1-1-2012

Integration and Segregation in Audition and Vision

Thomas Yao-Yu Lee
University of Pennsylvania, thomyle@sas.upenn.edu

Integration and Segregation in Audition and Vision

Abstract
Perceptual systems can improve their performance by integrating relevant perceptual information and segregating away irrelevant information. Three studies exploring perceptual integration and segregation in audition and vision are reported in this thesis. In Chapter 1, we explore the role of similarity in informational masking. In informational masking tasks, listeners detect the presence of a signal tone presented simultaneously with a random-frequency multitone masker. Detection thresholds are high in the presence of an informational masker, even though listeners should be able to ignore the masker frequencies. The informational masker's effect may be due to the similarity between signal and masker components. We used a behavioral measure to demonstrate that the amount of frequency change over time could be the stimulus dimension underlying the similarity effect.

In Chapter 2, we report a set of experiments on the visual system's ability to discriminate distributions of luminances. The distribution of luminances can serve as a cue to the presence of multiple illuminants in a scene. We presented observers with simple achromatic scenes with patches drawn from one or two luminance distributions. Performance depended on the number of patches from the second luminance distribution, as well as knowledge of the location of these patches. Irrelevant geometric cues, which we expected to negatively affect performance, did not have an effect. An ideal observer model and a classification analysis showed that observers successfully integrated information provided by the image photometric cues.

In Chapter 3, we investigated the role of photometric and geometric cues in lightness perception. We rendered achromatic scenes that were consistent with two oriented background context surfaces illuminated by a light source with a directional component. Observers made lightness matches to tabs rendered at different orientations in the scene. We manipulated the photometric cues by changing the intensity of the illumination, and the geometric cues by changing the orientation of the context surfaces. Observers' matches varied with both manipulations, demonstrating that observers used both types of cues to account for the illumination in the scene. The two types of cues were found to have independent effects on the lightness matches.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Psychology

First Advisor
David H. Brainard

Keywords
Auditory perception, Integration, Psychophysics, Segregation, Visual perception

INTEGRATION AND SEGREGATION IN AUDITION AND VISION

Thomas Y. Lee

A DISSERTATION

in

Psychology
Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
2012

Supervisor of Dissertation

_____
David Brainard, Professor of Psychology

Graduate Group Chairperson

_____
John Trueswell, Professor of Psychology

Dissertation Committee
Delphine Dahan, Associate Professor of Psychology
David Brainard, Professor of Psychology
Russell Epstein, Associate Professor of Psychology

INTEGRATION AND SEGREGATION IN AUDITION AND VISION

ABSTRACT

INTEGRATION AND SEGREGATION IN AUDITION AND VISION

Thomas Y. Lee

David H. Brainard

Perceptual systems can improve their performance by integrating relevant perceptual information and segregating away irrelevant information. Three studies exploring perceptual integration and segregation in audition and vision are reported in this thesis. In Chapter 1, we explore the role of similarity in informational masking. In informational masking tasks, listeners detect the presence of a signal tone presented simultaneously with a random-frequency multitone masker. Detection thresholds are high in the presence of an informational masker, even though listeners should be able to ignore the masker frequencies. The informational masker's effect may be due to the similarity between signal and masker components. We used a behavioral measure to demonstrate that the amount of frequency change over time could be the stimulus dimension underlying the similarity effect.

In Chapter 2, we report a set of experiments on the visual system's ability to discriminate distributions of luminances. The distribution of luminances can serve as a cue to the presence of multiple illuminants in a scene. We presented observers with simple achromatic scenes with patches drawn from one or two luminance distributions.

Performance depended on the number of patches from the second luminance distribution, as well as knowledge of the location of these patches. Irrelevant geometric cues, which we expected to negatively affect performance, did not have an effect. An ideal observer model and a classification analysis showed that observers successfully integrated information provided by the image photometric cues.

In Chapter 3, we investigated the role of photometric and geometric cues in lightness perception. We rendered achromatic scenes that were consistent with two oriented background context surfaces illuminated by a light source with a directional component. Observers made lightness matches to tabs rendered at different orientations in the scene. We manipulated the photometric cues by changing the intensity of the illumination, and the geometric cues by changing the orientation of the context surfaces. Observers' matches varied with both manipulations, demonstrating that observers used both types of cues to account for the illumination in the scene. The two types of cues were found to have independent effects on the lightness matches.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 2

Chapter 3

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

INTRODUCTION

Sensory systems are information-processing systems. Their task is to make sense of the world through stochastic and noisy sensory information. Two important and complementary functions that sensory systems can implement aid this task are *integration* and *segregation*. For example, imagine standing in a crowded room full of talking people. Many sounds from all over the room will reach the ears of a listener trying to understand a particular speaker. The auditory system of this listener can facilitate performance by grouping together the sounds belonging to the speaker's voice and ignoring all of the other sounds in the room. More generally, if a sensory system is trying to extract stable properties from a noisy process, it can improve the quality of its estimates by combining and averaging the available data. Conversely, if data from multiple processes were available, the sensory system would do well to separate data from each source and handle each case independently. The papers in this dissertation explore this theme of integration and segregation in audition and vision.

Paper 1 describes an experiment on the auditory phenomenon known as *informational masking* (Neff & Green, 1987). In traditional masking paradigms, the detection of a signal tone can be affected by the presence of a masker tone. The amount of interference from the masker tone depends on its frequency relative to that of the signal tone. The larger the frequency difference, the smaller the interference. This relationship appears to be a function of the architecture of the peripheral auditory system

(see Moore, 2003 for a review of auditory frequency filters). On the other hand, in informational masking paradigms this relationship seems to break down. In informational masking experiment, a number of masker tones are played simultaneously with a signal tone. On each presentation of the stimulus the masker frequencies are randomized, but always presented at frequencies far from the signal. Based on the architecture of the early stages of the auditory system, these compound masker tones should also have little effect on the detectability of the signal. However, many listeners find it very difficult to detect the signal in the presence of this informational masker. They behave as though they are unable to segregate the masker tones from the signal. Instead, it appears that they are integrating the masker with the signal where it is disadvantageous to do so. This interferes with their ability to detect the signal.

One aspect of the stimuli in informational masking paradigms that may explain the interference effects is the similarity between the masker tones and the signal tones. Grouping by similarity is often advantageous, and is implemented in various perceptual systems (e.g., Treisman, 1991). As both the masker and signal are pure tones, the auditory system may implement a similar automatic grouping process and integrate all the tones together. The failure to segregate the signal from the masker may come about because after similar sounds are grouped together, it becomes difficult to differentiate sounds within the group. An experiment by Durlach *et al.* (2003b) demonstrated that manipulating signal-masker similarity can improve listeners' ability to segregate the signal from the masker. For example, upward gliding masker tones interfered with the

detection of an upward gliding signal, but downward gliding tones did not. However, the exact stimulus dimensions of signal-masker "similarity" are not known. Across different studies, the choice of masker and signal was almost arbitrary, making it difficult to understand when such effects occur and why they occur.

To better understand how similarity might affect segregation of the masker from the signal, in Paper 1 we explored a behavioral method for discovering the underlying dimensions of the signal-masker similarity. A number of masker types from relevant experiments in the literature (Neff, 1995; Durlach et al., 2003b) were used in this experiment. The discriminability of these maskers from the signal was used as an index of similarity. This index was then correlated with the detectability of the same type of signal in a detection task with a random-frequency informational masker.

The discriminability measure correlated well with the amount of informational masking in the detection task — masker types that were easy to discriminate from the signal corresponded the maskers causing the least amount of informational masking. In particular, conditions with maskers comprised of frequency glides had the highest discriminability scores and the lowest detection thresholds. In a second experiment, we manipulated the similarity of the frequency glide to the tone by changing its frequency span. Glides that were more similar to the tone, both in terms of a smaller frequency span as well as poorer discriminability from the tone, caused more informational masking in the detection task. These results, along with other findings in the literature (e.g., Kidd Jr., Mason, Deliwala, Woods, & Colburn, 1994) suggested that the amount of frequency

modulation over time could be one underlying dimension behind the observed similarity effects. More separation along this dimension between the signal and masker seems to allow the auditory system to segregate the signal away from the masker more easily.

Papers 2 and 3 studied the response of the visual system to grayscale scenes with multiple illuminants. The information available to the visual system about achromatic surfaces in the world is confounded with information about the illumination falling on those surfaces. For the system to recover the reflectance properties of achromatic surfaces, it must somehow discount the illuminant information from the luminance signal. Integrating information from multiple surfaces seems to improve the visual system's ability to solve this problem (Arend & Goldstein, 1987). However, the problem is complicated in many real-world scenes because multiple illuminants of different intensities can be present within the same scene. Simply integrating information from more surfaces alone is not an appropriate solution for handling multiple illuminants. If the visual system assumes that surfaces under different illuminants all fall under the same illuminant, its resulting reflectance estimates will be erroneous.

Some researchers (e.g., Gilchrist, 1977; Adelson, 1993) have theorized that in order to process these more complex scenes, the visual system must identify the regions of different illumination before applying surface-recovery algorithms. That is, the visual system should first determine whether surfaces in the visual scene fall under the same illuminant or different illuminant. Then, it should process surfaces falling under one illuminant separately from those under a different illuminant.

For achromatic surfaces, there are at least two kinds of cues available in the visual scene that can indicate the presence of multiple illuminants. Surfaces have an estimated 30:1 reflectance range (Wyszecki & Stiles, 1982). If such a range of surfaces were present in a scene, the presence of multiple illuminants of substantially different intensities would lead to luminances larger than this range. Thus this *photometric cue* can signal the presence of different illuminant intensities

Previous studies (Hochberg & Beck, 1954; Gilchrist, 1977; Gilchrist, 1980; Knill & Kersten, 1991; Adelson, 1993; Boyaci, Maloney, & Hersh, 2003; Bloj et al., 2004; Kitazaki, Kobiki, & Maloney, 2008) have also demonstrated that *scene geometry* is used as a cue in the perception of surface lightness. Surfaces that directly face a light source, for example, receive more intense illumination than surfaces facing away from that light source. Changes in surface orientation that correlate with differences in surface luminance can be informative about the spatial layout of the illumination. Accounting for the illumination's layout can, in turn, affect the perception of lightness.

In Paper 2 we explicitly measured the visual system's ability to identify simple scenes with multiple illuminants. We first measured performance with scenes where photometric cues indicated the presence of multiple illuminants. Then we measured changes in this performance as a consequence of introducing geometric information. Observers were presented with checkerboard patterns consisting of computer-rendered matte grayscale squares with random reflectances. On every trial, observers were asked to pick which of two checkerboards contained simulated surfaces lit by two illuminants.

The difference between the intensity of the two illuminants needed to produce a threshold level of performance was measured and compared across conditions.

The more surfaces illuminated by the second illuminant, the better observers performed on the task. This improvement in performance with increasing numbers of surfaces suggests that observers integrated information across surfaces. Performance was modulated by uncertainty about the locations of the illuminated surfaces.

An ideal observer analysis was able to account for the data well. This kind of analysis describes how a perfect observer with the optimum strategy would use the information available in a task (Geisler, 2003). Observers' performance was similar to the ideal observer's performance, but less efficient. Further analyses also demonstrated that the trial-by-trial variability of the ideal observer was more similar to the human observers than models implementing a number of non-ideal heuristics.

The measurements in that experiment provided a baseline for additional measurements with more complex stimuli. The critical test was the introduction of geometric cues to segregation. We introduced geometric cues that suggested particular locations for the second illuminant. These locations were incongruent with the actual second illuminant locations. The irrelevant geometric cues were expected to interfere with observers' ability to perform the task. However, the data showed that the two geometric cue manipulations we employed had no effect. Performance was the same, regardless of whether geometric cues were present or not.

In Paper 3, we explored a different approach to understand the role of scene geometry in illuminant segregation and how geometric cues interact with photometric cues. The experiments used an indirect measure of observers' sensitivity to these cues. Observers viewed computer-rendered scenes of two contextual background surfaces ("contexts") at different orientations in three-dimensional space. The scene was consistent with two contexts lying in different orientations, one lit by a directional illuminant, the other lying in shadow. To gauge whether observers perceived different illumination regions in the scene, we presented probe tabs rendered at various orientations.

Observers were asked to match the tabs to a fixed set of grayscale chips. Differences in matches to the same physical tab in different locations would be a signature of various effects of cues in the scene to illuminant segregation. The photometric cues to the presence of two illuminant intensities were implemented via differences in the mean luminance across the two context surfaces. Observers' matches to a probe tab of fixed luminance changed depending on whether it was in front of the top or the bottom context surface. The magnitude of the match differences depended on the magnitude of the mean context luminance difference. Thus, observers were sensitive to the photometric cues and interpreted these cues as indicating two illumination regions.

Observers' matches to a fixed tab luminance also changed with the tab's orientation. This pattern of behavior suggested that the geometric cues in the scene were also interpreted as cues to different illumination regions, and that the illumination

was interpreted as at least partly directional. Moreover, the data suggested that the geometric cues affected observers' matches independently of the photometric cues. Varying the photometric cues for a fixed orientation of the contextual backgrounds affected performance as described above. Varying the orientation of the backgrounds while holding the photometric cues fixed, however, affected the pattern of matches across different tab orientations in a completely different manner that seemed to be independent of the effect of the photometric cues.

Taken together, the three papers establish conditions under which the auditory and visual systems demonstrate information integration and segregation, as well as some of the stimulus cues that affect these processes. The experiments provide interesting data that further our understanding of how our sensory systems function in complicated, information-rich environments.

CHAPTER 1: Evaluation of similarity effects in informational masking

This paper was published as: Lee, T.Y. and Richards, V.M. (2011). Evaluation of similarity effects in informational masking. *Journal of the Acoustical Society of America, 129*, EL280-EL285.

The paper has been reproduced in its entirety and reformatted. The figures have been removed from the body of the text and reformatted for the appropriate sections of the thesis.

Abstract

The degree of similarity between signal and masker in informational masking paradigms has been hypothesized to contribute to informational masking. The present study attempted to quantify "similarity" using a discrimination task. Listeners discriminated various signal stimuli from a multitone complex and then detected the presence of those signals embedded in a multitone informational masker. Discriminability negatively correlated with detection threshold in an informational masking experiment, indicating that similarity between signal and the masker quality contributed to informational masking. These results suggest a method for specifying relevant signal attributes in informational masking paradigms involving similarity manipulations.

PACS Numbers 43.66.Dc, 43.66.Ba

1. Introduction

The phenomenon of *informational masking* is characterized by difficulty in detecting a signal that cannot be accounted for by interfering energy patterns at relatively peripheral points in the auditory system (for a more complete discussion, see Durlach et al., 2003a). Investigations into the causes of informational masking have largely focused on the effects of uncertainty about the masker (e.g., Neff & Green, 1987), but an additional potential contributor to these masking effects is the perceptual similarity between the signal and the masker. In many informational masking experiments (e.g., Neff & Green, 1987; Oh & Lutfi, 1998), the signal is a tone, and the masker components are also tones with frequencies drawn at random prior to each presentation. For small numbers of masker components where the components are expected to be independently resolved, detection thresholds increase as the number of tones comprising the masker increase. The increase in thresholds may be attributed to the increase in number of masker elements sharing characteristics with the signal. For example, listeners may have difficulty segregating the signal from the masker due to their shared characteristics, confusing the signal tone with masker component tones.

Neff (1995) studied a variety of different stimuli as signals embedded in a random multitone masker to examine the effects of perceptual similarity on thresholds. Of the signals tested (tone, narrow band of noise, amplitude-modulated tone, quasi-frequency modulated tone), only the amplitude-modulated tone and narrow band of noise provided a consistent release from masking relative to tonal signals. Durlach *et al*. (2003b) also ran

a series of experiments where signals were constructed to be similar or dissimilar from the informational masker (e.g., a masker of upward frequency glides with signal frequency glides that swept upwards or downwards). These studies suggest that listeners can take advantage of differences in characteristics between the signal and masker in order to reduce the effects of informational masking. The most immediate problem in previous studies exploring the role of similarity, however, is that the conditions tested in different studies are not directly comparable. For example, in the Durlach *et al*. (2003b) study, as the authors acknowledged, stimuli tested in the various similarity-dissimilarity pairs were not held constant across pairs. Thus, conclusions about the mechanisms by which similarity affects informational masking are difficult to draw.

The present experiments aimed to establish an independent measure of similarity that could be used across different stimulus manipulations to predict the release from informational masking due to dissimilarity between signal and masker. A discrimination task was used to measure similarity. The logic behind the task is that similar stimuli should be harder to discriminate. Because random multitone complexes have been widely used as maskers in the informational masking literature, they were used in these experiments. For the discrimination task, targets that are more similar to the multitone complex should be more difficult to discriminate from the multitone complex. For a detection task, detection thresholds should be higher when the signal and masker are similar. Therefore, by a similarity argument, a negative correlation between

discriminability (target vs. multitone complex) and detection threshold (target masked by the multitone complex) is expected.

Two sets of target stimuli were tested using the proposed measure. Several targets were first discriminated from a random-frequency multitone complex, and then used as signals-to-be-detected in an informational masking experiment. The results show discriminability negatively correlates with thresholds in an informational masking task, provided informational masking is observed. These results support the hypothesis that similarity effects contribute to informational masking.

2. Experiment 1

The stimuli tested were selected from previous experiments examining the role of signal-masker similarity in informational masking (Neff, 1995; Durlach et al., 2003b). Four targets were tested: a tone, an amplitude-modulated (AM) tone, a rising frequency glide, and a sequence of tone pips (a tone gated on and off).

2.1 Discrimination Methods

For each target the center frequency was randomly selected from a range of 800 – 5000 Hz on a logarithmic scale except for the frequency glide target (the lowest and highest frequencies of which fell within this range — see following text). Frequencies were randomized to discourage listeners from learning unique frequencies for each target. The discriminability between each target and a random multitone stimulus was estimated. The multitone stimulus was composed of six tones with frequencies drawn randomly

from a logarithmic scale ranging from 800 – 5000 Hz on each presentation. The stimuli were turned on and off with 5-ms raised-cosine ramps, and had durations of 280 ms. To prevent perfect discrimination, the targets and multitone stimuli were presented concurrently with 50 dB SPL pink noise low-pass filtered at 7000 Hz.

The AM tone was 100% modulated by a raised 25 Hz sinusoid, resulting in seven modulation cycles per stimulus presentation. The frequency glide linearly spanned 50% of its lowest frequency (similar to Durlach et al., 2003b). The range for selecting the frequency glides' lowest frequency was from 800 to 3334 Hz so that the lowest and highest frequencies fell within the range of 800 – 5000 Hz occupied by the other stimuli. The tone pips consisted of seven short 40 ms tones played sequentially, each being turned on and off with 5-ms raised-cosine ramps (e.g., similar to the multiple-burst-same or MBS condition of Durlach et al., 2003b). For each listener the levels of all stimuli were adjusted to yield 90% detectability in the pink noise.

The stimuli were generated digitally with a two-channel, 16-bit DAC (TDT DA1) using a sampling rate of 20 kHz, and low-pass filtered at 7000 Hz (Stewart VBF 10 M Dual Variable Filter), individually attenuated, added, and presented diotically via Sennheiser HD410SL headphones. Listeners were tested individually in a double-walled sound-attenuated booth.

In a two-interval forced-choice (2IFC) procedure, listeners discriminated between a target and the multitone complex by indicating which of the two intervals contained the

multitone stimulus. The intervals were separated by 500 ms. Feedback was provided after every trial, as well as summary statistics after each block of 60 trials. Listeners first ran a set of five 60-trial blocks for each target in turn, and then repeated the process in the reverse order. The order in which the different targets were tested was chosen using a pseudo-Latin-squares design. If practice effects were apparent (one-tailed t-test), an additional set of five 60-trial blocks was collected and the first set discarded; however, practice effects were not observed for any of the listeners. Discriminability was measured as *d'*.

Three listeners (2 males, 1 female; age range 18 to 23) participated in this experiment. All listeners except L3 had thresholds in quiet under 20 dB HL at audiometric frequencies between 250 Hz and 8 kHz. Listener 3's threshold in quiet for 250 Hz was 20 dB HL at the right ear.

2.2 Discrimination Results

A repeated measures ANOVA indicates a significant main effect of target class ($F(3,6) = 139.44$, $p < 0.001$) on discriminability measured as *d'*. Averaging across all listeners, the frequency glide is most easily discriminated from the multitone complex (mean $d' \approx 2.1$, SEM ≈ 0.2); the discrimination between the other three targets and the multitone complex is much poorer (on average, $d' \approx 0.4$, SEM ≈ 0.1). This pattern was consistent for all three listeners. One listener's (L3) *d'* for the tone pips was a negative value. Because this value was close to zero, we assumed this to reflect variations associated with small sample size.

2.3 Detection Methods

The four targets in the discrimination task were next used as signals in a detection task with an informational masker. The target signal frequencies were fixed at 1000 Hz (the frequency glide started at 800 Hz and ended at 1200 Hz). To maintain consistency with prior informational masking experiments and to maximize the amount of informational masking, pink noise was not presented.

The masker was composed of six random-frequency equal-amplitude sinusoid tones with frequencies chosen at random from a logarithmically-spaced uniform distribution with a range of 200 – 5000 Hz. A "protected region" from 800 to 1200 Hz, in which none of the masking components could fall, reduced the impact of energetic masking. The overall level of the multitone masker was 70 dB SPL.

A three-down, 1-up algorithm was used to adjust the signal level (Levitt, 1971) in 60-trial sets. The signal level was initially 50 dB SPL, with a 4 dB step change for the first three reversals and a 2 dB step change thereafter. An even number of reversals (excluding the first three or four as required) was averaged to obtain an estimate of the 79% detection threshold level. In all other respects data collection followed the procedure described for the discrimination task, except listeners completed two sets of ten blocks of 60 trials rather than five blocks. Practice effects were not found across the two sets.

2.4 Detection Results and Comparison with Discrimination

For the tonal signal the average threshold (mean = 30.8 dB SPL; range = 22.5 – 35.8 dB SPL) falls in the lower range of those reported elsewhere in the literature, a range of 20 – 65 dB SPL (Neff & Dethlefs, 1995), but is well within the wide range of reported data. A repeated measures ANOVA indicates a significant effect of target type ($F(3,6) = 30.27$, $p < 0.001$) on threshold. For all listeners, the frequency glide had the lowest threshold (mean threshold = 25.0 dB SPL, SEM = 1.3). Relative to the tonal signal, thresholds decreased by ~9 dB for the frequency glide.

[---FIGURE 1 ABOUT HERE---]

The left panel of Figure 1 shows individual detection thresholds plotted as a function of discriminability. The most notable feature of this comparison is that there are two clusters of points: the points representing the results for frequency glide target are in the lower right hand corner, and the points representing the results for the other targets are in the upper left hand corner. The correlation coefficient was $r = -0.9$ for all data points, and the individual coefficients ranged from $r = -0.9$ to $r = -1$ across the three listeners, all of which were statistically significant or marginally significant ($p = 0.07$ for L2, $p < 0.05$ for L1 and L3). The stimulus that was the easiest to discriminate from the multitone complex had the lowest detection threshold.

2.5 Discussion of Exp.1

The large negative correlation between discriminability and detection thresholds suggests that discriminability may be a good measure of similarity effects in informational masking. However, because the correlation depends predominantly on the discriminability and detectability of the frequency glide target relative to the other targets, the data set more or less collapses into a categorical data set. Without systematic changes in *d'*, stronger conclusions regarding similarity and informational masking are difficult to draw. In a second experiment a new set of targets with a larger number of discrimination scores was used to provide a stronger test of the similarity hypothesis.

3. Experiment 2

3.1 Stimuli

The large negative correlation between discriminability and detection thresholds suggests that discriminability may be a good measure of similarity effects in informational masking. However, because the correlation depends predominantly on the discriminability and detectability of the frequency glide target relative to the other targets, the data set more or less collapses into a categorical data set. Without systematic changes in *d'*, stronger conclusions regarding similarity and informational masking are difficult to draw. In a second experiment a new set of targets with a larger number of discrimination scores was used to provide a stronger test of the similarity hypothesis.

3.2 Participants

Four listeners (2 males, 2 females, ages 19 to 24) participated in this experiment. All had thresholds in quiet less than 20 dB HL at audiometric frequencies between 250 Hz and 8 kHz. Listener 7 was the first author. None of the listeners in this experiment had participated in the previous one.

3.3 Results

The results for Experiment 2 are plotted in the right panel of Figure 1. Because the individual differences in detection thresholds are large, results are shown separately for each listener. The correlation between threshold (as dB SPL) and discrimination (as *d'*) is shown in the upper left hand corner of each panel. For the discrimination task, a repeated measures ANOVA shows a significant main effect of stimulus condition, $F(5,15) = 10.91$, $p < 0.0005$. Listeners' ability to discriminate a frequency glide from a random multitone complex improves with increases in the span of the frequency glide.

The results from the detection experiment indicate substantial individual differences. Thresholds for L4 are the lowest (mean threshold = 25.7 dB SPL), L5's thresholds are the highest (mean = 37.1 dB SPL), and L6 and L7's are intermediate (mean = 35.0 and 30.3 dB SPL, respectively). Even though there are large individual differences in thresholds, a repeated-measures ANOVA indicates a significant main effect of condition, $F(5,15) = 4.36$, $p < 0.05$.

Listener 4 exhibits little masking, and little release from masking with changes in the span of the target's frequency glide. As a result, the correlation for L4 is not significantly different from zero (r = -0.1, *n.s.*). Others have reported instances of observers who show little or no informational masking effects (e.g., Neff & Dethlefs, 1995; Oh & Lutfi, 1998).

Listeners 5-7 present larger amounts of informational masking than L4, as evidenced by the higher detection thresholds for the tone signal (thresholds of 49.4, 47.3, and 32.2 dB SPL, respectively). The correlations between discriminability and detectability for L5 and L6 are statistically significant in the expected direction (r = -0.8, $p < 0.05$ and r = -0.9, $p < 0.01$, respectively), and the correlation for L7 approaches statistical significance (r = -0.7, $p \approx 0.1$). For these three listeners, the amount of informational masking decreases when discriminability increases.

3.4 Discussion of Experiment 2

The results of these experiments indicate that the discrimination task measures similarity effects in informational masking. Targets that are more discriminable from a random-frequency multitone complex provide larger releases from informational masking. Therefore, the discrimination task can predict the amount of informational masking due to some change in similarity characteristics between signal and masker, at least to a rank-order approximation. There is one caveat, however: for listeners who are not susceptible to informational masking, the discrimination task is not a useful predictor of thresholds because there is little/no informational masking to measure.

4. General Discussion

The results from these experiments demonstrate the use of a method for predicting what sorts of similarity manipulations may lead to differences in detection in informational masking tasks. Considering the current study along with the narrow-band noise stimuli in Neff (1995), multiple-burst-different (MBD) stimuli in Kidd *et al.* (2002), and the glide stimuli in Durlach *et al.* (2003b), one possible stimulus quantity responsible for similarity-based effects in informational masking may be spectral differences across time. The "similar" stimuli in these experiments all have constant frequency differences between individual components and the signal as a function of time, whereas the "dissimilar" stimuli have components whose frequency differences change over time.

Alternatively, the frequency glide signals in Experiments 1 and 2 may have led to lower thresholds because they stimulated more auditory filters than tone signals and reduced the total entropy (Oh & Lutfi, 1998). Glides with greater frequency spans might then be more detectable because greater numbers of filters are stimulated, in contrast to the dynamic changes in spectral differences suggested above. While the data in these experiments do not offer evidence to test these alternatives, the glide stimuli from the Durlach *et al*. (2003b) study showed an improvement in threshold across conditions where the number of stimulated filters was comparable. Thus, a greater number of stimulated filters is at least not a necessary condition for the observed similarity-based effects.

While the discrimination measure described here has the advantage of providing estimates of similarity for the same stimuli tested in a detection task, there are limitations to the use of this method. The primary issue is the degree to which discrimination is a plausible measurement of similarity for the particular stimuli of interest. For example, onset differences have been shown to reduce informational masking (e.g., Neff, 1995; Durlach et al., 2003b). However, the discrimination measure requires masker and signal to be in different temporal intervals, negating any assessment of similarity for onset differences. Nonetheless, the proposed method provides an initial tool for evaluating similarity, and can form a starting point for efforts to understand the role of similarity for the perception of sounds in complex environments.

Acknowledgements

This work was supported by grant R01 DC002012 from NIH/NIDCD. We thank Drs. Daniel Shub and Robert Lutfi and two anonymous reviewers for comments on an earlier version of this manuscript.

Figures

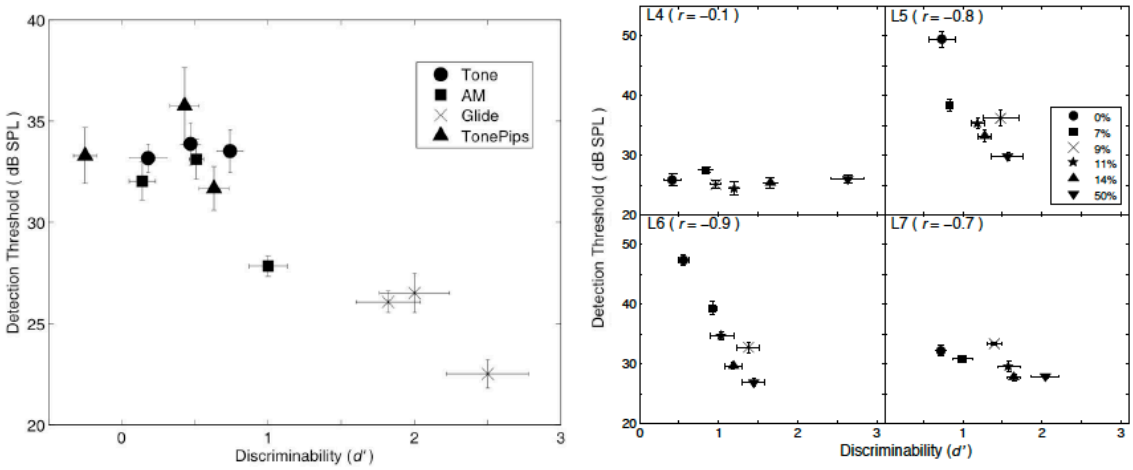

Figure 1: Left panel—for three listeners, detection thresholds for various signals are plotted as a function of discriminability for the corresponding stimuli. Different signal types are plotted using different symbols. Error bars are one standard error of the mean. Right panel—detection thresholds for frequency glides of different spans are plotted as a function of discriminability for the corresponding stimuli. Each panel shows data for one listener as well as the Pearson product correlation coefficient. Error bars are one standard error of the mean.

CHAPTER 2: Detection of changes in luminance distributions

This paper was published as: Lee, T.Y. and Brainard, D.H. (2011). Detection of changes in luminance distributions. *Journal of Vision, 11*, 1-16.

The paper has been reproduced in its entirety and reformatted. The figures and tables have been removed from the body of the text and reformatted for the appropriate sections of the thesis.

Abstract

How well can observers detect the presence of a change in luminance distributions? Performance was measured in three experiments. Observers viewed pairs of grayscale images on a calibrated CRT display. Each image was a checkerboard. All luminances in one image of each pair consisted of random draws from a single probability distribution. For the other image, some patch luminances consisted of random draws from that same distribution while the rest of the patch luminances (*test patches*) consisted of random draws from a second distribution. The observers' task was to pick the image with luminances drawn from two distributions. The parameters of the second distribution that led to 75% correct performance were determined across manipulations of 1) the number of test patches, 2) the observers' certainty about test patch location, and 3) the geometric structure of the images. Performance improved with number of test patches and location certainty. The geometric manipulations did not affect performance. An ideal observer model with high efficiency fit the data well and a classification image analysis showed a similar use of information by the ideal and human observers, indicating that observers can make effective use of photometric information in our distribution discrimination task.

1. Introduction

The visual system's representation of objects includes percepts that correlate with object surface reflectance. In general, these include color as well as perceptual correlates of object material properties, such as glossiness (Maloney & Brainard, 2010). The retinal image, however, does not provide an explicit representation of object reflectance. Rather, image intensities depend both on object reflectance and the properties of the illumination. To produce stable perceptual representations of object surface reflectance, the visual system must process the retinal image to minimize effects of variation in the illumination. Figure 1 shows an achromatic image where there is large spatial variation in the illumination.

A number of theorists have postulated that the stabilization of object appearance occurs in two stages (Kardos, 1934; Koffka, 1935; Gilchrist, 1977; Gilchrist et al., 1999; Adelson, 2000). The first stage segments the image into regions that each have roughly constant illumination. The second stage then, in effect, estimates the illuminant within each region and uses the estimate in its conversion between luminance and lightness for that region.

[---FIGURE 1 ABOUT HERE---]

What information could the visual system use to segment the image according to illumination? Photometric cues provide one source of information that can indicate illumination changes. Surface albedo is typically thought to vary over about a 30 to 1

range in natural scenes (see, for example, reflectance data summarized in Wyszecki & Stiles, 1982). Thus if two grayscale image regions vary in luminance by a factor much larger than 30, they are unlikely to share a common illuminant. In the image shown in Figure 1, it is easy to imagine that such a difference in image intensity helps mediate the impression that the floor is lit by two distinct illuminants.

On the other hand, a number of geometric factors may also correlate with illumination changes. One, for example is distance: the further apart two surface patches are in a scene, the less likely it seems that they will share a common illuminant. Accordingly, experiments have found a decreasing influence of contextual surfaces on target surface appearance with increasing distance (Reid & Shapley, 1988; Spehar, Debonet, & Zaidi, 1996; Shimozaki, Eckstein, & Abbey, 2005; Kurki, Peromaa, Hyvärinen, & Saarinen, 2009). Closely related is the idea that co-planar surfaces are more likely to share a common illuminant than surfaces oriented differently within a scene (Gilchrist, 1980). Various cues are available to indicate surface orientation in a scene (e.g. binocular disparity), as well as changes in orientation of groups of surfaces (e.g., Ψ-junctions, Sinha & Adelson, 1993). Finally, the luminance relations across certain geometric configurations may signal illumination boundaries (e.g., X- and T-junctions, Todorović, 1997).

Despite the centrality of segmentation in theories of lightness, little is known about how well observers can use the type of photometric information induced by changes of illumination to segregate scenes. For achromatic images, changing the

illumination changes the statistical distribution of the luminances reaching the observer, because the luminance distribution arises as the product of the illuminant intensity and the underlying distribution of surface albedos. In the present paper, then, we step back from the specifics of illuminant-based segmentation and ask the more basic question of how well observers can detect within-image changes in the distribution of image luminances. That is, we sought to study fundamental aspects of this ability, using simple stimuli that did not evoke percepts of illuminated surfaces. We used checkerboard stimuli and asked observers to judge which of two images contained a region where the luminance statistics differed from those in the rest of the scene. We also compared the data to predictions from an ideal observer model. Finally we asked whether manipulating the geometric structure of the images affected performance on our segregation task. The measurements provide baseline information that can be exploited in future experiments that study illumination segmentation in more complex scenes to determine the role of additional information sources.

Overview

Observers viewed displays consisting of two side-by-side grayscale checkerboards (e.g. Figure 2). The luminances of the patches in one of the checkerboards were drawn from a single probability distribution (a truncated Gaussian); for the other checkerboard two distributions were used. Observers indicated which of the two checkerboards contained patches with luminances drawn from two distributions. The

experiment embodies an abstracted version of the illuminant segmentation task that observers confront in natural viewing.

Performance was measured in three experiments. In Experiment 1 we varied the number of patches drawn from the second distribution, as well as observers' uncertainty about the spatial locations of these patches. Performance was compared to that of an ideal observer model, as well as a number of alternative simpler models.

In Experiments 2 and 3, geometric manipulations were introduced. In Experiment 2, these consisted of a) varying the contiguity of the patches drawn from the second distribution and b) changing the spatial arrangement of the images to introduce Ψ-junctions. This was done to test the idea that the geometric cues would limit integration of photometric information to patches grouped together by those cues. In Experiment 3, binocular depth cues were used to separate patches into two different depth planes, again with the idea that this might lead to processing grouped by depth.

2. Experiment 1 Methods

2.1 Observers

Four observers (one male, three females, mean age = 23) participated in this experiment. Each observer came to the lab for two sessions and was compensated for his or her time. The observers all had Snellen acuity of at least 20/40 (corrected) and scored at least 36/38 correct on the Ishihara color plates (Ishihara, 1998).

2.2 Stimuli and Setup

The stimuli on each trial consisted of two side-by-side grayscale checkerboards (Figure 2). These were presented on a calibrated ViewSonic G220fb CRT monitor. Observers viewed the monitor from a distance of 560 mm, with viewing position stabilized by a headrest-chinrest assembly. Each checkerboard consisted of five rows of five patches, with each patch 29x30 mm (2.97° x 3.07°).[1] The overall size of the 5x5 checkerboards was thus 145x150 mm (14.75° x 15.26°). The two checkerboards were presented against a dark gray background (4.4 cd/m^2) and were separated horizontally by 48 mm (4.91°). The CIE 1931 xy chromaticity of the background and checkerboard patches was held fixed at [0.30, 0.30].

[---FIGURE 2 ABOUT HERE---]

On each trial, one of the two checkerboards (left or right) was randomly designated the *test checkerboard* and the other the *standard checkerboard*. The luminances for the 25 patches in the standard checkerboard and almost all of the patches in the test checkerboard were randomly drawn from a truncated Gaussian distribution with mean 15.0 cd/m^2, standard deviation 5.0 cd/m^2, and truncation range [5.0, 50.0] cd/m^2. For the test checkerboard, the luminances of the remaining patches, which we refer to as the *test patches*, were drawn from a different distribution. The test patch distribution was a truncated Gaussian whose mean and standard deviation were larger

[1] Computation of visual angle does not take off-axis effects into account. Order of all dimensions specified is vertical first and then horizontal.

than the standard distribution by a multiplicative constant. Across trials, this constant varied between 1 (minimum) and 2.5 (maximum). These corresponded to test patch distributions with mean 15.0 cd/m^2, standard deviation 5.0 cd/m^2, and truncation range of [5.0, 50.0] cd/m^2 (minimum) and with mean 37.5 cd/m^2, standard deviation of 12.5 cd/m^2, and truncation range of [12.5, 125.0] cd/m^2 (maximum).

2.3 Procedure

The observer's task on each trial was to indicate via a button press which of the two checkerboards was the test checkerboard. We found during pilot experiments that the most effective instructions were to ask observers to identify the checkerboard that contained some patches drawn 'from a larger range of luminances than the rest of the patches'. These instructions were used. The full instructions are provided in the supplemental material available at http://color.psych.upenn.edu/supplements/distribdiscrim/. Feedback was provided by a tone whenever the observer made an error.

The multiplicative constant for the test distribution parameters was adjusted trial-by-trial based on whether the observer's response was correct, using a 2-down 1-up staircase procedure (Levitt, 1971): the test distribution parameters decreased after every two correct responses and increased after every incorrect response. The threshold test distribution, parameterized by its mean, at which observers were correct on 75% of the trials was estimated by fitting a Weibull psychometric function to all of the data, using the psignifit toolbox version 2.5.6 (Wichmann & Hill, 2001; see http://bootstrap-

software.org/psignifit). The psignifit software implements a maximum-likelihood fit of the Weibull parameters along with a lapse rate parameter.

Two variables were manipulated across conditions: the number of test patches in the signal checkerboard and the observers' certainty about the location of these patches. The number of test patches varied between 1 and 5. The test patches were either at fixed locations known to the observer (in the center row of the signal checkerboard; location-known condition; top panel of Figure 2) or were randomly selected on each trial from the 25 checkerboard patches (location-unknown condition; bottom panel of Figure 2). This resulted in a total of 10 conditions (5 patch numbers x 2 levels of certainty). Conditions were blocked and observers were informed beforehand which condition was being tested. The order of conditions was randomized for each observer. For each condition, observers ran five blocks of 100 trials before moving onto the next condition.

3. Experiment 1 Results

Figure 3 plots average thresholds as a function of number of test patches. Two broad effects are apparent. First, in both the location-known and location-unknown conditions, thresholds fall with increasing number of test patches, with the largest drop occurring between one and two test patches. Second, knowledge of test patch location decreased thresholds for all test patch numbers. On average, the location-known condition thresholds were 4.9 cd/m^2 lower than the location-unknown thresholds.

Ideal observer thresholds are also plotted in Figure 3. The ideal observer calculations are described in the appendix. The broad patterns visible in the experimental data are also apparent for the ideal observer: thresholds decrease with increasing numbers of test patches and thresholds are greater for the location-unknown case. In addition, the ideal observer thresholds are slightly lower than the human observer thresholds. We scaled the ideal observer data to fit the experimental data (red lines in Figure 3, separate scaling for location-known and location-unknown conditions). In each case, the ideal observer predicts the dependence of threshold on number of test patches. The scale factor required was slightly larger for the location-unknown case (1.03 for location known; 1.09 for location unknown.) Thus human performance is close to ideal in the location-known case, but uncertainty in test patch location adds an additional cost, beyond what would be experienced by an ideal observer faced with the same uncertainty.

[---FIGURE 3 ABOUT HERE---]

4. Intermediate Discussion 1

Although the data are consistent with an ideal observer model that efficiently integrates information from the test patch locations to judge which image contained the distributional change, it is possible that similar performance could be obtained for our stimuli using simpler strategies. We thus considered models based on three such strategies: a) a *mean luminance model* that chooses the checkerboard with the larger mean luminance on each trial, b) a *highest luminance model* that chooses the checkerboard containing the highest luminance patch on each trial, and c) a *highest range*

model that chooses the checkerboard with the highest luminance range on each trial. We constructed variants of these models for the location-known and location-unknown conditions. For the location-known condition, the mean luminance and highest luminance were evaluated only over the known test patch locations in each checkerboard, while the range was obtained by subtracting the lowest non-test patch location luminance from the highest test patch location luminance. For the location unknown condition, the mean, highest luminance, and luminance range were computed over all the locations in each checkerboard.

[---FIGURE 4 ABOUT HERE---]

Figure 4 replots the data from Experiment 1 along with the predictions from each of these models. The data from the location-unknown condition clearly falsify the mean luminance model, as that model's dependence on test patch number has a very different form from the measurements. Both the highest luminance model and luminance range models, however, make predictions quite similar to those of the ideal observer model and are not ruled out by the data. Of note is that these models require only the most primitive form of integrating information across test patch locations: identifying which patches in each checkerboard images have the highest (and lowest) luminances.

To investigate further, we conducted a classification image analysis of the relationship between the trial-by-trial variation in the stimulus and the trial-by-trial responses, both for our human observers and simulations of performance based on the

models. Such analyses can reveal the relative weighting of various sources of stimulus information that lead to the same level of overall performance (Ahumada & Lovell, 1971; Gold, Murray, Bennett, & Sekuler, 2000; Murray, Bennett, & Sekuler, 2002; Abbey & Eckstein, 2006). The intuition behind this kind of analysis is that the correlations between aspects of the stimulus that contribute to the decision and the response will be high, while the corresponding correlations for aspects of the stimulus that are irrelevant will be low.

To implement the analysis, we fit the coefficients of a logistic regression model (Alexander & Lutfi, 2004) to estimate the weights placed on patch luminance differences from trial-by-trial responses:

$$\mathbf{r} = \frac{1}{1 + e^{-\mathbf{P}\cdot\mathbf{w}}} \qquad \textbf{(1)}$$

In equation (1), **r** is a binary column vector coding responses (left/right) as 1s and 0s, **w** is a row vector with the weights found by the regression, and **P** is a matrix whose rows were per-trial vectors of luminances obtained from the stimulus. The values in the rows of **P** were obtained as follows: on each trial, for each checkerboard (left and right), we took the luminances from the test patch locations, the two most luminous non-test patches, and the two least luminous non-test patches. We sorted the luminances for the test and non-test patches in descending order separately, for each checkerboard. Then, we took the differences between the corresponding luminance-ranked patches (i.e. the most intense test patch on the left minus the most intense test patch on the right, the second most intense test patch on the left minus the second most intense test patch on the

right, etc.). The regression thus tells us how much weight is assigned to the luminance differences between corresponding rank-ordered test patches across the two checkerboards, and to the luminance differences of corresponding rank-ordered non-test patches at the high and low luminance end of the non-test patch range. We analyzed only the data from the location-known conditions in this way, as we found that our dataset did not have sufficient power to provide reliable estimates of the weights when all checkerboard squares were considered.

[---FIGURE 5 ABOUT HERE---]

The estimated weights are plotted in Figure 5 for the human observers, the ideal observer model simulation, and the highest luminance model simulation. Because our interest is in the relative importance of the luminance-ordered test patch differences, we normalized the weights within model/condition by the weight assigned to the most luminous test patch.

For both model simulations, the weights for the non-test patches cluster around zero, indicating minimal influence on the decision, consistent with the fact that these test patches have no influence on the model's decision. The scatter of the weights around zero provides a visual sense of how precisely the weights are determined, given the number of simulated trials. For the highest luminance model, only the most luminous test patch receives a high weight; the weights for the remaining test patches resemble those of the non-test patches. For the ideal observer, which integrates information from all test

patches, the weights fall as a function of number of test patches. We also estimated weights for the other two models (plots provided in the supplemental materials available at http://color.psych.upenn.edu/supplements/distribdiscrim/). The test patches weights for the mean luminance model are equally high, as all test patches contribute equally to the average. The test patch weights for the highest range model resemble those of the highest luminance model, but with an equally large negative weight on the least luminous non-test patch, as both these patches are necessary in the calculation of the luminance range.

If the human observers were using one of the strategies implemented in the models, then the pattern of their estimated weights should resemble the pattern from that model. The human observers' average weights more closely resemble those of the ideal observer than the other models. In particular, the human observers assign positive weight to multiple test patches. As with the weights from the ideal observer model, the human weights for the test patches decrease with test patch luminance. This analysis suggests that humans integrate photometric information over multiple test patches. The fact that overall the weights obtained for the human observer decrease more rapidly with ordinal test patch luminance than for the ideal observer is consistent the fact that human efficiency with respect to the ideal observer is less than 1.

5. Experiment 2 Methods

5.1 Purpose

Experiment 1 established that observers can perform the luminance distribution discrimination task and the classification image analysis showed they are able to integrate

information across patch locations. In Experiment 2, we explored the effect of other display manipulations on performance. Geometric cues were introduced in an attempt to negatively affect performance, by inducing segregation of the checkerboards into regions that were spatially incongruent with the different luminance distribution regions. The hypothesis we sought to test was that such geometric cues impose a mandatory segregation on the scene and prevent the use of photometric information from both sides of the geometrically-cued boundary.

Suppose, for example, that a geometric cue grouped one of two test patch locations separately from the other test patch location. If the observer were unable to use information from the two regions, performance for the two test patch case with the geometric cue would then resemble performance for one test patch without the geometric cue. To put it another way, thresholds would be predicted to increase across the geometric manipulation.

Two types of geometric cues were tested: separation and the presence of Ψ-junctions. For the separation manipulation, the test patches were located non-contiguously with each other. For the Ψ-junctions, the shapes of the checkerboards were manipulated in a way consistent with folding part of the checkerboard in depth.

5.2 Observers

Three observers participated in this experiment. Each observer came to the lab for one session and was compensated for his or her time. All three had participated in Experiment 1.

5.3 Stimuli and Procedure

[---FIGURE 6 ABOUT HERE---]

The stimuli were presented using the same apparatus as for Experiment 1, and the luminance statistics were also the same.

Observers were tested in three main conditions, all using two test patches. The first was a replication of the location-known condition of Experiment 1 using two test patches ("*center*"). The test patches were the center patch and the patch to its right within each checkerboard.

The second condition presented two test patches in a spatially-separated configuration ("*sep*"). In this condition the two test patches were the center patch and the lower-right hand corner patch of each checkerboard (top panel of Figure 6).

The third condition ("*psi*") presented two test patches on a square + parallelogram checkerboard (bottom panel of Figure 6). To generate this checkerboard, the rightmost vertices of the square checkerboard were shifted down a distance equal to two side lengths of each square (5.94°) and the next rightmost vertices down by one side length (2.97°). This manipulation introduced Ψ-junctions between the third and fourth columns

of each checkerboard. Interpreted as a three-dimensional object, this checkerboard would appear to have its rightmost two columns folded in depth. Note, however, that this was a purely monocular manipulation: no binocular depth cues were used. The test patches were again the center patch and the patch immediately to its right.

Three control conditions were also run, all with one test patch. The first was also a replication of the location-known condition of Experiment 1, but with only one test patch ("*center*"). The other two used the Ψ-junction checkerboard, with the test patch in the center ("*psiC*") or immediately to the right of center ("*psiR*"). These control conditions tested the effect of introducing parallelograms on thresholds for a single square patch and for a single parallelogram patch.

In all conditions of Experiment 2, observers knew the locations where the test patch or patches could appear (location-known). The observer's task was the same as in Experiment 1—to pick the test checkerboard containing the test patches. Here, however, observers ran one block of trials from each of the conditions in random order, then repeated all the conditions in a different order, until five blocks of each condition were obtained. As in Experiment 1, observers were informed before each block of trials as to what condition was being run for that block. Psychometric functions were again fit to the data, and the threshold value estimated as a measure of performance.

6. Experiment 2 Results

Mean thresholds are plotted in Figure 7 for the various conditions tested. A repeated measures ANOVA with observer as a random factor indicated a significant effect of condition, $F(5,10) = 10.4$, $p \leq 0.001$. Examination of the plot suggests that this result was driven by the effect of test patch number: conditions with two test patches led to lower thresholds than conditions with one test patch. This difference is comparable to the difference between the 1- and 2-test patch location-known conditions in Experiment 1, replotted in Figure 7 with X symbols.

[---FIGURE 7 ABOUT HERE---]

According to the ideal observer model, only the number of test patches and knowledge of their locations should affect performance. If thresholds for the *psi*/*sep* two-patch conditions are greater than those for the *center* two-patch condition, then those manipulations have a detrimental effect on human performance not accounted for in the model. However, the data in the left panel of Figure 7 show a minimal effect on threshold for those manipulations: a repeated measures ANOVA with observer as a random factor was not significant, $F(2,4) = 1.27$, $p = 0.37$. In addition, note that the non-significant trend towards an elevated threshold for the *sep* condition is small relative to the effect of test patch number.

[---TABLE 1 ABOUT HERE---]

The thresholds from the control conditions using one test patch were also not different from each other, $F(2,4) = 4.08$, $p = 0.68$. Simply introducing parallelograms into the image does not appear to affect performance for a single test patch, regardless of its shape.

The weights for the second test patch in the three 2-patch conditions were also estimated, and the means across observers are listed in Table 1. The weights were all positive, and not different from each other, $F(2,6) = 0.80$, $p = 0.49$. This pattern suggests that similar strategies were used regardless of scene geometry, and that both test patch luminances affected the decision.

7. Intermediate Discussion 2

We did not find evidence for an effect of the geometric manipulations in Experiment 2. Interestingly, this suggests that for the location-unknown condition of Experiment 1, the performance decrease may have been driven primarily by uncertainty and not by the non-contiguity of the target patches. That is, the fact that a random draw of patches within a checkerboard in the location-unknown condition of Experiment 1 often resulted a noncontiguous configuration of test patches may not have affected performance directly, at least if the results from the single separation manipulation of Experiment 2 generalize to more test patches.

One possibility for our failure to find an effect of geometry in Experiment 2 is that the geometric manipulations we used were either ineffective at introducing

segmentation, or at least not effective enough to overcome a larger effect of the provided photometric cues. In Experiment 3 we introduced binocular disparity as a different cue to geometric segregation. The logic was the same as in Experiment 2—would performance for a 2 test patch condition resemble performance for a 1 test patch condition in the presence of disparity cues that segregated the test patches into separate depth planes?

8. Experiment 3 Methods

8.1 Observers

Three observers (1 male, 2 female, mean age = 20) participated in this experiment. None had participated in Experiments 1 or 2. Each was screened using the same screening procedure and criteria as in Experiment 1. An additional screening test for stereopsis was also used. This test was performed using the same apparatus as for the experiment (see below). A 2.62° square patch appeared either in front or in back of a fixed background plane due to binocular disparity, and observers were asked to indicate where the patch appeared ("front" or "back"). The background plane was rendered to be 764 mm away from the observer. The simulated depth (i.e. amount of disparity) was adjusted via a 2-down 1-up staircase procedure (Levitt, 1971) to estimate 71% accuracy on this task. All observers had thresholds below a simulated depth change of 10 mm.

8.2 Setup and Stimuli

The stimuli were presented on a different apparatus from the previous two experiments. This stereo apparatus, illustrated in Figure 8, consisted of two calibrated hp p1230 CRTs controlled by a single computer. Observers sat with their heads stabilized

via a chinrest in front of a black felt-covered face plate. They viewed the stimuli through two 30 mm x 30 mm square openings in the plate. The distance between the centers of the two openings was 40 mm. A black cardboard divider sat perpendicular to the face plate, preventing overlap between the visual input to the two eyes.

Each eye received input from a CRT whose light was reflected off an angled mirror before reaching the eye. The optical distance of the CRTs to the eyes was approximately 764 mm. The apparatus was aligned by replacing the mirrors with beamsplitters and aligning a grid image on each monitor to a physical grid located 764 mm from the eyes.

The stimuli closely resembled the square checkerboards from the location-known condition in Experiment 1, with the following changes. The checkerboards were rendered to be 764 mm away from the observer (11.21° x 11.21°), and the space between checkerboards was 28 mm (2.10°). Owing to a smaller monitor gamut, the standard distribution had a mean of 6.0 cd/m^2 and a standard deviation of 3.0 cd/m^2, truncated on the interval [1.5, 15] cd/m^2. The test distribution had minimum parameters equal to the standard distribution parameters. Its maximum parameters were a mean of 17.7 cd/m^2, standard deviation of 8.8 cd/m^2, and truncation range [4.4, 44.2] cd/m^2.

[---FIGURE 8 ABOUT HERE---]

The number of test patches was always one or two, in different conditions. For the 1-test patch conditions, the test patch was the center patch of the checkerboard; for

the 2-test patch conditions, the test patches were the center patch and the patch immediately to its right.

The experimental condition was a 2-test patch condition ("*Mixed*"). However, the center patch was rendered with a binocular disparity, so that it appeared to float 100 mm in front of the rest of the checkerboard. The size of the floating patch was increased to 34.5 mm x 34.5 mm (2.59° x 2.59°) so that with the disparity it appeared to be approximately the same size as it was while coplanar with the checkerboard, 30 mm x 30 mm (1.12° x 1.12°). The test patch to the right remained in the same plane as the checkerboard.

A series of control conditions were also tested. The 1- and 2-test patch cases from the location-known condition of Experiment 1 were replicated on this stereo apparatus ("*BackPlane*", for 1 and 2 test patches). Additionally, two control conditions using 1 and 2 test patches brought forward in depth using binocular disparity were also run ("*FrontPlane*", for 1 and 2 test patches). These conditions were used to check for changes in performance due to the presence of binocular disparity in the stimulus, without the segregation of test patches into two depth planes. Finally, a random depth plane condition ("*RandomPlanes*") was tested. This was a 2-test patch condition where the potential test patches on both checkerboards could each randomly appear at either depth plane. The depth arrangement of the two checkerboards on every trial was yoked, e.g. if both potential test patches on one were closer in depth, so were the corresponding test patches on the other.

8.3 Procedure

The procedure was similar to the one used in Experiment 2. Observers completed five blocks of all six conditions in random order, and knew in advance of each block which condition was being run. Each block consisted of 100 trials.

9. Experiment 3 Results

The mean data for three observers are plotted in Figure 9 as a function of condition. A repeated-measures ANOVA with observer as a random factor indicated a significant main effect of condition, $F(5,10) = 4.4$, $p \leq 0.02$. As with Experiments 1 and 2, thresholds for the one-test patch conditions are higher than those for the two test patch conditions.

[---FIGURE 9 ABOUT HERE---]

The critical comparisons are for the two-patch conditions. Here thresholds did not differ across the conditions we ran. A repeated-measures ANOVA with observer as a random factor revealed no effect of condition, $F(3,6) = 0.6$, $p = 0.66$. That is, separating the two test patches in depth did not affect threshold, nor was threshold different in the condition where the depth of the two test patches was randomized on every trial.

As in Experiment 2, the weights for the second test patch in the four 2-patch conditions were also estimated and listed in Table 1. Again, the weights were not found to be different across condition, $F(3,8) = 1.08$, $p = 0.41$.

10. General Discussion

10.1 Summary

In Experiment 1, we investigated how well observers use photometric information to detect changes in luminance distributions. We found that observers perform this task with high-efficiency, relative to an ideal observer. This was true both when the test patch locations were known and when there was uncertainty about these locations. Efficiency was somewhat higher, however, in the location-known case. In both cases, the dependence of threshold on the number of test patches was also well-modeled by the ideal observer. Our classification image analysis of the trial-by-trial responses showed that although high-efficiency for our stimuli could be achieved with a simple strategy that only relied on the highest luminance in the two checkerboard images, observers appeared to follow the ideal observer in that they integrated photometric information from multiple locations. In Experiments 2 and 3, we showed that simple geometric manipulations did not affect performance on our task.

Taken together, our findings suggest that the visual system is quite sensitive to the sorts of changes in luminance distributions that might indicate changes of illumination within a scene in the presence of uncertainty about surface albedo. That is, the visual system is fairly efficient at using such photometric cues to perform a task that models illuminant segregation. Although for low-level perceptual tasks human efficiency is generally low compared to that of ideal observers (~5%, Banks, Geisler, & Bennett, 1987), higher efficiencies have been reported previously for tasks such as symmetry detection (Barlow, 1980).

Overall, thresholds in our experiments changed only with test patch number and uncertainty about test patch location, but not as a result of manipulations of test patch separation or geometric cues that might segment the image. Our data did not reveal a significant effect of introducing geometric segregation cues on performance. A caveat, of course, is that conclusions in this regard hold only up to the power of our data. Our data did, however, contain sufficient power to reveal changes in performance between presentation of one and two test patches.

The lack of geometric effects in our experiments is perhaps not surprising, given that our task was structured so that photometric cues provided the only information available to perform the task. What our results do show is that when photometric information is available, the visual system can integrate this information across spatial boundaries created by geometric factors. An interesting question, but one different from that we studied, is how geometric and photometric information interact when both types of information are task-relevant.

10.2 Relation to Other Studies

Our work makes contact with a number of related threads in the literature. We touch on these below.

Illuminant perception

The literature on illumination perception is much smaller than that on surface perception, but there are a number of studies that relate to our current work. Koenderink

et al. (2007) asked observers to adjust the illuminant impinging on one object in a scene so that it matched the illumination field of the scene as a whole. The fact that observers could do this with reasonable precision indicates that they could discriminate local changes in illumination within a single image, broadly consistent with our results. Earlier studies (e.g., Beck, 1959; Beck, 1961; Oyama, 1968; Noguchi & Masuda, 1971; Kozaki & Noguchi, 1976; Noguchi & Kozaki, 1985; Rutherford & Brainard, 2002) also employed illuminant matching or explicit judgments of the illuminant, but considered between rather than within scene variation.

Gerhard & Maloney (2010a) showed that observers can differentiate between illumination changes common to all surfaces in a scene and illumination changes that vary from one scene location to another. In a second paper, they found that observer performance in a task that required estimating the motion of a collimated light source was well predicted by an ideal observer model that interpreted photometric changes in the context of noisy knowledge about surface geometry (Gerhard & Maloney, 2010b). In their case, however, the focus was on geometric changes in the three-dimensional location of an illumination source, rather than on the efficacy of photometric cues for illumination discrimination. Similarly, Khang et al. (2006) found that observers could match illuminant source directions across scenes. In the color domain Craven & Foster's (1992) work has a similar flavor, although they cast their measurements in terms of discriminations between a spatially global illumination change and changes in the surface reflectances of the objects within a scene (see also Nascimento & Foster, 2000).

Role of geometric cues

A number of studies show that changes in perceived geometry affect perceived surface lightness (Hochberg & Beck, 1954; Gilchrist, 1977; Gilchrist, 1980; Knill & Kersten, 1991; Boyaci, Maloney, & Hersh, 2003; Ripamonti et al., 2004; Radonjić, Todorović, & Gilchrist, 2010). These effects are often interpreted as resulting from an effect of geometry on the (perhaps implicitly) perceived illumination (see e.g., Brainard & Maloney, 2011). A recent result in this tradition, however, suggests that when strong photometric cues are available, they can dominate geometric information (Gilchrist & Radonjić, 2010).

In related work, a number of lightness illusions (Adelson, 1993; Todorović, 1997; Adelson, 2000; Anderson & Winawer, 2005) also implicate a key role for geometric cues in the perception of surface lightness. In much of this work, the emphasis has been on understanding how the luminance relationships across junctions support particular scene interpretations. Similar themes are found in the literature on transparency (Beck, Prazdny, & Ivry, 1984; Metelli, 1985; Anderson, 1997; Singh & Anderson, 2002; Anderson & Winawer, 2005).

Our results do not contradict the conclusion that geometric factors play a role in the perception of scene illumination. They do, however, emphasize the need to understand in detail how information carried by photometric and geometric cues interact.

Formal connections

The formal structure of our task is that the observers had to identify which of two images contained patches drawn from a mixture of two luminance probability distributions, and which contained luminance patches drawn from a single distribution. At this formal level, our task is thus closely related to work on the perception of texture (for a review, see Landy & Graham, 2004), where textures are defined in terms of the statistical properties of their luminance distributions. Despite the formal similarity, however, there are important content differences between most texture experiments and ours. For example, texture work often holds the mean luminance and variance (contrast) constant across distributional manipulations, so as to allow investigation of the structure carried by higher-order statistical regularities. In addition, these textures are typically generated using small, spatially-contiguous micro patterns, rather than the more macroscopic spatial structure of interest when one considers illumination discrimination.

Our task also shares formal features with a contour integration task introduced by Field et al. (1993), where observers were asked to detect the presence of a coherently oriented contour of Gabor patches embedded in a field of randomly-oriented Gabor patches.

Future directions

Our work represents an initial foray towards understanding how photometric information enables illuminant segregation. To make progress, we employed simple stimuli and studied performance using a simple psychophysical task. These

simplifications allowed us to observe a number of clear regularities within the laboratory model we studied. Nonetheless, it is worth keeping in mind some of the limitations of this model. First, the test checkerboards did not produce a strong perceptual sense of a collection of surfaces seen under two illuminants, and thus may not have engaged all of the mechanisms that normally subserve illuminant segregation. Second, our use of a two-alternative forced-choice procedure simplified the task demands, relative to the case of viewing a single image that might contain multiple regions of illumination. Third, the only task relevant information in our stimuli was photometric and this may have weakened any potential geometric effects. Expanding the research to richer stimuli and tasks, so as to overcome these limitations is a clear direction for future research.

Appendix

Ideal Observer Calculations

An ideal observer was developed for each condition in Experiment 1 and its performance characterized through simulation. For each condition, 100 trials each of 38 multiples of the test distribution parameters were simulated, and for each simulated trial the ideal observer calculation indicated which checkerboard contained the test patches, based on the luminance of all 50 checkerboard patches. Each condition was repeated three times in a run of the simulation; the results were analyzed in the same way as the human data. For each simulation, the mean threshold from the three repetitions was taken as the measure of performance.

In the location-known conditions, means from ten simulations were averaged together for each value of test patch number. This was also done for the 1-, 2-, and 3-test patch conditions for the location-unknown conditions. The computations for the 4- and 5-test patch location-unknown conditions were lengthy, and for these only a single simulation mean (of three repetitions) was obtained.

The ideal observer's choice was based on the log likelihood ratio

$$\ell(\mathbf{x}) = \log\left(\frac{p(\mathbf{x} \mid TestOnLeft)}{p(\mathbf{x} \mid TestOnRight)}\right) \tag{A1}$$

where the vector $\mathbf{x}$ represents the luminances of the 50 checkerboard patches presented on a particular trial. If $\ell(\mathbf{x})$ was greater than 0, the ideal observer indicated that the test was on the left; if $\ell(\mathbf{x})$ was less than or equal to 0, the ideal observer indicated that the test was on the right. For a given trial, the vector $\mathbf{x}$ can be thought of as the concatenation of the vectors $\mathbf{x}^{Left}$, representing the luminances of the 25 checkerboard patches on the left checkerboard, and $\mathbf{x}^{Right}$, the luminances of the 25 checkerboard patches on the right checkerboard.

Location-known condition

For the location known conditions, the log likelihood of the data given that the test was on the left is:

$$\log(p(\mathbf{x} \mid TestOnLeft)) = \sum_{i \in \{t\}} \log(p_t(x_i^{Left})) + \sum_{i \in \{s\}} \log(p_s(x_i^{Left})) + \sum_{i=1}^{25} \log(p_s(x_i^{Right})) \quad . \qquad \text{(A2)}$$

In this expression, *i* indexes patch location within a single checkerboard (left or right), $\{t\}$ represents the indices of the test patches within the test checkerboard, and $\{s\}$ represents the remaining indices within the test checkerboard. The probability $p_t(x)$ is the probability of observing luminance x at a single patch under the test distribution, and $p_s(x)$ is the probability of observing luminance x at a single patch under the standard distribution. The corresponding expression when the test is on the right is

$$\log(p(\mathbf{x} \mid TestOnRight)) = \sum_{i \in \{t\}} \log(p_t(x_i^{Right})) + \sum_{i \in \{s\}} \log(p_s(x_i^{Right})) + \sum_{i=1}^{25} \log(p_s(x_i^{Left})) \quad . \qquad \text{(A3)}$$

The probability $p_s(x)$ was evaluated using the probability density function of a truncated Gaussian distribution with mean μ = 15 cd/m^2 and standard deviation σ = 5 cd/m^2 (the parameters of the luminance distribution for stimulus patches under the standard illuminant). The Gaussian density function was truncated between the range [5, 50] and renormalized so that the total probability was 1, to match how the stimuli were generated for the experiments.

For the case where the parameters of the test distribution are known to the observer, the probability $p_t(x)$ would be evaluated using the same basic method as described for the standard distribution, but with the truncated Gaussian having a mean, variance, and truncation range computed for the test distribution.

Because the experiments were run using a staircase method, there was uncertainty about the parameters of the test distribution. To model this uncertainty, in a separate simulation $p_t(x)$ was evaluated as a weighted sum of the likelihood for a given set of test distribution parameters, with the weights given by the probability of those parameters. The distribution of the parameters, $p(TestLevel)$, was estimated by creating a histogram of the parameters used in the observers' experimental runs corresponding to the condition being simulated. With this,

$$p_t(x) = \int_{TestDistParam\mu=15,\sigma=5}^{TestDistParam\mu=37.5,\sigma=12.5} p_t(x \mid TestDistParam)\, p(TestDistParam) \quad . \qquad \text{(A4)}$$

We found that adding this uncertainty had little effect on the predictions, and in interest of computational efficiency ran our main simulations with the test distribution parameters known.

Location-unknown condition

Before providing the general equations for the location-unknown conditions, we first develop the ideas for a simplified example. Suppose that there are only four patches in each checkerboard, and that there are two signal patches to be detected. Let $\mathbf{x}$ be the vector of eight luminances concatenated from the vectors $\mathbf{x}^{Left}$, the luminances of patches $x_1^{\text{Left}}, \ldots x_4^{\text{Left}}$ belonging to the left checkerboard, and $\mathbf{x}^{Right}$, the luminances of the four patches $x_1^{\text{Right}}, \ldots x_4^{\text{Right}}$ belonging to the right checkerboard.

Suppose that the test is on the left. In this case, there are $\binom{4}{2} = 6$ possible combinations of the standard illuminant and test illuminant patch locations in the left checkerboard, each equally likely. The log likelihood of the observed luminance vector can be written as:

$$\log(p(\mathbf{x} \mid TestOnLeft)) = \log(p(\mathbf{x}^{Left} \mid TestOnLeft)) + \log(p(\mathbf{x}^{Right} \mid TestOnLeft)). \quad \text{(A5)}$$

Here

$$\begin{aligned} p(\mathbf{x}^{Left} \mid TestOnLeft) = {} & (1/6)p_t(x_1^{Left})p_t(x_2^{Left})p_s(x_3^{Left})p_s(x_4^{Left}) + (1/6)p_t(x_1^{Left})p_s(x_2^{Left})p_t(x_3^{Left})p_s(x_4^{Left}) + \\ & (1/6)p_t(x_1^{Left})p_s(x_2^{Left})p_s(x_3^{Left})p_t(x_4^{Left}) + (1/6)p_s(x_1^{Left})p_t(x_2^{Left})p_t(x_3^{Left})p_s(x_4^{Left}) + \\ & (1/6)p_s(x_1^{Left})p_t(x_2^{Left})p_s(x_3^{Left})p_t(x_4^{Left}) + (1/6)p_s(x_1^{Left})p_s(x_2^{Left})p_t(x_3^{Left})p_t(x_4^{Left}) \end{aligned} \quad \text{(A6)}$$

is the weighted sum of the likelihoods for all six equally-likely possible combinations of standard and test distribution locations. The expression

$$p(\mathbf{x}^{Right} \mid TestOnLeft) = p_s(x_1^{Right})p_s(x_2^{Right})p_s(x_3^{Right})p_s(x_4^{Right}) \quad \text{(A7)}$$

provides the likelihood of the observed luminances under the standard distribution.

For notational simplicity, the combinations of locations in each term on the right-hand-side of Equation (A6) can be represented as a set of combinations of two *t* and two *s* characters, where *t* in the ith location indicates that the *i*th patch on the left is drawn from the test distribution, $p_t(x_i^{Left})$, and *s* indicates that that the *i*th patch on the left is drawn from the standard distribution, $p_s(x_i^{Left})$. Let C be a matrix of the set of combinations of *t*

and s in (A6) whose rows are C_a for the six combinations a = 1, 2, … 6. Within a given row C_a the columns are indexed by i = 1, 2, 3, 4. We write

$$C = \left\{ \begin{array}{cccc} t & t & s & s \\ t & s & t & s \\ t & s & s & t \\ s & t & t & s \\ s & t & s & t \\ s & s & t & t \end{array} \right\} \tag{A8}$$

where the individual entries may be denoted by C_{ai}. Equation (A6) can then be represented as

$$p(\mathbf{x}^{Left} \mid TestOnLeft) = (1/6) \sum_{C_a \in C} \prod_{i=1}^{4} p_{C_{ai}}(x_i^{Left}) \tag{A9}$$

Combining (A7) and (A9) into (A5), we obtain

$$\log(p(\mathbf{x} \mid TestOnLeft) = \log(1/6) + \log(\sum_{C_a \in C} \prod_{i=1}^{4} p_{C_{ai}}(x_i^{Left})) + \sum_{i=1}^{4} \log(p_s(x_i^{Right})) \tag{A10}$$

We can generalize Equation (A10) to the case of k test distribution patches displayed at n possible patch locations. In this case, there are $N = \left(\begin{array}{c} n \\ k \end{array} \right)$ possible test patch arrangements and we obtain

$$\log(p(\mathbf{x} \mid TestOnLeft)) = \log(1/N) + \log(\sum_{C_a \in C} \prod_{i=1}^{n} p_{C_{ai}}(x_i^{Left})) + \sum_{i=1}^{n} \log(p_s(x_i^{Right})) \tag{A11}$$

A similar equation can be written for $\log(p(\mathbf{x} \mid TestOnRight))$. The log likelihoods are then compared as in Equation (A1).

In a separate simulation, Equation (A11) was modified by using Equation (A4) for every instance of $p_t(x)$ to model observer uncertainty in the parameters of the test distribution. The weights for signal level were again taken from observers' empirical test distribution parameter probability histograms from the corresponding experimental condition. Because of computational limitations, this simulation was done only for the 1-, 2- and 3-test patch cases. As with the corresponding simulations for the location-known case, we found little effect of test level uncertainty and we report results for the case where there was no uncertainty about test illuminant level.

Acknowledgements

Supported by NIH RO1 EY10016, NIH P30 EY001583, NIH T32 EY007035, and NIH T90 DA022763. We thank C. Broussard for technical assistance, and two anonymous reviewers for commenting on an earlier version of this manuscript. A preliminary version of this work was presented as a poster at the 2010 Optical Society of America Fall Vision Meeting.

Figures

Figure 1: Image containing regions with different illumination. The parts of the garden seen through the windowpanes in direct sunlight are adjacent to shadowed walls inside the room. However, the two lighting environments are very different. Image taken from http://www.flickr.com/photos/molinarius/3585205048/, and used with permission of the photographer.

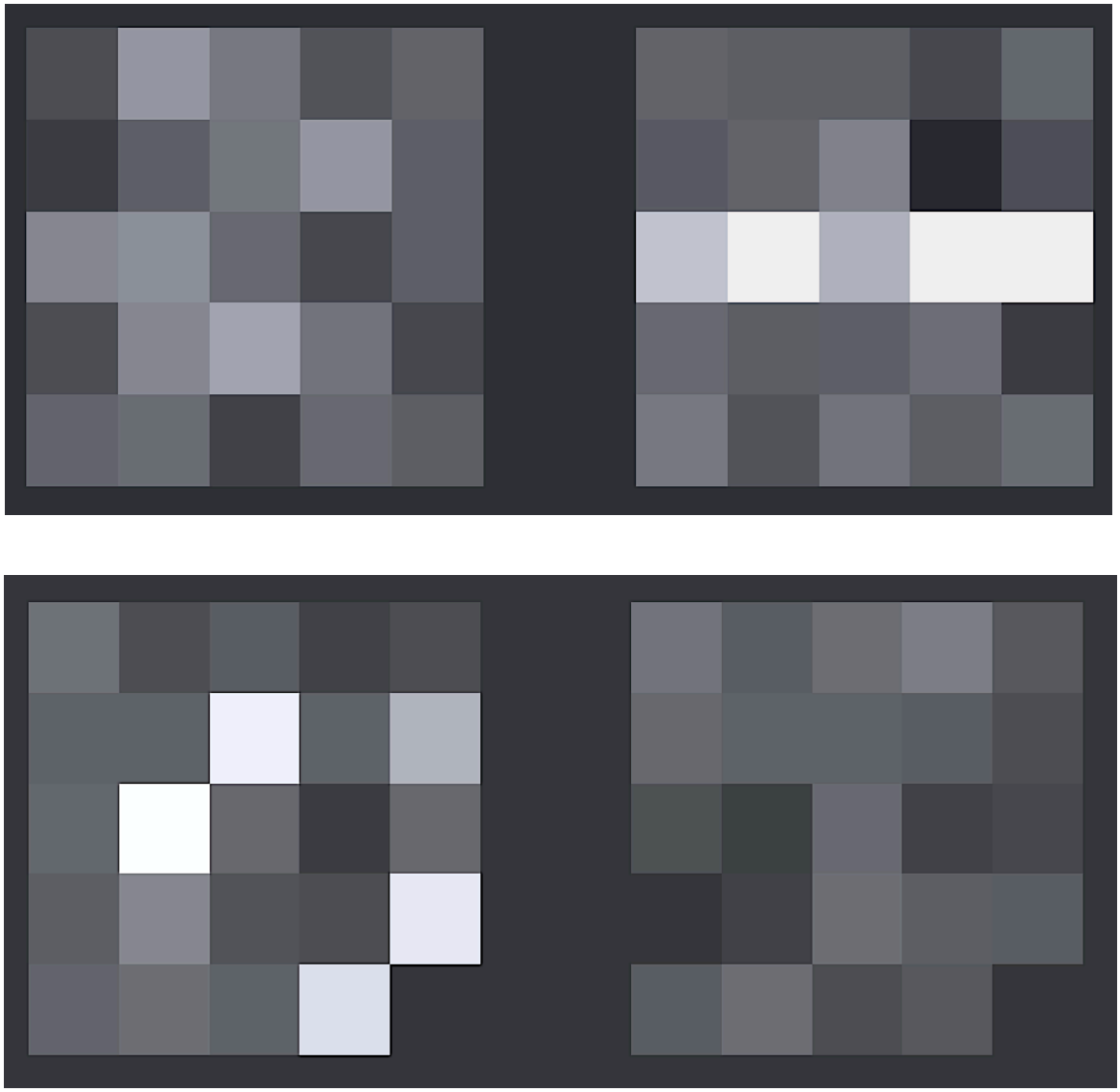

Figure 2: Examples of Experiment 1 stimuli. Top panel, location-known condition; bottom panel, location-unknown condition. In both examples, there are 5 test patches in one of the two checkerboards. For the top panel, they are in the center row on the right; for the bottom panel, they are scattered in the left checkerboard.

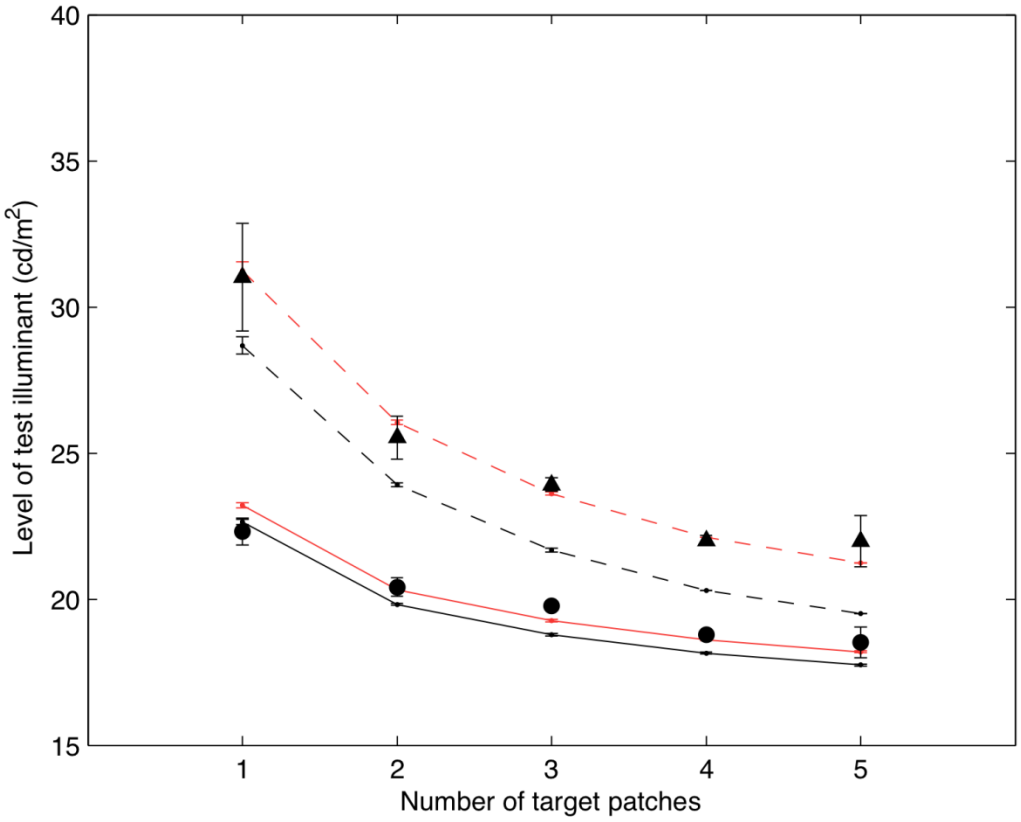

Figure 3: Average (across observers, *n*=4) threshold plotted as a function of number of test patches, for the location-known (solid circles) and location-unknown (solid triangles) conditions. Error bars show +/- 1 SEM. Ideal observer simulation data (dots connected by solid black lines) are also shown (solid line, location-known, dashed line location-unknown). Error bars show +/- 1 SEM over multiple simulation runs, except for the 4 and 5 test patch points for the location-unknown condition where only a single run was done. Red lines/red solid dots show the ideal observer data scaled by a multiplicative constant to best fit the experimental data. Individual observer data for this experiment, as

well as for Experiments 2 and 3, are provided in the supplemental material available at http://color.psych.upenn.edu/supplements/distribdiscrim.

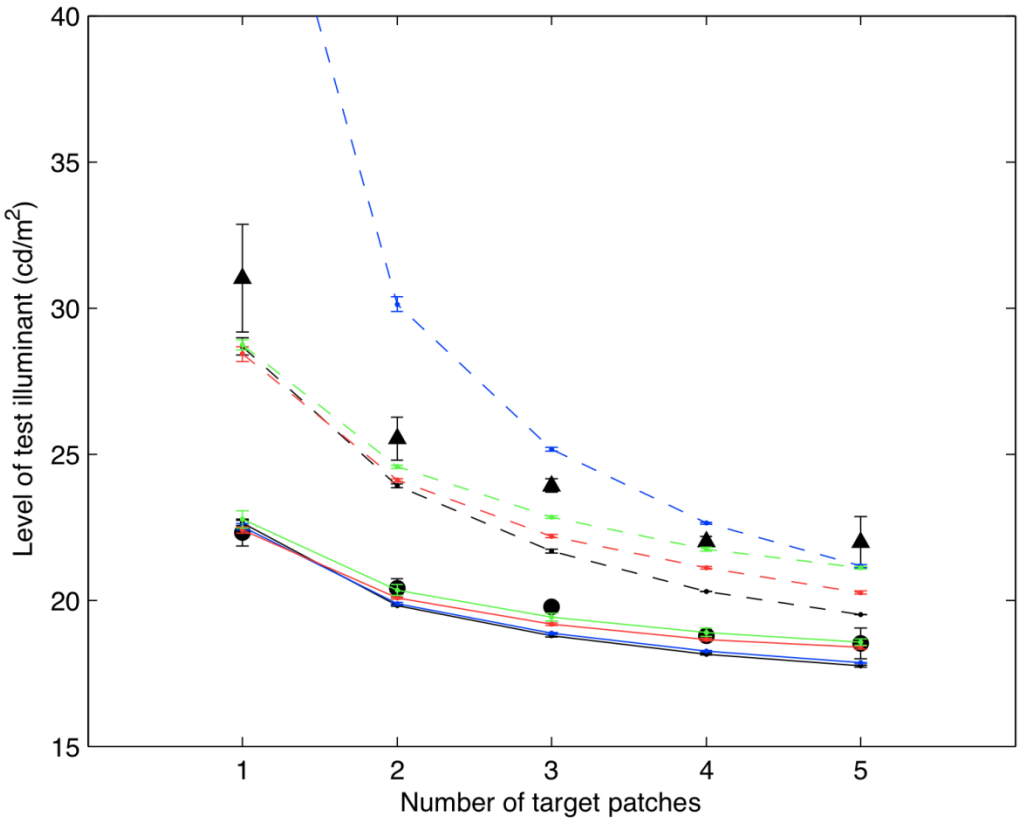

Figure 4: Data from Experiment 1, plotted with simulated thresholds from ideal observer model (black), mean luminance model (blue), highest luminance model (red), and highest range model (green). Circles and triangles are replotted from Figure 3. Solid lines connect points from location-known conditions; dashed lines connect points from location-unknown conditions. Error bars show +/- 1 SEM.

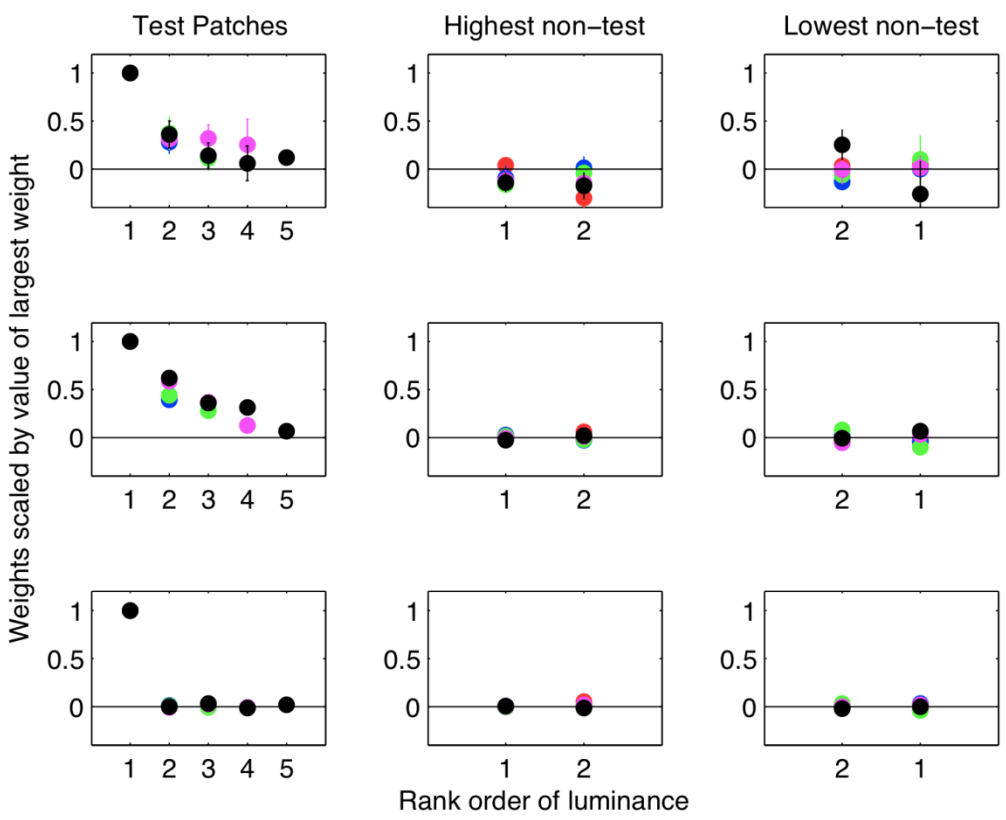

Figure 5: Estimated relative weights from classification analysis for human observers (top row), ideal observer model (middle row), and highest luminance model (bottom row). The leftmost column graphs show the weights for the luminance rank-ordered test patches, the middle column plots graphs show the weights for the two most luminous non-test patches, and the rightmost column plots show the weights for the two least luminous non-test patches. Only the weights for the location-known conditions are plotted. For 1 to 5 test patches, the color code is: red, blue, green, purple, black. The weights for human observers are the mean of four observers, and error bars are one standard error of the mean. For the model simulations, we matched the number of trials to

that used in the human experiments, and ran the simulations four times to match the number of observers. The weights shown for the simulations are the mean of these four runs.

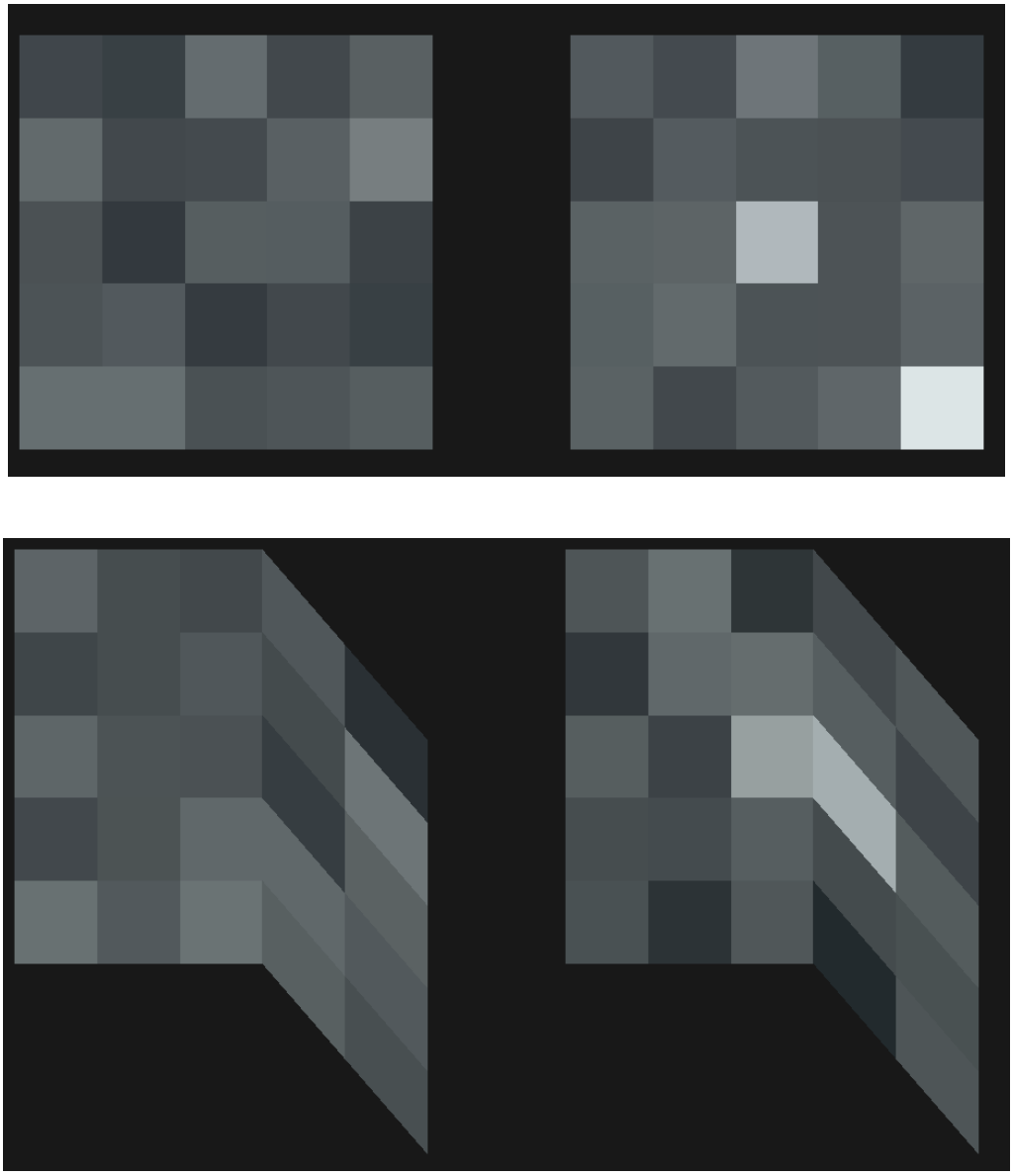

Figure 6: Examples of noncontiguous and Ψ-junction stimuli in Experiment 2. Top panel, the target patches lie in the right checkerboard, center patch and lower right hand corner. Bottom panel, the target patches lie in the right checkerboard, center patch and parallelogram contiguous with it to the right.

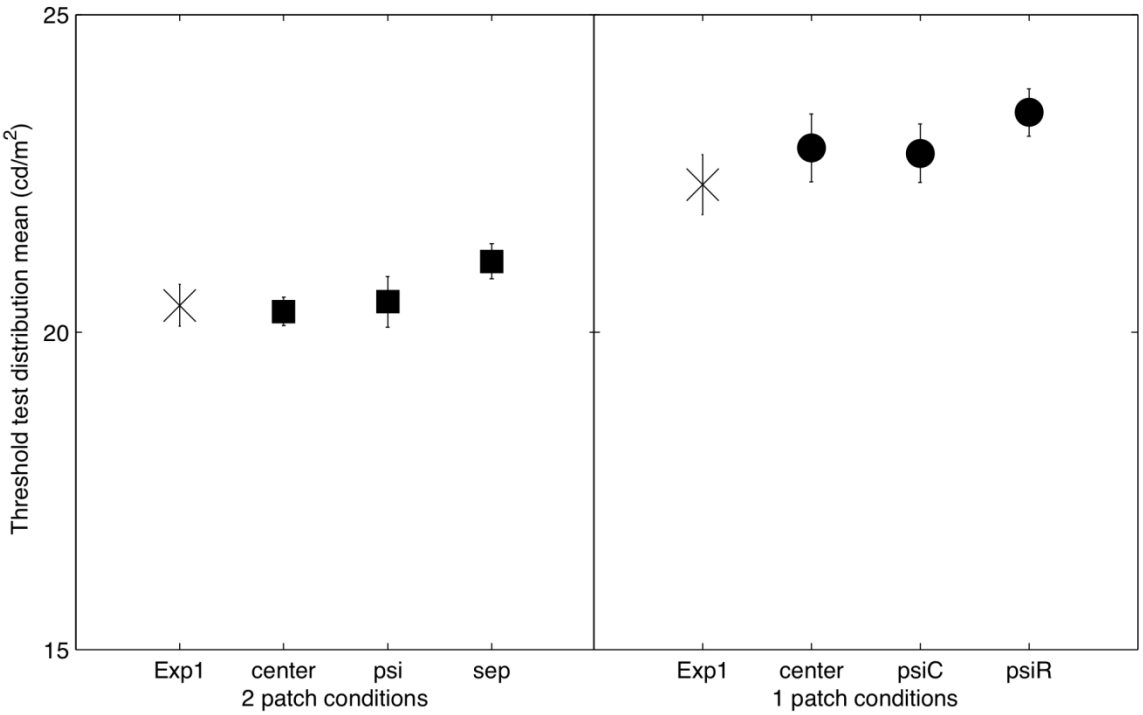

Figure 7: Mean thresholds from three observers in Experiment 2. Left panel, 2 test patch conditions: center row, Ψ-junction checkerboard, separated. Right panel, 1 test patch conditions: center row, and two control conditions with Ψ-junction checkerboards. The 1 and 2 test patch means from the location-known condition in Experiment 1 are plotted with the X-symbols for comparison. Error bars are one standard error of the mean.

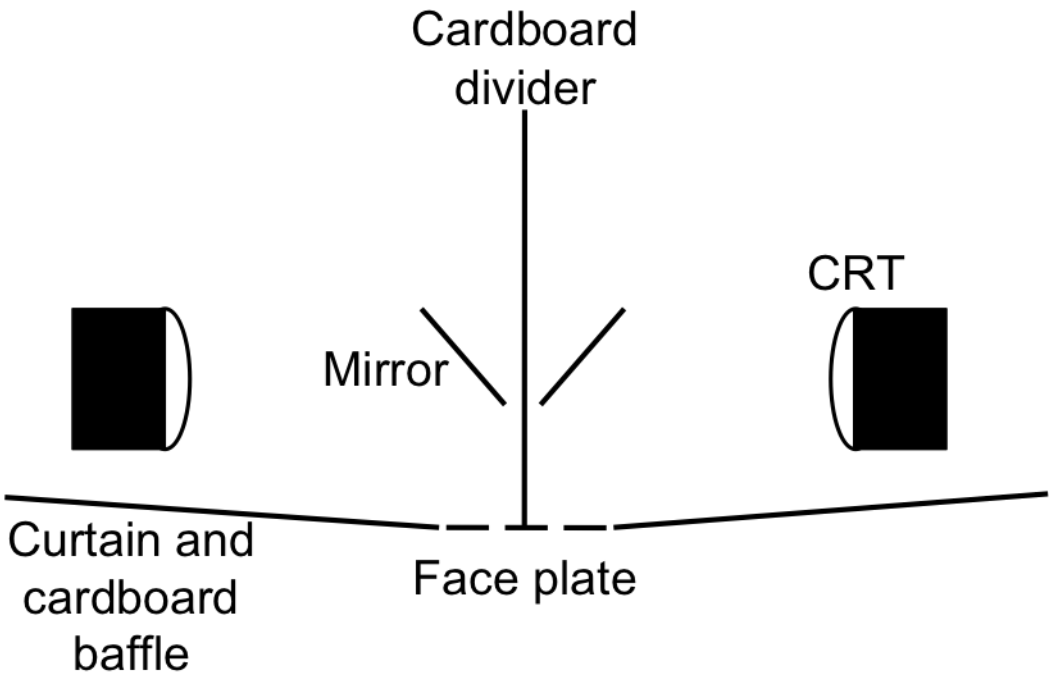

Figure 8: Bird's-eye-view schematic of stereo apparatus. Two CRTs and two mirrors were separated by a black cardboard divider. Each mirror reflected the light from one CRT to one eye of the observer, who was seated in front of the face plate. Curtains hid the CRTs from the observer's view.

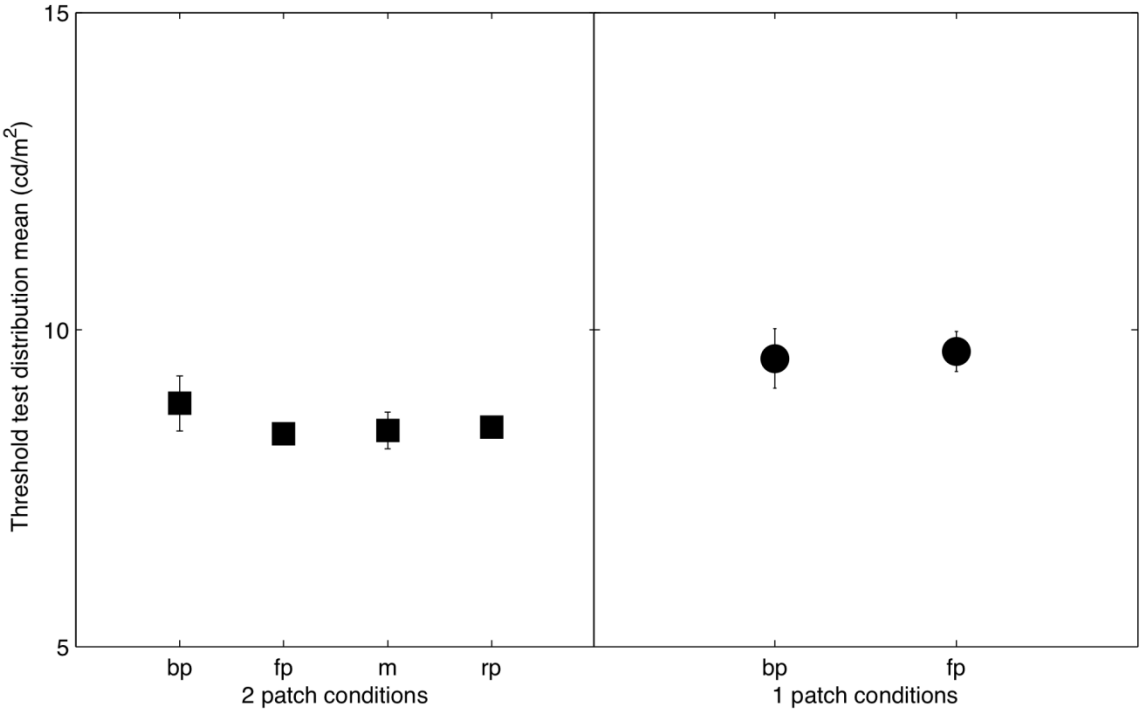

Figure 9: Mean thresholds for three observers in Experiment 3. Left panel, two test patch conditions; right panel, one test patch conditions. Error bars are one standard error of the mean. "*BackPlane*" (bp) refers to the test patches being in the same plane as the checkerboard; "*FrontPlane*" (fp) refers to the test patches in a different plane from the checkerboard. The "*Mixed*" (m) condition has one test patch in the same plane and one in a different plane from the checkerboard. The depth planes for "*RandomPlanes*" (rp) were randomized across trials as described in the text.

Tables

Experiment 2

Test patch 2 weights

	center	*psi*	*sep*
Mean weight	0.57	0.46	0.17
SEM	0.32	0.21	0.13

Experiment 3

Test patch 2 weights

	bp	*fp*	*m*	*rp*
Mean weight	0.61	0.30	0.43	0.27
SEM	0.20	0.14	0.16	0.06

Table 1: Mean estimated weights (n = 3) for second test patch in 2-patch conditions, Experiments 2 and 3. Conditions are labeled as described in the text.

CHAPTER 3: The Role of Photometric and Geometric Cues in Lightness Perception

Abstract

How does the visual system account for illumination in a scene containing both photometric and geometric cues to variation in the illuminant? To answer this question, we measured observers' responses to rendered achromatic stimuli that were consistent with two contextual planes ("contexts") illuminated by a directional light source. In Experiment 1, we presented the contexts oriented at 90° relative to each other. Their luminances were consistent with one of the contexts' being illuminated by lighting with a directional component, and the other receiving only indirect lighting. Observers made lightness matches to a probe tab rendered at a number of orientations in the scene. Photometric cues were manipulated by changing the difference in illuminant intensity across the contexts. Observers' matches depended on both the illuminant manipulation and the tab orientation. In Experiment 2, we varied both photometric and geometric cues, by manipulating the mean context luminance as well as the angle between the contextual planes. Observers' matches to the probe tabs depended on both the photometric and geometric cues independently. Overall, the data suggest that photometric cues and geometric cues affect different aspects of how the visual system represents scene illumination.

1. Introduction

The visual system extracts information about the properties of surfaces in the visual scene. From the light at the eye, the visual system determines quantities such as surface color, texture, and glossiness. To do so, it must disentangle the properties of the surface from the properties of the illumination. This is because the visual signal at the eye confounds both of these properties.

The relevant surface property for a matte achromatic surface is its *reflectance*, a single number that represents the proportion of incident light that the surface reflects. The amount of light reaching the eye is the reflected *luminance*, and is the product of the *illuminant* intensity and surface reflectance. The visual system represents surface reflectance through a perceptual quantity called *lightness*. For lightness to be a reliable indicator of surface identity, the visual system must account for the effect of the illuminant intensity on the reflected luminance. This computational problem is underdetermined, however, because many combinations of illuminant and reflectance can produce the same luminance. The ability of the visual system to stabilize lightness is known as *lightness constancy*.

Typical real-world scenes contain an additional source of complexity—the illuminant is rarely constant across a scene. Figure 1 (of Chapter 2) shows a scene containing multiple illumination regions: areas in shadow and areas receiving direct illumination. Some theories of lightness constancy (Kardos, 1934; Koffka, 1935; Gilchrist et al., 1999; Adelson, 2000) posit that the visual system handles this complexity

by first identifying these regions before accounting for the illumination intensity within each region.

One source of information available to the visual system for segregating illuminant regions is *photometric* and depends on the distribution of luminances in the scene. Surfaces in the world have a reflectance range of about 30:1 (Wyszecki & Stiles, 1982). Consequently, under a single illuminant, images with that distribution of reflectances would be expected to refelct about a 30:1 luminance range. Images with luminances that exceed this range probably result from scenes with substantial variation in illuminant intensity. In a related paper, we showed that the visual system is able to detect differences in the distribution of luminances in simple scenes (Lee & Brainard, 2011).

A different type of information for parsing the visual scene into separate illumination regions is *geometric* in nature (e.g., Hochberg & Beck, 1954; Beck, 1965; Mershon, 1972; Gilchrist, 1977; Schirillo, Reeves, & Arend, 1990; Adelson, 1993; Boyaci, Maloney, & Hersh, 2003; Ripamonti et al., 2004; Snyder, Doerschner, & Maloney, 2005; Kitazaki, Kobiki, & Maloney, 2008). We limit our discussion to the effect of surface orientations: surfaces that face a directional light source receive stronger illumination than surfaces facing away. Knowledge about the orientation of surfaces, in addition to their luminances, can modulate how the visual system accounts for the illumination and thus affect lightness.

In a series of studies using illuminated paper displays, Gilchrist (1977; 1980) demonstrated the effect of perceived surface coplanarity on lightness perception (see also Radonjić, Todorović, & Gilchrist, 2010). A schematic of a particular set of Gilchrist's stimuli is shown in Figure 1. Two background context surfaces were oriented at a 90° angle, forming a roof-like ridge. One context surface was white and one was black. The white context was directly illuminated by a light bulb. The black context received only weak ambient illumination and appeared to be in shadow. Two equiluminant tabs extended in orthogonal directions from the ridge of the stimulus. When the stimulus was viewed binocularly, the perceived orientation of the tabs was veridical. When the stimulus was viewed monocularly, however, the tabs appeared to lie flat on the two contexts. Gilchrist (1977; 1980) found that the coplanar relationship between the tabs and the contexts was the primary factor affecting lightness. When the equiluminant tabs were viewed monocularly, they appeared to lie flat in the planes of the two contexts. The tab lying flat on the black context appeared lighter. When the tabs were viewed binocularly, they appeared to be rotated to lie in the plane of the other context. In this case the tab that was coplanar with the black context appeared lighter.

[---FIGURE 1 ABOUT HERE---]

One can interpret Gilchrist's results in terms of the visual system's representation of the lighting environment (see also Howe, 2006). For two surfaces under two different illuminants to be equiluminant, their reflectances must be different. In this situation, given photometric or geometric cues indicating two different illuminant intensities, a

lightness constant system should represent the equiluminant surfaces with different lightnesses. The photometric cues (the luminances of the background context surfaces) and geometric cues (the orientation of the contexts and the tabs) cues in Gilchrist's (1980) experiment are consistent with a directional light source directly illuminating the white context (and indirectly illuminating the black context), so that observers' lightness reports could have resulted from a lightness constancy process that was using these cues. Indeed, the change in lightness with tab orientation could be a signature of a lightness constant system that used both types of cues in accounting for the illumination.

The goal of the present study was to better understand the interaction between the effects of photometric and geometric cues, using a stimulus similar to that in Gilchrist's studies (1977; 1980). By independently manipulating the luminance statistics and the scene geometry, we could study how the two types of cues interacted in eliciting a lightness percept.

Overview

In Experiment 1, observers viewed stimuli rendered on a stereoscopic display. The stimuli, shown in Figure 2, consisted of two articulated background context surfaces ("contexts") oriented at 90° from each other. The ridge where the two contexts meet was pointed towards the observer. The luminances of the contexts were consistent with a light source (with a directional component) illuminating the top context, and a shadow falling on the bottom context. A probe tab extended out from the center of the ridge and

was rendered at each of four orientations. We measured the perceived lightness of the tab, and used the results to understand observers' use of image cues to the illumination.

[---FIGURE 2 ABOUT HERE---]

We first varied the mean luminance across the two contexts, to study the effect of the photometric cues. By obtaining matches to probe tabs of fixed luminance at different orientations, we could determine whether effects of the geometric cues could be measured using this stimulus. We theorized that if observers interpret the changes in these cues as a change in the illumination, their lightness matches to the probe tabs should vary with the cues. We expected that the overall pattern of matches made to tabs in front of one context and tabs in front of the other context should depend on each context's mean luminance. We also expected that matches to the same tab luminance at different orientations should change with the orientation of the tab.

In Experiment 2 we also varied geometric cues, allowing us to study the interaction between the two types of cues. We varied both the difference in mean luminance across background contexts and the angle between the contexts. In addition, we rendered the probe tab at a larger number of orientations for a more sensitive measure of the effects of the scene geometry. By analyzing the pattern of matches as a function of the photometric cues and the geometric cues, we could determine whether each type of cue had an independent effect on observers' lightness perception.

2. Experiment 1 Methods

2.1 Observers

Five observers (1 male, 4 female, mean age = 23.4; labeled Observers 1-5) participated in this experiment. Each observer came to the lab for 3-5 sessions and was compensated for his or her time. The observers all had Snellen acuity of at least 20/40 (corrected) and scored at least 36/38 correct on the Ishihara color plates (Ishihara, 1998). In addition, the observers were all screened for stereopsis. This test was performed using the same apparatus used in the experiment (described below). By means of binocular disparity, a 2.62 DVA square patch appeared either in front or in back of a fixed background plane[2]. Observers were asked to indicate whether the patch appeared in "front" or in "back" of this plane. The background plane was rendered to be 764 mm away from the observer. The distance of the patch from the background plane was adjusted via a 2-down 1-up staircase procedure (Levitt, 1971) to estimate 71% performance accuracy. All observers' depth discrimination thresholds were below 20 mm of rendered depth.

2.2 Setup

The stimuli were presented on a stereo apparatus, illustrated in Figure 3. Two calibrated NEC PA241W LCD monitors were controlled by a single computer. Observers sat with their heads stabilized via a chinrest in front of a black anodized metal

[2] Degrees of Visual Angle are calculated with regard to the central cyclopean eye, for all shapes as if rendered fronto-parallel to the observer in the fixation plane.

faceplate. They viewed the stimuli through two 25 mm x 27 mm openings[3] in the plate. The distance between the centers of these openings was 64 mm. A black cardboard divider sat perpendicular to the faceplate, preventing visual input to one eye from reaching the other.

[---FIGURE 3 ABOUT HERE---]

Each eye received input from an LCD monitor whose light was reflected off an angled mirror before reaching the eye. The optical distance from the LCD to the eye was approximately 764 mm. The apparatus was aligned by replacing the mirrors with beamsplitters and aligning a grid image on each monitor to a physical grid located 764 mm from the eyes.

For the matching experiments, observers viewed a palette of Munsell papers that sat inside a plywood chamber (400 mm long x 405 mm wide x 405 mm deep). The chamber was painted matte gray and was illuminated by a fluorescent bulb. Inside the chamber, an additional LCD screen was placed on the back wall. The Munsell palette had values between 0.5 and 9.5 at 0.5 value intervals. These values corresponded to reflectance values ranging between 0.9% and 91% (details of the reflectance measurement procedure can be found in Allred, Radonjić, Gilchrist, & Brainard, 2012). The LCD screen in the chamber was used during experimental trials to display a number corresponding to one of the palette chip values.

---

[3] Convention for physical dimensions is vertical, then horizontal dimension.

2.3 Stimuli

The stimuli were modeled after the stimuli in Gilchrist's (1980) earlier study. They consisted of two 200 mm x 200 mm (14.91 DVA x 14.91 DVA) square background *context* planes, rendered at a 90° angle to each other (Figure 2 top panel, front view of stimuli as seen by observer; bottom panel, schematic of stimuli viewed from the side). The contexts each contained 100 individual polygons—a selection of squares, rectangles, and L-shapes. The arrangement of these polygons was the same in both contexts. A background 350 cm x 550 cm (25.80 x 39.59 DVA) dark gray rectangle was rendered behind the contexts at a simulated distance of 984 mm from the observer. At this distance, its size on the screen was 272 mm x 427 mm (20.19 x 31.23 DVA). Its luminance was 0.78 cd/m^2.

The 100 polygons in each context were simulations of reflectance values between .07 and 1.00 in equal log relative steps. Within each block of trials, 100 reflectances were randomly assigned to the 100 polygons in each context. This randomization occurred independently in each experimental block. The luminances of each context were then calculated by multiplying all the reflectances by a simulated illuminant.

The simulated illuminants for the two contexts were different in most conditions. This manipulation changed the mean luminance of each of the two contexts. The following top context-bottom context pairs of illuminants were used in different conditions: [45.44, 45.44], [78.96 26.15], [137.20 15.05], [238.39 8.66] cd/m^2. The stimuli resulting from two of these pairs are shown in Figure 2. The illuminant values

were chosen so that each increment/decrement in illuminant intensity between adjacent pairs was an equal log step.

The surfaces bordering the ridge joining the two contexts were adjusted to appear to be continuous surfaces across the ridge. This was achieved by two additional manipulations. First, the polygons along the top and bottom sides of the ridge were arranged such that their edges along the ridge aligned with each other. For example, in the right panel of Figure 2, the L-shape in the lower right hand corner of the top context is aligned with a rectangle in the upper right hand corner of the bottom context. The edge they share along the ridge is the same for both polygons. This alignment was made for all the polygons across the ridge. Second, after the initial random assignment of reflectances, the reflectances for these polygons were adjusted so that corresponding polygons across the ridge had the same reflectances. This was done by swapping reflectances with polygons within a context.

A trapezoidal probe tab was rendered near the center of the stimulus. The base of the tab lay along the border connecting the two contexts. The tab could be rendered in each of four orientations. In the two *in-plane* orientations, the tab lay flat in the plane of each context. In the two *out-plane* orientations, the tab was rotated 90° from each in-plane position. Rotating each in-plane tab 90° preserved the local surround information of the tab, but changed its orientation to be coplanar with the opposite context.

[---TABLE 1 ABOUT HERE---]

The shape of the tab was adjusted so that the two pairs of in-plane and out-plane tabs had the same retinal size and shape for the right eye, using the following procedure. First, we rendered two square checkerboard contexts. When the tab was rendered at the in-plane orientation, it overlaid one of the squares on the checkerboard exactly. We then rendered the tab at the out-plane orientation. While viewing this out-plane tab with the right eye alone, we adjusted the vertices of the tab until the tab projection overlaid the same square on the checkerboard. Given these new vertices in the scene space, the OpenGL routines ("glFrustum") used to create the left and right eye projections then adjusted the left eye image. As a consequence of this procedure, in the left eye image the tabs were two different trapezoids at the two orientations. The vertices of the probe tabs in rendered 3D scene space are listed in Table 1. The probe tab could take on any of nine luminance values between 0.58 and 238.39 cd/m2, with equal log spacing.

2.4 Procedure

Observers indicated on each trial which palette chip most closely matched the probe tab, using a slider that changed the number appearing on the LCD screen in the matching chamber. Observers were given instructions to match the lightness of the tab; specifically, they were told to match "what color paint the tab is coated with". There were three special values that could appear on the LCD screen for judgments of values outside the range of the palette. The first was a value for surfaces "darker than 0.5". Observers chose this value if the target patch appeared to be a surface darker than the darkest palette chip. The second was a value for surfaces "lighter than 9.5", but still

appeared to be surfaces. Observers choose this value if the target patch appeared to be lighter than the lightest palette chip, but still appeared to be an illuminated surface. Finally, the third was a value for surfaces that appeared to be "glowing". If the target patch no longer resembled an illuminated surface, but rather appeared to be generating light, observers chose this value.

In each experimental block, observers viewed the stimuli either binocularly or monocularly. In the monocular conditions, the stimuli were rendered exactly the same as in the binocular conditions, but a piece of felt was placed over the left or the right eye openings in the faceplate. This prevented light from one of the monitors from reaching the eye. In these conditions, the absence of binocular disparity cues was expected to reduce or eliminate the percept that the probe tab was oriented differently from the background context immediately surrounding it.

The pair of simulated illuminants was held constant within each block and varied across blocks. Within each block, observers saw each target luminance at each of the four orientations three times in a random order, resulting in a total of 108 trials. For each observer, the 12 blocks (3 viewing conditions x 4 illuminant pair conditions) were presented in random order.

3. Experiment 1 Results

3.1 Binocular Conditions

The group mean matches are plotted in Figure 4.[4] Data from different pairs of context illuminants are plotted in each panel. The different symbols represent the mean matches for the four orientations. Out-of-range responses were excluded from the calculation of these means. Across observers, the number of trials excluded ranged from 0 to 20, with 9 trials excluded on average. The excluded trials always corresponded to the highest and lowest target luminances.

[---FIGURE 4 ABOUT HERE---]

The effect of the photometric cues can be seen by taking the difference between the means of the two in-plane curves (circle symbols) in each panel. As the difference between context illuminants grows, so does the difference between the two in-plane curves. The magnitudes of the mean differences are plotted with red symbols in Figure 5 as a function of the context illuminant difference. These data are in agreement with previous studies investigating the role of local contrast effects (Heinemann, 1955). We take this effect as a consequence of the visual system's taking the photometric cues to the illuminant into account.

[---FIGURE 5 ABOUT HERE---]

---

[4] Individual observer data for Experiments 1 and 2 can be found in the supplemental material online at http://color.psych.upenn.edu/supplements/lightness_photo_geo/.

The presence of an additional geometric effect can be seen by taking the difference between the means of the in-plane (circle symbols) and out-plane curves (triangle symbols) in Figure 4. If there were no effect of changing the tab orientation, the in-plane curves should lie on the out-plane curves. As the illuminant difference grows, the effect of the tab orientation grows, increasing the deviation between the in- and out-plane curves. The magnitudes of the mean differences between in- and out-plane curves are plotted with black and white symbols in Figure 5 as a function of the context illuminant difference. The geometric effects are smaller than the local contrast effects, but both grow with increasing differences between the mean context luminances.

We fit lines through the origin for each effect in the group data. This model choice was validated via the following cross-validation algorithm. The models under consideration were a horizontal line, a line through the origin, a quadratic function, and a cubic function. The parameters of each model were fit to the group data with one observer's data removed, for all observers. The predictions of the model using the fitted parameters were then compared to the removed observer's data, and the prediction error summed across observers.

The summed prediction error was lowest for both the linear and quadratic functions. Summing across the local contrast effect and both geometric effects, the summed prediction error across observers was 5.99 for the horizontal line, 4.01 for the line through the origin, 3.97 for the quadratic function, and 4.71 for the cubic function. We fit the simpler model to each effect separately using the mean data across observers.

The resulting slopes were .007 for the local contrast effect and .003 and .002 for the geometric effects. To estimate the variability of these slopes, we bootstrapped the match data across observers 200 times. For each bootstrap replication, we redrew the observers' data five times with replacement to obtain the mean matching curves. We took the differences between the means of the curves as before and fit lines to each effect. For the local contrast effect, the range of slopes was [.006, .008]. For the dark context geometric effect, the range was [.002, .004], and for the light context geometric effect, [.001, .002]. In all three cases, the range of slopes does not include zero, and so all three fitted lines are unlikely to be flat, horizontal lines. Because lines fit the data well, the data suggest that the two effects covary—as the effect of local contrast on tab lightness increases, so does the effect of the tab orientation.

3.2 Monocular Conditions

[---FIGURE 6 ABOUT HERE---]

The corresponding matching functions for the monocular conditions are plotted in Figures 6 and 7, and the effect sizes are plotted in Figure 8. As in the binocular condition, the local contrast effect increases with the increase in mean context luminance. Unlike the binocular data, however, the geometric effect is essentially zero in all conditions. There are no binocular cues to the tab orientation in the monocular images, so any differences in tab orientation can only be perceived through linear perspective cues. But because the tabs were adjusted to have the same retinal projection at both in-plane and out-plane orientations when viewed with the right eye, there are no differences

in the linear perspective cues to the tab's orientation for the right eye images. This lack of geometric information explains the lack of a geometric effect in the right eye conditions. However, the corresponding lack of a geometric effect in the left eye conditions suggests that any linear perspective cues available were not effective as geometric cues.

[---FIGURE 7 ABOUT HERE---]

The linear fits to the monocular data support these observations. The fitted slopes for the local contrast effects were .006 and .007 for the left and right eye data, respectively.[5] These values are close to the slope for the local contrast effect data in the binocular condition. The slopes for all the geometric effects were essentially zero,[6] ranging between $-2x10^{-4}$ and $5x10^{-4}$. The local contrast effect slopes are comparable in both the binocular and monocular conditions, but the geometric effect slopes differed.

[---FIGURE 8 ABOUT HERE---]

4. Intermediate Discussion

The results in Experiment 1 establish that the effect of the photometric cues on lightness perception could be reliably measured using our stereoscopic display, and that

---

[5] For the local contrast effect, the range of slopes from the bootstrap replications was [.005, .007] for the left eye and [.005, .008] for the right eye. Thus the slope of the lines fit to the local contrast effects are unlikely to be zero.

[6] The range of slopes from the bootstrap replications for the geometric effects always included zero, and ranged between -.001 to .002.

there is a geometric aspect to the effects for our stimulus configuration. Lightness matches to a given tab luminance depended on both the mean luminance surrounding the tab as well as on the tab's orientation.

The effects of the photometric cues were independent of viewing condition—the local contrast effects were comparable across binocular and monocular conditions. On the other hand, the geometric effects could be attributed to the manipulations of binocular disparity, rather than other monocular cues to orientation. This finding is in agreement with the conclusions from Gilchrist's studies (1977; 1980).

Interestingly, the magnitudes of the effects of local contrast and geometry covaried, as revealed by the linear fits. A parsimonious way to interpret the data would be that the photometric cues are informative about the intensities of the illumination in the scene, and the geometric cues are informative about the spatial layout of the illumination. In the conditions where there are differences in the photometric cues across contexts, the visual system accounts for two illuminant regions. When geometric cues are present, the visual system's representation adds some kind of directional lighting, as demonstrated by the fact that lightness changes with differences in the orientation of the tab. The visual system's representation of the intensity of this directional lighting, however, depends on the photometric cues. This can be seen by the effect of the mean context luminances on the magnitude of the geometric effect.

To test this hypothesis, in Experiment 2 we varied both the intensity of the illumination as well as the geometry of the scene. Across different conditions we independently manipulated both the mean context luminances and the angle between the two context planes. We also rendered the probe tab at a larger number of orientations. If the photometric and geometric cues affect different aspects of the visual system's representation of the illumination in the scene, then the pattern of matches should depend on both kinds of cues in different ways.

5. Experiment 2

5.1 Observers

Five new observers (1 male, 4 female, mean age = 22.6; labeled Observers 6-10) participated in this experiment. Each observer came to the lab for 5-6 sessions and was compensated for his or her time. The observers were screened using the same procedures and criteria as in Experiment 1. None of the observers had participated in Experiment 1.

5.2 Setup and Stimuli

The same apparatus from Experiment 1 was used to display stimuli and record responses in Experiment 2. Observers viewed contextual planes like those presented in Experiment 1, but with a number of changes. First, the angle between the context planes ("context angle") could be 45°, 90°, or 180° in different conditions (Figure 9). For the 180° conditions, the two contexts lay in the same fronto-parallel plane relative to the observer.

[---FIGURE 9 ABOUT HERE---]

Second, the probe tab was rendered at a larger number of orientations. We define as 0° the horizontal plane perpendicular to the observer's fronto-parallel plane, and represent orientation as an angle relative to this plane. Angles above this horizontal plane are defined as negative, and angles below this plane are defined as positive (see Figure 10 for a diagram of this convention). Note that the context angle refers to the angle between the two context planes, and does not follow the naming convention for the tab angles. For the 45° contexts, the tab angles were: [±157.5°, ±135.0°, ±90.0°, ±45.0°, ±22.5°]. The tab angles used in the 90° context condition were the same as those in the 45° context condition, excluding the ±157.5° angles. Those tabs would have to have been rendered behind the contexts. For the same reason, the tab angles used in the 180° context condition were those used in the 90° context condition, excluding the ±135.0° angles. The shapes of the probe tabs were all adjusted so that in the right eye image, they were the same as tabs at the ±157.5° angles, using the same procedure as in Experiment 1. The vertices of the tabs in rendered 3D scene space are also listed in Table 1.

[---FIGURE 10 ABOUT HERE---]

The context surfaces were illuminated by three pairs of illuminants in different experimental blocks: [78.96 26.15], [137.20 15.05], [238.39 8.66] cd/m^2. The tab luminance could take on any of five values: [1.23, 3.79, 11.73, 36.30, 112.28] cd/m^2. These changes were made to reduce the number of trials per experimental block.

5.3 Procedure

Observers were told to match the lightness of the tabs as in Experiment 1. In each experimental block, observers viewed one context angle illuminated by one set of illuminants. Within each block, observers made three matches to each tab luminance at each tab angle.

In a set of control conditions, observers viewed the 45° and 180° context angles monocularly as in Experiment 1. For these monocular conditions, the same pairs of illuminants were used as in the binocular conditions, but with a reduced set of tab angles. For the 45° contexts, these were [±157.5°, ±22.5°]. For the 180° contexts, these were [±90°, ±22.5°].

The order of the 9 binocular conditions (3 context angles x 3 illuminant pairs) and 12 monocular conditions (2 context angles x 3 illuminant pairs x 2 monocular viewing conditions) was randomized for each observer. The number of trials per condition varied between 60 and 150, depending on how many different tab angles were presented.

6. Experiment 2 Results

6.1 Binocular Conditions

The data were first averaged together as in Experiment 1. For each observer, the matches to each probe tab luminance at a particular angle were averaged together, for each experimental block. Out-of-range values were excluded from these averages. For the 85 blocks (across 5 observers x 17 conditions), the number of excluded trials ranged

from 0 to 21 (mean number of excluded trials per block = ~2), and corresponded to matches for the lowest and highest luminance values. The data were then averaged across observers. An example of these data is plotted in the top panel of Figure 11.

[---FIGURE 11 ABOUT HERE---]

A subset of the binocular conditions from Experiment 2 was a direct replication of the binocular conditions in Experiment 1. Using the analysis from Experiment 1, we measured the effect sizes of the photometric and geometric effects for these data. The effect sizes and the best fit lines through the origin are plotted in the bottom panel of Figure 11. The slopes of these lines were .006 for the local contrast effect, and .002 for the two geometric effects. These slopes are similar to those from Experiment 1, and are all unlikely to be flat.[7]

As in Experiment 1, to visualize the data more clearly we took the mean of each matching curve. Within each condition, the group mean data for each tab angle were averaged together. These means are plotted in Figure 12 as a function of the tab angle. Each column shows data for one pair of illuminants. Each row shows data for one context angle. To estimate the variability of these means, the observer matching curves

[7] The analysis was repeated for each bootstrapped replication of the data. The range of the slopes for the local contrast effect was [.0053, .0077]. The range of the slopes for the dark geometric effect was [.0004, .0032], and the range for the light geometric effect was [.0002, .0029]. Because none of these ranges include zero, we conclude that the fitted lines are unlikely to be flat.

were bootstrapped 200 times with replacement, and the error bars are +/- 1 standard deviation of the means of the bootstrapped matching curves.

[---FIGURE 12 ABOUT HERE---]

The data in each panel of Figure 12 show systematic changes in the mean matches with tab angle. For some of the panels, there appears to be a roughly linear relationship; for others, the pattern of matches is more complicated. To a first approximation, however, a simple straight line fit to the data can account for the trend in each panel. The best fits are plotted in blue in Figure 12.

We wanted to test whether there was more structure in the parameters of the best fit lines across panels. Specifically, we wanted to test whether the parameters could be expressed as the sum of two independent effects corresponding to the photometric and geometric cues.

The best fit lines in Figure 12 are of the form:

$$a_{ij} \times Angles + b_{ij} \tag{1}$$

where the *ij* subscripts refer to the panel in row *i*, column *j*. We write the parameters a_{ij} and b_{ij} as column vectors $\mathbf{a}$ and $\mathbf{b}$, such that

$$\mathbf{a} = \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \\ a_{21} \\ \dots \\ a_{33} \end{bmatrix} \tag{2}$$

$$\mathbf{b} = \begin{bmatrix} b_{11} \\ b_{12} \\ b_{13} \\ b_{21} \\ \dots \\ b_{33} \end{bmatrix} \tag{3}$$

For the slope parameters (a corresponding equation can be written for the intercepts), we write

$$\mathbf{a} = \mathbf{M} * \mathbf{s} \tag{4}$$

The * symbol denotes matrix multiplication. In Equation 4, **s** is a 5x1 vector whose elements are

$$\mathbf{s} = \begin{bmatrix} Slope_{12} - Slope_{22} \\ Slope_{21} - Slope_{22} \\ Slope_{22} \\ Slope_{23} - Slope_{22} \\ Slope_{32} - Slope_{22} \end{bmatrix} \tag{5}$$

The center element of $\mathbf{s}$ is the slope of the line in the center panel. The other elements are the change in slope from the center panel to each panel in the four cardinal directions. By constructing the matrix $\mathbf{M}$ so that

$$\mathbf{M} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix} \tag{6}$$

we are expressing $\mathbf{a}$ as a sum of independent row and column effects. The choice of the panels expressed in $\mathbf{s}$ is for conceptual ease, and any five panels from Figure 12 could have been chosen given the appropriate construction of $\mathbf{M}$. We solve for $\mathbf{s}$ in Equation 4. Because $\mathbf{s}$ is overdetermined, we find the least squares solution. Then we multiply $\mathbf{M}$ by $\mathbf{s}$ to get the slopes of the best fit lines with the independence constraint.

Repeating this procedure for the intercept parameter gives us a set of line parameters, which we plot in red in Figure 12. To test whether the red lines are a reasonable fit for the data, we bootstrapped the data 200 times across observers. For each bootstrap, we fit the best line to the bootstrapped data, and these lines are plotted as the light blue spread in Figure 12. The red lines reasonably fall within this spread. This suggests that a more complicated model including an interaction term between the row

and column effects is not necessary. A model of independent effects can account for all the lines in the figure.

6.2 Monocular Conditions

The data from the monocular conditions were analyzed in the same manner as the monocular conditions from Experiment 1. The graphs in Figure 13 plot the magnitudes of the local contrast effect and geometric effects for both eyes and both context angles. The best linear fits through the origin are also plotted. The slopes of these lines for the local contrast effect are .004 for both 45° conditions, and .003 and .004 for the left and right eye 180° conditions, respectively. These values are slightly smaller than the corresponding slopes from the binocular condition, but are unlikely to be zero.[8] For the geometric effects, the slopes are all close to zero, ranging from $-1x10^{-4}$ to .001. The lines fitted to the geometric effects are unlikely to be different from a flat line.[9]

[---FIGURE 13 ABOUT HERE---]

7. GENERAL DISCUSSION

The results from Experiment 2 build on those from Experiment 1. In both experiments, the stimuli consisted of context planes that were consistent with surfaces lit by lighting with a directional component. They elicited patterns of lightness matches that

---

[8] Across bootstrap replications, the slopes for the local contrast effects ranged from .002 to .005, for all monocular viewing conditions.

[9] Across bootstrap replications, the range of slopes for the geometric effects always included 0, from a minimum of -.002 to a maximum of .002, for all monocular viewing conditions.

indicated that observers accounted for the presence of such lighting. Observers' matches also depended on the mean luminance of the contexts, suggesting they were accounting for differences in the intensity of the illumination.

In Experiment 2, the geometric cues in the stimuli were varied by changing the angle between the contexts. In each experimental condition, there was a systematic change in the mean lightness match with a change in the tab orientation, which we summarized with fitted straight-line models. Our analysis showed that the parameters of these lines could be written as the sum of two independent effects. This pattern of results is consistent with the hypothesis that geometric cues and photometric cues independently affect different aspects of the illumination representation.

Equivalent illumination models

An important issue to consider is the choice of model for the matching data in Experiment 2. The data show systematic deviations from a line, suggesting that a linear fit is an approximation. Moreover, the luminance of an achromatic surface does not linearly depend on the angle between the surface normal and a purely directional light source, but varies with the cosine of the angle. A more accurate way to account for the data would be to fit the parameters of an equivalent illumination model (EIM) (Brainard, Brunt, & Speigle, 1997; Boyaci, Maloney, & Hersh, 2003; Bloj et al., 2004; Brainard & Maloney, 2011). This class of models provides a compact way to describe the relationship between the illumination, surface properties, and perception of the surfaces. Typically, a limited set of surfaces is presented to observers who report the appearance of

these surfaces. The EIM is developed, and the parameters of the model are fit to the data. These parameters describe aspects of the illumination. Importantly, the illumination derived from these models may not reflect the actual illumination of the stimuli, but rather the visual system's hypothetical representation. This representation can then be validated with responses to different stimuli.

Boyaci *et al*. (2003) and Bloj *et al*. (2004) successfully developed EIMs to account for the perception of lightness of slanted surfaces. In their models, the illumination consisted of a diffuse ambient illuminant and a directional illuminant. This kind of a model could in principle account for the data in Experiment 2. Preliminary analysis of fitting EIMs with this parameterization of the illumination suggests that for our data, two ambient illuminant terms are necessary to account for the effect of the mean context luminance. Changing the angle of the directional light source relative to the contexts can account for some of the deviations of the data from the linear fit. Fitting the parameters of this kind of model could be informative about the interaction between photometric and geometric cues. For example, imagine if a single value of a parameter corresponding to the directional component of the illumination can be used to fit the data within a context angle. Conversely, imagine if a single parameter value corresponding to the effects of the illumination intensity can be used to fit the data for a pair of illuminants. Such an analysis would support the independence of the two cues' effects. Specifically, it could demonstrate that the photometric cues affect the illumination intensity, while the geometric cues affect the spatial layout of the illumination.

Relation to other studies

Because our experimental stimuli were closely related to those in Gilchrist's (1980) experiment, some comparison between the two is warranted. In that study, matches made to equiluminant tabs also depended on their orientation. However, the magnitude of that geometric effect was on the order of 4 Munsell steps, compared to the ~.5 Munsell step effect in our Experiment 1. The stimuli in Gilchrist's experiment consisted of illuminated grayscale papers, while our stimuli were rendered on a computer screen. It is possible that some image cues available in the illuminated paper stimuli modulate the effect of the tab orientation, and these image cues were not present in our rendered stimuli.

Our effect sizes are more comparable to those found in a study Beck (Beck, 1965), which replicated and extended an earlier study by Hochberg and Beck (1954). In this experiment, a directional light source was placed perpendicular to and above a table. There was also trapezoidal card standing perpendicular to the table. When viewed monocularly, this card appeared to lie flat on the table. By waving a rod behind the card, the card appeared to change orientation and was perceived veridically. The change in orientation produced a change in lightness of about .5 Munsell steps. Beck also measured the effect of local contrast by changing the table cover from black to white. The effect of this manipulation was on the order of about 1 Munsell step.

Future directions

More work on explaining the mechanisms that process photometric and geometric cues is needed. Developing and testing an EIM will help clarify the behavior of these mechanisms and how they function. Collecting data using different context angles and illumination, perhaps with asymmetric pairs, would provide a dataset to test the validity of such a model.

A closer comparison of the differences between cues available in the rendered stimuli and cues in the colored paper and light bulb stimuli is also needed. Applying the EIM to data collected from the paper stimuli could help explain the difference by relating the difference in image cues to a difference in the representation of the illumination.

8. Appendix 1

8.1 Overview

In Experiments 1 and 2, we used a matching paradigm to elicit observers' responses. Observers viewed a particular stimulus and adjusted a slider to indicate which Munsell chip most closely resembled the probe tab. This response method is intuitive and easy to understand, and observers can perform many trials in a session, allowing us to collect data across different conditions.

In a separate control experiment, we tested a different group of observers using a two-alternative forced-choice (2AFC) version of the experiment. As with Experiments 1 and 2, we found that 1) under binocular viewing, the orientation of the tabs affected the perceived lightness of the tabs, and 2) under monocular viewing, the orientation of the tabs and the differences in their shapes did not produce a comparable effect on lightness.

Both the 2AFC and matching versions of the experiment yielded similar results. We conclude that the effect of the geometric cues is robust across measurement paradigms.

8.2 Methods

8.2.1 Observers

5 observers (1 male, 4 female, mean age = 24; labeled Observers A1-A5) participated in this experiment. Each observer came to the lab for 3-5 sessions and was compensated for his or her time. All observers were screened using the same methods

and criteria as those reported in Experiments 1 and 2. None of the observers participated in either Experiment 1 or 2.

8.2.2 Setup and Stimuli

The stimuli were presented on the same stereo apparatus used in Experiments 1 and 2. Observers viewed the stimuli under both binocular and monocular viewing conditions. For the monocular conditions, a piece of felt was taped over one of the openings in the faceplate, and the stimuli were rendered just as in the binocular conditions. The stimuli were as in Experiments 1 and 2, except for the changes detailed below.

The contexts were 210 mm x 200 mm (30.74 x 29.34 DVA) and rendered at a 90° angle. They consisted of 116 individual polygons each. The top context[10] consisted of luminances in the range [15.89, 228.39] cd/m^2. The bottom context's luminances were in the range [0.58, 8.66] cd/m^2. The 116 luminances per context were drawn with equal log spacing.

[---FIGURE A1 ABOUT HERE---]

Instead of a probe tab, two 35 mm x 35 mm (5.25 x 5.25 DVA) square patches were rendered near the center of the stimulus along the ridge between the two contexts (Figure A1). One patch was always rendered to lie flat in the plane of a context, and the

[10] Due to a programming error, these values were not the same for Observer A4. For that observer, the top context luminances were in the range [17.46, 261.86] cd/m^2, and the bottom context luminances were in the range [0.63, 9.45] cd/m^2.

other was rotated 90° out of the plane of the same context. The patches could take on any value[11] between 0.58 and 238.39 cd/m^2. The centers of the patches were 82.6 mm apart from each other in the horizontal direction.

8.2.3 Procedure

On each trial, one of the two patches was designated the *reference* patch. Its luminance[12] was fixed at one of four values: [1.23, 5.53, 24.91, 112.28] cd/m^2. The luminance of the other *match patch* was adjusted via a 1-up 1-down staircase procedure (Levitt, 1971) to estimate a luminance that was perceived to be equal to the reference patch luminance (point of subjective equality, or PSE). This resulted in a total of 16 interleaved staircases in each block (2 contexts x 2 reference patch orientations x 4 reference luminances). Whether the reference or the match was the patch on the left was randomly chosen on every trial.

There were 800 trials per block (16 staircases x 50 trials per staircase). Observers were tested in three blocks: one binocular block and two monocular blocks, in random order.

8.3 Results

The data from each staircase were fit with a logistic function using routines from the Palamedes MATLAB toolbox (Prins & Kingdom, 2009). PSEs were estimated from

---

[11] These values were 0.63 and 261.86 cd/m^2 for Observer A4.
[12] For Observer A4, these values were [1.34, 6.04, 27.29, 123.23] cd/m^2.

the fitted function as the level of the match patch required for a 50% probability of responding “match lighter than reference”.

PSEs for two observers are plotted in Figure A2, and all the data can be found in the supplemental materials online at http://color.psych.upenn.edu/supplements/lightness_photo_geo/. Each point plots the reference patch luminance and corresponding PSE from each staircase. The blue symbols represent data from the binocular condition. The red and green symbols represent the data from the monocular conditions. The different shapes within a group of colored symbols represent the orientation of the reference patch. Points that fall along the diagonal line show conditions where there was no effect of the geometric cues. For each observer, the data were combined across the top and bottom contexts and plotted so that points lying above the diagonal show an effect of the geometry cues on lightness in the expected direction.

[---FIGURE A2 ABOUT HERE---]

The differences between conditions were small, as can be seen by the tight clustering of the data. We quantified the effect of the geometric cues by taking the median Euclidian distance of the data points from the diagonal line. The median distance for the binocular data ranged from 0.07 to 0.18 across observers. The median distance for the monocular data ranged from -.005 to 0.02. A Wilcoxon signed-rank test showed

that the binocular data was significantly different from the mean monocular data, $p < 0.05$ for all observers. The monocular data was not different from the diagonal.

9. Appendix 2

9.1 Overview

The magnitudes of the geometric cue effects on lightness were on the order of half a Munsell step. This effect size is about 20% of those found in an previous experiment using similar stimuli by Gilchrist (1980; but see Beck, 1965). The stimuli in the Gilchrist (1980) experiment were illuminated papers, whereas ours were rendered on a computer screen, and this difference in the stimuli might account for the difference in effect size. Differences in the perception of slanted surfaces have been reported in the literature, depending on whether the surfaces were made of paper/cardboard or rendered on a stereoscopic display (e.g., Van Ee, Banks, & Backus, 2001). Typically, observers' perceived surface slants deviate from their actual values and are rotated toward the plane spanned by the fronto-parallel surface of the monitor.

If this misperception of surface slant affects our stimuli, then all the contexts and probe tabs are being "flattened" into the plane of the monitor. This would result in smaller perceived angles between the contexts and the tabs. For example, the angle between in-plane and out-plane tabs in Experiment 1 was rendered to be 90°. If both the contexts and the tabs are pulled toward the monitor plane, this angle will be smaller than 90°. A smaller angle might be the cause of a smaller effect of the scene geometry on lightness.

To check whether the misperception of surface slant could account for the difference in effect size, we tested observers on a slant-matching task. In this experiment, the observers from Experiment 2 indicated their perceived slants of the two contexts and various probe tab orientations. In most cases, observers' perception of slant was not veridical, and was pulled in the direction of the fronto-parallel plane. However, the slants were perceived to be 70-80% of their actual values. This small deviation from the rendered values is unlikely to account for the large difference in lightness. In addition, flattening the stimuli into the monitor plane produces stimuli that are more similar to the 180° context stimuli. The pattern of lightness matches for the other context angles does not closely resemble the matches in the 180° condition. Thus we conclude that the misperception of slant due to the rendered stimuli's presentation on a monitor cannot account for the difference in effect size.

9.2 Methods

9.2.1 Observers

The five observers from Experiment 2 participated in this experiment. The conditions of this experiment were randomly interleaved with the conditions in Experiment 2.

9.2.2 Setup and Stimuli

The stimuli were presented on the same stereo apparatus used in Experiments 1 and 2. The same context angles and probe tabs used in Experiment 2 were presented in

this experiment. The luminance of the probe tab was always set to its maximum value, 238.39 cd/m^2. The rendered illuminants for the top and bottom contexts were [238.39 8.66] cd/m^2, corresponding to the largest difference in mean context luminance.

The monitor in the plywood chamber located to the right of the observer displayed a green circle, 50 mm in diameter. This diameter was marked with a thin white line, which could be rotated around the center of the circle. At the beginning of each trial, this line was reset to its horizontal position. Instructional text appeared above the circle.

9.2.3 Procedure

Observers were asked to view the stimuli and imagine what they would look like when viewed from the left side. Their task was to match the slant of the various surfaces in the stimulus. On each trial, the observer read the text at the top of the monitor in the plywood chamber. The text indicated whether the observer should match the slant of the top context, bottom context, or probe tab. Observers rotated the line in the green circle until it matched the slant of the relevant surface. They used a Logitech gamepad to rotate the line and enter their response. Lines rotated counterclockwise were recorded as positive angles, and lines rotated clockwise were recorded as negative angles.

Observers were tested in three experimental blocks. In each block, only one context angle was presented (45°, 90°, or 180°). Each probe tab was presented three

times at each possible orientation. The order of surfaces to be matched was randomized within a block.

9.3 Results

The naming convention for the orientations of the surfaces is shown in Figure A3. Angles are defined relative to the horizontal zero plane. Unlike the naming convention in Experiment 2, the zero plane is not defined as a vector extending out from the center of the stimuli. Rather, it is the entire horizontal plane through the center of the stimuli.

[---FIGURE A3 ABOUT HERE---]

Before further analysis, the data were first processed in the following manner to be consistent with the angle naming convention. First, coplanar surfaces are defined as having the same angle. As a result, rotations greater than +/- 90° were rotated 180°. Thus a recorded response of 120° became -60°. Second, on trials where the tab or context to-be-matched was rendered to lie in the fronto-parallel plane, the sign of the responses were fixed. If the tab/context was below the zero plane, the sign of the response was made to be positive, and if the tab/context was above the zero plane, the sign of the response was made to be negative.

Matches for two observers from the 90° context angle condition are plotted in Figure A4, and all the data can be found in the online supplemental material at http://color.psych.upenn.edu/supplements/lightness_photo_geo/. The black diagonal line is the unity line. Points that fall on this line are veridical matches. The blue line is the

best straight-line fit to the data through the origin. The slope of the blue line was taken as a measure of the observer's slant misperception. Deviations of this slope from 1 indicate some misperception of the surfaces. The closer the blue line is to a horizontal line, the more the surfaces are perceived to lie in the fronto-parallel plane.

[---FIGURE A4 ABOUT HERE---]

The slopes of the best fit lines ranged from 0.32 to 0.98. Across observers, the mean slopes for the 45°, 90°, and 180° context angle conditions were 0.67, 0.77, and 0.86. The mean slopes for each observer across context angles ranged from 0.66 to 0.87. On average, observers perceived the slants of the surfaces in the stimuli at about 77% of the actual rendered slants.

FIGURES

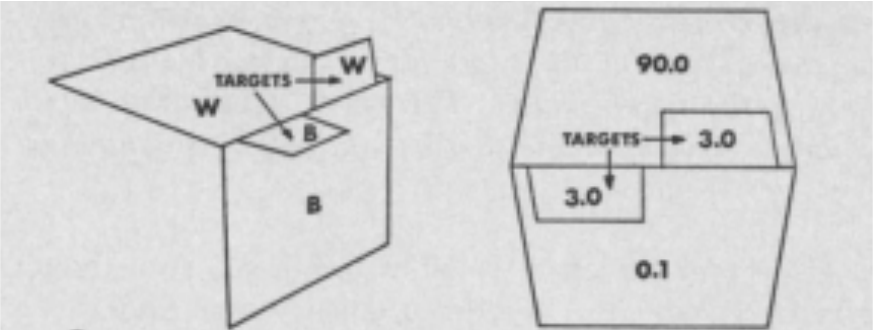

Figure 1: Figure reproduced from Gilchrist (1977). The left panel shows the arrangement of the papers in perspective. A black tab lies coplanar with a white sheet of paper, and a white tab lies coplanar with a black sheet of paper. The right panel shows the arrangement of the papers when viewed monocularly from the front, as well as the relative luminance of each surface. The tabs are illuminated to be equiluminant.

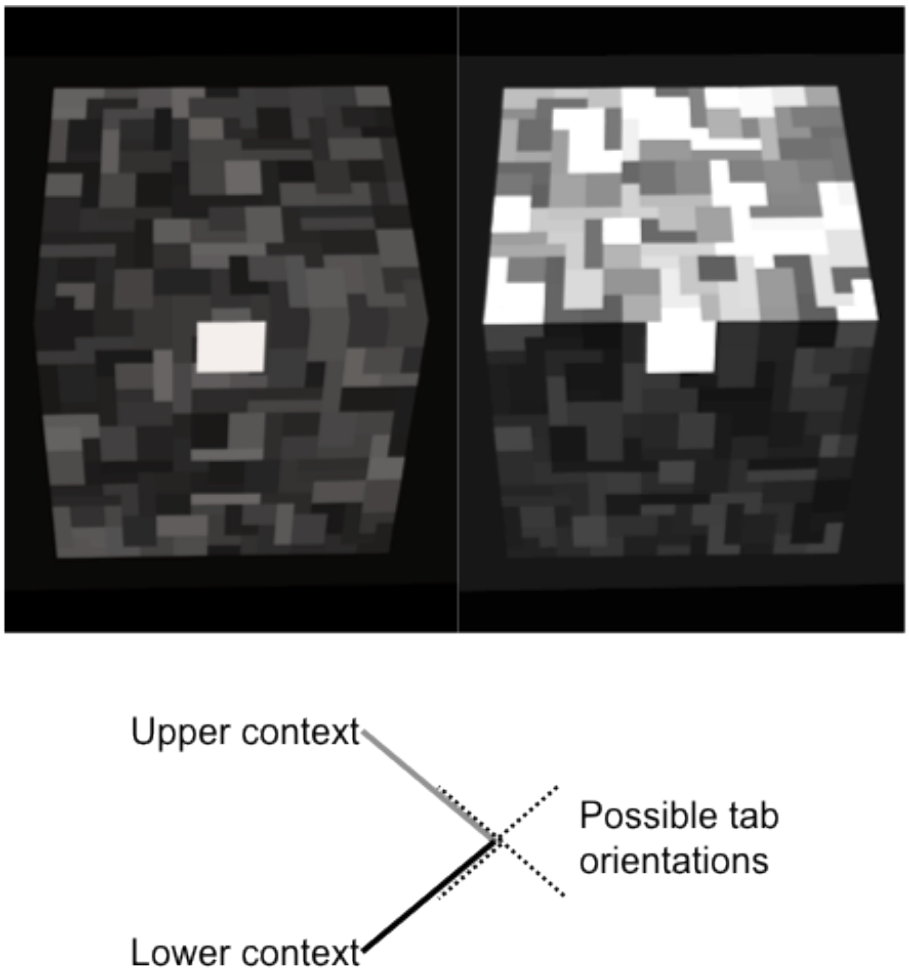

Figure 2: Example stimuli from Experiment 1. Top panel, front view of stimuli. The white tab is the probe tab. The left image shows the contexts illuminated by the [45.44 45.44] cd/m^2 illuminant pair. The right image shows the contexts illuminated by the [238.39 8.66] cd/m^2 illuminant pair. Bottom panel, schematic side view of stimuli. The

solid lines represent the context planes, and the dotted lines the possible orientations of the tab. Note that the colors in the figure have been adjusted to increase visible contrast and are not accurate replications of the stimuli.

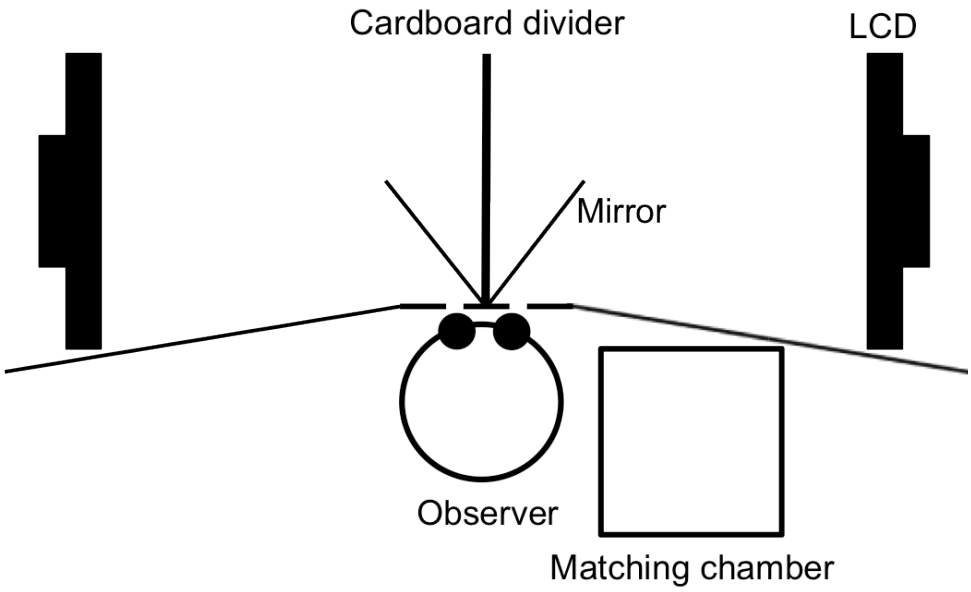

Figure 3: Bird's-eye-view schematic of stereoscopic setup. Two LCDs independently project light via mirrors to the observer's eyes. The LCDs are hidden from the observer via baffles and a faceplate. Observers turn to their right to make a response, using a Munsell scale inside the plywood chamber.

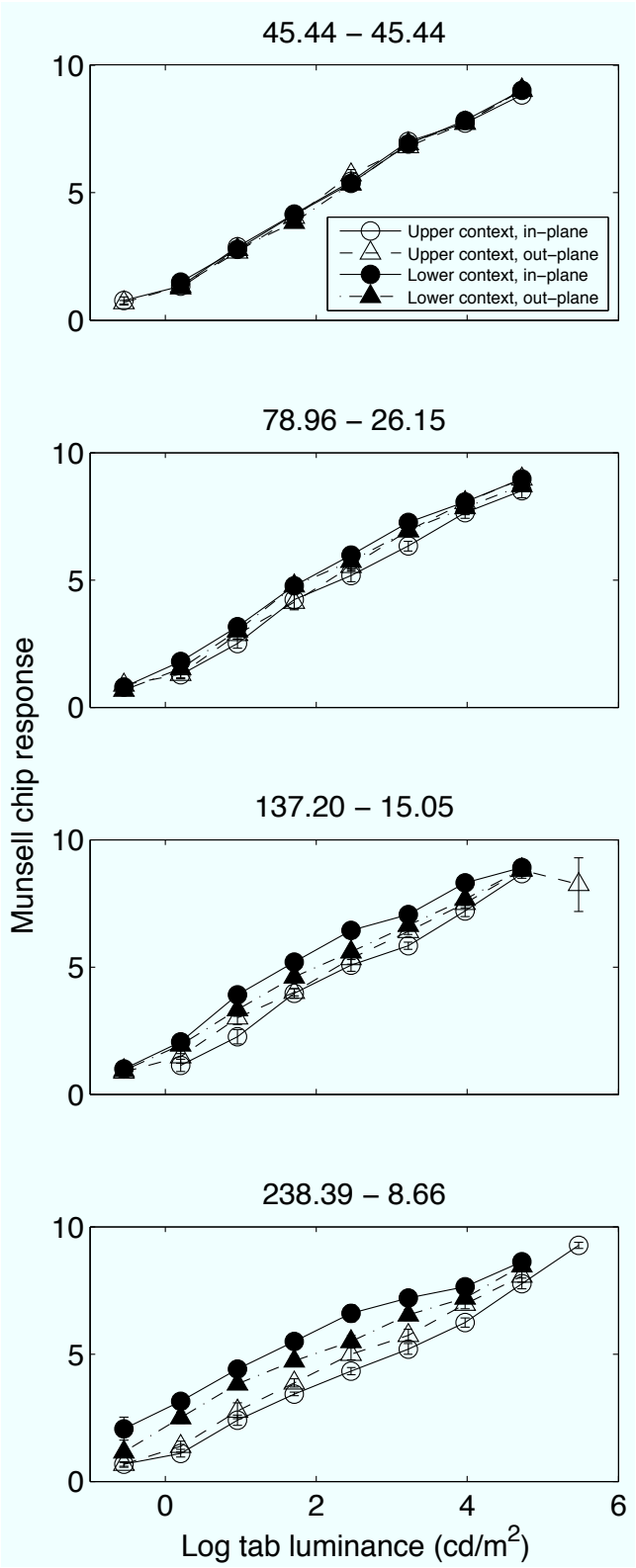
45.44 – 45.44
78.96 – 26.15
137.20 – 15.05
238.39 – 8.66
Upper context, in–plane
Upper context, out–plane
Lower context, in–plane
Lower context, out–plane
Munsell chip response
Log tab luminance (cd/m2)
0
5
10
0
2
4
6

Figure 4: Group matching curves from the binocular viewing conditions of Experiment 1. Each panel shows data for one set of context illuminants, whose intensities in cd/m^2 are given by the numbers above each panel. Matches to the probe tab are plotted as a function of the tab luminance. Open symbols are for in- and out-plane tabs relative to the upper light context, closed symbols are for in- and out-plane tabs relative to the lower dark context. Circles show the data for in-plane tabs, triangles show data for the out-plane tabs. Error bars are +/- 1 SEM.

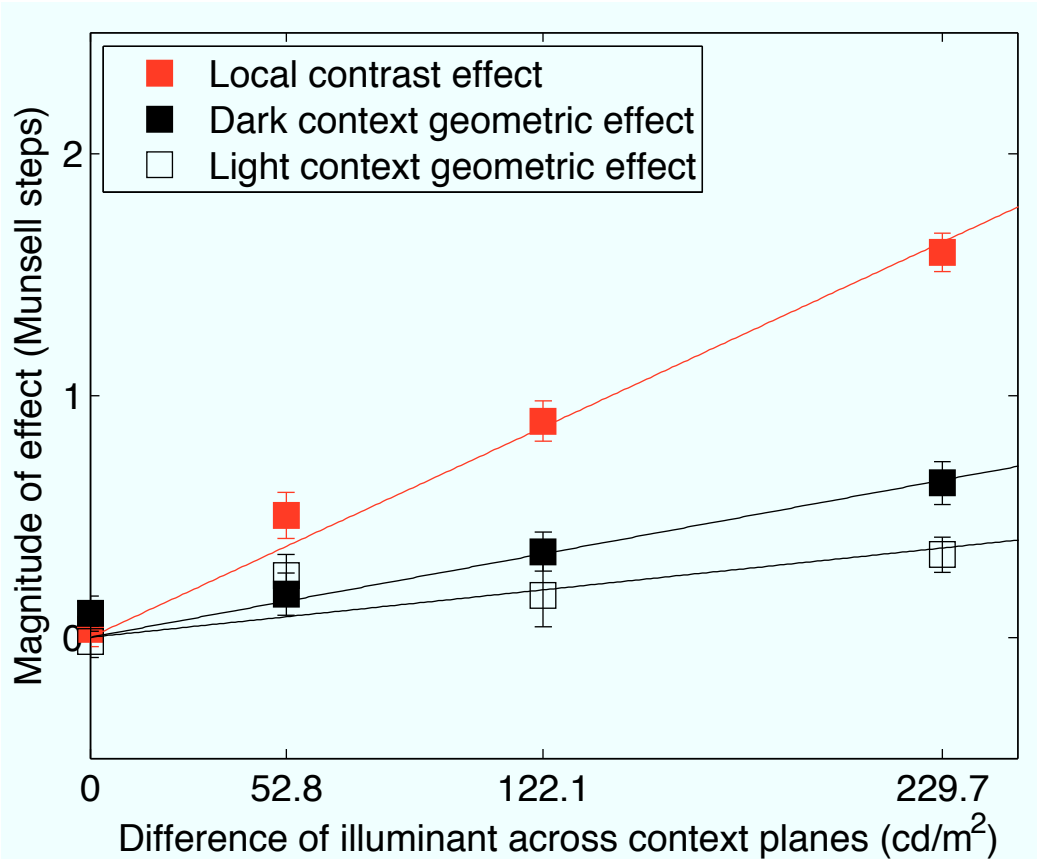

Figure 5: The magnitudes of the photometric contrast effect and the geometric effects from both contexts are plotted as a function of the difference of illuminant across context planes. The data come from the binocular viewing conditions of Experiment 1. The different colors represent different effects. The lines are the best linear fits through the origin for each effect. Error bars are +/- 1 standard deviation from the 200 bootstrap replications of the data.

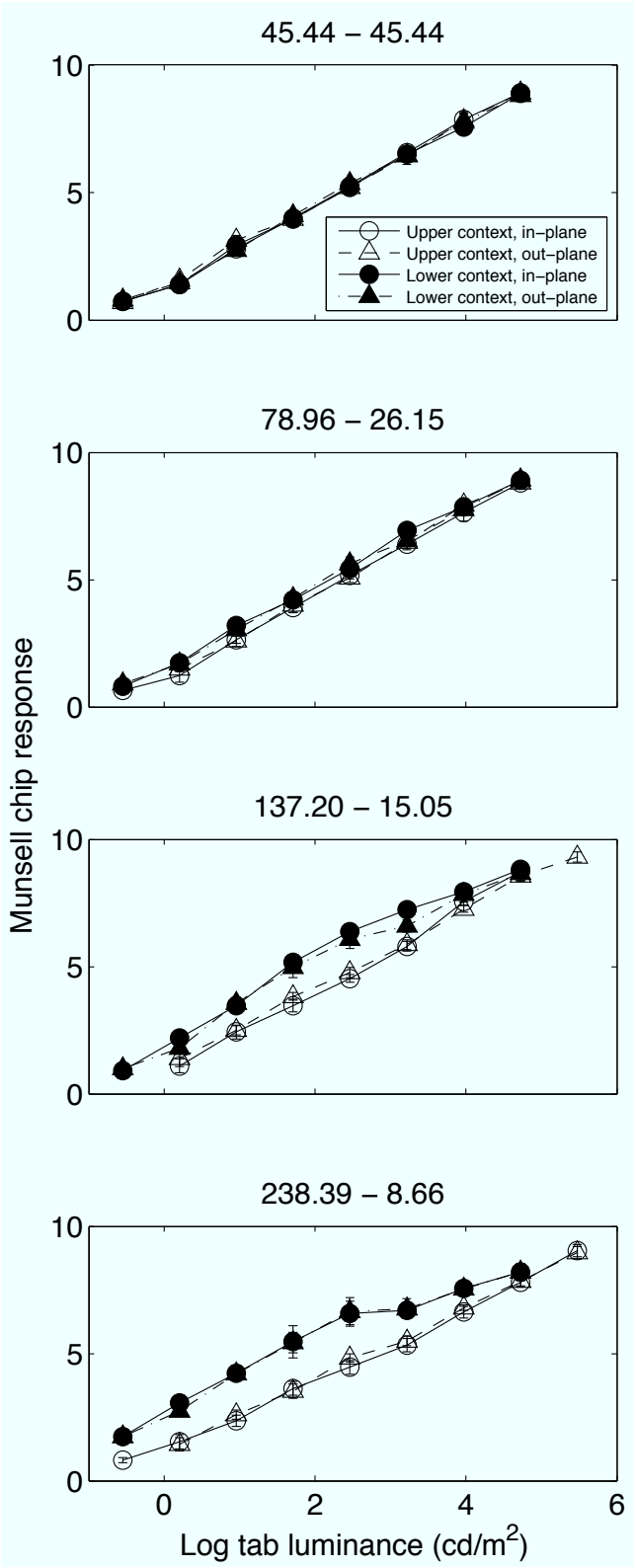
45.44 – 45.44
78.96 – 26.15
137.20 – 15.05
238.39 – 8.66
Upper context, in–plane
Upper context, out–plane
Lower context, in–plane
Lower context, out–plane
Munsell chip response
Log tab luminance (cd/m2)
0
5
10
0
2
4
6

Figure 6: Group matching curves from the left eye monocular viewing conditions of Experiment 1. Plot conventions follow those from Figure 4.

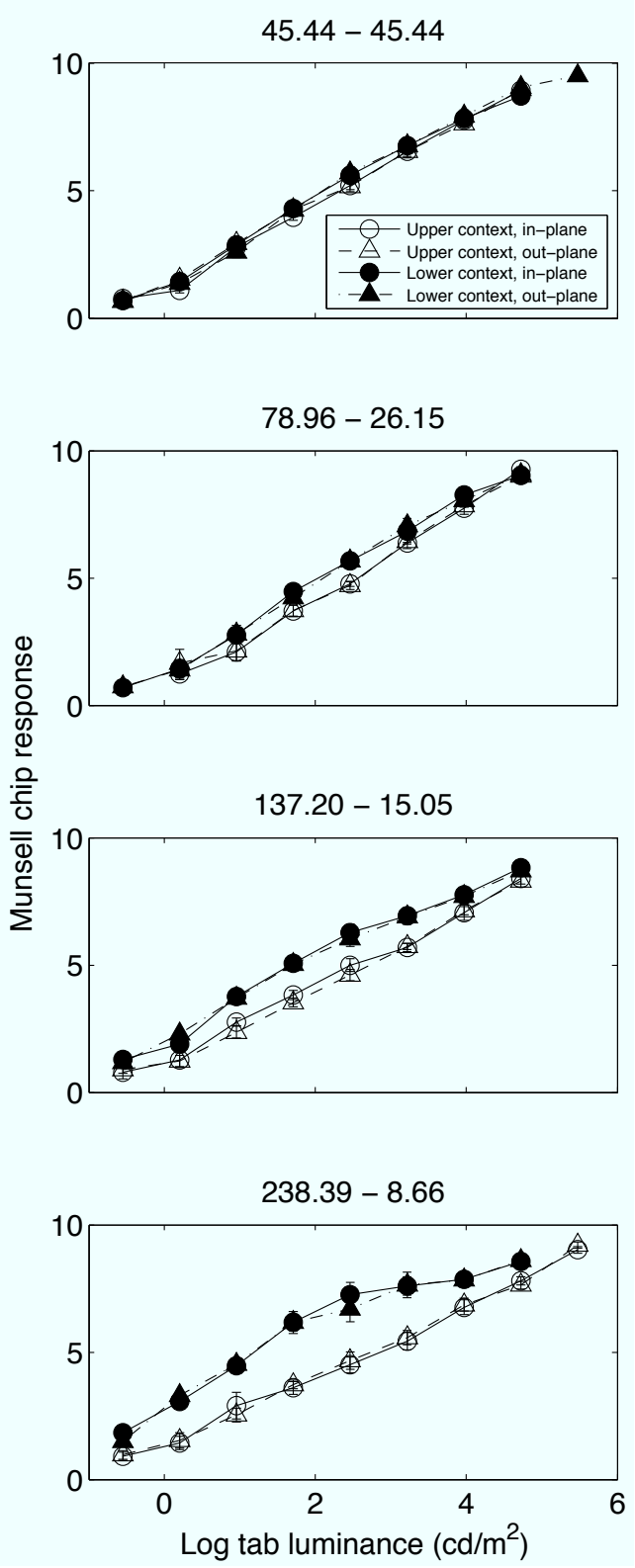
45.44 – 45.44
78.96 – 26.15
137.20 – 15.05
238.39 – 8.66
Upper context, in–plane
Upper context, out–plane
Lower context, in–plane
Lower context, out–plane
Munsell chip response
Log tab luminance (cd/m2)
0
5
10
0
2
4
6

Figure 7: Group matching curves from the right eye monocular viewing conditions of Experiment 1. Plot conventions follow those from Figure 4.

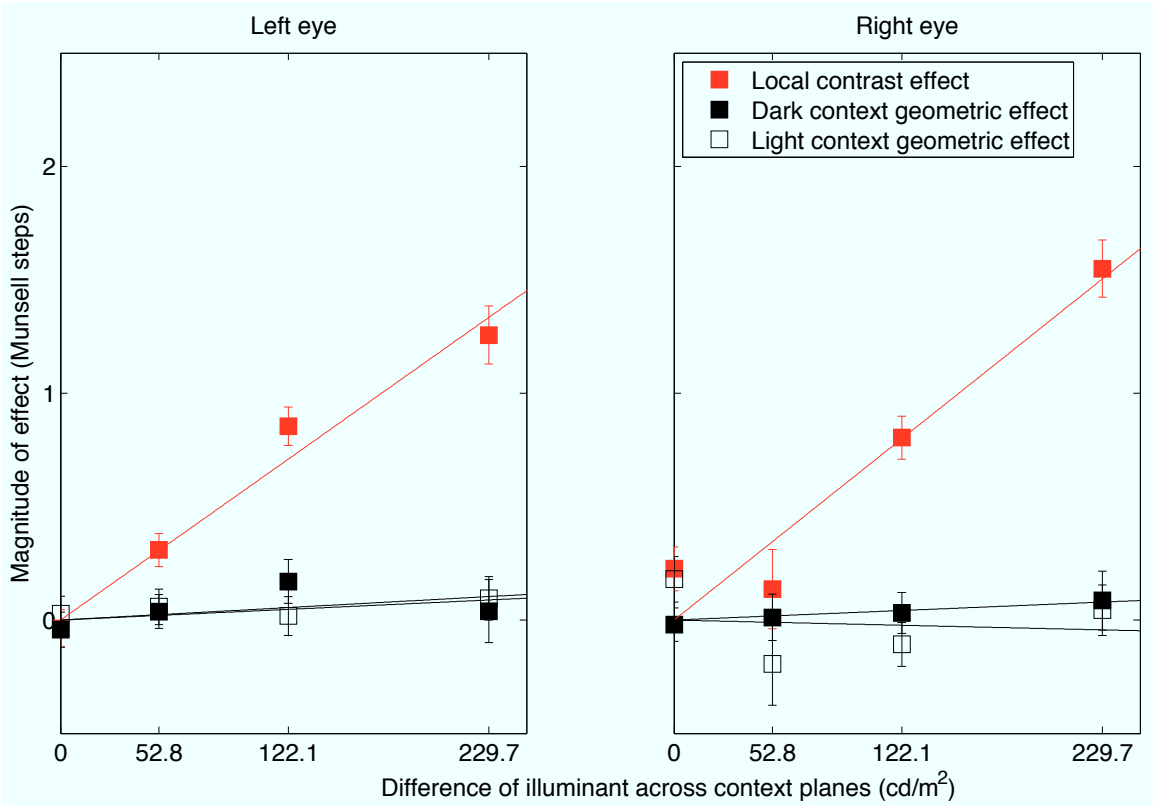

Figure 8: The magnitudes of various effects from the monocular viewing conditions of Experiment 1 are plotted as a function of the difference of illuminant across context planes. Left panel, left eye condition; right panel, right eye condition. The plot conventions follow those in Figure 5.

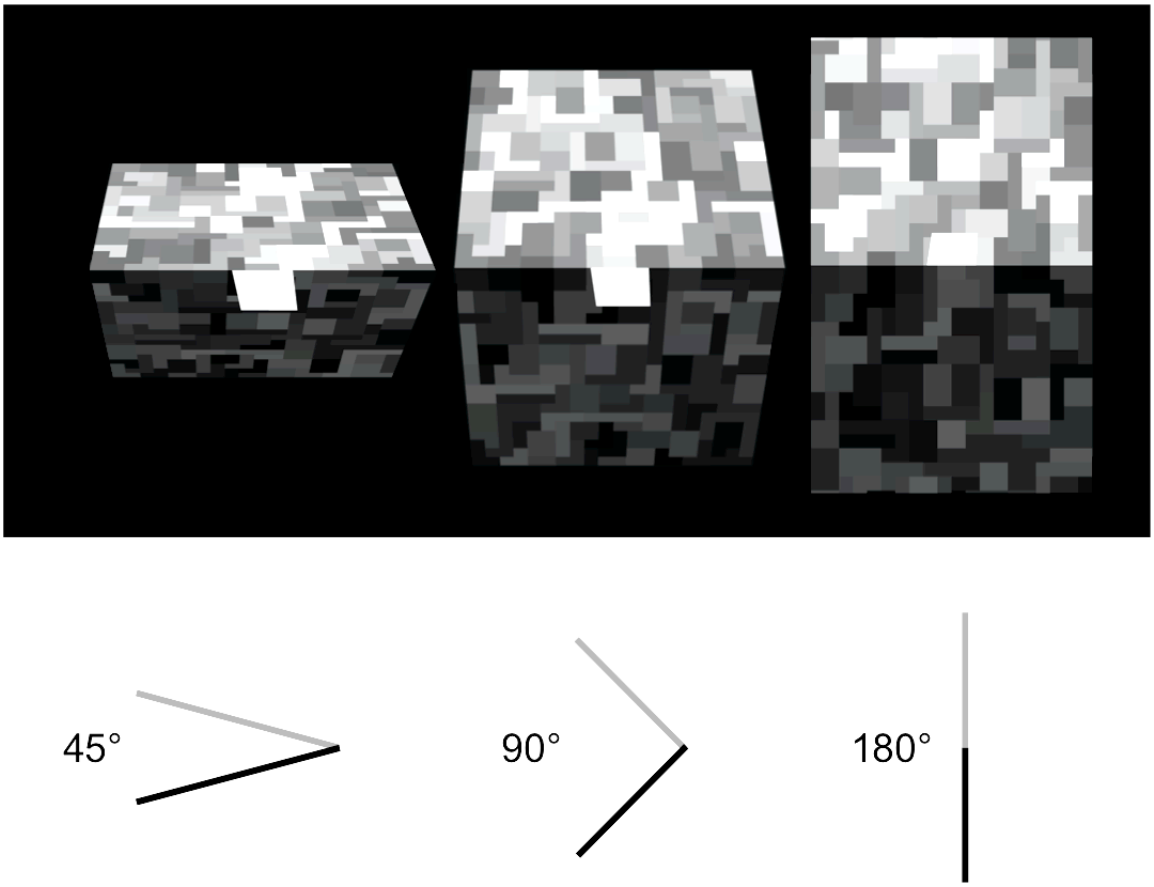

Figure 9: Example of stimuli from Experiment 2. Top row, front view of stimuli. The probe tab is the white trapezoid in each stimulus. Bottom row, side view of stimuli. Solid lines represent the contexts. The angle between each set of contexts is noted next to each side view panel. Note that the colors in the figure have been adjusted to increase visible contrast and are not accurate replications of the stimuli.

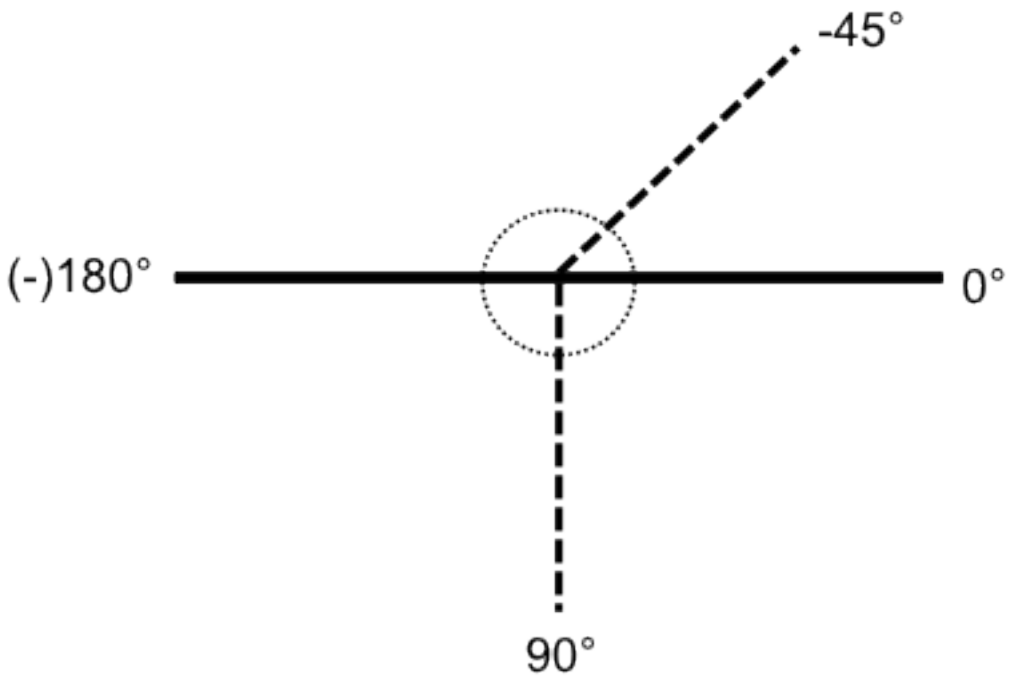

Figure 10: Diagram of the naming convention for tab angles, side view. 0° is defined as the horizontal plane pointing towards the observer. Tabs (dashed lines) rotated above this plane have a negative angle. Tabs rotated below this plane have a positive angle. Note that this convention only applies to the tab angles. The context angle refers to the angle between the two context planes and does not follow this convention.

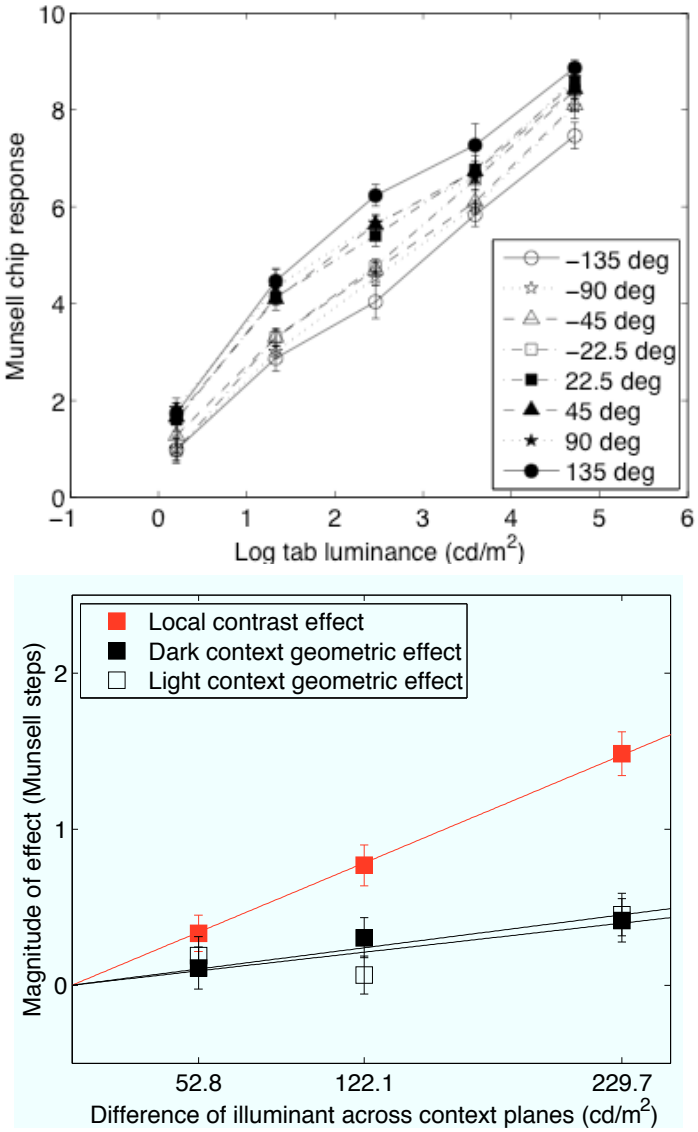

Figure 11: Top panel, an example of group mean matching curves (N = 5) from Experiment 2. Matches are plotted as a function of the tab luminance. The data in this figure come from the 90° context angle, with illuminant intensities [238.39 8.66] cd/m^2.

Each symbol represents a different tab angle. Open symbols are for tabs that lie in front of the upper light context, and closed symbols are for tabs that lie in front of the lower dark context. Error bars are +/- 1 SEM. Bottom panel, photometric and geometric effect sizes for the binocular conditions in Experiment 2 that are analogous to conditions in Experiment 1. The lines are the best fit lines through the origin. The match data were bootstrapped 200 times across observers with replacement. For each set of bootstrapped data, the effect sizes were calculated, and the error bars are +/- 1 standard deviation of these bootstrapped effect sizes.

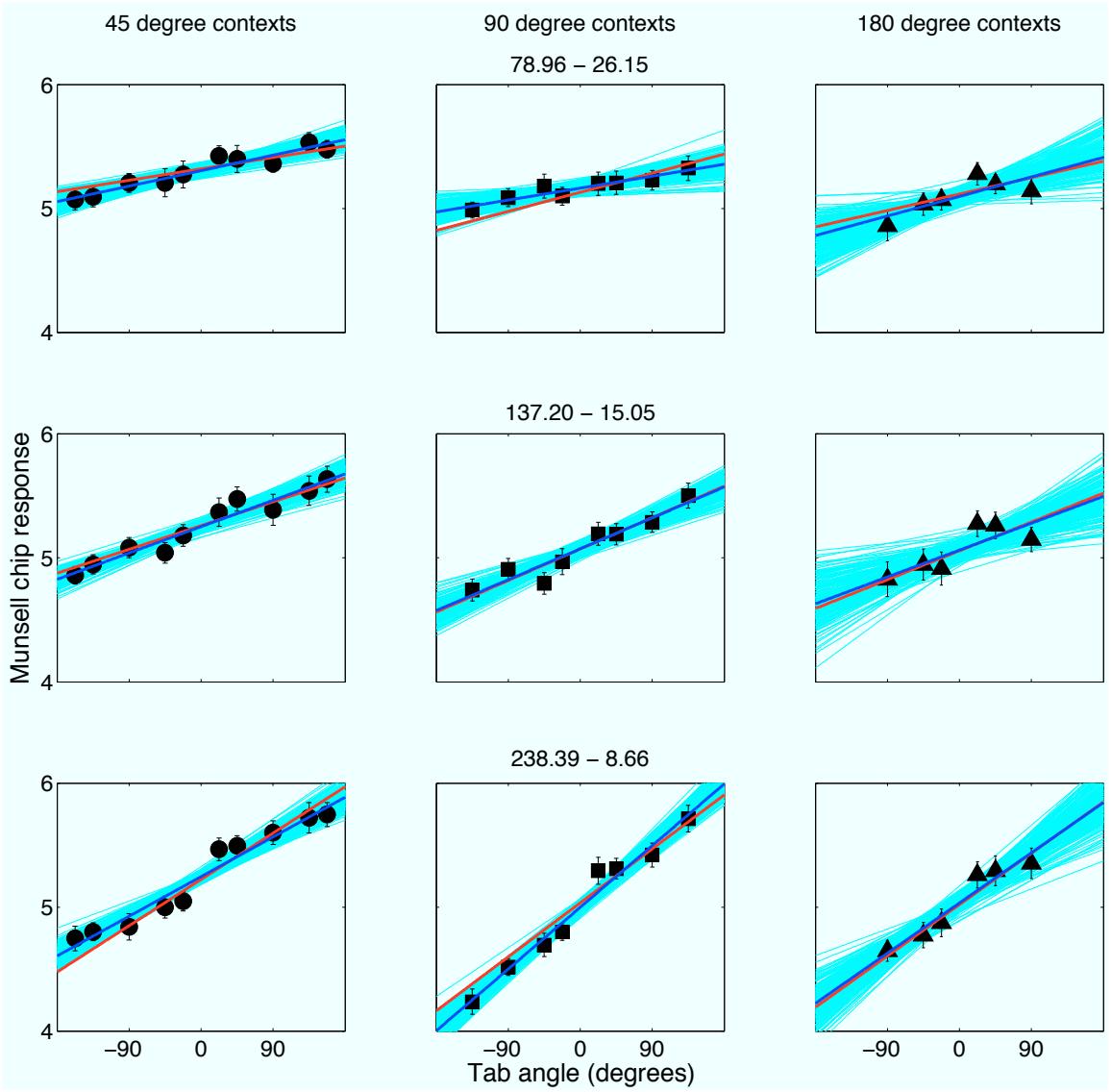

Figure 12: Mean matches from Experiment 2 as a function of tab angle. Each column shows data from one context angle. Each row shows data from one pair of illuminants, and the numbers at the top of each row are the intensities of the illuminants. The match data in were bootstrapped across observers with replacement, and the error bars are +/- 1 standard deviation of the mean matches for each tab angle. The dark blue lines are the

best fit lines for each panel. The red lines are the best fit lines whose parameters are the sum of a row effect and a column effect. The light blue lines are the best fit lines to the bootstrapped data of each panel, and provide a sense of the variability of the individual panel best fit lines.

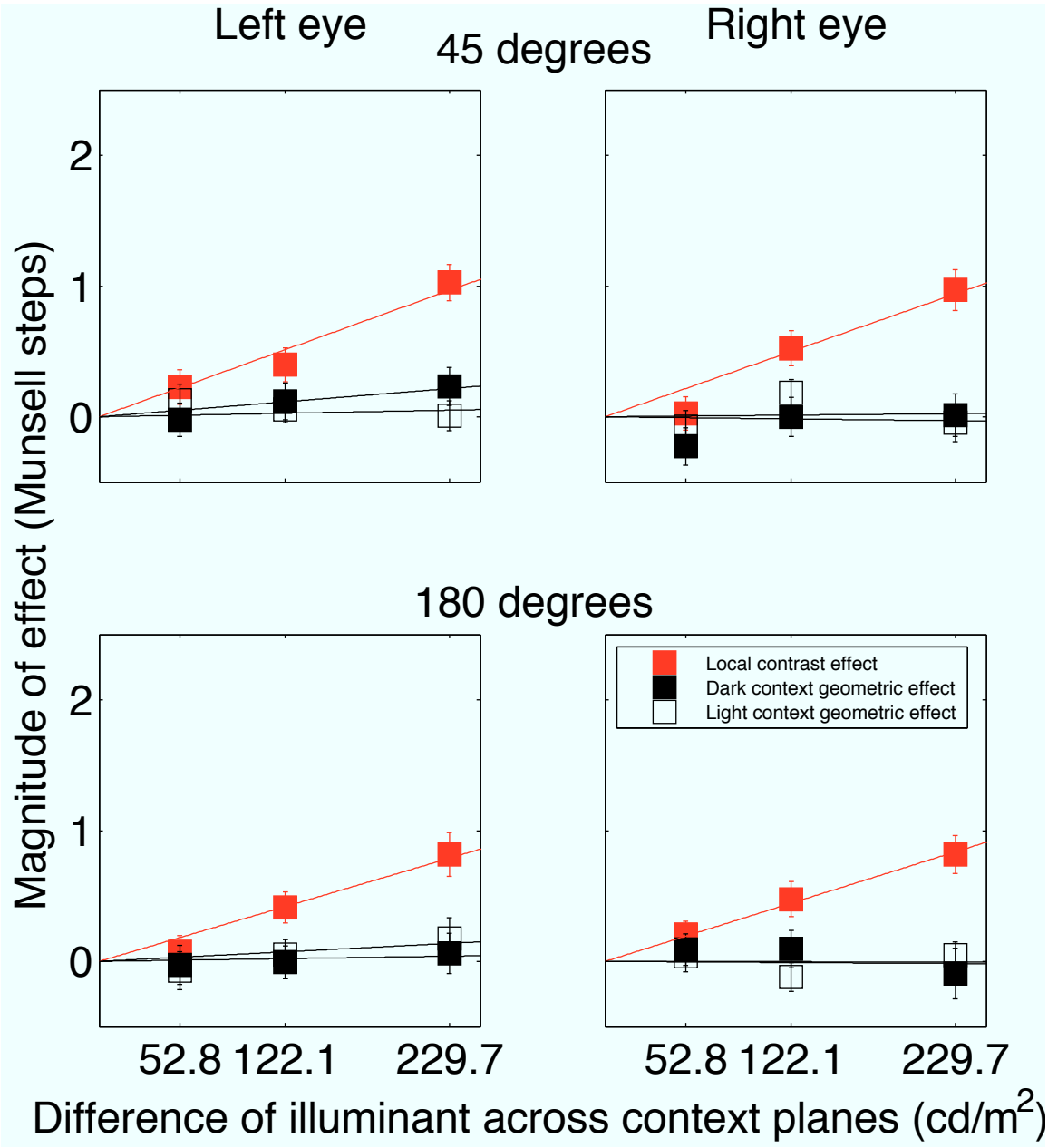

Figure 13: Photometric and geometric effect sizes for the 45° and 180° context angles in the monocular viewing conditions are plotted as a function of the difference of illuminant across the context planes. The lines are the best fit lines through the origin. Error bars are

+/- 1 standard deviation across the computed effect size values from the 200 bootstrapped replications of the data.

Figure A1: Example of stimuli from 2AFC experiment. The two tabs are the large white and black squares that lie in the upper half of the stimulus, near the ridge between the upper and lower contexts. Note that the colors of the stimuli have been adjusted to increase visible contrast, and are not indicative of the actual luminances used.

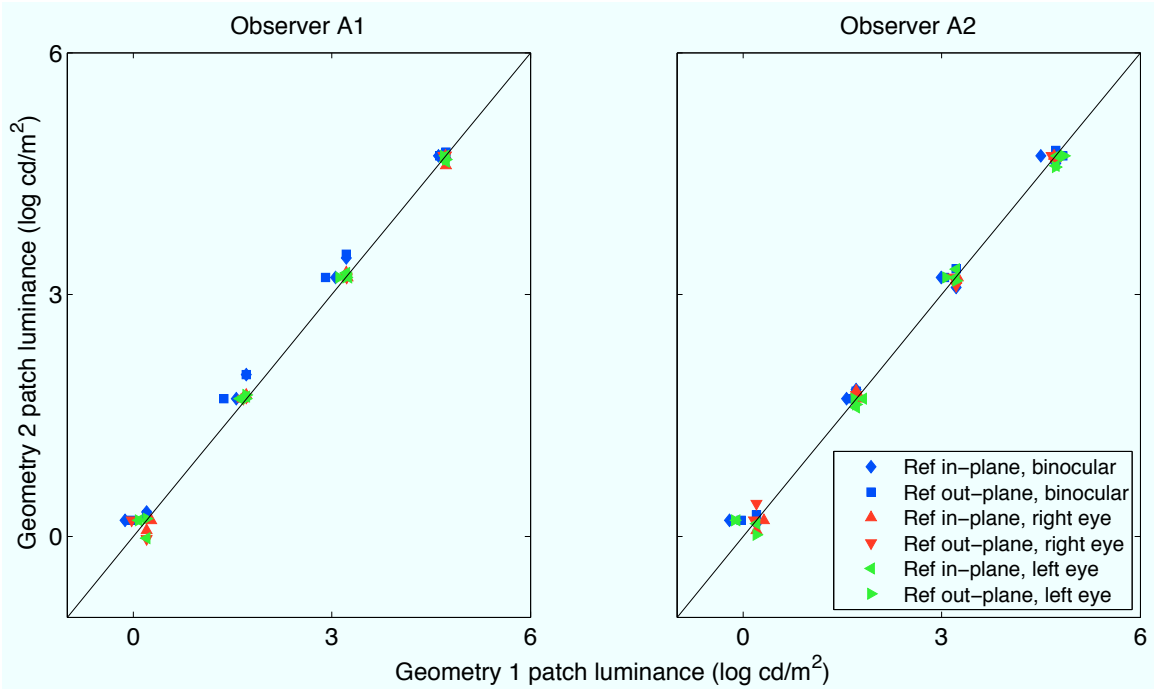

Figure A2: Two observers' data from the 2AFC experiment. Points that lie on the diagonal line in each panel are veridical matches between the adjustable match patch and the reference patch. Points that lie above the diagonal indicate an effect of the scene geometry in the expected direction. For all observers, the blue symbols (for the binocular conditions) were above the diagonal, but the red and green symbols were not.

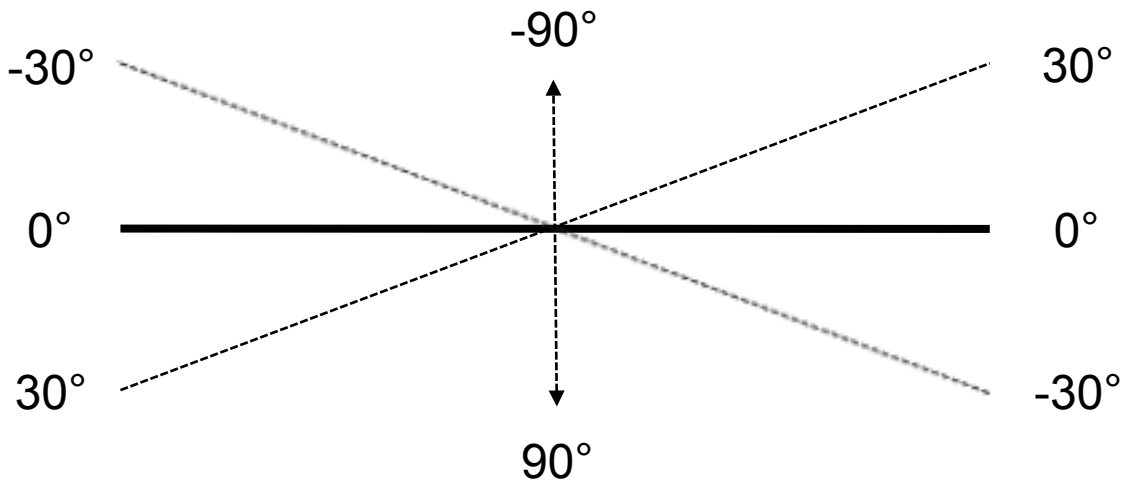

Figure A3: Angle convention for angle match experiment. Observers rotated a line that, at the beginning of each trial, was lined up with the thick horizontal line in the figure (0°). If they rotated the line counter clockwise, the response was recorded as a positive angle up to 90°. For clockwise rotations, the response was negative. Rotations past 90° were rotated 180° in the analysis.

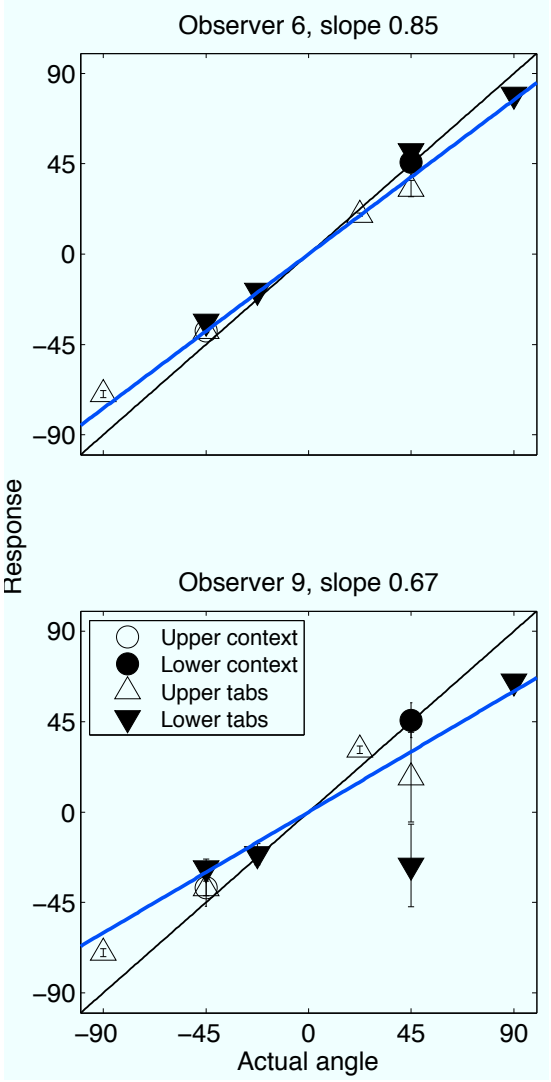

Figure A4: Angle matching data from two observers in the 90° context angle condition. In each panel, points that fall on the black diagonal line are veridical matches to the angle of the tab or context. The blue line is the best fit line through the origin, and its slope is displayed above each panel. Error bars are +/- 1 SEM across trials.

Tab vertices in rendered 3D scene space			
Experiment 1			
	X	Y	Z
Top context, in-plane	-17.5 -17.5 17.5 17.5	24.7 0.0 0.0 24.7	-24.7 0.0 0.0 -24.7
Top context, out-plane	-14.8 -17.5 17.5 18.3	24.7 0.0 0.0 24.7	24.7 0.0 0.0 24.7
Bottom context, out-plane	-14.5 -17.5 17.5 18.3	-24.7 0.0 0.0 -24.7	24.7 0.0 0.0 24.7
Bottom context, in-plane	-17.5 -17.5 17.5 17.5	-24.7 0.0 0.0 -24.7	-24.7 0.0 0.0 -24.7
Experiment 2			
-157.5°	-17.5 17.5 17.5 -17.5	0.0 0.0 24.7 24.7	0.0 0.0 -59.6 -59.6
-135.0°	-17.5 17.5 18.1 -15.5	0.0 0.0 23.7 23.7	0.0 0.0 -23.7 -23.7
-90.0°	-17.5 17.5 18.5 -14.0	0.0 0.0 23.0 23.0	0.0 0.0 0.0 0.0
-45.0°	-17.5 17.5 18.8 -12.8	0.0 0.0 22.3 22.3	0.0 0.0 22.9 22.5
-22.5°	-17.5 17.5	0.0 0.0	0.0 0.0

(-22.5°, continued)	19.2 -11.0	21.4 21.4	52.5 52.5
22.5°	-11.0 19.3 17.5 -17.5	-21.5 -2.5 0.0 0.0	52.5 52.0 0.0 0.0
45.0°	-12.9 18.7 17.5 -17.5	-22.3 -22.4 0.0 0.0	22.5 22.5 0.0 0.0
90.0°	-14.1 18.4 17.5 -17.5	-23.0 -23.1 0.0 0.0	0.0 0.0 0.0 0.0
135.0°	-15.5 -18.0 17.5 -17.5	-23.7 -23.8 0.0 0.0	-23.7 -23.7 0.0 0.0
157.5°	-17.5 17.5 17.5 -17.5	-24.7 -24.7 0.0 0.0	-59.6 -59.6 0.0 0.0

Table 1: XYZ coordinates of probe tabs in rendered 3D scene space, Experiments 1 and 2. The coordinates are in mm, with respect to the origin located in the center of the screen at the plane of fixation.

General Discussion

The three experiments in this thesis explored perceptual integration and segregation in audition and vision. In Chapter 1, we investigated the effects of similarity between target tones and noise tones in an informational masking paradigm. In a masking paradigm where listeners should be able to segregate the target from the masker, listeners seem to be unable to do so. The role of perceptual similarity in this integration was clarified by two experiments that systematically measured and manipulated the stimulus dimensions of similarity. Rather than a basic stimulus dimension like amplitude or frequency, similarity appeared to be related to a higher order stimulus dimension—the amount of frequency change over time.

In Chapter 2, we explored the segregation of luminance distributions and their interaction with spatial cues. Observers identified scenes consisting of patches drawn from one or two distributions. Observers successfully integrated information from multiple patches, and randomization in the location of the patches affected performance, but irrelevant geometric cues did not. We were interested in using these data to test theories about the use of photometric and geometric image cues in illuminant region segregation. Evidently, the specific choice of task has a large effect on how observers process these cues.

Finally in Chapter 3 we continued to examine the use of photometric and geometric cues in illuminant region segregation through an indirect task. Observers

made lightness matches to patches in scenes rendered to be consistent with lighting that had a directional component, in a scene with two illumination regions. Observers accounted for changes in the illumination, and their data suggested that the two types of cues affected different aspects of how the visual system accounts for the illumination in the scene.

These experiments provide new insights into how perceptual systems handle complex stimuli. Perceptual integration and segregation can depend on many aspects of the stimuli and task. The experiments in this thesis also demonstrate some of the difficulties in designing experiments involving these complex functions.

REFERENCES

Abbey, C. K., & Eckstein, M. P. (2006). Classification images for detection, contrast discrimination, and identification tasks with a common ideal observer. *Journal of Vision, 6*, 335-355.

Adelson, E. H. (1993). Perceptual organization and the judgment of brightness. *Science, 262*, 2042-2044.

Adelson, E. H. (2000). Lightness perception and lightness illusions. In M. Gazzaniga (Ed.), *The New Cognitive Neurosciences, 2nd ed.* (pp. 339-351). Cambridge, MA: MIT Press.

Ahumada, A. J., & Lovell, J. (1971). Stimulus features in signal detection. *Journal of the Acoustical Society of America, 49*, 1751-1756.

Alexander, J. M., & Lutfi, R. A. (2004). Informational masking in hearing-impaired and normal-hearing listeners: Sensation level and decision weights. *Journal of the Acoustical Society of America, 116*, 2234-2247.

Allred, S. R., Radonjić, A., Gilchrist, A. L., & Brainard, D. H. (2012). Lightness perception in high dynamic range images: Local and remote luminance effects. *Journal of Vision, 12(2)*, 1-16.

Anderson, B. L. (1997). A theory of illusory lightness and transparency in monocular and binocular images: The role of junctions. *Perception, 26*, 419-453.

Anderson, B. L., & Winawer, J. (2005). Image segmentation and lightness perception. *Nature, 434*, 79-83.

Arend, L. E., & Goldstein, R. (1987). Simultaneous constancy, lightness, and brightness. *Journal of the Optical Society of America A, 4*, 2281-2286.

Banks, M. S., Geisler, W. S., & Bennett, P. J. (1987). The physical limits of grating visibility. *Vision Research, 27(11)*, l915-1924.

Barlow, H. B. (1980). The absolute efficiency of perceptual decisions. *Philosophical Transactions of the Royal Society (London) B, 290*, 71-82.

Beck, J. (1959). Stimulus correlates for the judged illumination of a surface. *Journal of Experimental Psychology, 58(4)*, 267-274.

Beck, J. (1961). Judgments of surface illumination and lightness. *Journal of Experimental Psychology, 61*, 368-373.

Beck, J. (1965). Apparent spatial position and the perception of lightness. *Journal of Experimental Psychology, 69*, 170-179.

Beck, J., Prazdny, K., & Ivry, R. (1984). The perception of transparency with achromatic colors. *Perception and Psychophysics, 35(5)*, 407-422.

Bloj, M., Ripamonti, C., Mitha, K., Greenwald, S., Hauck, R., & Brainard, D. H. (2004). An equivalent illuminant model for the effect of surface slant on perceived lightness. *Journal of Vision, 4*, 735-746.

Boyaci, H., Maloney, L. T., & Hersh, S. (2003). The effect of perceived surface orientation on perceived surface albedo in binocularly viewed scenes. *Journal of Vision, 3*, 541-553.

Brainard, D. H., Brunt, W. A., & Speigle, J. M. (1997). Color constancy in the nearly natural image. 1. asymmetric matches. *Journal of the Optical Society of America A, 14*, 2091-2110.

Brainard, D. H., & Maloney, L. T. (2011). Surface color perception and equivalent illumination models. *Journal of Vision, 11(5)*, 1-18.

Craven, B. J., & Foster, D. H. (1992). An operational approach to colour constancy. *Vision Research, 32*, 1359-1366.

Durlach, N. I., Mason, C. R., Kidd, J., G., Arbogast, T. L., Colburn, H. S., & Shinn-Cunningham, B. G. (2003a). Note on informational masking. *Journal of the Acoustical Society of America, 113*, 2984-2987.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., & Kidd Jr., G. (2003b). Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *Journal of the Acoustical Society of America, 114(1)*, 368-379.

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system - Evidence for a local association field. *Vision Research, 33(2)*, 173-193.

Geisler, W. S. (2003). Ideal observer analysis. In L. Chalupa & J. Werner (Eds.), *The visual neurosciences* (pp. 825-837). Cambridge, MA: MIT Press.

Gerhard, H. E., & Maloney, L. T. (2010a). Detection of light transformations and concomitant changes in surface albedo. *Journal of Vision, 10(9)*, 1-14.

Gerhard, H. E., & Maloney, L. T. (2010b). Estimating changes in lighting direction in binocularly viewed three-dimensional scenes. *Journal of Vision, 10(9)*, 1-22.

Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., Spehar, B., Annan, V., & Economou, E. (1999). An anchoring theory of lightness perception. *Psychological Review, 106*, 795-834.

Gilchrist, A. L. (1977). Perceived lightness depends on perceived spatial arrangement. *Science, 195*, 185.

Gilchrist, A. L. (1980). When does perceived lightness depend on perceived spatial arrangement? *Perception and Psychophysics, 28*, 527-538.

Gilchrist, A. L., & Radonjić, A. (2010). Functional frameworks of illumination revealed by probe disk technique. *Journal of Vision, 10(5)*, 1-12.

Gold, J. M., Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2000). Deriving behavioural receptive fields for visually completed contours. *Current Biology, 10*, 663-666.

Heinemann, E. G. (1955). Simultaneous brightness induction as a function of inducing- and test-field luminances. *Journal of experimental Psychology, 50*, 89-96.

Hochberg, J. E., & Beck, J. (1954). Apparent spatial arrangement and perceived brightness. *Journal of Experimental Psychology, 47*, 263-266.

Howe, P. D. L. (2006). Testing the coplanar ratio hypothesis of lightness perception. *Perception, 35*, 291-301.

Ishihara, S. (1998). *Ishihara's Tests for Colour Deficiency*. Tokyo: Kanehara & Co., Ltd.

Kardos, L. (1934). Ding und Schatten: Eine experimentelle Untersuchung über die Grundlagen des Fabensehens. *Zeitschrift für Psychologie, 23*, 1-184.

Khang, B., Koenderink, J. J., & Kappers, A. M. L. (2006). Perception of illumination direction in images of 3-D convex objects: Influence of surface materials and light fields. *Perception, 35*, 624-645.

Kidd, J., G., Mason, C. R., & Arbogast, T. L. (2002). Similarity, uncertainty and masking in the identification of nonspeech auditory patterns. *Journal of the Acoustical Society of America, 111*, 1367-1376.

Kidd Jr., G., Mason, C. R., Deliwala, P. S., Woods, W. S., & Colburn, H. S. (1994). Reducing informational masking by sound segregation. *Journal of The Acoustical Society of America, 95(6)*, 3475-3480.

Kitazaki, M., Kobiki, H., & Maloney, L. T. (2008). The effect of pictorial depth cues, binocular disparity and motion parallax depth cues on lightness perception in three-dimensional virtual scenes. *PLoS ONE, 3(9)*, 1-9.

Knill, D. C., & Kersten, D. (1991). Apparent surface curvature affects lightness perception. *Nature, 351*, 228-230.

Koenderink, J. J., Pont, S. C., van Doorn, A. J., Kappers, A. M., & Todd, J. T. (2007). The visual light field. *Perception, 36(11)*, 1595-1610.

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

Kozaki, A., & Noguchi, K. (1976). The relationship between perceived surface-lightness and perceived illumination. *Psychological Research, 39*, 1-16.

Kurki, I., Peromaa, T., Hyvärinen, J., & Saarinen, J. (2009). Visual features underlying perceived brightness as revealed by classification images. *PLoS One, 4(10)*, 1-8.

Landy, M. S., & Graham, N. (2004). Visual perception of texture. In L. M. Chalupa & J. S. Werner (Eds.), *The Visual Neurosciences* (pp. 1106-1118). Cambridge, MA: MIT Press.

Lee, T. Y., & Brainard, D. (2011). Detection of changes in luminance distributions. *Journal of Vision, 11*, 1-16.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49*, 467-477.

Maloney, L. T., & Brainard, D. H. (2010). Color and material perception: Achievements and challenges. *Journal of Vision, 10(9:19)*.

Mershon, D. H. (1972). Relative contributions of depth and directional adjacency to simultaneous whiteness contrast. *Vision Research, 12*, 969.

Metelli, F. (1985). Stimulation and perception of transparency. *Psychological Research, 47*, 185-202.

Moore, B. C. J. (2003). *An Introduction to the Psychology of Hearing* (5th Edition ed.). San Diego, CA: Academic Press.

Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2002). Optimal methods for calculating classification images: Weighted sums. *Journal of Vision, 2*, 79-104.

Nascimento, S. M. C., & Foster, D. H. (2000). Relational color constancy in achromatic and isoluminant images. *Journal of the Optical Society of America, 17(2)*, 225-231.

Neff, D. L. (1995). Signal properties that reduce masking by simultaneous, random-frequency maskers. *Journal of The Acoustical Society of America, 98(4)*, 1909-1920.

Neff, D. L., & Dethlefs, T. M. (1995). Individual differences in simultaneous masking with random-frequency, multicomponent maskers. *Journal of the Acoustical Society of America, 98*, 125-134.

Neff, D. L., & Green, D. M. (1987). Masking produced by spectral uncertainty with multicomponent maskers. *Attention, Perception, & Psychophysics, 41(5)*, 409-415.

Noguchi, K., & Kozaki, A. (1985). Perceptual scission of surface lightness and illumination: An examination of the Gelb effect. *Psychological Research, 47(1)*, 19-25.

Noguchi, K., & Masuda, N. (1971). Brightness changes in a complex field with changing illumination: A re-examination of Jameson and Hurvich's study of brightness constancy. *Japanese Psychological Research, 13(2)*, 60-69.

Oh, E. L., & Lutfi, R. A. (1998). Nonmonotonicity of informational masking. *Journal of The Acoustical Society of America, 104*, 3489-3499.

Oyama, T. (1968). Stimulus determinants of brightness constancy and the perception of illumination. *Japanese Psychological Research, 10(3)*, 146-155.

Prins, N., & Kingdom, F. A. A. (2009). Palamedes: Matlab routines for analyzing psychophysical data.

Radonjić, A., Todorović, D., & Gilchrist, A. L. (2010). Adjacency and surroundedness in the depth effect on lightness. *Journal of Vision, 10(9)*, 1-16.

Reid, C. R., Jr., & Shapley, R. (1988). Brightness induction by local contrast and the spatial dependence of assimilation. *Vision Research, 28(1)*, 115-132.

Ripamonti, C., Bloj, M., Hauck, R., Mitha, K., Greenwald, S., Maloney, S. I., & Brainard, D. H. (2004). Measurements of the effect of surface slant on perceived lightness. *Journal of Vision, 4*, 747-763.

Rutherford, M. D., & Brainard, D. H. (2002). Lightness constancy: a direct test of the illumination estimation hypothesis. *Psychological Science, 13*, 142-149.

Schirillo, J., Reeves, A., & Arend, L. (1990). Perceived lightness, but not brightness, of achromatic surfaces depends on perceived depth information. *Perception and Psychophysics, 48(1)*, 82-90.

Shimozaki, S. S., Eckstein, M. P., & Abbey, C. K. (2005). Spatial profiles of local and nonlocal effects upon contrast detection/discrimination from classification images. *Journal of Vision, 5*, 45-57.

Singh, M., & Anderson, B. L. (2002). Toward a perceptual theory of transparency. *Psychological Review, 109(3)*, 492-519.

Sinha, P., & Adelson, E. H. (1993). *Recovering reflectance and illumination in a world of painted polyhedra.* Paper presented at Proceedings of the 4th International Conference on Computer Vision, 156-163.

Snyder, J. L., Doerschner, K., & Maloney, L. T. (2005). Illumination estimation in three-dimensional scenes with and without specular cues. *Journal of Vision, 5*, 863-877.

Spehar, B., Debonet, J. S., & Zaidi, Q. (1996). Brightness induction from uniform and complex surrounds: A general model. *Vision Res, 36(13)*, 1893-1906.

Todorović, D. (1997). Lightness and junctions. *Perception, 26*, 379-394.

Treisman, A. (1991). Search, Similarity, And Integration Of Features Between And Within Dimension. *Journal Of Experimental Psychology-Human Perception And Performance, 17(3)*, 652-676.

Van Ee, R., Banks, M. S., & Backus, B. T. (2001). An analysis of binocular slant contrast. *Perception, 28*, 1121-1145.

Wichmann, F. W., & Hill, N. J. (2001). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception and Psychophysics, 63(8)*, 1314-1329.

Wyszecki, G., & Stiles, W. S. (1982). *Color Science - Concepts and Methods, Quantitative Data and Formulae* (2nd ed.). New York: John Wiley & Sons.