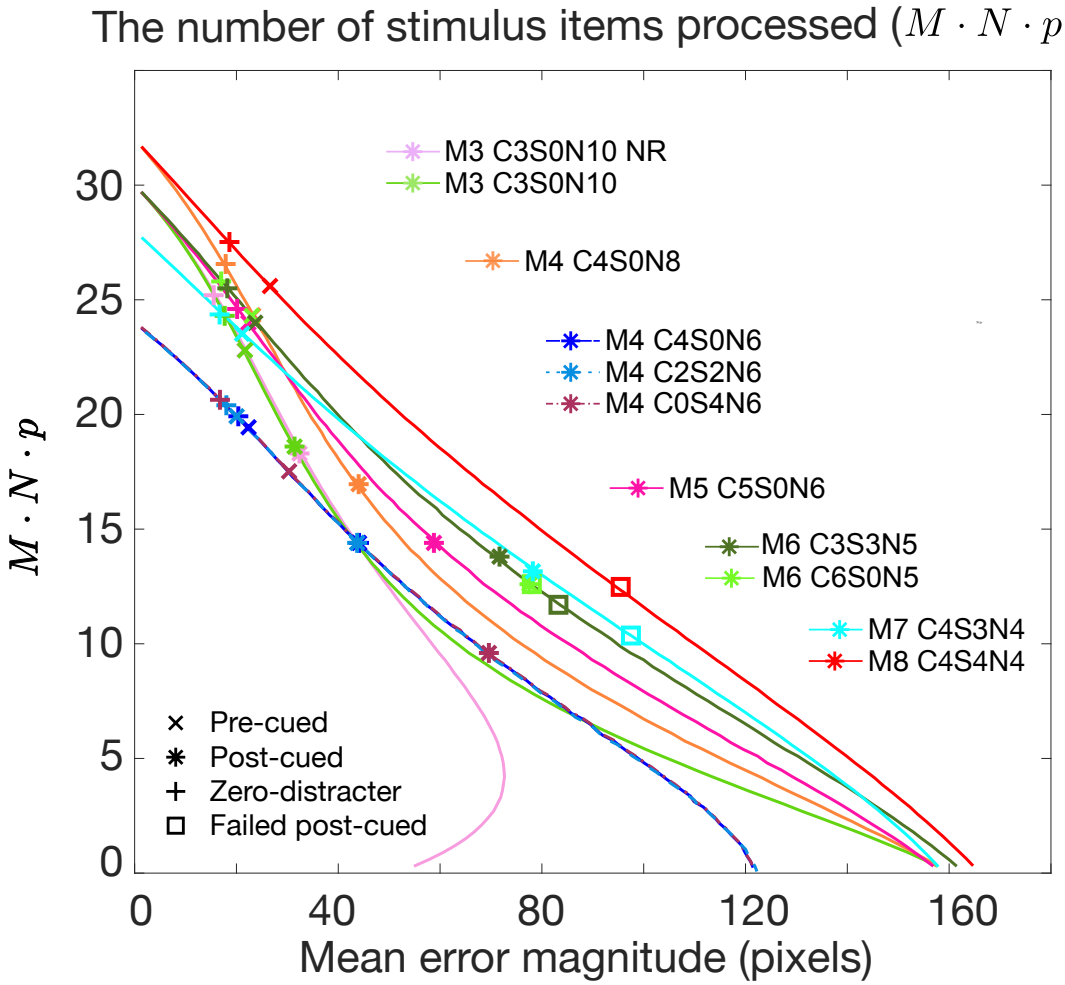



Figure B.2: The number of stimulus items processed by an ideal observer (ordinate) as a function of the mean response error magnitude (abscissa) for 3 kinds of trials in 11 experiments. Text figure 4b shows $N \cdot p$; figure S2 shows $M \cdot N \cdot p$ . The data points show $M \cdot N \cdot p$, the least number of stimulus items that must be processed by an ideal observer with a perfect target filter to match a subject's performance. The colored lines represent the number of stimulus items processed by an ideal observer as a function of the mean response error magnitude for each of the 11 experiments that used 8 different stimulus configurations. Data points indicated by 'x', '$*$', and '$+$' represent the mean error magnitude averaged over subjects' data from an $M$-feature experiment in which performance was better than could be predicted from the corresponding *(M-1)*-feature experiment. Squares represent post-cue data from an $M$-feature experiment that failed to exceed the prediction from the corresponding *(M-1)*-feature experiment.

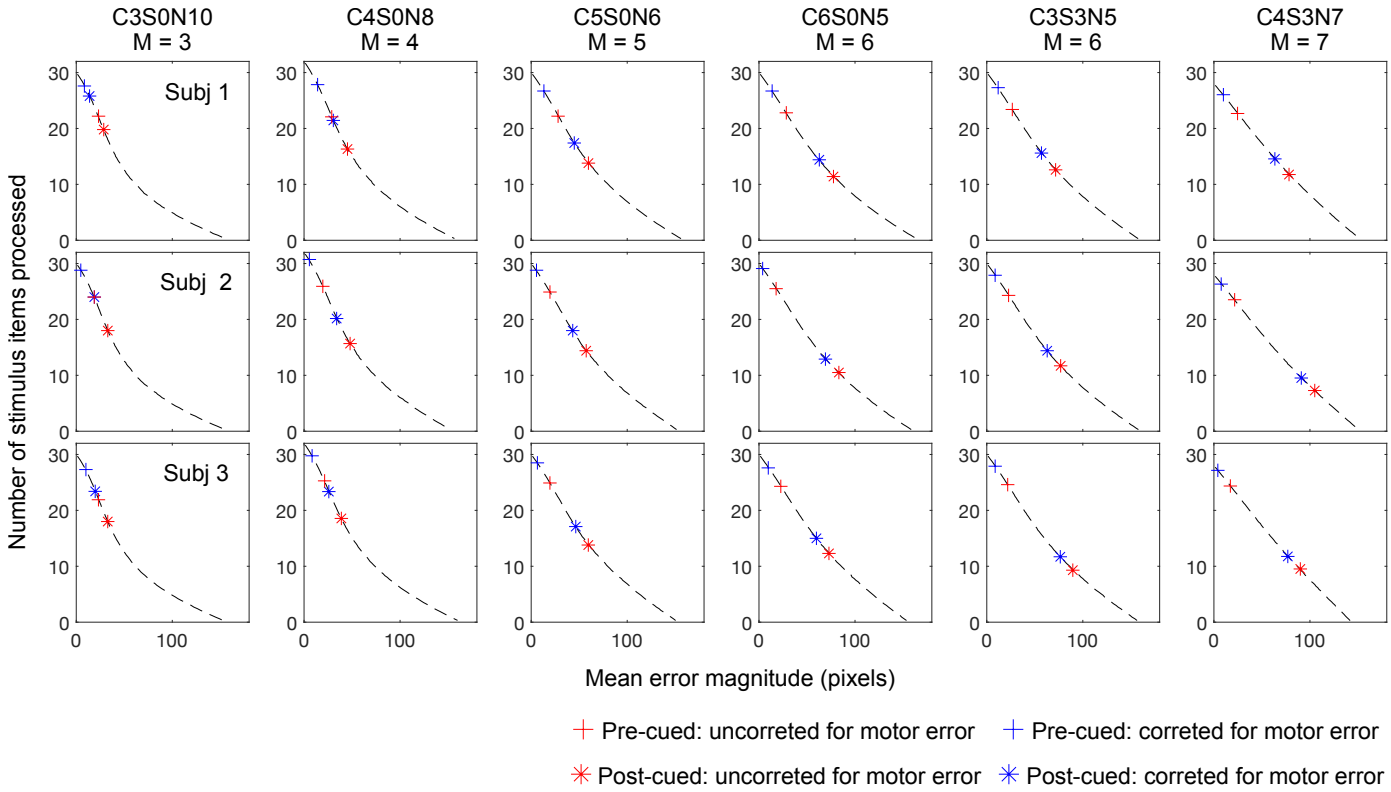



Figure B.3: Number of stimulus items processed in multi-item pre-cued and post-cued trials, uncorrected and corrected for motor error. Each row represents a subject and each column represents one experiment. The dashed black curve in each panel represents the number of stimulus items processed by an ideal observer (ordinate) as a function of the mean response error magnitude (abscissa). The data points show $M \cdot N \cdot p$, the least number of stimulus items that must be processed by an ideal observer with a perfect target filter to match a subject's performance. Red points represent the data uncorrected for motor error Blue points represent data corrected for motor error. '$*$' and '$+$' represent multi-item post-cued and pre-cued trials.

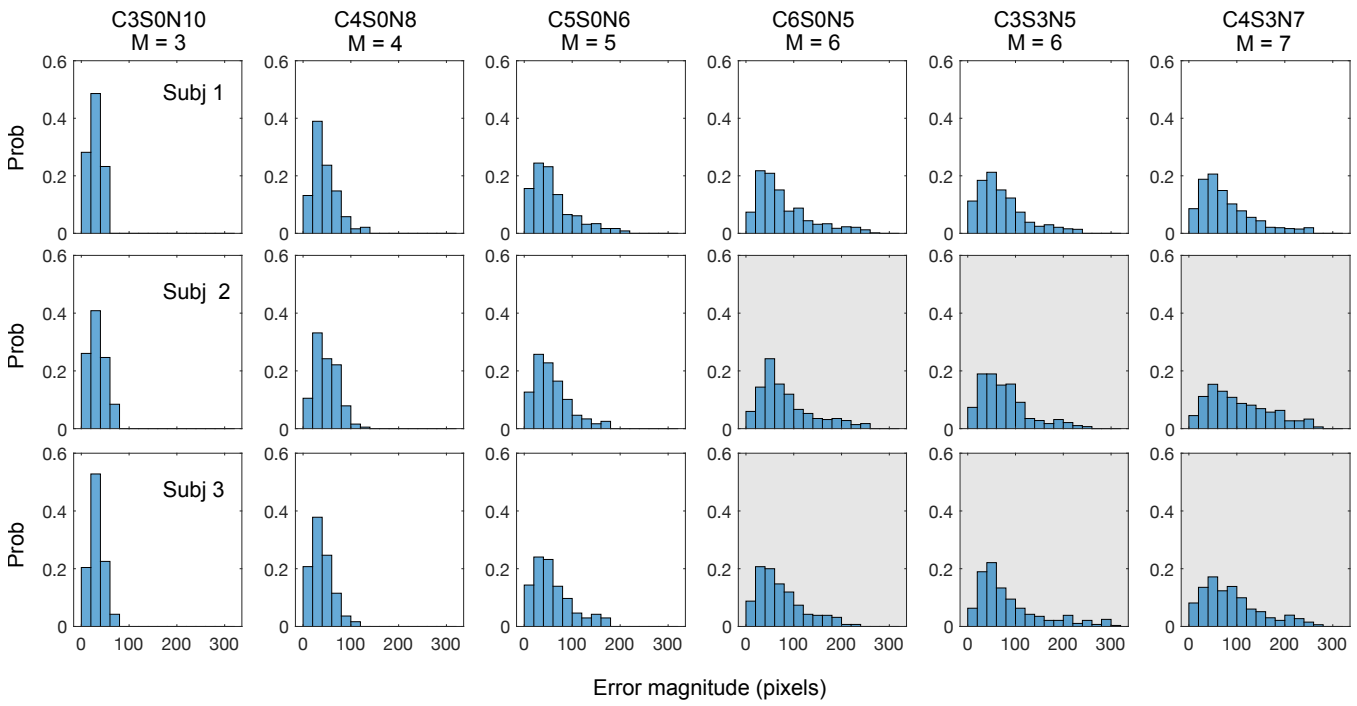



Figure B.4: Histogram plots of error magnitude of multi-item post-cued trials. Each row represents a subject and each column represents one experiment. A gray background indicates data failing to exceed the subsampling strategy prediction.

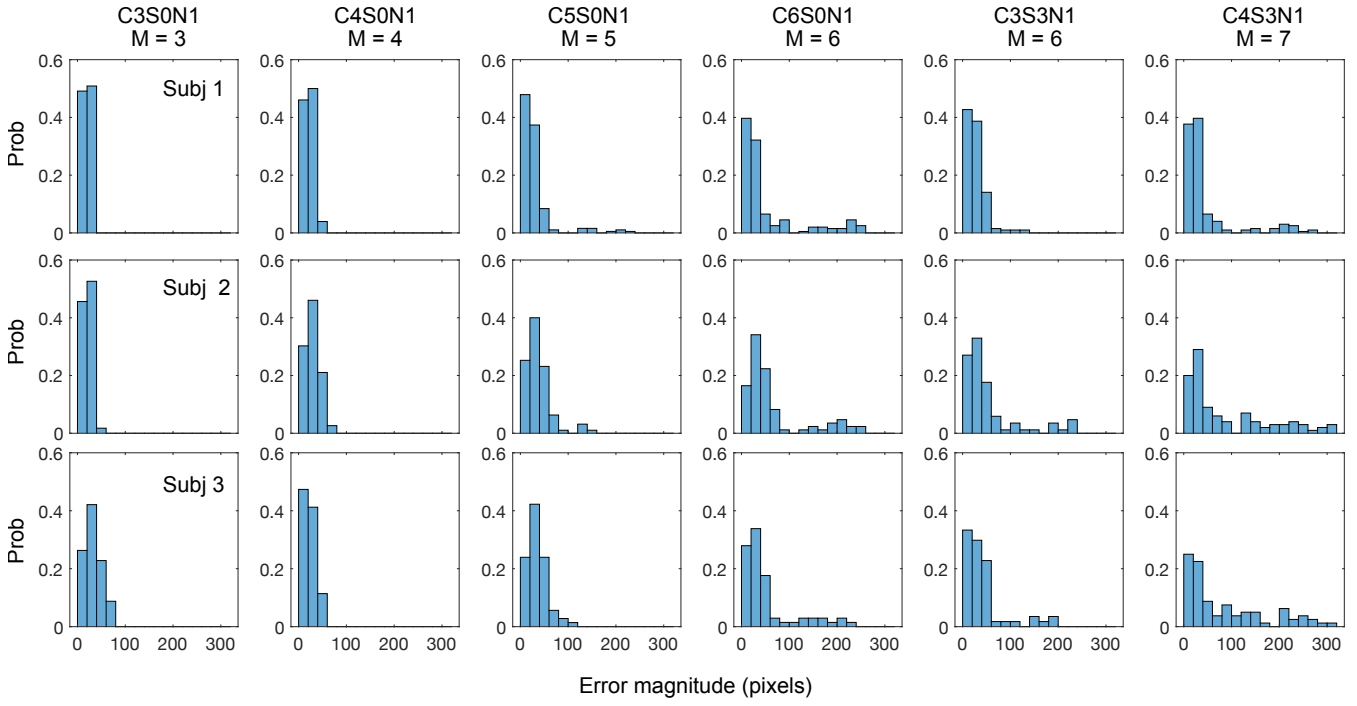



Figure B.5: Histogram plots of error magnitude of singletons post-cued trials for 3 subjects in six experiments. Each row represents a subject and each column represents one experiment.

# Chapter 3

# Chapter 3: Central fovea isoluminance

# Abstract

Luminance, introduced by the CommissionInternationale de l'Éclairage (CIE), is a perceptual analog of radiance that describes the visual effectiveness of a light. Isoluminance, a condition in which lights of different wavelength compositions have equal luminance, offers a way to evaluate the importance of luminance by studying alternative visual processing capabilities when luminance information is unavailable. The visual system has two early processing systems: the magnocellular pathway, which involves neurons having large receptive fields and being sensitive to rapid luminance changes, and the parvocellular pathway, which involves neurons with small receptive fields and poor temporal resolution but excel in high spatial resolution and color sensitivity. Typically, isoluminance has been determined almost exclusively under conditions that primarily stimulate the magno processing system (e.g. flicker photometry) although it often is applied in situations that activate the parvo system. This study exclusively measures parvo isoluminance. The stimuli are extremely fine-gratings focused on a tiny area in the most central fovea. This area is highly sensitive to both color and luminance and has the highest spatial frequency resolution, which is resolvable only by the parvo system. Subjects viewed high spatial-frequency yellow/red and yellow/green gratings and were instructed to judge the orientation of the gratings. The intensity of the red and green components varied over the full available range. This guarantees finding a red and a green intensity that is isoluminant with yellow. At isoluminant intensities, the luminance system could not distinguish yellow from red or from green and, therefore, could not contribute to the discrimination of grating orientation. Results: (1) For 10 male and 10 female subjects, there was no point at which grating performance was reduced to chance. (2) The minimal visibility points in the central fovea were not necessarily the isoluminant points. (3) Subjects were nearly always more sensitive to color contrasts than to luminance contrasts. (4) There were major between-subject accuracy differences in judging the orientation of yellow/red versus yellow/green gratings, with subject data falling into three distinctive

patterns. Conclusion: (1) For grating acuity in the central fovea, the luminance system has no measurable advantage over the color system. (2) Isoluminance for the multichromatic parvo system is completely different from the isoluminance for the monochromatic magno system which is determined by the common measures of isoluminance.

## 3.1 Introduction

### 3.1.1 What is isoluminance and why isoluminance is important?

Luminance, introduced by the CommissionInternationale de l'Éclairage (CIE), is a perceptual analog of radiance that describes the visual effectiveness of a light [1, 2, 3]. It is widely assumed that luminance is the primary source for representing spatial visual information. Isoluminance is a condition in which two lights of different compositions of wavelengths have equal luminance. Under isoluminant conditions, the visual system cannot rely on luminance differences to distinguish between the lights, forcing it to use other visual information such as color or texture. Isoluminance offers a way to evaluate the importance of luminance by studying visual processing capabilities when luminance information is unavailable and a way of studying alternative systems for processing visual information.

**How is isoluminance conventionally produced and evaluated?**

In 1924, CIE introduced the first version of the luminous efficiency as a function of wavelength, $V(\lambda)$, intended to describe the human eye's relative sensitivity to different wavelengths of light in terms of luminance [4]. The $V(\lambda)$ function represents the averaged data from many observers measured using heterochromatic flicker photometry.

In heterochromatic flicker photometry, a visual field is flickered between two lights at a frequency typically between 10 and 20 Hz [5]. The task for subjects is to adjust the relative radiance intensity of the two lights to minimize the perception of flicker. When the flicker is minimally visible, it is assumed that the two lights are isoluminant. Flicker at a frequency above 10 Hz is predominantly detected by the magnocellular pathway, which is highly sensitive to rapid changes in luminance and insensitive to color [6, 7, 8]. Therefore, when the luminance difference between the two lights is minimized, the magnocellular path-

way is ineffective at detecting flick. The ratio of the radiance of the two lights generating the minimally visible flicker is inversely proportional to the relative sensitivity of the subject's eyes to these different lights, providing a measure of their luminance sensitivity as a function of wavelength.

One major advantage of heterochromatic flicker photometry is that it adheres to Abney's law, meaning the luminance of a multi-wavelength light is the weighted sum of the luminance of its individual wavelengths. In addition, the luminous efficiency function measured by heterochromatic flicker photometry can be approximately by the weighted sum of the Stockman-Sharpe's L-cone and M-cone cone fundamentals[9, 3, 10]. Several adjustments have been made since 1924 primarily to correct $V(\lambda)$ at the short-wavelength end of the spectrum [11, 12].

Aside from heterochromatic flicker photometry, minimum motion, and minimally distinct border are the most often used methods to measure isoluminance.

In minimum motion, subjects view two superimposed gratings that shift in phase 90 deg at a frequency between 2 Hz to 15 Hz [13]. One grating, for example, is made of red and green, and the other is made of light gray and dark gray. Subjects adjust the relative radiance of the red and green lights to minimize the perception of motion. When the minimal motion is achieved, the red and green are defined to be isoluminant for that subject [13]. The advantage of the minimum motion compared to heterochromatic flicker photometry is that it can be applied across a wider range of temporal frequencies, down to 1 Hz and up to 20 Hz, and its measurements are invariant to temporal frequency [13]. At higher temporal frequencies, the magnocellular pathway dominates motion perception. Since the magnocellular pathway is now thought to be exclusively sensitive to luminance rather than color [7, 8], the minimal motion method effectively isolates luminance differences. At lower temporal frequencies, both magnocellular and parvocellular pathways contribute. However, the task of minimizing motion still predominantly relies on luminance contrast. Heterochromatic flicker photometry,

on the other hand, is not suitable for temporal frequencies below 7.5 Hz because, at low temporal frequencies, the parvocellular pathway gets involved and detects chromatic flickers. As a result, flicker perception is influenced by the combined effect of chromatic and luminance differences.

In minimally distinct border, two patches composed of different lights are placed side by side. The task in this paradigm for subjects is to adjust the relative radiance of the two lights until the boundary between them is minimized. When this minimal boundary is achieved, the two lights are thought to be of equal luminance. lights[14]. In this paradigm, the two patches usually have different chromaticity, and adjusting the radiance of the lights does not change their chromaticity. In other words, the chromatic difference between the patches always exists and remains relatively constant; it is the luminance difference between the two patches that dominates the perception of the boundary. When the luminance of the two patches matches, the boundary visibility is minimized.

### 3.1.2 What is the problem with conventional measurements of isoluminance?

Isoluminance measures vary widely across subjects, and different measurement methods yield different isoluminant settings[1, 15, 16]. This variation is attributed to several factors, including but not limited to the dramatic variation in the ratio of L and M cones and spatial arrangement among individuals[17, 18, 19, 20], the different contributions of L, M, and S cones to various tasks, different visual fields [21] and different chromatic adaptations[22].

He et al.[16] compared the heterochromatic flicker photometry, minimal motion, and minimally distinct border methods on the same group of 24 subjects, including 2 experienced and 22 naive subjects. They found that heterochromatic flicker photometry and minimal motion had low within-subject variability and high test-retest reliability. In contrast, mini-

mally distinct border showed higher within-subject variability and low test-retest reliability. Heterochromatic flicker photometry has low between-subject variability. In addition, the L:M contrast ratio was lower in heterochromatic flicker photometry and minimal motion compared to minimally distinct border.

Koenderink et al.[15] compared 8 different paradigms, including heterochromatic flicker photometry, minimally distinct border, and a modified version of minimal motion, on 17 subjects (10 naive subjects and 7 subjects with some knowledge of colorimetry). Their results showed that minimal motion and minimally distinct border yielded similar measures, but were significantly different from the measures from heterochromatic flicker photometry.

The results from both studies demonstrate that minimally distinct border and heterochromatic flicker photometry yield different isoluminant measures even when tested on the same subjects, indicating that these two paradigms assess the luminance of distinct visual pathways. These empirical results are consistent with the physiological basis of visual input processing from the retina to the cortical areas. The physiological basis suggests the existence of at least two parallel luminance mechanisms: one associated with the parvocellular pathway and the other with the magnocellular pathway.

The typical frequency of stimuli used in heterochromatic flicker photometry ranges from 10 to 20 Hz[5], which is thought to be processed by the magnocellular pathway [6]. Due to the characteristics of the neurons in the magnocellular pathway, it is sensitive to rapid changes in luminance but insensitive to color. Therefore, when the stimulus in heterochromatic flicker photometry or minimal motion is at high temporal frequencies, these paradigms predominantly measure the luminance mechanism associated with the magnocellular pathway.

In contrast, in minimally distinct border, the stimuli are static, which are primarily processed by the parvocellular pathway[7, 8], thereby measuring parvo isoluminance. The inherent complexity of the task in minimally distinct border can also lead to large within- and between-

subject variabilities. One significant source of this variability is the vagueness of the minimal boundary definition. There is no well-defined criterion that all subjects can consistently follow, and even the same subject may change their "minimal boundary" criterion from trial to trial. Consequently, the lack of a consistent criterion can result in significant within- and between-subject variabilities in minimally distinct border measurements.

### 3.1.3 Parvocellular pathway VS Magnocellular pathway

The magnocellular and parvocellular pathways process visual information in parallel, each with distinct characteristics[7, 8, 6]. The magnocellular pathway involves neurons of large receptive fields and high temporal resolution. The parvocellular pathway involves neurons with small receptive fields, poor temporal resolution, but high spatial resolution and color sensitivity. The next section includes a brief overview of how signals travel from the cones to neurons in the lateral geniculate nucleus (LGN) and describes the characteristics of neurons in the parvocellular and magnocellular pathways.

#### Retinal Processing: from cone to ganglion cells via bipolar cells

Humans with normal color vision have three classes of light-sensitive cones, each peaking in sensitivity at different wavelength regions. L cones peak at 558.4 $\pm$ 5.2 nm, M cones peak at 530.8 $\pm$ 3.5 nm, and S cones peak at 419.0 $\pm$ 3.6 nm [23].

In the retina, the output of the cones is relayed to several classes of bipolar cells, with this output being modulated by horizontal cells. Horizontal cells receive input nonselectively from nearby L and M cones and provide lateral inhibition to bipolar cells[24, 25]. Consequently, the receptive field of bipolar cells consists of an excitatory region that receives direct input from cones and an antagonistic surround region that receives inhibitory input from horizontal

cells. These two regions are arranged concentrically, with the excitatory region at the center surrounded by the antagonistic region.

Bipolar cells project to different classes of ganglion cells, which in turn primarily project to distinct layers in the lateral geniculate nucleus (LGN). Both ganglion cells and LGN cells have similar center-surround receptive field characteristics as bipolar cells [26, 27, 28, 29, 30].

Most bipolar and ganglion cells can be categorized as either ON or OFF cells. ON cells respond vigorously when the center of their receptive field is illuminated, while OFF cells respond vigorously when the light is removed from the center of their receptive field. ON bipolar cells only project to ON ganglion cells, and so do the OFF bipolar cells. For both bipolar cells and ganglion cells, the dendritic fields and receptive fields increase as the increase of retina eccentricity [31].

The ON and OFF bipolar can be further categorized into diffuse bipolar cells and midget bipolar cells. Midget bipolar cells have a smaller receptive field than diffuse bipolar cells at any eccentricity. The receptive field center of midget bipolar cells receives input from one to a few cones, while the receptive center of diffuse bipolar cells connects to multiple cones.

Two major classes of ganglion cells are midget ganglion cells and parasol ganglion cells, both of which have ON and OFF subtypes. Midget ganglion cells receive input from midget bipolar cells, and parasol ganglion cells receive input from diffuse bipolar cells.

Parasol ganglion cells have approximately three times larger receptive field in diameter than midget ganglion cells at any retinal eccentricity[32, 33]. As the distance from the central fovea increases, the number of cones that converge on both midget ganglion cells and parasol ganglion cells increases. Parasol ganglion cells connect to tens to hundreds of diffuse bipolar cells. By contrast, midget ganglion cells connect to one to a few midget bipolar cells. In the central fovea, a one-to-one connectivity exists wherein each cone connects to one ON midget bipolar cell and one OFF midget bipolar cell. Each ON midget bipolar cell connects to one

ON midget ganglion cell, and each OFF midget bipolar cell connects to one OFF midget ganglion cell. In other words, in the central fovea, midget ganglion cells get excitatory input from a single cone through midget bipolar cells. Parasol ganglion cells receive excitatory input from 40 to 140 cones [31].

Besides the difference in the size of receptive fields, parasol ganglion cells have larger axons, hence having faster conduction velocities than the midget ganglion cells.

**From ganglion cells to LGN**

Parasol ganglion cells send projection to the magnocellular layers of the LGN, where the neurons have larger cell bodies, while midget ganglion cells project to the parvocellular layers of the LGN, where the neurons have smaller cell bodies [7]. The receptive field characteristics of LGN neurons closely resemble those of the ganglion cells that project to them.

**Different characteristics of Parvo and Magno pathway**

Due to differences in receptive fields and conduction velocities, the parvocellular pathway and the magnocellular pathway exhibit varying sensitivities in both spatial and temporal dimensions.

**Temporal sensitivity** Parasol ganglion cells have larger axons and faster conduction velocities, making the magnocellular pathway more sensitive to stimuli with high temporal frequencies. Magnocellular cells have the highest sensitivity when the stimulus has a temporal frequency near 20 Hz [6]. On the contrary, parvocellular cells respond strongly to static stimuli and stimuli with a temporal frequency of up to 10Hz. Parvocellular cells show a steep sensitivity decline when the stimuli's temporal frequency is above 10 Hz[6, 8].

**Spatial sensitivity** Diffuse bipolar cells and parasol ganglion cells have large receptive fields, with both their excitatory centers and inhibitory surrounds receiving nonselective input from both L and M cones. This configuration makes the magnocellular pathway sensitive to low spatial frequency stimuli with luminance contrast and insensitive to color. In contrast, midget bipolar cells and midget ganglion cells have much smaller receptive fields at any eccentricity, receiving excitatory input from one to a few cones. In the fovea, the excitatory center of midget bipolar cells and midget ganglion cells receives input from a single L cone or M cone, while the inhibitory surround receives nonselective input from nearby L and M cones [34, 35]. This configuration renders the parvocellular pathway highly sensitive to both color and luminance in the fovea[36]. The high sampling rate of the cone mosaic in the fovea allows the parvocellular pathway to resolve stimuli with a high spatial frequency of up to 40 cycles per degree [6]. The sensitivity to color and high spatial frequency stimuli decreases because as moving away from the fovea, the receptive fields of midget bipolar cells and midget ganglion cells increase.

### 3.1.4 A new paradigm to measure parvo isoluminance

In the literature, it is not uncommon for isoluminance measures derived from minimal motion or heterochromatic flickers to be used as the isoluminant settings for tasks primarily involving the parvocellular pathway (e.g. [37, 38, 39]). However, as discussed earlier, at least two distinct luminance mechanisms exist in the visual system. The isoluminant settings for the magnocellular pathway, which is sensitive only to luminance, not color, do not necessarily match those for the parvocellular pathway, which is sensitive to both color and luminance. This distinction is supported by the empirical results from He et al. [16] and Koenderink et al. [15] which show significant variations in isoluminance settings depending on whether the task engages the magnocellular, parvocellular pathway or both.

Minimally distinct border preferentially measures parvo isoluminance; however, several factors affect its reliability. First, the task in minimally distinct border is a type 2 task, meaning the subject's responses cannot be defined as correct or incorrect[40]. Second, the lack of a clear definition of the minimal boundary further complicates the measurements, leading to significant variability and reduced reliability in the results.

In this study, an extremely fine-grating paradigm was used to exclusively measure parvo isoluminance, focusing specifically on the most central fovea which is highly sensitive to both color and luminance and has the highest spatial frequency resolution. In this paradigm, subjects foveally viewed high spatial-frequency yellow/red and yellow/green gratings. Subjects were instructed to judge the orientation of the gratings. The intensity of the red and green components was varied over the full available range to determine the points of minimal visibility. For the yellow/red gratings, the red components of the grating were modulated only by the red primary of the monitor, while the yellow components remained constant. Similarly for the yellow/green gratings, only the green primary was varied. Thus, varying the intensity of the red and green components only modulated their luminance, leaving their chromaticity unchanged, except when the components were extremely dim, and their chromaticity could not be distinguished.

Given that the yellow used in the experiment was composed of 20% of the maximum output of the green primary and 30% to 50% of the maximum output of the red primary (subjects had different unique yellow), varying the intensity of the red and green components over the full available range guarantees finding a red and a green intensity that is isoluminant with the yellow. At isoluminant intensities, the luminance system can not distinguish yellow from red or from green and, therefore, can not contribute to the discrimination of grating orientation. In the central fovea, if luminance were the primary source of representing visual information, a dramatic drop in performance would be expected when the luminance of red or green components matches the luminance of yellow as the luminance system can not

contribute, and only chromatic information would be available for orientation judgments. However, if the luminance were not the dominant system in the central fovea, the drop in accuracy would occur at a point that is not the isoluminance point. For the green/yellow gratings, the isoluminant green intensity should be greater than the intensity of the green primary that composed the yellow. Same for the red/yellow gratings.

The foveally-viewed-fine-gratings paradigm offers a way to evaluate the relative importance of luminance and chromatic systems in the extreme central fovea. In addition, in the fine-grating paradigm, the task of identifying a grating's orientation is a type 1 task, in which the subject's response can be clearly defined as correct or not [40]. Therefore the measurements are more objective and less prone to subjective interpretation. This approach aims to reduce variability and provide a more accurate assessment of parvo isoluminance.

The stimulus is designed to exclusively involve the parvocellular pathway and to target the most color-sensitive region with the highest spatial resolution in the retina. The choice of this stimulus is based on the following factors:

(1) Ganglion Cell Distribution: In the fovea, midget ganglion cells constitute 90% to 95% of all ganglion cells [32], while parasol ganglion cells account for about5% [33]. In the very center of the fovea, parasol ganglion cells are nearly absent[6, 41]. As eccentricity increases, the proportion of parasol ganglion cells increases, with midget ganglion cells comprising 40% to 45% of total ganglion cells [32] and parasol ganglion cells making up 20% in the periphery[33].

(2) Connectivity in the fovea: In the fovea, nearly all midget ganglion cells receive excitatory input from a single midget bipolar cell, which in turn receives excitatory input from a single cone [42, 43, 44]. This one-to-one connectivity results in high color sensitivity and spatial resolution of the parvocellular pathway in the foveal region.

(3) Spatial Frequency Resolution: The grating's spatial frequency of 28.6-36.8 cycles per

degree (depending on the subject) exceeds the resolving capability of the magnocellular pathway but falls within the range of the parvocellular pathway, which can resolve stimuli with spatial frequencies up to 40 cycles per degree in the central fovea [6].

(4) The dominant view holds that S cones have little impact on luminance perception [45, 46, 47]. However, substantial evidence suggests that S cones may have a negative effect on luminance [48]. To avoid complications related to the potential impact of S cones on luminance, only the red and green primaries were used in the experiment. Instead of using plain gray as the standard light, each subject's unique yellow was used, measured individually for each subject (details provided in section 3.2.3). Additionally, the grating spanned only 0.15 to 0.19 degrees of visual angle (dva) in the central fovea, where very few S cones are present[17].

(5) To minimize the effect of temporal transients, the stimulus was presented with a raised sinusoidal temporal profile (Fig 2d).

In addition to measuring parvo isoluminance, the corresponding magno isoluminance was measured by a motion reversal paradigm described in Ortiz, Chubb, Wright, Sun, and Sperling [49].

## 3.2 Method and Procedure

### 3.2.1 Subjects

Twenty subjects (10 females and 10 males) participated in this study, including the first author. All subjects had normal or corrected-to-normal vision and normal color vision.

Except for two subjects (LG and JR), all subjects had limited knowledge of colorimetry

and were unaware of the study's hypothesis. Five subjects (LG, JR, DH, MB, and ZZ) had prior experience with psychophysical experiments, while the remaining subjects had no such experience.

All subjects viewed the displays binocularly in a room with controlled and consistent lighting conditions to produce surface luminances not too different from the average monitor luminance.

This study was approved by the University of California, Irvine's Institutional Review Board.

### 3.2.2 Apparatus

A Dell Ultrasharp UP 2720Q LCD monitor (3840 x 2160, 60 HZ) with Adobe 1998 D65 colorspace was used to present stimuli in the experiment. Figure 3.1a shows the relative spectral power of the primaries, which was measured by an Ocean Optics USB2000 spectrometer setup with a collimating lens (74-VIS Single lens) positioned 25 cm from the surface of the monitor.

The maximum output of red, green, and blue primary guns of the monitor had CIE 2012 xyY color coordinates of Red = (0.6357, 0.3386, 226.67), Green = (0.2080, 0.7163, 406.93), and Blue = (0.1387, 0.0709, 63.39). The white point's xyz coordinates are (0.313,0.327,0.360). Figure 3.1b shows the xy chromaticity diagram of the monitor in the CIE 2012 color space. [1] The experiment was programmed in MATLAB using PsychToolBox [50, 51] on a Mac Pro 2015.

[1]The calibration and measurements of the spectral power of monitor primaries were conducted by Dr. Jordan Rashid.

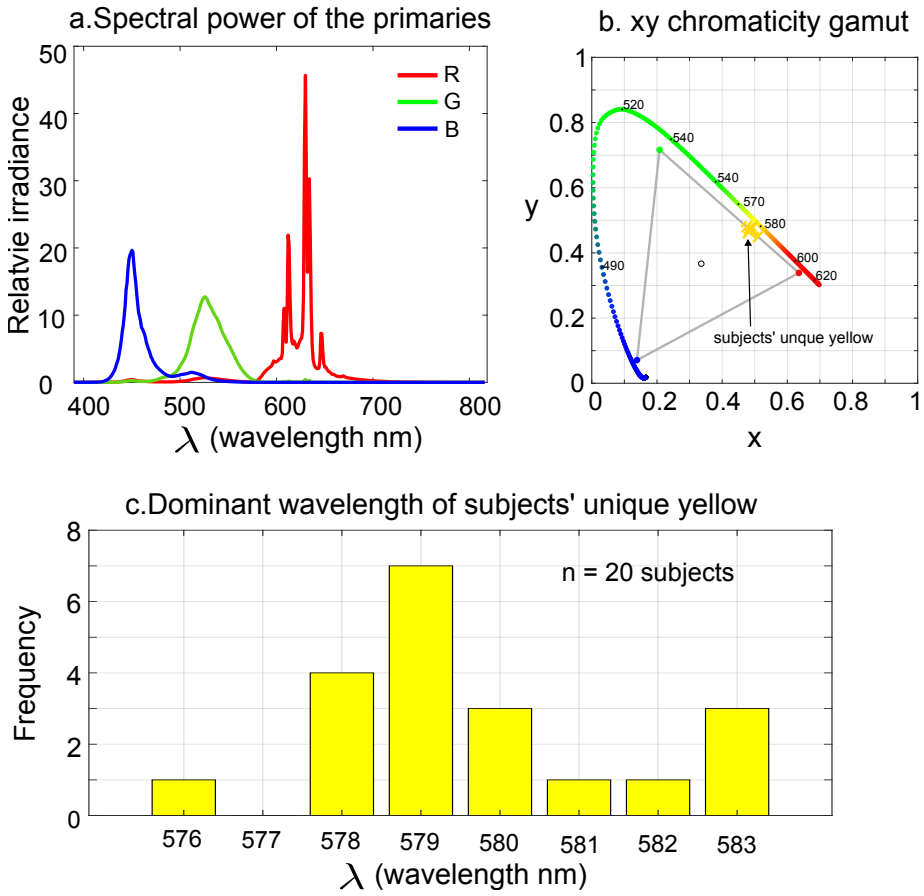



Figure 3.1: Display parameters of the monitor and equivalent spectral wavelengths of subjects' unique yellow. (a) The relative spectral power distribution as a function of wavelength for the primaries of the Dell Ultrasharp UP 2720Q LCD monitor. (b) xy chromaticity diagram of the monitor in the CIE 2012 color space and 20 subjects' unique yellow. (c) Frequency plot of the dominant wavelength of subjects' unique yellow.

### 3.2.3 Calibration- Measuring subjects' unique yellow

Before the formal experiment, a unique yellow was measured individually for each subject such that the yellow appeared neither red nor green. To measure a subject's unique yellow, the subject was presented with a green circular field centered on the screen (with only the green primary on, fixed at 20% of the full power of the green primary), while the rest of the screen remained black. The subject was instructed to adjust the intensity of the red primary until the center circular field appeared neither green nor red. The size of the circular field was the same as the yellow circular field in the formal experiment, which spanned 12.6 dva when viewed from a distance of 153 cm. This adjustment was repeated 3 times.

To obtain a rough estimate of the intensity that generates minimal visibility of the chromatic gratings in the central fovea, four rows of Snellen letters were displayed against the unique yellow background, with the size of the letters increasing from the top row to the bottom. All letters had the same color setting, with either only the green primary gun or only the red primary gun on. Subjects were instructed to adjust the intensity of the green primary (or the red primary) to find the intensity level at which the letters on the second row were most difficult to read. The size of the letters on the second row was 30 by 30 pixels (0.18 dva), similar to the size of the gratings in the formal experiment. The rough estimates of intensities that minimize the visibility of the letters later were tested in the formal experiment along with other intensity levels in a mixed-list design.

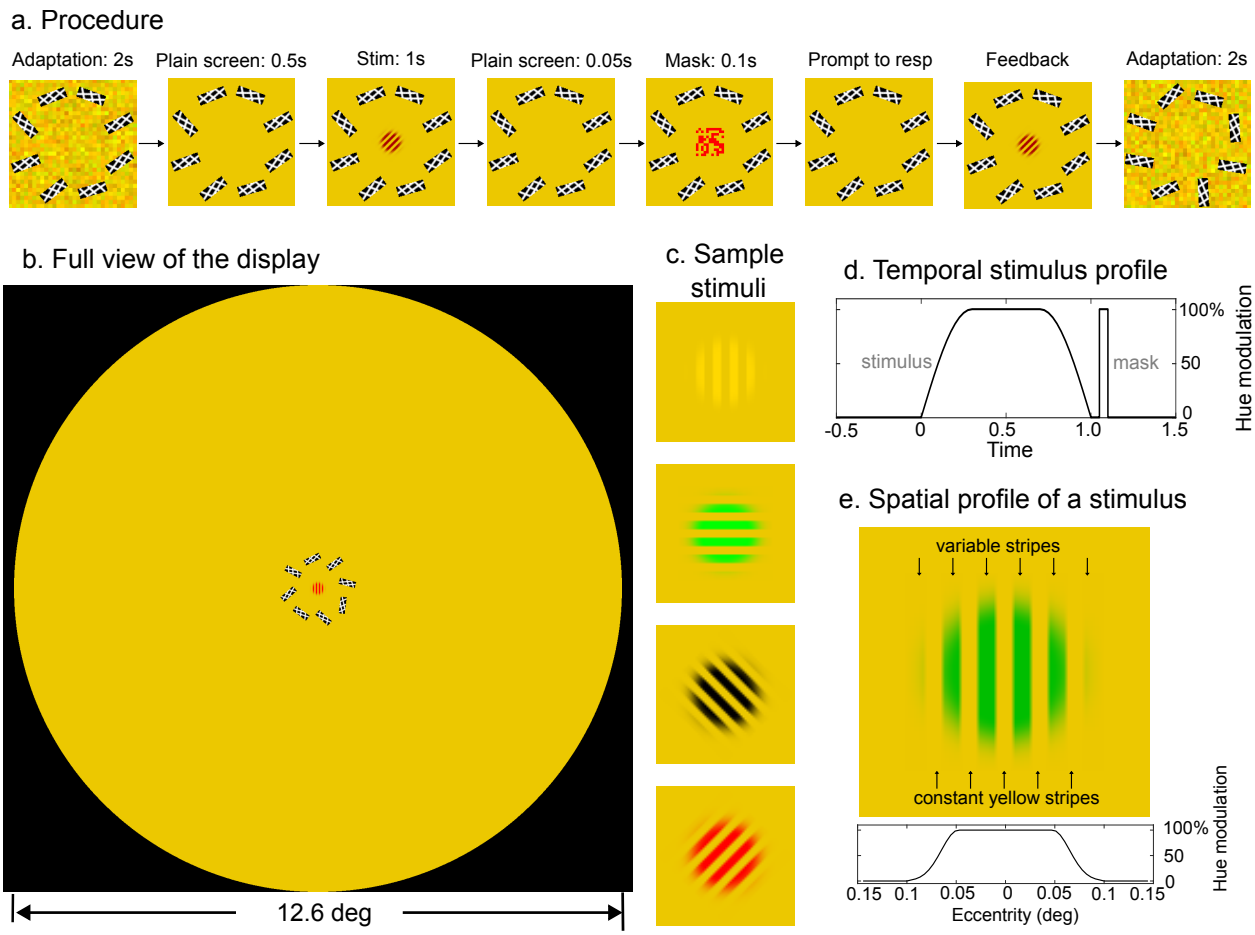



Figure 3.2: Procedure and stimuli for the foveal hyperacuity grating-orientation judgments. (a) Experimental procedure: A yellow disk and 8 fixation bars were continuously presented at the center of the screen, while the rest of the screen remained black. The yellow was previously measured individually for each subject such that it appeared neither reddish nor greenish to the subject. To help the subject focus on the plane of the stimuli, 8 randomly oriented textured bars encircled the critical display area. The orientations of the fixation bars were randomly rotated at the beginning of each trial. The subject initiated a trial by pressing any key. Immediately after initiation, a 2-second adaptation screen was shown, with random noise (N(0,40)) added to the yellow field. A 0.5-sec plain screen (without noise) immediately followed the adaption screen. Then, the stimulus, a grating, was shown for 1 sec at the center of the yellow field. A 0.05-second plain screen and 0.1-second mask followed the stimulus. The subject then was prompted to report the orientation of the grating (vertical, horizontal, tilted clockwise or titled counterclockwise) and their confidence level in their judgment (on a 1-to-5 scale, where 1 is not confident at all and 5 is very confident) by pressing keys on a keyboard. Feedback was provided at the end of each trial. b) Full view of the display. The yellow field spanned 12.6 dva and had a diameter equal to the width of the monitor screen. c) Four sample stimuli. d) Temporal profile of the stimulus. It took 0.3 sec to reach the full modulation, stayed at the full modulation for 0.4 sec, and decayed over 0.3 sec. e) Spatial profile of a stimulus. Each stimulus had a modified Gaussian spatial window, with maximum hue modulation in the central 0.1 dva.

### 3.2.4 Parvo isoluminance: Judging the orientation of fine gratings in the central fovea

Throughout the entire experiment session, a yellow circular field filled with the subject's unique yellow [2] along with eight fixation bars were continuously presented at the center of the screen, while the rest of the screen remained black. The centers of fixation bars were displayed at equal distances on an imaginary circle centered at the yellow circle field. The imaginary circle had an eccentricity of 0.46 degrees of visual angle (dva), and the yellow circle field spanned 12.6 dva when viewed from a distance of 153cm. Each fixation bar was randomly rotated for each trial. From this point forward, we will refer to the display consisting of the yellow circular field and fixation bars as the "plain screen."

#### Trial procedure

The subject pressed a key to initiate a trial. Immediately after initiation, a 2-second adaptation screen was displayed. The adaptation screen was identical to the plain screen except that random noise (N(0,40)) was added to the yellow field.

The purpose of having an adaptation screen with random noise appearing at the beginning of each trial was to have a stable adaptation state and prevent shifts in unique yellow from prolonged exposure to the spectrally altered environment. Supplement Fig. C.1 shows the accuracy of trials done in the first half of each session (solid lines) and the accuracy of trials done in the second half (dashed lines).

A 0.5-sec plain screen immediately followed the adaption screen. Then, the stimulus, a grating, was shown for 1 sec at the center of the yellow field. Each stimulus had a raised sinusoidal temporal profile to minimize the temporal transients, depicted in Fig. 3.1d. It took

[2] For illustration purposes, the yellow in Fig.3.2 is brighter than the yellow used in the experiment.

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Correct | 0.1 | 0.3 | 1 | 3 | 10 |
| Incorrect | 0 | -0.1 | -1 | -9 | -90 |

Table 3.1: Contingency of the confidence judgment

0.3 sec to reach the full modulation, stayed at the full modulation for 0.4 sec, and decayed over 0.3 sec. A 0.05-sec plain screen and a 0.1-sec plain screen with a mask followed the stimulus. The mask was presented over the stimulus region. After the mask, a plain screen with prompts appeared, instructing the subjects to report the orientation of the grating (vertical, horizontal, tilted clockwise or titled counterclockwise) and their confidence level in their judgment (on a 1-to-5 scale, where 1 is not confident at all and 5 is very confident) by pressing keys on a keyboard. The contingency of the confidence judgment is described in Table 3.1. Feedback was provided after the confidence judgment. On the feedback screen, the stimulus, the points for the current trial, and the accumulated points since the beginning of the session were displayed on the plain screen. subjects were instructed to try to maximize their total points for a possible reward.

**Stimuli**

The stimulus was a 5.5-cycle square wave grating modulated by a modified Gaussian envelope (Fig. 3.1e), presented at the center of a yellow circular field. The grating had an eccentricity of 0.10 dva when viewed from a distance of 153 cm from the screen. The even stripes of the grating were always the subject's constant yellow stripes, while the odd stripes were variable stripes and were modulated on a trial-by-trial basis.

There were three types of trials, one type of achromatic trials and two types of chromatic trials: (1) Yellow achromatic trials: The variable stripes had the same chromaticity as the subject's yellow, but of different contrasts. (2) Green chromatic trials: Only the green primary was turned on for the variable stripes. (3) Red chromatic trials: Only the red

primary was turned on for the variable stripes. Each type of trial included eight different intensities and four orientations, resulting in a 3x8x4 design. To minimize the potential complications that would be induced by possible attention modulation between conditions, a mixed-list design was used so that subjects had no cue about what the upcoming stimulus would look like.

In the yellow achromatic trials, the luminance of the variable stripes had eight different levels, ranging from very dark to the same luminance as the subject's yellow, and then to the brightest achievable yellow that matched the chromaticity of the subject's yellow. In the green and red chromatic trials, the primary color's intensity for the variable stripes took eight different values, ranging from very dark to very bright. On a linearized scale from 0 to 1, these intensity levels spanned the full range from 0.01 to 1 (maximum output).

Figure 3.3 shows sample gratings for each of the three types of trials. The top row, middle row, and bottom row represent the achromatic yellow trials, green chromatic trials, and red chromatic trials, respectively. Within each row, the eight different intensity levels are sorted in ascending order based on the sum of L and M cone excitation of the variable stripes. The variable stripes of the achromatic yellow/yellow gratings and of chromatic green/yellow gratings were paired so that for each pair, the sum of L and M cone excitation was the same. The monitor used in our experiment has a higher intensity output compared to standard monitors. We can use the normal monitor range to represent the yellow/yellow and yellow/green gratings, but it is impossible to accurately represent the two most intense yellow-red gratings, hence the diagonal black lines in Fig. 3.3.

**Representing stimuli in terms of their activation of retinal cones**

Cone excitation was calculated based on the cone fundamentals from Stockman and Sharpe (2000)[9]. The spectral sensitivity of L cones was rescaled for each subject so that L and

M cones have equal activation at the subject's unique yellow. The activation of L cones ($A_{L,S_j}$) and M cones ($A_M$) for an arbitrary color stimulus of intensity $I(\lambda)$ for a subject j is calculated as follows:

$$\begin{aligned} A_{L,S_j}(I(\lambda)) &= \sum_{\lambda} I(\lambda) \cdot L_{S_j}(\lambda) \\ A_M(I(\lambda)) &= \sum_{\lambda} I(\lambda) \cdot M(\lambda) \end{aligned} \tag{3.1}$$

where $L_{S_j}(\lambda)$ is the rescaled L cone fundamental for a subject j and $M(\lambda)$ is M cone fundamental, and $I(\lambda)$ is the spectral power of any arbitrary stimulus, defined as:

$$I(\lambda) = i_R \cdot R(\lambda) + i_G \cdot G(\lambda) + i_B \cdot B(\lambda) \tag{3.2}$$

where $R(\lambda)$, $G(\lambda)$, and $B(\lambda)$ represent the spectral power distributions of the red, green, and blue display device primaries (Fig. 3.1a), and $i_R$, $i_G$, and $i_B$ are the intensities of the red, green, and blue display device primaries, ranging on a 0 to 1 linearized scale. A value of 0 means the primary is off completely, and a value of 1 means the primary is set to its maximum output.[3]

Expanding the expressions for $A_{L,S_j}(I)$ and $A_M(I)$ yields:

[3]The information of the maximum output of each primary is in section 3.2.2

$$
\begin{aligned}
A_{L,S_j}(I(\lambda)) &= \sum_{\lambda} I(\lambda) \cdot L_{S_j}(\lambda) \\
&= \sum_{\lambda} (i_R \cdot R(\lambda) + i_G \cdot G(\lambda) + i_B \cdot B(\lambda)) \cdot L_{S_j}(\lambda) \\
&= i_R \sum_{\lambda} L_{S_j}(\lambda) \cdot R(\lambda) + i_G \sum_{\lambda} L_{S_j}(\lambda) \cdot G(\lambda) + i_B \sum_{\lambda} L_{S_j}(\lambda) \cdot B(\lambda) \\
&= A_{L,S_j}(i_R) + A_{L,S_j}(i_G) + A_{L,S_j}(i_B)
\end{aligned}
\tag{3.3}
$$

where

$$
A_{L,S_j}(i_c) = i_C \sum_{\lambda} L_{S_j}(\lambda) \cdot C(\lambda) \tag{3.4}
$$

$A_{L,S_j}(i_c)$ stands for the excitation of L cones at intensity i of $C$ for a subject j, where $C$ represents the red, green, or blue primary ($R$, $G$, or $B$).

Similarly, the excitation of M cones ($A_M$) is calculated as:

$$
\begin{aligned}
A_M(I(\lambda)) &= \sum_{\lambda} I(\lambda) \cdot M(\lambda) \\
&= \sum_{\lambda} (i_R \cdot R(\lambda) + i_G \cdot G(\lambda) + i_B \cdot B(\lambda)) \cdot M(\lambda) \\
&= i_R \sum_{\lambda} M(\lambda) \cdot R(\lambda) + i_G \sum_{\lambda} M(\lambda) \cdot G(\lambda) + i_B \sum_{\lambda} M(\lambda) \cdot B(\lambda) \\
&= A_M(i_R) + A_M(i_G) + A_M(i_B)
\end{aligned}
\tag{3.5}
$$

where

$$A_M(i_C) = i_C \sum_{\lambda} M(\lambda) \cdot C(\lambda) \tag{3.6}$$

and $C \in (R, G, B)$.

The intensity of the components of subject j's unique yellow $Y_{S_j}$ is $(i_{R,Y_{S_j}}, i_{G,Y_{S_j}}, 0)$. The L cone and M cone excitation for the subject j's unique yellow are denoted as $A_{L,S_j}(i_{Y_{S_j}})$ and $A_M(i_{Y_{S_j}})$.

In this experiment, the ratio between the cone excitation of the variable stripes and the constant yellow stripes matters. Table 3.2 lists the important symbols and their descriptions.

| **Symbols** | **Description** |
|---|---|
| $S_j$: | a subject j |
| $L_{S_j}(\lambda)$: | a subject j's re-scaled L cone spectral sensitivity |
| $M(\lambda)$: | M cone spectral sensitivity |
| $i_c$: | intensity i of the color primary C; $C \in (R, G, B)$, $0 \leq i \leq 1$ |
| $A_{L,S_j}(i_c)$ | L cone excitation at intensity i of C for a subject j |
| $A_M(i_c)$ | M cone excitation at intensity i of C |
| $Y_{S_j}$ | a subject j's unique yellow |
| $i_{R,Y_{S_j}}$ | the intensity of the red primary of subject j's unique yellow |
| $i_{G,Y_{S_j}}$ | the intensity of the green primary of subject j's unique yellow |
| $A_{L,S_j}(i_{Y_{S_j}})$ | L cone excitation of the subject j's unique yellow |
| $A_M(i_{Y_{S_j}})$ | M cone excitation of the subject j's unique yellow |
| $i_{R,mr}$ | the intensity of the red primary that cancels the motion perception in the motion reversal paradigm. |
| $i_{G,mr}$ | the intensity of the green primary that cancels the motion perception in the motion reversal paradigm. |

Table 3.2: Glossary of the important symbols

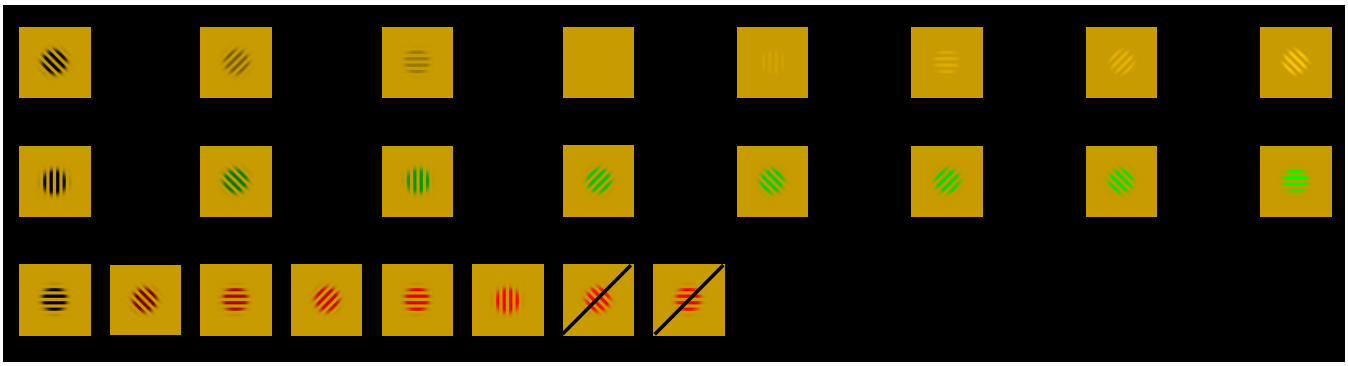


Figure 3.3: Examples of the 3 types of trials at eight different intensity levels. The top row, middle row, and bottom row represent the achromatic yellow trials, green chromatic trials, and red chromatic trials, respectively. Within each row, the eight different intensity levels are sorted in ascending order based on the sum of L and M cone excitation of the odd stripes. The monitor used in our experiment has a higher intensity output than standard monitors. The yellow color setting was amplified in Fig. 3.3 to better present it on standard monitors. Due to the monitor's limitations, we couldn't achieve the increased brightness within the available range for the red stripes of the last two red/yellow gratings. This limitation is indicated by the diagonal black lines.

### 3.2.5 Magno isoluminance: a motion reversal paradigm

A motion reversal paradigm was used to measure subjects' magno-isoluminant settings. The description of this particular version paradigm is in the method section of [49], with more details and the theoretical basis in Lu and Sperling (2001) [52] and Van Santen and Sperling (1984) [53].

In this paradigm, the intensity of the red primary was adjusted to match the subject's unique yellow, achieving a state in which a rotating wheel appeared to wobble without rotating distinctly clockwise or counterclockwise. Similarly for the green primary. We denote the intensity of the red primary and green primary that cancel motion-direction perception (and therefore are isoluminant with the neutral yellow) as $i_{R,mr}$ and $i_{G,mr}$.

| Subj Initial | x | y | Dominant wavelength (nm) | Subj Initial | x | y | Dominant wavelength (nm) |
|---|---|---|---|---|---|---|---|
| LG | 0.486 | 0.471 | 579 | ZZ | 0.477 | 0.479 | 578 |
| DH | 0.466 | 0.488 | 576 | AGK | 0.475 | 0.480 | 578 |
| MB | 0.486 | 0.471 | 579 | RM | 0.503 | 0.455 | 582 |
| BH | 0.485 | 0.472 | 579 | XL | 0.482 | 0.474 | 579 |
| MM | 0.498 | 0.461 | 581 | JM | 0.486 | 0.471 | 579 |
| JN | 0.478 | 0.478 | 578 | LZ | 0.493 | 0.465 | 580 |
| KN | 0.488 | 0.469 | 580 | JK | 0.489 | 0.468 | 580 |
| JR | 0.491 | 0.466 | 580 | SM | 0.510 | 0.450 | 583 |
| SA | 0.510 | 0.450 | 583 | NH | 0.482 | 0.474 | 579 |
| SX | 0.510 | 0.450 | 583 | DD | 0.475 | 0.480 | 578 |

Table 3.3: xy chromaticity coordinates and dominant wavelength of subjects' unique yellow

## 3.3 Results

### 3.3.1 Subjects' unique yellow

Table 2 lists the xy chromaticity coordinates and the corresponding dominant wavelength of each subject's unique yellow that was determined prior to the main experiment (Section 3.2.3). In Fig. 3.1b, each subject's unique yellow is indicated by a tiny yellow rectangle. The length of each bar represents the frequency of occurrence. Figure 3.1c shows the frequency distribution of the corresponding dominant wavelength of each subject's unique yellow. The dominant wavelength was determined on the CIE 2012 chromaticity diagram by extending a line from the white point to the locus of the unique yellow and continuing that line to intersect the spectral locus. The xy coordinates of the spectral locus of subjects' unique yellows spanned from 576 nm to 583 nm, with a mean of 579.5 nm and a standard deviation of 2.74 nm.

### 3.3.2 Main experiment: Orientation judgments of foveally viewed fine gratings

Except for 2 subjects (ZZ and XL) who completed 6 sessions, the other 18 subjects completed 9 sessions. Each session consisted of 288 experimental trials, where each of 3 grating types at 8 intensities and 4 orientations was tested 3 times in random order.

Subjects' performances were measured by the accuracy of their grating orientation judgments, averaged over the four orientations. Each panel of Fig. 3.4 shows data for just one subject. The three curves plot the accuracy for yellow-yellow, green-yellow, and red-yellow gratings as a function of the intensity of the variable color. More specifically, the y-axis is the fraction of correct orientation judgments averaged over the four orientations. The x-axis represents the ratio $r$ of the luminance of the variable color stripe divided by the luminance of the individual subject's constant yellow. Luminance for each subject is defined as the sum of their L and M cone excitations. The ratio of variable stripe luminance to constant yellow stripe luminance, $r$, is given by

$$r = \frac{A_{L,S_j}(i_C) + A_M(i_C)}{A_{L,S_j}(i_{Y_{S_j}}) + A_M(i_{Y_{S_j}})} \tag{3.7}$$

where $C$ is defined as follows: For green chromatic trials, $C = G$; for red chromatic trials, $C = R$; for yellow achromatic trials, $C = R + G$. $r < 1$ means that the variable stripes have less L+M excitation than the subject's unique yellow, indicating they were dimmer. Conversely, $r > 1$ means that the variable stripes have more L+M excitation than the subject's unique yellow, indicating they were brighter.

Luminance is defined as the sum of L and M cone excitations, with equal weights assigned to both L and M cones. This definition differs with that of Sharpe and Stockman [10], where nearly twice the weight is assigned to L cones compared to M cones. This difference in

exactly how luminance is computed does not change any results, it affects the alignment of the data in the graphs.

The vertical dashed red and green lines indicate where the intensities of the red ($i_{R,Y_{S_j}}$), and green ($i_{G,Y_{S_j}}$) primaries that composed the subject j's unique yellow fall on the x-axis.

The solid red line and green lines represent the measures of the motion reversal paradigm. Specifically, the red line represents the luminance ratio between the red primary point $i_{R,mr}$ that canceled motion perception and the subject's yellow. Similarly for the green solid line.

### Yellow achromatic trials

For the yellow achromatic trials, the odd and even stripes had the same chromaticity, so the only information available for orientation judgments was luminance contrast. The yellow achromatic trials enabled us to evaluate, in the central fovea, how sensitive subjects were to luminance contrasts when only luminance information was available. For all subjects, the yellow curves exhibited a typical V shape, with an expected minimum accuracy of 0.25 when $r = 1$. As $r$ approached 1, the subjects' performance dropped. When $r = 1$, the even and odd stripes were identical, resulting in invisible gratings, and the accuracy of 0.25 corresponds to the guess rate. The results of the yellow achromatic trials demonstrate that in the central fovea, when there was no chromatic information, the parvocellular pathway utilized luminance to distinguish different lights. The width of the yellow curve is inversely proportional to the efficiency of luminance information.

### Green and red chromatic trials

For the chromatic trials, the variable and constant yellow stripes differed in both chromaticity and luminance contrast. The variable stripes were modulated in such a way that only their

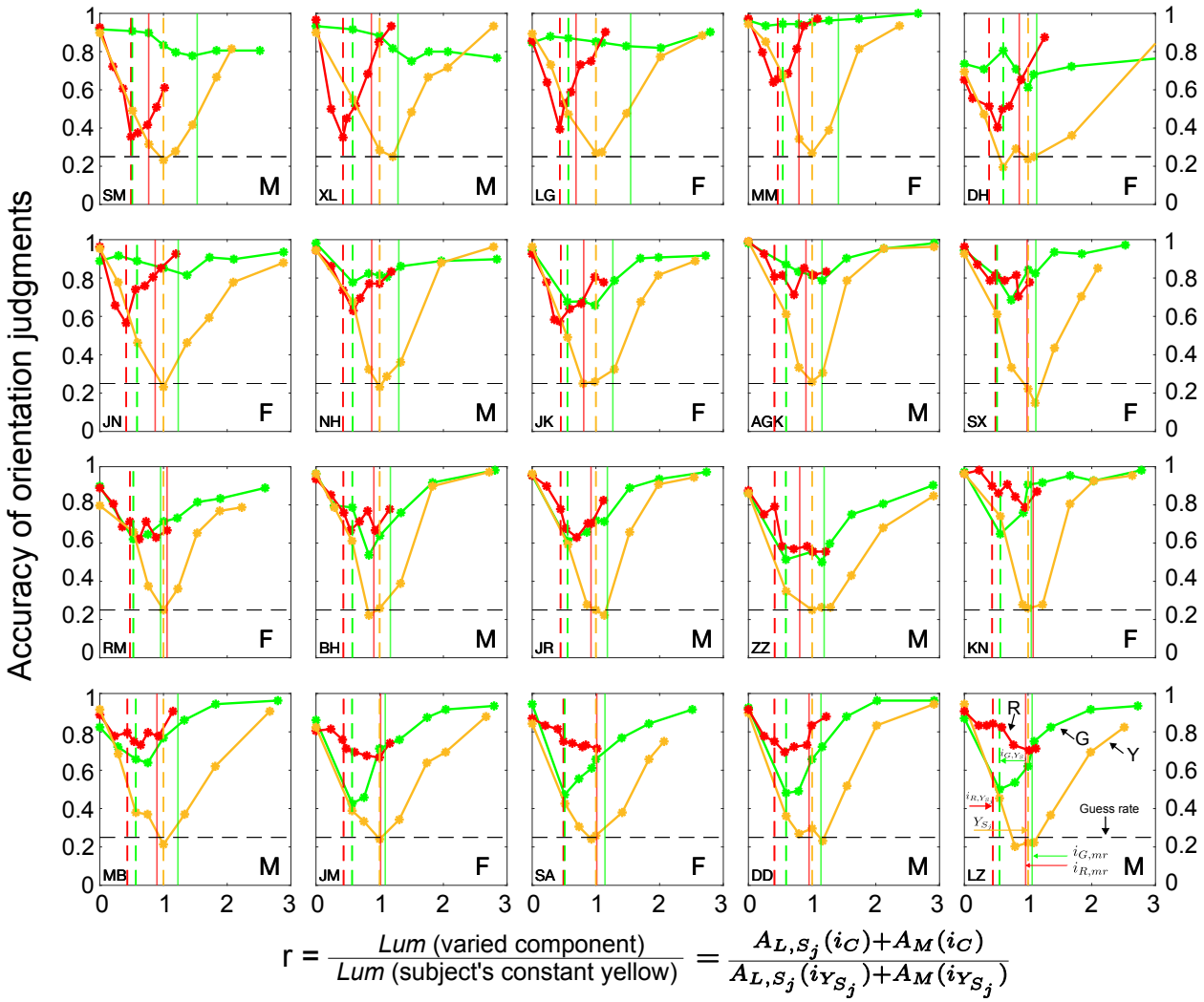



Figure 3.4: Results of orientation judgments of foveally viewed fine gratings and of judgments from the motion reversal paradigm. Each panel represents one subject, the subject's initials and gender are displayed at the bottom left and right corners respectively. All panels share the same legends and x and y axes. The x-axis represents the ratio of the luminance of variable stripes to the luminance of the subject's constant yellow stripes, where luminance is calculated as the sum of L and M cone excitation. $A_{L,S_j}(i_C)$ and $A_M(i_C)$ stand for L and M cone excitation at intensity i of C for a subject j. For the green chromatic trials, C = G; for the red chromatic trials, C = R; for the yellow achromatic trials, C = R+G. $A_{L,S_j}(i_{Y_{S_j}})$ and $A_M(i_{Y_{S_j}})$ stand for the L and M cone excitations of subject j's unique yellow. The y-axis represents the accuracy of the orientation judgments. The three curves plot the accuracy for yellow-yellow, green-yellow, and red-yellow gratings as a function of the variable color. Symbols in the bottom right panel: Y, yellow achromatic trials; R, red chromatic trials; G, green chromatic trials; solid lines $i_{R,Y_{S_j}}$ and $i_{G,Y_{S_j}}$ are the intensities of the red and green primaries that composed $Y_{S_j}$, subject j's unique yellow; dashed lines $i_{G,mr}$, and $i_{R,mr}$: the intensity of the green and red primary that canceled motion perception in the motion reversal paradigm and are thereby motion-equiluminant with the subject's unique yellow.

luminance changed, while their chromaticity remained constant for intensities that were sufficient to communicate color.

As shown in Fig. 3.4, for all 10 male and 10 female subjects, there was no point at which performance for the chromatic red and green gratings was reduced to chance. For all subjects, the accuracy of the chromatic trials was no worse than the accuracy of the yellow achromatic trials. For most subjects, when r is between 0.5 and 2, the accuracy of the red and green chromatic trials was greater than the achromatic yellow trial. It indicates that subjects were nearly always more sensitive to color contrasts than to luminance contrasts.

**Three distinctive data patterns**

There were major between-subject accuracy differences in judging the orientation of yellow/red versus yellow/green gratings, with subject data falling into three distinctive patterns. Figure 3.5 shows a typical example of each of the three patterns :

1. Pattern I: a nearly flat green curve + a V-shape red curve:
   (a) Subject SM, XL, LG, MM, DH, and JN exhibited Pattern I
   (b) The accuracy of green chromatic trials was independent of the luminance of the green component in the green/yellow gratings. A one-way ANOVA performed on the pooled data from these 7 subjects across 8 different green intensities showed no significant difference in accuracy ($F(7, 48) = 0.96, p = 0.474$).
   (c) The accuracy of green chromatic trials was significantly better than the paired yellow achromatic trials except for the two extreme intensities (the darkest ones and the brightest ones) where the accuracy was similar.
   (d) The average accuracy of green trials was better than the red trials at comparable $r$ values.

(e) The curves of the red chromatic trial exhibited a V-shape, but the minimal visibility points did not necessarily reflect isoluminance. For Subject SM, XL, LG, minimal motion, and JN, the intensity of the red primary that produced minimal visibility was at the intensity of the red primary that composed the yellow, which had less luminance than yellow no matter how luminance is calculated. For the other 2 subjects, the point of the red primary producing minimal visibility has greater luminance than the red primary composed yellow. DH: in between the luminance of the red and green primary that composed yellow. NH: at the point that has the same luminance as the green primary composed of yellow.

2. Pattern II: a V-shape green curve + a V-shape red curve

   (a) Subject JK, AGK, SK, RM, BH, JR, and ZZ exhibited Pattern II.

   (b) Both the red curve and green curve exhibited a V-shape.

   (c) The average accuracy of green trials was similar to the red trials at comparable $r$ values.

   (d) The intensities of the green and red primaries that produced minimal visibility of the gratings varied significantly across subjects, mostly occurring beyond the corresponding green and red primaries that composed the unique yellow for each subject.

   (e) The accuracy of chromatic trials was significantly better than the achromatic trials. The minimum of both curves was above 0.5.

3. Pattern III: a V-shape green curve + a shallow U-shape red curve

   (a) Subject KN, MB, DD, SA, LZ, and JM exhibited Pattern III

   (b) The green curves exhibited a V-shape. For 4 out of 6 subjects (KN, SA, LZ, and JM), the intensity of the green primary that produced minimal visibility of the

green/yellow gratings was at the intensity of the green primary that composed the yellow, which had less luminance than yellow.

(c) The red curves exhibited a shallow U-shape and for all 6 subjects, the intensity of red primary producing minimal visibility was around the intensity that had the same L+M excitation as the yellow.

(d) The average accuracy of green trials was worse than the red trials at comparable $r$ values. The green curves had lower minimum values compared to the red curves, but none fell below 0.45.

(e) The accuracy of chromatic red trials was significantly better than the yellow achromatic trials at comparable $r$ values.

Table 3.4 lists each subject's mean accuracy and the difference between the maximum and minimum accuracy for both green and red chromatic trials. As the patterns transition from Pattern I to Pattern III, the difference between the maximum and minimum accuracy for green chromatic trials increases, while the difference for red chromatic trials decreases. Subjects in Fig. 3.4 were sorted based on the difference in the mean accuracy of the green and red chromatic trials at comparable values of $r$.

**Comparing results between grating orientation judgments and the motion reversal paradigm**

The solid red and green lines in each panel of Fig. 3.4 represent the intensities of the red and green primaries that match the luminance of the subject's unique yellow in the motion reversal paradigm. As shown in Fig. 3.4, the points that generated minimal visibility of the chromatic gratings rarely coincided with the solid red or green lines. Figure 3.6 plotted the luminance ratios between the points that generated minimal visibility and yellow in both tasks. This clearly shows that the isoluminant settings obtained from the motion reversal

| | Subj Initial | Green chromatic trials | | Red chromatic trials | |
|---|---|---|---|---|---|
| | | Mean | Max-Min | Mean | Max -Min |
| Pattern I | SM | 0.83 | 0.19 | 0.54 | 0.58 |
| | XL | 0.83 | 0.18 | 0.65 | 0.62 |
| | LG | 0.86 | 0.08 | 0.67 | 0.51 |
| | MM | 0.96 | 0.06 | 0.81 | 0.33 |
| | DH | 0.72 | 0.19 | 0.58 | 0.47 |
| | JM | 0.89 | 0.12 | 0.78 | 0.40 |
| | NH | 0.86 | 0.20 | 0.78 | 0.31 |
| Pattern II | JK | 0.81 | 0.29 | 0.72 | 0.35 |
| | AGK | 0.89 | 0.19 | 0.84 | 0.28 |
| | SX | 0.87 | 0.29 | 0.81 | 0.26 |
| | RM | 0.77 | 0.28 | 0.71 | 0.27 |
| | BH | 0.80 | 0.44 | 0.77 | 0.27 |
| | JR | 0.81 | 0.36 | 0.77 | 0.32 |
| | ZZ | 0.69 | 0.40 | 0.66 | 0.32 |
| Pattern III | KN | 0.88 | 0.33 | 0.89 | 0.19 |
| | MB | 0.80 | 0.32 | 0.80 | 0.18 |
| | DD | 0.73 | 0.46 | 0.72 | 0.23 |
| | SA | 0.72 | 0.47 | 0.77 | 0.16 |
| | LZ | 0.74 | 0.44 | 0.80 | 0.20 |
| | JM | 0.76 | 0.53 | 0.79 | 0.25 |

Table 3.4: Mean accuracy and accuracy difference between the maximum and minimum accuracy of the green and red chromatic trials.

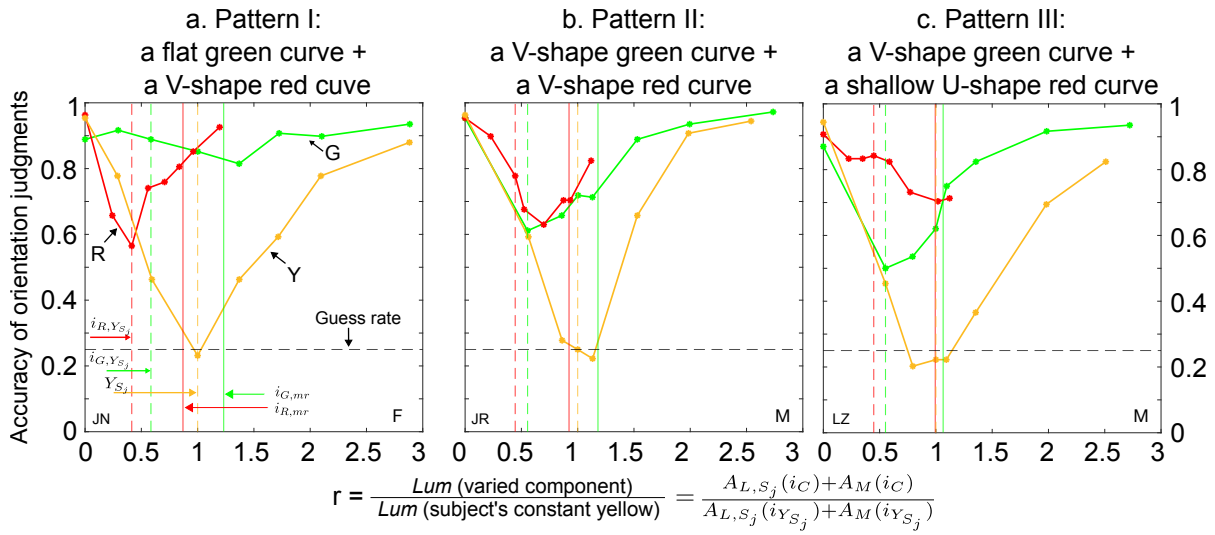



Figure 3.5: Three typical examples of the three patterns of results. The y-axis represents accuracy and the x-axis represents the luminance ratio between the modulated stripes (odd stripes) and the subjects' unique yellow (even stripes). Legends: Y: yellow achromatic trials; R: red chromatic trials; G: green chromatic trials; G: green chromatic trials; $i_{R,Y_{S_j}}$ and $i_{G,Y_{S_j}}$: the intensities of the red and green primaries that composed subject j's unique yellow f; $Y_{S_j}$ subject j's unique yellow. $i_{G,mr}$, and $i_{R,mr}$: the intensities of the green and red primary that generated motion reversal. (a) Pattern I: a flat green curve + a v-shape red curve; the performance for green chromatic trials was better than red chromatic trials when luminance ratios were comparable; (b) Pattern II: a v-shape green curve + a v-shape red curve; the performance for green chromatic trials was similar to red chromatic trials when luminance ratios were comparable; (c) Pattern III: a v-shape green curve + a shallow u-shape red curve; the performance for green chromatic trials was worse than red chromatic trials when luminance ratios were comparable;

paradigm targeting the mango pathway are distinct from those derived from the fine grating orientation judgment targeting the parvo pathway.

## 3.4 Discussion

In this study, each subject's unique yellow was used as the reference light. Two paradigms were employed to determine the corresponding intensities of the red and green primaries that match the luminance of subjects' unique yellow: one paradigm assessed the luminance mechanism associated with the parvocellular pathway, and the other assessed the luminance

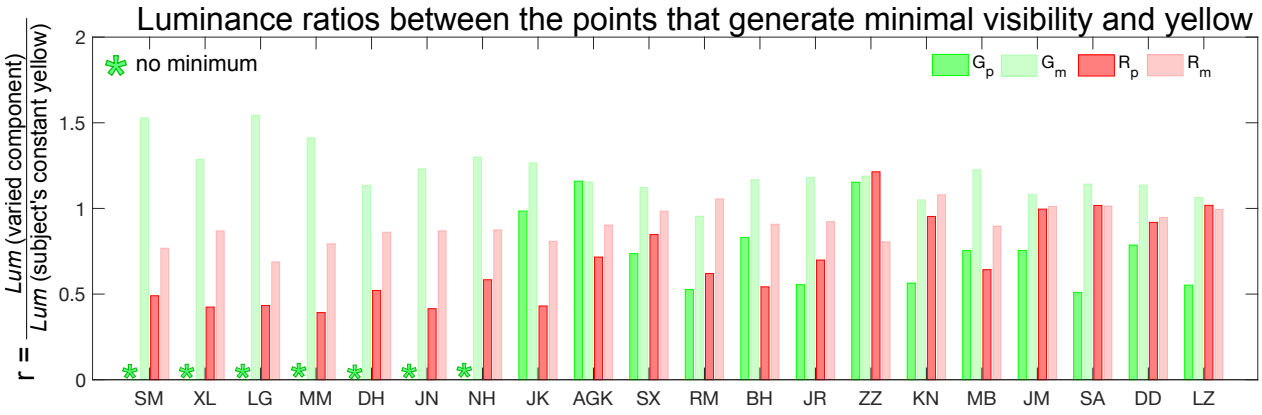



Figure 3.6: The luminance ratios between the points that generated minimal visibility and yellow in both tasks. The x-axis represents subjects and the y-axis represents the luminance ratio. Luminance is defined as the sum of L and M cone excitation.$G_p$ and $R_p$ represent the green and red primary points that generate minimal visibility of the green/yellow and red/yellow gratings targeting the parvo pathway. $G_m$ and $R_m$ represent the green and red primary points that generate motion reversal in the motion reversal paradigm targeting the mango pathway. For the first 7 subjects, there was no minimum point for the green/yellow gratings.

mechanism associated with the magnocellular pathway.

To measure the magno isoluminance, a motion reversal paradigm at a temporal frequency of 15 Hz was used, which exceeds the resolving capability of the parvo pathway. In this paradigm, isoluminance is achieved when the rotating wheel appears to wobble, neither rotating to the left nor the right. The intensity of the red (or green) primary was adjusted until the wheel reached this wobbling state, indicating that the luminance of the red or green primary matched the subject's unique yellow for the mango pathway.

To measure the parvo isoluminance in the central fovea, an extremely fine grating paradigm was used, in which high spatial-frequency yellow/red and yellow/green gratings were projected to the subjects' central fovea. The intensity of the red and green components was varied to determine the points of minimal visibility. The grating spanned 0.15 to 0.19 dva of the central fovea and had a spatial frequency of 28.6 to 35.8 cycles/degree, exceeding the resolving capacity of the magno pathway.

### 3.4.1 Does minimal visibility represent isoluminance in the central fovea?

There was significant between-subject variability in the results of the chromatic trials (red/yellow and green/yellow gratings). Not only where the minimal visibility occurred, but also the shape of curves varied greatly among subjects.

As discussed earlier, in Pattern I, 5 out 7 subjects' minimal visibility of the red/yellow gratings occurred at the intensity of the red primary that composed the yellow, which had less luminance than the yellow. For 7 out of 7 subjects, the visibility of the green/yellow gratings was independent of the changes in the luminance of the green component of the green/yellow grating. In pattern III, 4 out 6 subjects' minimal visibility of the green/yellow gratings occurred at the intensity of the green primary that composed the yellow, which again has less luminance than the yellow.

Altogether, the results show that in the central fovea, the minimal visibility of the chromatic gratings does not necessarily reflect the equal luminance of the two components of the grating.

### 3.4.2 Color sensitivity vs Luminance sensitivity

The experiment of grating orientation judgment was conducted using a mixed list design. All three types of trials were randomly mixed, and subjects had no prior knowledge of what the stimuli would be for the upcoming trial. By keeping the trial types unpredictable, it ensures that the responses are based purely on the visual information presented in each trial, rather than on any pre-conceived strategies or expectations. For all subjects, curves for the chromatic trials were consistently above the curve for the achromatic yellow curve. Excluding the darkest and brightest conditions, the accuracy of chromatic trials was generally better than that of achromatic trials for the majority of subjects. There was no point at

which chromatic grating performance was reduced to chance. For 15 out of 20 subjects, the lowest accuracy for chromatic trials was greater than 0.5 (the guess rate is 0.25). The significant difference between the accuracy of the chromatic trials and the achromatic trials indicates that in the central fovea, subjects were nearly always more sensitive to color than to luminance.

### 3.4.3 Is isoluminance measurable in the central fovea?

Current methods, such as heterochromatic flicker photometry, minimally distinct border, minimal motion, and motion reversal, enable accurate measurement of isoluminance only when luminance is the dominant factor in a task. The typical tasks to which these paradigms rely primarily on luminance, minimize the influence of chromatic cues, and have a clear indication of luminance cancellation, thus enabling a clear measurement of isoluminance. However, when both chromatic and achromatic cues are of the same or greater magnitude than luminance cues, isoluminance becomes difficult or impossible to determine.

In the present grating orientation judgment paradigm, the design guaranteed a red and a green point that matched the luminance of the yellow. However, the data did not show any subjects' performance in the chromatic trials dropping to chance levels. Additionally, for most subjects, the points of minimal visibility clearly had less luminance than yellow, i.e., they were not points of isoluminance. These observations clearly show that there was no particular reduction in acuity at the point of physically defined isoluminance, nor any definitive evidence that the luminance system was inactive when chromatic stimuli were present in the central fovea. Consequently, although there was guaranteed to be a point of isoluminance in each of the two chromatic procedures, there was no clear indication when this point was achieved.

### 3.4.4 Are parvo- and magno- measures interchangeable?

In our study, the intensities of the red and green primaries that motion-direction perception in the motion reversal paradigm were generally different from the intensities of the red and green primaries that produced minimal visibility of the foveally viewed fine red/yellow and green/yellow gratings. Additionally, in the central fovea, minimal visibility points do not necessarily represent isoluminance. Our findings, alongside those of He et al. [16] and Koenderink [15], show that the parvo isoluminance and magno isoluminance are different. Isoluminance is influenced by the specific mix of neurons engaged, which varies based on the nature of the task, retinal location, and other factors.

### 3.4.5 A possible cause of the significant between-subject variability

There were major between-subject accuracy differences in judging the orientation of yellow/red versus yellow/green gratings, with subject data falling into three distinctive patterns. Pattern I and Pattern III are nearly the opposite of each other; Pattern II falls between Pattern I and III.

For the 7 subjects in Pattern I, their performance on the green chromatic trials was independent of the modulation of the green variable stripes in the green/yellow gratings. This suggests that the subjects were not using the green stripe information in their judgments. Conversely, their performance on the red chromatic trials varied significantly with the modulation of the red variable stripes in the red/yellow gratings. Minimal visibility occurred when the red variable stripes matched the red component in yellow. Together, these findings indicate that the decision mechanism of these subjects gave a high weight to the red component in yellow. As a result, the performance on green/yellow gratings remained constant

(a flat curve) due to the absence of red in the green variable stripes.

Pattern III is the symmetric opposite of Pattern I. The data suggest that Pattern 3 subjects gave a high weight to the green component in yellow. Consequently, their performance on the green/yellow grating varied greatly with the modulation of the green stripes and achieved minimal visibility when the green variable stripes matched the green component in yellow. In contrast, their performance on the red/yellow gratings varied only slightly due to the absence of green in the red variable stripes.

For subjects in Pattern II, the performance on both green chromatic and red chromatic trials was similar, indicating that their decision mechanism gave similar weights to the red and green components in yellow. As a result, their performance in both types of chromatic trials varied with the contrast modulation of the variable stripes.

## 3.5 Conclusions

The present experiment attempted to determine whether it was possible to achieve isoluminance with chromatic stimuli in the central fovea where the cone receptors connect almost exclusively to the parvo system. In fact, the design of the experiment guaranteed to achieve isoluminance between the variable red or green stripes and constant yellow stripes in an acuity grating by modulating the intensity of the variable stripes across the full range of the primaries. However, the data show no particular acuity impairment at the point of physically defined isoluminance, nor is there any specific signature to indicate when the luminance system is inactive in the presence of chromatic stimuli in the central fovea. Conclusion: Isoluminance for the multichromatic parvo system is completely different from the isoluminance for the monochromatic magno system which is determined by the common measures of isoluminance.

## 3.6 Acknowledgements

Thanks to Dr. Jordan Rashid for calibrating and measuring the spectral power of monitor primaries.

# Bibliography

[1] Peter Lennie, Joel Pokorny, and Vivianne C Smith. Luminance. *JOSA A*, 10(6):1283–1293, 1993.

[2] János Schanda. Cie colorimetry and colour displays. In *Color and Imaging Conference*, volume 4, pages 230–234. Society of Imaging Science and Technology, 1996.

[3] Lindsay T Sharpe, Andrew Stockman, Wolfgang Jagla, and Herbert Jägle. A luminous efficiency function, v*($\lambda$), for daylight adaptation. *Journal of vision*, 5(11):3–3, 2005.

[4] Commission Internationale de l'Éclairage. *Principales décisions de la CIE, 6e Session, 1924, Commission Internationale de l'Éclairage Compte Rendu des Séances, Sixième Session, Genève, Juillet 1924*. Bureau Central de la Commission, the NPL, Teddington, the University Press, Cambridge, Cambridge, 1926.

[5] BB Lee, PR Martin, and A Valberg. The physiological basis of heterochromatic flicker photometry demonstrated in the ganglion cells of the macaque retina. *The Journal of physiology*, 404(1):323–347, 1988.

[6] AM Derrington and P Lennie. Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. *The Journal of physiology*, 357(1):219–240, 1984.

[7] Margaret S Livingstone and David H Hubel. Psychophysical evidence for separate chan-

nels for the perception of form, color, movement, and depth. *Journal of Neuroscience*, 7(11):3416–3468, 1987.

[8] Margaret Livingstone and David Hubel. Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science*, 240(4853):740–749, 1988.

[9] Andrew Stockman and Lindsay T Sharpe. The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision research*, 40(13):1711–1737, 2000.

[10] Lindsay T Sharpe, Andrew Stockman, Wolfgang Jagla, and Herbert Jägle. A luminous efficiency function, vd65*($\lambda$), for daylight adaptation: a correction. *Color Research & Application*, 36(1):42–46, 2011.

[11] D. B. Judd. Report of us secretariate committee on colorimetry and artificial daylight. Technical Report Part 7, CIE Central Bureau, 1951.

[12] Commission Internationale de l'Éclairage. Cie 1988 2° spectral luminous efficiency function for photopic vision. Technical Report CIE 86-1990, CIE Central Bureau, Vienna, Austria, 1990.

[13] Stuart Anstis and Patrick Cavanagh. *A minimum motion technique for judging equiluminance*. York University Toronto, 1983.

[14] Robert M Boynton and Peter K Kaiser. Vision: the additivity law made to work for heterochromatic photometry with bipartite fields. *Science*, 161(3839):366–368, 1968.

[15] Jan Koenderink, Andrea van Doorn, and Karl Gegenfurtner. Color weight photometry. *Vision Research*, 151:88–98, 2018.

[16] Jingyi He, Yesenia Taveras Cruz, and Rhea T Eskew. Methods for determining equiluminance in terms of l/m cone ratios. *Journal of vision*, 20(4):22–22, 2020.

[17] JD Mollon and JK Bowmaker. The spatial arrangement of cones in the primate fovea. *Nature*, 360(6405):677–679, 1992.

[18] Austin Roorda and David R Williams. The arrangement of the three cone classes in the living human eye. *Nature*, 397(6719):520–522, 1999.

[19] JK Bowmaker, JWL Parry, and JD Mollon. The arrangement of l and m cones in human and a primate retina. *Normal and defective colour vision*, pages 39–50, 2003.

[20] Heidi Hofer, Joseph Carroll, Jay Neitz, Maureen Neitz, and David R Williams. Organization of the human trichromatic cone mosaic. *Journal of Neuroscience*, 25(42):9669–9679, 2005.

[21] Lynda Bilodeau and Jocelyn Faubert. Isoluminance and chromatic motion perception throughout the visual field. *Vision research*, 37(15):2073–2081, 1997.

[22] Alvin Eisner and DIA MacLeod. Flicker photometric study of chromatic adaptation: selective suppression of cone inputs by colored backgrounds. *JOSA*, 71(6):705–718, 1981.

[23] Herbert JA Dartnall, James K Bowmaker, and John Dixon Mollon. Human visual pigments: microspectrophotometric results from the eyes of seven persons. *Proceedings of the Royal society of London. Series B. Biological sciences*, 220(1218):115–130, 1983.

[24] Brian Blundell Boycott, JM Hopkins, and HG Sperling. Cone connections of the horizontal cells of the rhesus monkey's retina. *Proceedings of the Royal society of London. Series B. Biological sciences*, 229(1257):345–379, 1987.

[25] Dennis M Dacey, Barry B Lee, Donna K Stafford, Joel Pokorny, and Vivianne C Smith. Horizontal cells of the primate retina: cone specificity without spectral opponency. *Science*, 271(5249):656–659, 1996.

[26] Stephen W Kuffler. Neurons in the retina: organization, inhibition and excitation

problems. In *Cold Spring Harbor Symposia on Quantitative Biology*, volume 17, pages 281–292. Cold Spring Harbor Laboratory Press, 1952.

[27] Stephen W Kuffler. Discharge patterns and functional organization of mammalian retina. *Journal of neurophysiology*, 16(1):37–68, 1953.

[28] David H Hubel and TN Wiesel. Receptive fields of optic nerve fibres in the spider monkey. *The Journal of physiology*, 154(3):572, 1960.

[29] Torsten N Wiesel and David H Hubel. Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey. *Journal of neurophysiology*, 29(6):1115–1156, 1966.

[30] Gregg E Irvin, Vivien A Casagrande, and Thomas T Norton. Center/surround relationships of magnocellular, parvocellular, and koniocellular relay cells in primate lateral geniculate nucleus. *Visual neuroscience*, 10(2):363–373, 1993.

[31] Ann K Goodchild, Krishna K Ghosh, and Paul R Martin. Comparison of photoreceptor spatial density and ganglion cell morphology in the retina of human, macaque monkey, cat, and the marmoset callithrix jacchus. *Journal of Comparative Neurology*, 366(1):55–75, 1996.

[32] Dennis M Dacey. The mosaic of midget ganglion cells in the human retina. *Journal of Neuroscience*, 13(12):5334–5355, 1993.

[33] Dennis M Dacey. Physiology, morphology and spatial densities of identified ganglion cell types in primate retina. In *Ciba Foundation Symposium 184-Higher-Order Processing in the Visual System: Higher-Order Processing in the Visual System: Ciba Foundation Symposium 184*, pages 12–34. Wiley Online Library, 2007.

[34] Peter Lennie, P William Haake, and David R Williams. The design of chromatically opponent receptive fields. 1991.

[35] Kathy T Mullen and Frederick AA Kingdom. Losses in peripheral colour sensitivity predicted from "hit and miss" post-receptoral cone connections. *Vision research*, 36(13):1995–2000, 1996.

[36] Andrew M Derrington, John Krauskopf, and Peter Lennie. Chromatic mechanisms in lateral geniculate nucleus of macaque. *The Journal of physiology*, 357(1):241–265, 1984.

[37] Yeon Jin Kim, Alexandre Reynaud, Robert F Hess, and Kathy T Mullen. A normative data set for the clinical assessment of achromatic and chromatic contrast sensitivity using a qcsf approach. *Investigative ophthalmology & visual science*, 58(9):3628–3636, 2017.

[38] Karl R Gegenfurtner and Daniel C Kiper. Contrast detection in luminance and chromatic noise. *JOSA A*, 9(11):1880–1888, 1992.

[39] Kathy T Mullen and M Angeles Losada. Evidence for separate pathways for color and luminance detection mechanisms. *JOSA A*, 11(12):3136–3151, 1994.

[40] George Sperling. Type 1 and type 2 experiments. `https://sites.socsci.uci.edu/~whipl/Type_1_and_Type_2_Expts.pdf`. Accessed: 2024-07-29.

[41] Peter H Schiller. Parallel information processing channels created in the retina. *Proceedings of the National Academy of Sciences*, 107(40):17087–17094, 2010.

[42] David J Calkins, Stanley J Schein, Yoshihiko Tsukamoto, and Peter Sterling. M and l cones in macaque fovea connect to midget ganglion cells by different numbers of excitatory synapses. *Nature*, 371(6492):70–72, 1994.

[43] Helga Kolb and David Marshak. The midget pathways of the primate retina. *Documenta Ophthalmologica*, 106:67–81, 2003.

[44] Chi Zhang, Yeon Jin Kim, Ana R Silverstein, Akina Hoshino, Thomas A Reh, Dennis M

Dacey, and Rachel O Wong. Circuit reorganization shapes the developing human foveal midget connectome toward single-cone resolution. *Neuron*, 108(5):905–918, 2020.

[45] Alvin Eisner and Donald IA MacLeod. Blue-sensitive cones do not contribute to luminance. *JOSA*, 70(1):121–123, 1980.

[46] Brian W Tansley and RM Boynton. Chromatic border perception: the role of red-and green-sensitive cones. *Vision Research*, 18(6):683–697, 1978.

[47] Wayne Verdon and Anthony J Adams. Short-wavelength-sensitive cones do not contribute to mesopic luminosity. *JOSA A*, 4(1):91–95, 1987.

[48] Caterina Ripamonti, Wen Ling Woo, Elizabeth Crowther, and Andrew Stockman. The s-cone contribution to luminance depends on the m-and l-cone adaptation levels: silent surrounds? *Journal of Vision*, 9(3):10–10, 2009.

[49] Christian Herrera Ortiz, Charles Chubb, Charles E Wright, Peng Sun, and George Sperling. Color scrambles reveal red and green half-wave linear mechanisms plus a mechanism selective for low chromatic contrast. *Vision Research*, 191:107964, 2022.

[50] David H Brainard and Spatial Vision. The psychophysics toolbox. *Spatial vision*, 10(4):433–436, 1997.

[51] Denis G Pelli. The videotoolbox software for visual psychophysics: transforming numbers into movies. *Spatial vision*, 10(4):437–442, 1997.

[52] Zhong-Lin Lu and George Sperling. Sensitive calibration and measurement procedures based on the amplification principle in motion perception. *Vision Research*, 41(18):2355–2374, 2001.

[53] Jan PH Van Santen and George Sperling. Temporal covariance model of human motion perception. *JOSA A*, 1(5):451–473, 1984.

# Appendix for Chapter 3

## C1 Supplementary figures

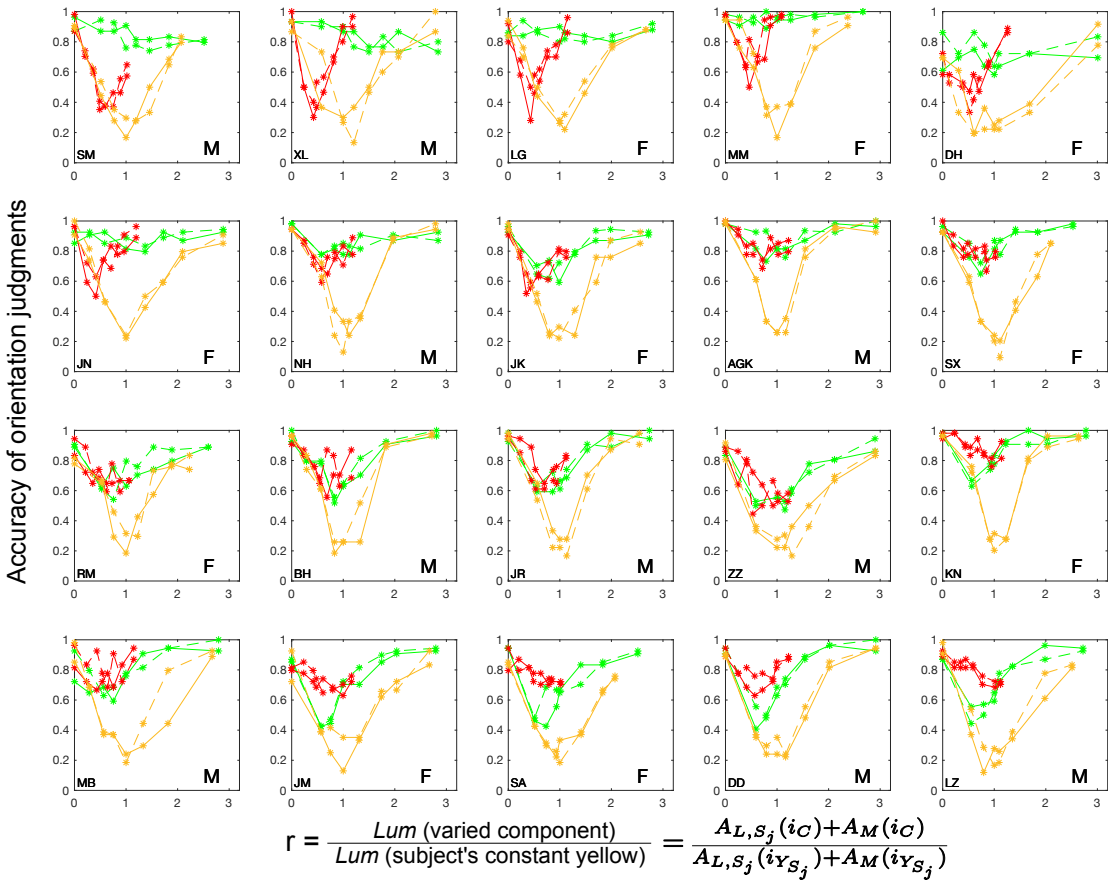



Figure C.1: Accuracy of orientation judgments: comparison between the first and second halves of sessions. The solid line represents the trials in the first half of the sessions and the dashed line is for the trials done in the second half of the sessions. Legend: $i_{R,Y_{S_j}}$ and $i_{G,Y_{S_j}}$: the intensities of the red and green primaries that composed subject j's unique yellow f; $Y_{S_j}$ subject j's unique yellow. Y: yellow achromatic trials; R: red chromatic trials; G: green chromatic trials; G: green chromatic trials

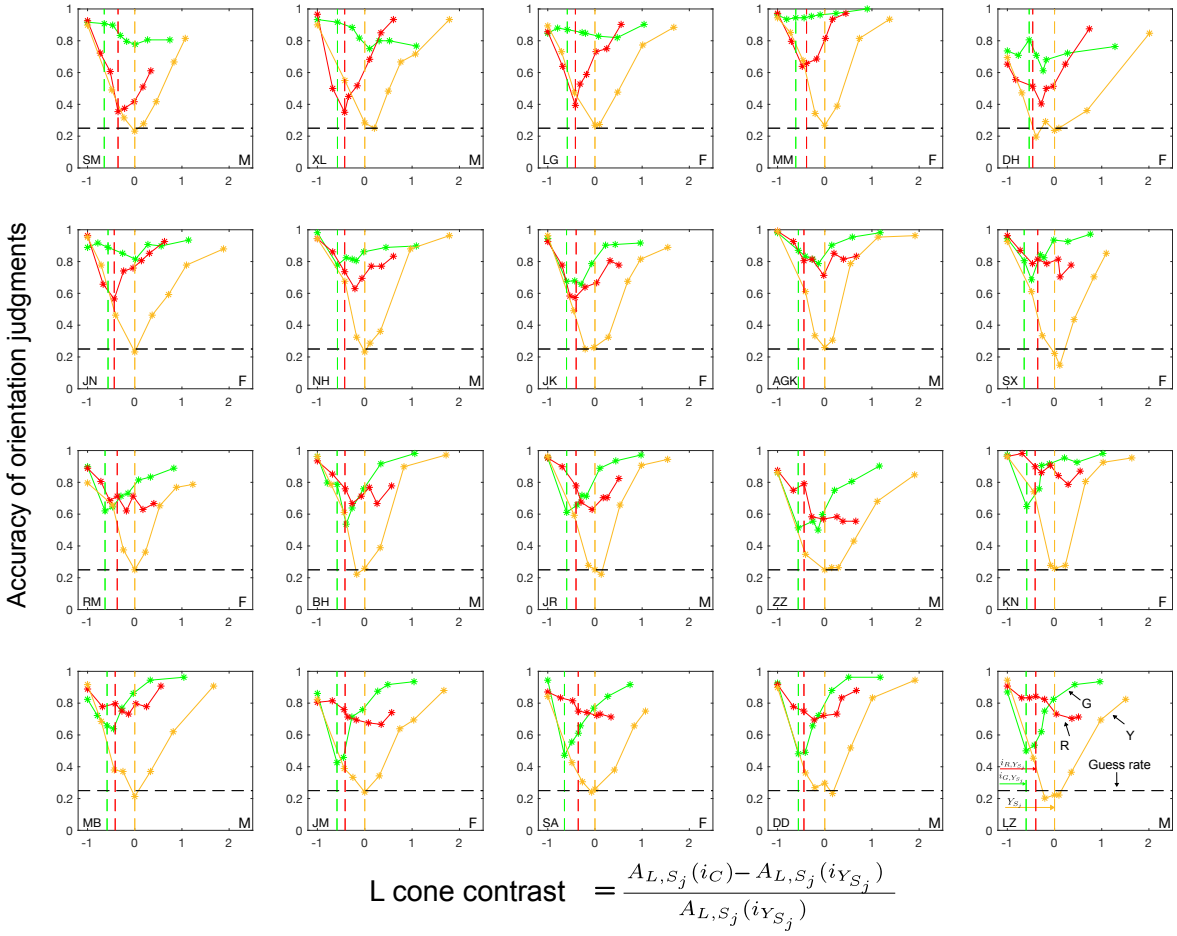



Figure C.2: Accuracy of orientation judgments of the three types of trials as a function of the L cone contrast of the variable stripes. $A_{L,S_j}(i_C)$ stands for L cone excitation at intensity i of C for a subject j. For the green chromatic trials, C = G; for the red chromatic trials, C = R; for the yellow achromatic trials, C = R+G. $A_{L,S_j}(i_{Y_{S_j}})$ stands for the L cone excitation of subject j's unique yellow. Legend: $i_{R,Y_{S_j}}$ and $i_{G,Y_{S_j}}$: the intensities of the red and green primaries that composed subject j's unique yellow f; $Y_{S_j}$ subject j's unique yellow. Y: yellow achromatic trials; R: red chromatic trials; G: green chromatic trials; G: green chromatic trials

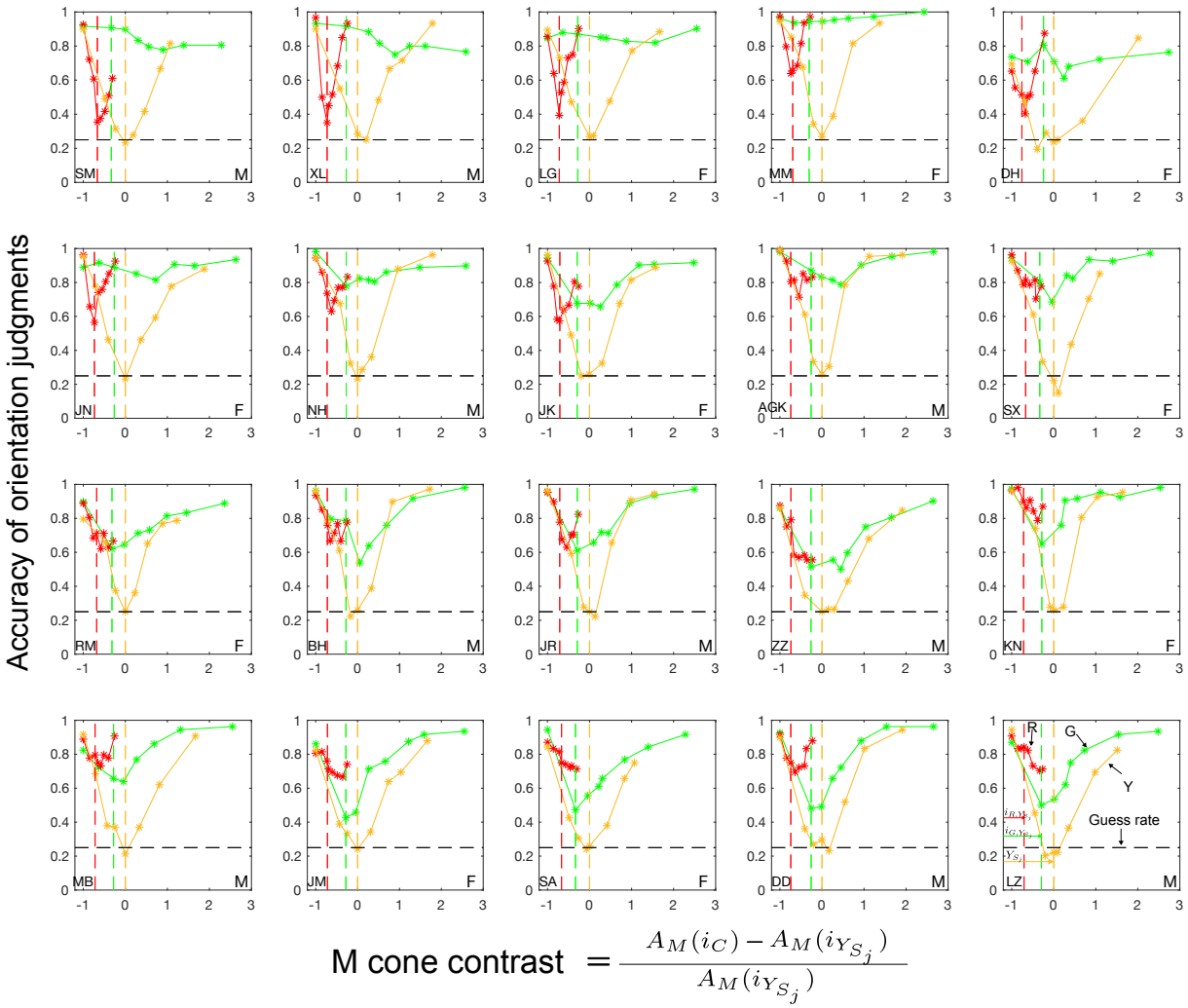



Figure C.3: Accuracy of orientation judgments of the three types of trials as a function of the M cone contrast of the variable stripes. $A_M(i_C)$ stands for M cone excitation at intensity i of C. For the green chromatic trials, C = G; for the red chromatic trials, C = R; for the yellow achromatic trials, C = R+G. $A_M(i_{Y_{S_j}})$ stands for the M cone excitation of subject j's unique yellow.Legend: $i_{R,Y_{S_j}}$ and $i_{G,Y_{S_j}}$: the intensities of the red and green primaries that composed subject j's unique yellow f; $Y_{S_j}$ subject j's unique yellow. Y: yellow achromatic trials; R: red chromatic trials; G: green chromatic trials; G: green chromatic trials