The Aesthetics of Color Combinations

by

Karen B. Schloss

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Psychology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Stephen E. Palmer, Chair

Professor Karen K. De Valois

Professor William Prinzmetal

Professor Katherine Sherwood

Spring 2011

The Aesthetics of Color Combinations

by

Karen B. Schloss

Abstract

The Aesthetics of Color Combinations

by

Karen B. Schloss

Doctor of Philosophy in Psychology

University of California, Berkeley

Professor Stephen E. Palmer, Chair

The experiments described here were aimed at characterizing people's aesthetic responses to color pairs, both in terms of which colors people prefer in combination and how the spatial organization of the component colors influences pair preference. Previous studies of preference for and harmony of color combinations have produced conflicting results. For example, some claim that harmony increases with hue similarity, whereas others claim that it decreases. In the first set of experiments, we argue that such conflicting results are resolved by distinguishing among three types of judgments about color pairs: (a) preference for the pair as a whole, (b) harmony of the pair as a whole, and (c) preference for its figural color when viewed against its colored background. Empirical support for this distinction shows that pair preference and harmony both increase as hue similarity increases, but preference relies more strongly on component color preference and lightness contrast. Although pairs with highly contrastive hues are generally judged to be neither preferable nor harmonious, figural color preference ratings increase as hue contrast with the background increases. The present results thus refine and clarify some of the best-known and most contentious claims of color theorists. In the second set of experiments, we investigated how spatial organization influences *color-pair preference asymmetries*: differential preference for one color pair over another when the pairs contain the same colors in opposite spatial configurations. We found robust preference asymmetries, in which participants strongly preferred pairs with yellower, lighter figures on bluer, darker grounds. We also investigated which spatial factors influence these preference asymmetries. Relative area of the two component regions is clearly important, and relative *surface-based area* (i.e., after amodal figure-ground completion) is more influential than relative *image-based area*. Surroundedness is not required, because yellowness-blueness effects were comparable for figure-ground pairs in which the figure was surrounded by the ground and for mosaic arrangements in which the regions were adjacent and separated by a gap). Lightness-darkness effects, however, were in the opposite direction for figure-ground versus mosaic organizations: people prefer figure-ground organizations in which the smaller regions are lighter, but prefer mosaic organizations in which the smaller regions are darker. We provide possible phenomenological and ecological explanations for the reported results.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would first and foremost like to thank my parents, Nina and Lou Schloss, for always encouraging me to follow my dreams and teaching me that no goal is too great to accomplish. I thank Lori, Jill, Unc, Melody, and Grandma for their unconditional love and support. I thank Joseph Austerweil being my sunshine, loving partner, and programming/statistics knight in shinning armor. I thank Ani Flevaris, Amy Finn, and Francesca Fortenbaugh for being my personal cheerleaders and my best Berkeley friends.

I thank Steve Palmer for being an outstanding academic advisor and so much more. He has given me the tools to spread my research wings and is always ready to talk about anything, be it new research ideas, pretty data, or boys. I thank Bill Prinzmetal for his unconditional encouragement and smoothie outings. I thank Karen De Valois for teaching me about color, for dinner dates, and being my Berkeley mom. I thank Katherine Sherwood for giving me enthusiastic feedback from an artist's perspective. I thank Robert Remez for shaping a young undergraduate who wanted to study color into a young scientist who was ready to embark upon graduate school. I thank Mike Webster for teaching me how to accurately produce colors and for answering my countless questions over email from Reno. I thank Tom Wickens for his enthusiastic help with statistical analyses. I thank Ted Crum for his outstanding technical support.

I thank everyone I have worked with in the Palmer Lab for making the lab feel like my second family. Especially, I thank Joseph Brooks for being my role model for what it means to be a great graduate student and Jonathan Sammartino for being my awesome "brother" with whom I got to share this journey through graduate school. I thank Rosa Poggesi for being my right-hand woman and close friend and Will Griscom always being willing to talk about ideas. Finally, I thank my fantastic flock of ducklings for their help with data collection and analysis: Eli Strauss, Christie Nothelfer, Lily Lin, Patrick Lawler, Mathilde Heineman, Laila Kahn, Christopher Lau, Divya Ahuja, Jing Zhang, Cat Stone, Gary Hackett, Zoe Xu, Matt Barker-Benfield, Madison Zeller, and Arielle Younger.

1. Introduction

People make decisions about how to combine colors to achieve desired aesthetic effects nearly every day. Some decisions are as simple as choosing which color of shirt to wear with a chosen pair of slacks, whereas others are as complex as deciding on this week's color scheme for their personalized webpage interface. Those whose jobs require attention to the aesthetic impact of color combinations – e.g., interior decorators, graphic designers, architects, and artists – face many more such decisions that have even more important consequences.

When evaluating the aesthetic value of color combinations scientifically, there are two primary factors to consider: a) the relations among the colors in color space and b) the relative sizes of the colored regions. Color theorists in art have proposed many rules that prescriptively describe which colors should be combined to produce "preferable" and/or "harmonious" combinations, terms that they typically conflate. Some of the best known of these prescriptions include: Chevreul's (1838) theory that there are both harmonies of analogous colors and harmonies of contrasting colors, Itten's (1961/1973) theory that two or more colors are harmonious if they compose neutral gray when mixed together, Munsell (1921/1969) and Ostwald's (1931) theories that colors are harmonious when they lie on certain paths in color space (e.g., varying in lightness while holding hue and saturation constant). Still others have been proposed by Nemscics (1993), Goethe (1810/2006), and Moon and Spencer (1944a; 1944b).

There has been considerably less work on how the relative size of different colored regions influences the aesthetic value of the combination. For Munsell (1921/1969), images appear harmonious when the colors are centered on middle gray in Munsell color space and are equal in the product of their area times their value (lightness/darkness) times their chroma (saturation). Moon and Spencer's (1944c) theory is similar to Munsell's, but suggests that the center point of the color composition should be the adaptation point (rather than simply middle gray). For Moon and Spencer (1944c), a combination is pleasing if area-1 times distance 1 = area-2 times distance 2, where distance is a Euclidean distance in Munsell space between a color and the adaptation point. Finally, Itten's (1961/1973) theory of color proportions is based on Goethe's (1810/2006) notion of color intensities, where yellow is most "intense," followed by orange, green, red, blue and violet. Itten suggested that colors should be combined in a spatial ratio that is reciprocal to their intensities. For example, if yellow is a "9" in intensity and violet is a "3" they should be combined in a spatial ratio of 1:3.

The present research is an extended empirical investigation of aesthetic preferences for color combinations. Chapter 2 examines preferences for color pairs and how they relate to color harmony, color similarity, and single color preferences. The goal is to clarify the confusing and conflicting claims of previous theories, such as the claim that harmony increases with hue similarity versus the claim it decreases with hue similarity. Such confusions are largely resolved by distinguishing among three judgments about color pairs: (a) preference for the pair as a whole, (b) perceived harmony of the pair, and (c) preference for its figural color when viewed against its colored background. Empirical support for this distinction shows that pair preference and harmony both increase as hue similarity increases, but preference relies more strongly on component color preference and lightness contrast. Although pairs with highly contrastive hues are generally judged to be neither preferable nor harmonious, figural color preference ratings increase as hue contrast with the background increases. Thus, the present results refine and clarify some of the best-known and most contentious claims of color theorists.

Chapter 3 investigates how spatial factors influence preference for color combinations. Evidence for preference asymmetries is presented, where people prefer one color pair to another even though the two pairs differ only in having opposite assignment of the colors to different spatial regions. Participants preferred two-color figure-ground displays (a small square centered on a large square) when the color of the larger region was bluer, darker, and more preferred than that of the smaller region. Different spatial arrangements were then tested to isolate which factors most strongly influence these asymmetries (e.g., relative area and surroundedness). The results show that the relative surface-based area (after amodal ground completion behind the figure) is more important than image-based area.

Many of the experiments discussed in this dissertation are part of the Berkeley Color Project (BCP), a massive repeated measures (MRM) design aimed at understanding color aesthetics within the context of color perception and various color associations (Palmer & Schloss, 2010). All participants completed the same set of 30 tasks, divided over eight experimental sessions, using the same set of colors (see below) so that direct comparisons could be drawn across data sets.

The colors tested in all of the experiments were from the 32 BCP chromatic colors (Figure 1 and Table 1). The colors were sampled according to the dimensional structure of the Natural Color System (NCS) (Hård & Sivik, 1981), although they were actually chosen from the Munsell Book of Colors, Glossy Series (Munsell, 1966), and translated into CIE xyY coordinates to generate them on our computer using the Munsell Renotation Table (Wyszecki & Stiles, 1967). The sample included highly saturated colors of the four so-called Hering primaries approximating the unique hues: red (R), green (G), blue (B), and yellow (Y), (Munsell hues 5R, 5Y, 3.75G, and 10B, respectively). There were also four well-balanced binary hues that contained approximately equal amounts of the adjacent pair of unique hues: orange (O) between Y and R, purple (P) between R and B, cyan (C) between B and G, and chartreuse (H) between G and Y (Munsell hues 5YR, 5GY, 5BG, and 5P, respectively).

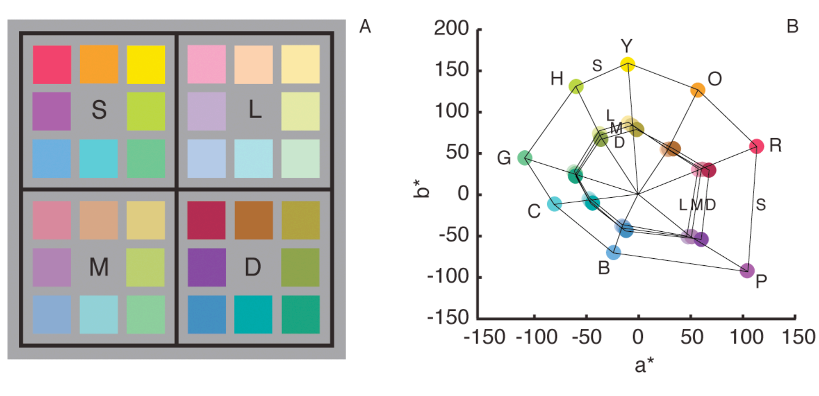

Figure 1. (A) The 32 chromatic colors of the BCP as defined by eight hues, consisting of four approximately unique hues (*Red, Green, Yellow, Blue*) and their approximate angle bisectors (*Orange, cHartreuse, Cyan, Purple*), at four "cuts" (saturation-brightness levels) in color-space (*Saturated, Light, Muted, and Dark*) and (B) the projections of these 32 colors onto an isoluminant plane in CIELAB color-space (Palmer & Schloss, 2010).

Then four "cuts" through color space were defined that differed in their saturation and lightness levels, as follows. Colors in the "saturated" (S) cut were defined as the most saturated color of each of the eight hues that could be produced on our monitor. Eight colors in the "muted" (M) cut were those that were approximately halfway between the S color and the Munsell value of 5 and chroma of 1 for the same hue. Eight colors in the "light" (L) cut were those that were approximately halfway between each S color and the Munsell value of 9 and chroma of 1 for the same hue. Eight colors in the "dark" (D) cut were those that were approximately halfway between each S cut and Munsell value of 1 and chroma of 1 for the same hue. The L, M, and D colors within each Munsell hue were equivalent in Munsell chroma (saturation).

Table 1. CIE 1931 values and Munsell values for the 32 chromatic colors (from Palmer & Schloss, 2010).

Color		x	y	Y	Hue	Value/Chroma
Red	Saturated	0.549	0.313	22.93	5 R	5/15
	Light	0.407	0.326	49.95	5 R	7/8
	Muted	0.441	0.324	22.93	5 R	5/8
	Dark	0.506	0.311	7.60	5 R	3/8
Orange	Saturated	0.513	0.412	49.95	5 YR	7/13
	Light	0.399	0.366	68.56	5 YR	8/6
	Muted	0.423	0.375	34.86	5 YR	6/6
	Dark	0.481	0.388	10.76	5 YR	3.5/6
Yellow	Saturated	0.446	0.472	91.25	5 Y	9/12
	Light	0.391	0.413	91.25	5 Y	9/6.5
	Muted	0.407	0.426	49.95	5 Y	7/6.5
	Dark	0.437	0.450	18.43	5 Y	5/6.5
Chartreuse	Saturated	0.387	0.504	68.56	5 GY	8/11
	Light	0.357	0.420	79.90	5 GY	8.5/6
	Muted	0.360	0.436	42.40	5 GY	6.5/6
	Dark	0.369	0.473	18.43	5 GY	4.5/6
Green	Saturated	0.254	0.449	42.40	3.75 G	6.5/11.5
	Light	0.288	0.381	63.90	3.75 G	7.75/6.25
	Muted	0.281	0.392	34.86	3.75 G	6/6.25
	Dark	0.261	0.419	12.34	3.75 G	3.75/6.25
Cyan	Saturated	0.226	0.335	49.95	5 BG	7/9
	Light	0.267	0.330	68.56	5 BG	8/5
	Muted	0.254	0.328	34.86	5 BG	6/5
	Dark	0.233	0.324	13.92	5 BG	4/5

Blue	Saturated	0.200	0.230	34.86	10 B	6/10
	Light	0.255	0.278	59.25	10 B	7.5/5.5
	Muted	0.241	0.265	28.90	10 B	5.5/5.5
	Dark	0.212	0.236	10.76	10 B	3.5/5.5
Purple	Saturated	0.272	0.156	18.43	5 P	4.5/17
	Light	0.290	0.242	49.95	5 P	7/9
	Muted	0.287	0.222	22.93	5 P	5/9
	Dark	0.280	0.181	7.60	5 P	3/9

2. Dissociating Preference, Harmony, and Similarity of Color Pairs

2.1. Introduction to the Aesthetics of Color Pairs

Colors are rarely experienced in isolation. In nature, yellow daffodils are seen against green grass; in the built environment, a dark brown couch is viewed against a light beige wall; in Van Gogh's *Starry Night,* the golden moon is highlighted against a deep blue sky. In all of these examples, the aesthetic experience of any given color is strongly influenced by its participation in combinations of two or more colors. In discussing color aesthetics it is therefore essential to consider not only how much people like individual colors (e.g., Hurlbert & Ling, 2007; Palmer & Schloss, 2010), but also how colors interact in more complex chromatic compositions.

2.1.1. A New Framework for the Aesthetics of Color Pairs

In this section, a new framework for discussing the aesthetic preferences for color combinations is outlined, in which there are three distinct ways of evaluating perceptual responses to color combinations: (a) people's *aesthetic preference* for a given combination, (b) their perception of *harmony* for that combination, and (c) their preference for its *figural color* when viewed against a colored background. These concepts have often been confused and/or confounded in the literature on color combinations, as explained below. It is argued that distinguishing among these three concepts and show that they are demonstrably different when they are clearly defined and appropriately measured. Moreover, results show that making these distinctions clarifies many previous confusions and resolves existing conflicts in the literature.

Pair preference is defined as how much an observer *likes* a given pair of colors as a Gestalt, or whole. *Pair harmony* is defined as how strongly an observer experiences the colors in the combination as *going or belonging together*, regardless of whether the observer likes the combination or not. These two judgments will be quite similar for an observer who likes harmonious color combinations (e.g., dark blue and light blue), but they can be arbitrarily different for an observer who likes contrastive color combinations (e.g., dark blue and saturated yellow). The distinction we draw between preference and harmony for colors is most easily understood by analogy to music. Nearly everyone who hears representative works by Mozart and Stravinsky agrees that Mozart's music is more harmonious (or consonant) and Stravinsky's music is more disharmonious (or dissonant). Nevertheless, some people prefer Stravinsky, whereas others prefer Mozart. There will be a positive correlation between average judgments of musical harmony and musical preference if people generally prefer harmonious to disharmonious music, but that does not constitute evidence that they are conceptually the same. Because preference and harmony are so clearly different concepts in music perception, it seems unlikely that they are the same concept in color perception. Finally, *figural preference* is defined as *how much the observer likes the figural color itself, when viewed against its background color*. Figural color preference is only indirectly a measure of perception of the color combination because the observer is specifically asked to respond only to the figural color. It is nevertheless relevant to aesthetic response to color combinations because the same color can look quite different when viewed against different background colors, as documented in the well-known phenomenon of simultaneous color contrast (e.g., da Vinci, 1492; Chevreul, 1839; Helmholtz, 1866/1925; Walraven, 1976; Shevell, 1978).

2.1.2. Art Theorists on Aesthetics of Color Combinations

Previous analyses of the aesthetics of color combinations have not clearly distinguished among the three aforementioned types of judgments. "Preference" and "harmony" are often used interchangeably, and preference for a combination taken as a whole is frequently confused with preference for a figural color against a background color. For example, in one of the most influential art-based theories of color aesthetics, Chevreul (1839) used the terms 'preference' and 'harmony' as though they were equivalent, and further claimed that there are harmonies of both analogous colors and contrasting colors. His *harmony of analogous colors* includes: (a) *harmony of scale* for colors that are similar in lightness and the same in hue and (b) *harmony of hues* for colors that are the same in lightness and similar in hue. *Harmony of contrast* includes: (a) *harmony of contrast of scale* for colors that differ significantly in lightness and are the same in hue, (b) *harmony of contrast of hues* for colors that differ in lightness and are similar in hue, and (c) *harmony of contrast of colors* for colors that are different in hue and different in lightness (although the lightness difference is claimed to be auxiliary). Other theories of harmony include Itten's (1961/1973) theory that two or more colors are harmonious if they produce neutral gray when mixed together as paints, Munsell's (1921) and Ostwald's (1931) theories that colors are harmonious when they have certain relations in color space (e.g., when they vary in lightness but are constant in hue and saturation), as well as other theories proposed by Nemscics (1993), Goethe (1810/2006), and Moon and Spencer (1944a; 1944b). (See Westland, Laycock, Cheung, Henry, and Mahyar (2007) and Burchett (2001) for a review). These theories are different enough that, if all their predictions were pooled, nearly every color pair could be considered harmonious!

2.1.3. Empirical Work on Aesthetics of Color Pairs

The art theoretical literature is thus riddled with confusions and contradictions. Not surprisingly, these carry over to the empirical literature as well. For example, Granger (1952; 1953; 1955a; 1955b; 1955c) conducted an extensive series of experiments on color combinations but used "preference" and "harmony" interchangeably. Indeed, he inexplicably changes terminology from one article to another in the same issue of the same journal, referring to "harmony" judgments he reported in two of these articles (Granger, 1955a; 1955b) as "preferences" in the third (Granger, 1955c). Even so, it is useful to consider his tasks and results in light of the distinctions we raise among judgments of pair preference, pair harmony, and figural preference.

Granger (1955a) found that perception of what he called "harmony" increased as hue difference increased. In Chevreul's terminology, this result appears to indicate that people perceive harmony of contrastive hues but not harmony of analogous hues. The task Granger (1955a) used, however, was ambiguous about what aspect of the color combinations were to be judged. He gave participants a color wheel with 20 removable hue wedges. Their task was to move one of the wedges (the "standard") around the circle until they found the hue "with which it made the best combination." When a hue was chosen, it was removed from the circle and the selection process was repeated until all of the remaining hues were chosen, defining a rank ordering of the "harmonies" of each figural color against all background colors. In light of our three-fold distinction, it is manifestly unclear what criterion his observers should use to define the "best combination." Is it how well the colors go together (pair harmony), how preferable the combination is as a whole (pair preference), or which accompanying color made the standard color look best (figural preference)? Granger's (1955a) finding that "harmony" increased with increasing hue contrast resembles the pattern that we find when we ask observers to make ratings of figural preference (see Experiment 4) and the pattern Helson and Lansford (1970) found when they asked observers to rate "object colors" on different colored backgrounds. This suggests that Granger's (1955a) participants may actually have judged what we are calling figural preference: which accompanying (background) color made the standard (figural) color look best.

In the same journal issue Granger (1955c) measured preferences and/or harmony again by asking participants to rank order single color preferences and all pair-wise combinations of 20 hues. He then modeled color combination preferences in terms of individual color preferences and hue distance. He found that harmony/preference increased as hue distance increased in this task as well, suggesting that his subjects may actually have liked and/or found the *combinations* more harmonious when they differed greatly in hue. However, more recent empirical results (e.g., Ou & Luo, 2006, and those reported in Experiment 1 below) have found the opposite. To make matters worse, Allen and Guilford (1936) measured the "affective value" of color combinations (presented side-by-side) and found no clear overall effect of hue similarity, although there was some evidence that very small or very large differences in hue were more pleasing than moderate differences. There has been additional empirical work on color harmony (e.g., Nemcsics, 2007; 2008; 2009a; 2009b), but it does not seem to settle the conflicting results.

2.1.4. Evidence for a Distinction between Preference and Harmony

A few previous art theorists (e.g., Albers, 1971) and perceptual researchers (e.g., Ou, Luo, Woodcock, Wright (2004b; 2004c)) have made a distinction similar to the one we advocate between pair preference and pair harmony. Albers (1971), for example, argued against Chevreul's idea that people necessarily prefer harmonious combinations, suggesting that dissonance can be as desirable as consonance. One can find evidence of this belief in many of his well-known color studies entitled "Homage to the Square."

More recently, Ou et al. (2004a; 2004b) measured both preference and harmony for 190 color pairs by asking subjects to report two binary judgments: whether each pair was liked or disliked and whether it was harmonious or disharmonious. They found that average harmony and average preference judgments were indeed highly correlated ($r = +.85$), but emphasized that even if an observer finds a pair to be harmonious, there is a moderate (31%) chance that he or she will dislike the color pair. However, Ou et al. (2004b) neglected to describe which types of combinations are harmonious yet disliked and to investigate whether there are individual differences in preference for harmony. In the present thesis we address both issues.

2.1.5. Empirical Work on Figural Color Preference

Thus far, we have focused on judgments of color combinations as a whole, either in terms of experiences of preference or harmony. Distinct from both of these judgments is preference for a figural color against a background color. Simultaneous color contrast is a well-known phenomenon: The color of the surround can strongly influence the appearance of the surrounded color (da Vinci, 1492; Chevreul, 1839; Helmholtz, 1866/1925; Walraven, 1976; Shevell, 1978). Presumably, this implies that the color of the background can also influence an observer's preference for the figural color. Helson and Lansford (1970) studied the effects of background color on preference for "object" (figural) colors by asking participants to rate (from 1-9) 125 object colors against 25 different colored backgrounds. Object colors were more preferred against backgrounds with contrasting lightness and, to a lesser extent, contrasting saturation. The effects of hue difference were more ambiguous, but generally speaking, object colors were more preferred on backgrounds with contrasting hues. It is noteworthy that although Helson and Lansford (1970) framed their research question in terms of preference for "object colors" against different backgrounds – a clear example of *figural preference* in our terminology – they actually discussed their results in terms of pair preference and pair harmony without making a principled distinction among these types of judgments. Even so, it is clear from their description of the task that, in our terms, they were actually studying what we term figural preference for a foreground color against a colored background.

Camgöz, Yerner, and Güvenç (2002) also studied how background color influenced object color preference, but they reported finding no effects of similarity or contrast. This might have occurred because they only measured each participant's single most preferred color on each of eight background hues, which is unlikely to have provided sufficiently detailed data to observe figural preference effects, even if they exist.

2.2. Aim of Experiments 1-4: Distinguishing Pair Preference, Harmony, Similarity, and Figural Color Preference

Experiments 1-4 examine the same participants' judgments of pair preference, pair harmony, pair similarity, and figural preference against colored backgrounds, drawing also on their single color preference ratings (Palmer & Schloss, 2010), as assessed within the MRM design. Evidence shows that the three kinds of judgments distinguished above are empirically as well as conceptually distinct and that a principled analysis of their interrelations clarifies much of the confusion in the literature on perception of color combinations.

2.3. Experiment 1: Preference for Color Pairs

Experiment 1 investigated preference for all pair-wise combinations of the BCP 32 chromatic colors studied in the BCP described in the Introduction (see Figure 1). Participants saw all possible pairs of the 32 chromatic colors in a figure-ground organization: a small square centered within a larger square, displayed against a neutral gray background. Both figure-ground organizations of each pair of colors were tested: A on B and B on A. For each pair, participants rated their aesthetic preference (how much they *liked* the pair as a whole) by selecting the appropriate point along a continuous line-mark response scale.

2.3.1. Methods

Participants. There were 48 participants (24 males and 24 females) who completed all 30 tasks of the BCP. All participants were screened for color deficiency using the Dvorine Pseudo-Isochromatic Plates, and none of them were found to be color deficient. All participants gave informed consent, and the Committee for the Protection of Human Subjects at the University of California, Berkeley, approved the experimental protocol (#2006-7-38).

Design. All pair-wise combinations of the 32 chromatic colors described above (see Figure 1 and the Appendix) were used to generate 992 figure-ground color combinations.

Displays. Test configurations were figure-ground pairs consisting of a small square (100px x 100px) centered on a larger square (300px x 300px). A continuous rating scale (400px long), containing demarcated center and endpoints, was located below the figure-ground pair. The rating scale was used to indicate how much each participant liked each display, ranging from "not at all" (written below the left endpoint) to "very much" (written below the right end point). Participants viewed the computer screen from approximately 70 cm. The monitor (Dell M990) was 18" diagonally with a resolution of 1024 x 768px. The background of the display was always a neutral gray (CIE $x = 0.312$, $y = 0.318$, $Y = 19.26$). The chromaticity and luminance functions of the red, green, and blue guns were measured as each gun ranged in voltage from 0-255 in equal steps of 17 using a Minolta CS100 Chroma Meter. The chromaticity and luminance functions for each gun were used to calculate the appropriate RGB values to ensure that we accurately presented the CIE xyY values for our colors. The displays were generated and presented using Presentation (www.neurobs.com).

Procedure. The 992 figure-ground combinations were displayed one at a time in a random order. The participants' task was to indicate how much they *liked* each combination on a scale from "not at all" to "very much." To respond, they used the mouse to move the cursor along the response scale and click on the point that best represented their degree of preference. Participants were informed that the vertical mark crossing the center of the scale represented a neutral (or zero) point. The recorded datum on a given trial corresponded to the x-coordinate (in pixels) at which the participant clicked on the scale for that trial, where 0 was the center of the scale. The response scale thus ranged from -200 (left endpoint of the 400 px scale) to +200 (right endpoint of the 400 px scale) and was normalized to range from -100 to +100 in the reported data. Trials were preceded by a 500ms inter trial interval (ITI) and lasted until participants made a response. Participants were allowed to take a break after each set of 60 trials.

2.3.2. Results and Discussion

Mean preference ratings for color pairs as a function of figural hue and ground hue are plotted in Figure 2A, averaged over S, L, M, and D cuts. The data show main effects of figural hue ($F(7,329) = 8.32$, $p < .001$) and ground hue ($F(7,329) = 10.70$, $p < .001$) as well as a powerful interaction between them ($F(49, 2303) = 25.42$, $p < .001$). The pattern of results, although complex, is highly regular, with three primary features. First, the peaks in the functions of Figure 2A show that figure-ground combinations for each ground hue are most preferred when the ground hue and figure hue are the same.[1] Second, pair preferences decrease monotonically as a function of the difference in hue between figure and ground. For example, the green ground-hue function in Figure 2A peaks when the figural color is another shade of the same green hue and decreases systematically as the figural color becomes less similar to green on both sides of the peak. (The reader is reminded that hue is a circular dimension, such that purple on the right end of the graph is perceptually similar to red on the left end of the graph.)

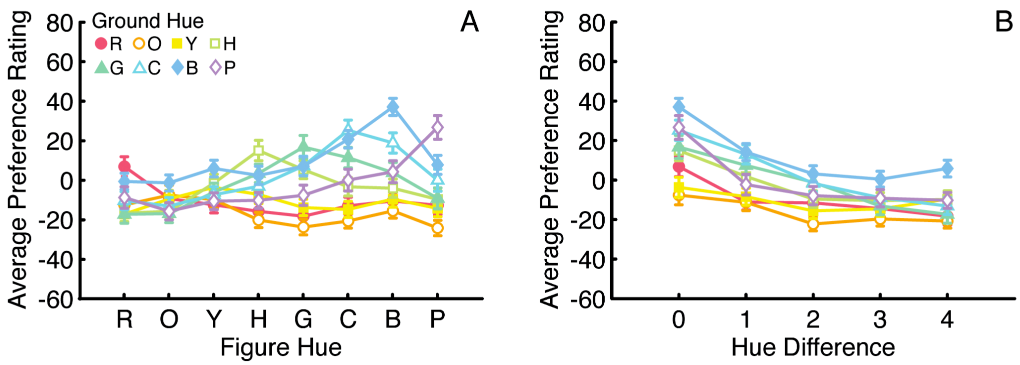

Figure 2. Preference ratings for color pairs (A) as a function of figural hue (x-axis) and ground hue (separate lines) and (B) as a function of the hue difference (in terms of steps in the BCP design) between the figure and ground. Error bars represent the standard errors of the means (SEM).

Figure 2B shows the same data as in Figure 2A, but re-plotted as a function of the hue difference between the figure and ground colors (in terms of the number of hue steps in the BCP color sample). This plot emphasizes that pair preferences are highest when the figure and ground have the same hue (but differ in saturation and/or lightness levels) and decrease monotonically as hue difference between the figure and ground increases. It also provides clear evidence that people like Chevreul's (1839) "harmonies of analogous colors" but virtually no evidence in favor of corresponding effects for contrastive hues. If the latter were present, the functions would curve upward toward the right end, where the figure and ground hues are maximally contrasting. No increases in preference for complementary colors are evident when the Bonferroni correction is applied to adjust for the eight t-tests, one for each ground hue ($\alpha = .006$).

Although this definition of "maximally contrasting" uses the perceptual complementary colors (red-green and yellow-blue), there is also little evidence of preference for contrastive hue combinations using paint-complementary colors: yellow-purple, blue-orange, and red-green. This was tested by comparing preference for pairs of paint complements versus the average of the pairs containing the two hues adjacent to their paint-complements ($F(1,47) = 1.53$, $p > .05$), after accounting for the variance explained by figure and ground color preferences (when judged singly on a neutral gray background, see Palmer and Schloss, 2010). The only paint-

[1] Notice that there are no data from the conditions in which the figural hue is the same as the ground hue when both are in the same cut (lightness and saturation level), because there would be zero contrast between them. The statistical tests are therefore computed from the averages of all pairwise combinations of the four cuts for a given hue pair (16 pairs when the hues in the

complementary pair that was more preferred than its nearest neighbors (after applying the Bonferroni correction) was orange-blue compared with the average of orange-cyan and orange-purple ($F(1,47) = 11.17$, $p < .008$).

The third salient feature of the results is the systematic variation in pair preferences with hue. Both the maxima of the ground-hue functions and their overall level generally increase as the hues become bluer and decrease as they become yellower. The strong correlation ($r = +.94$) between the level of the curves in Figure 2B (mean preference across hue differences of 0 to 3) and the sharpness of their decline (slope of the best-fitting line between hue differences of 0 to 3) indicates that grounds containing more preferable hues (e.g., blue, cyan, and purple) get a larger preference increment when paired with figures of the same or similar hues than do grounds containing less preferable hues (e.g., yellow and orange).

Figure 3A isolates main effects of figural hue and ground hue. The shape of these functions, showing preference for cooler over warmer colors, closely resembles the shape of the hue preference function for single color preference ratings (Palmer & Schloss, 2010) from the same participants (Figure 3B). This resemblance strongly suggests that preferences for color pairs are influenced to some degree by preferences for the component colors.

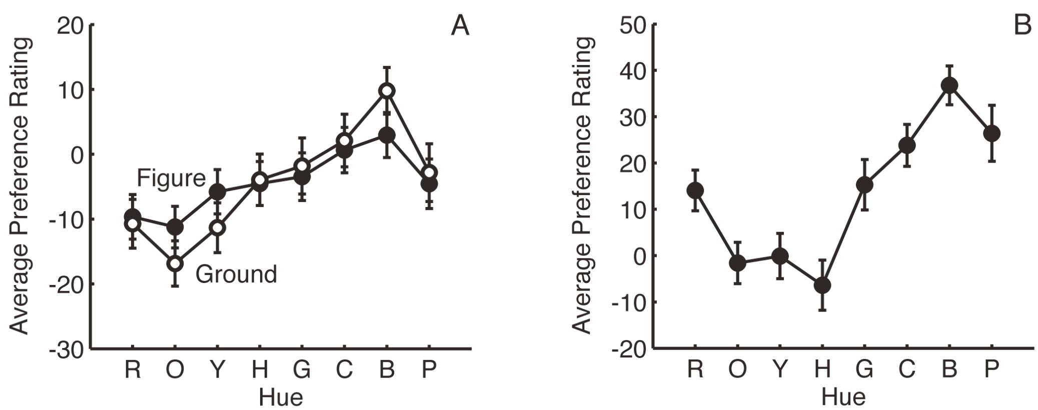

Figure 3. (A) Main effects of ground hue (open circles) and figure hue (closed circles) for pair preference ratings, and (B) for single color preferences of the same participants (Palmer and Schloss, 2010). Error bars represent the standard errors of the means (SEM).

A multiple linear regression model was used to determine the degree to which the same participants' preferences for the component ground and figure colors (when judged singly on a neutral gray background, see Palmer and Schloss, 2010) could account for pair preferences. Only 21.7% of the variance in pair preferences for all 992 color pairs could be explained by single color preferences: 15% from ground color preference and an additional 6.7% from figural color preference. Ground color preference influences pair preference more than figural color preference, as indicated by the facts that ground color preference accounts for more variance than figural color preference and that the ground curve in Figure 3A is more extreme than the figural curve. This somewhat surprising result may simply reflect the fact that the ground color covers more area than the figural color. Even so, this additive model based on single color preferences accounts for relatively little variance in the overall pattern of results because it cannot, by definition, explain the complex figural-hue x ground-hue interaction so clearly present in Figure 2A. One or more relational factors are required. Below we attempt to identify what those relational factors might be using various predictors derived from Munsell dimensions.

The ten Munsell factors considered in this analysis were the hue difference (the number of Munsell hue steps by which the figural and ground colors differed), the sum, the signed difference, and the absolute value of the figure-ground difference in hue coolness (the number of Munsell hue-steps removed from Munsell hue 10R)[2], the value (or lightness) and the chroma (or saturation) of the figural and ground colors. All possible combinations of factors were tested for all possible numbers of factors (i.e., all pairs of factors were tested in 2-factor models and all triplets were tested in 3-factor models, and so on up to 10 factors). The model we report as the "best" model was the model that explained the largest percentage of variance that also explained at least 1% more variance than the next best model with the same number of factors. We also report the results of the "full model" that includes all factors, but we do not name or give the order of entry for the factors included beyond those in the best model as just defined.

The left-most bar in Figure 4 shows the best fitting model for pair preference ratings, where each factor's increment in percentage of variance explained is represented by a corresponding increment in the height of the bar, with the lowest segment being the factor that was entered first. The best fitting model explained 53.5% of the variance in pair preference ratings, showing that more preferred pairs contained cooler colors that were similar in hue and contrasting in Munsell value (lightness). An additional 7% of the variance can be explained in the full model when all 10 factors are included, but there is no clearly defined "best" model (see above) in any of the regressions containing more than 3 factors.

When figure and ground color preference are added to the 3-factor Munsell regression model shown in Figure 4, they account for an additional 9.4% of the variance (6.9% from ground color preference and 2.5% from figural color preference). This brings the total amount of variance explained to 62.9%, which shows that component color preferences are still important after the variance due to the relational factors in the Munsell model has been removed. In discussing the results of Experiment 2, however, we report an even better model, which explains 80.8% of the variance, based on rated color harmony as a relational factor.

[2] 10R (red-orange) was chosen because it was closest to the minimum of the coolness function obtained from participants' ratings of this dimension.

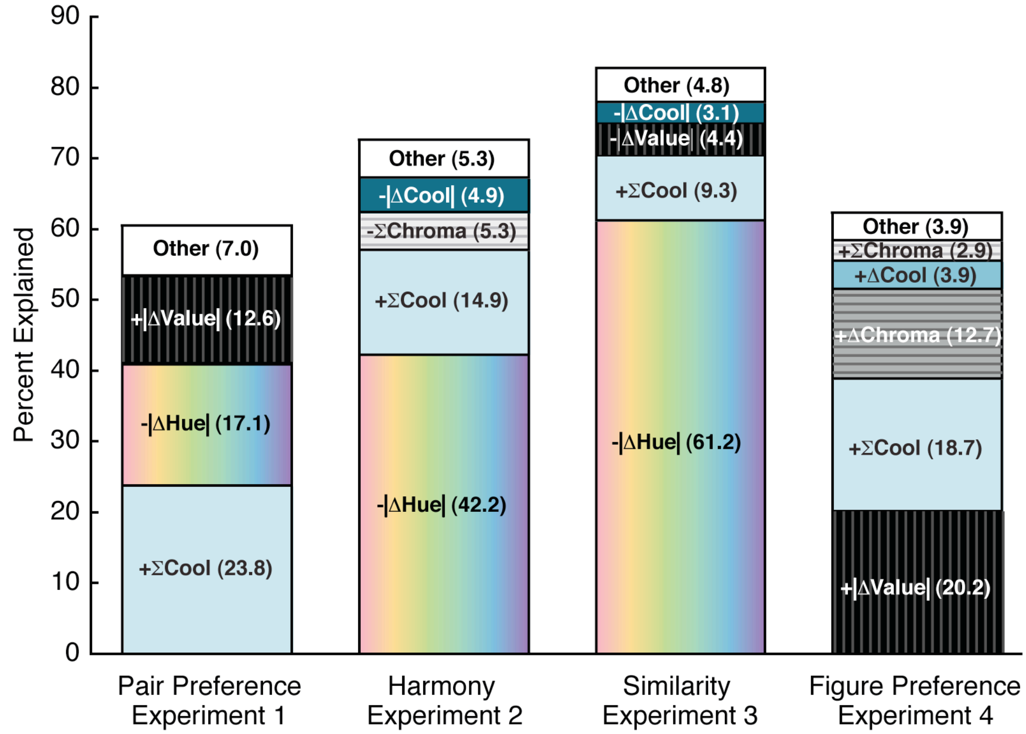

Figure 4. Bars show the percentages of variance explained by the best-fitting Munsell models for pair preference (Experiment 1), pair harmony (Experiment 2), two-color similarity (Experiment 3), and figural color preference against colored backgrounds (Experiment 4). Stripes within each bar show the percentage of variance explained by each factor in the order with which they were entered in to the regression model (bottom to top). The sign before each term indicates whether the factor was positively or negatively weighted in the corresponding regression equation (e.g., "+∑Cool" indicates that the sum of the coolnesses of the component colors was positively related to rated preference, harmony, and similarity, whereas "-|ΔHue|" indicates that the absolute value of their difference in hue was negatively related to these ratings).

To further understand the nature of pair preferences, we also examined the effects of figural and ground cut: saturated (S), light (L), muted (M) and dark (D). The means that were analyzed (see Figure 5) only included pairs with hue-difference steps of 1 through 4 because there were no zero hue-difference data for same-cut pairs. The results show no main effects of figural cut ($F(3,141) = 2.90$, $p > .05$) or ground cut ($F < 1$), but there was a reliable interaction between them ($F(9,423) = 7.66$, $p < .001$). Pair-wise comparisons of cut combinations showed that the only effects of cut occurred for the saturated ground conditions: Combinations with saturated figures on saturated grounds were preferred to those with light, muted and dark figures on saturated grounds ($t(47) = 3.74, 6.33, 3.56$, $p < .002$), and those with light figures on saturated grounds were preferred to those with muted figures on saturated grounds ($t(47) = 3.64$, $p < .002$). (A critical value of .002 was used after applying the Bonferroni correction to compensate for the 24 comparisons.)

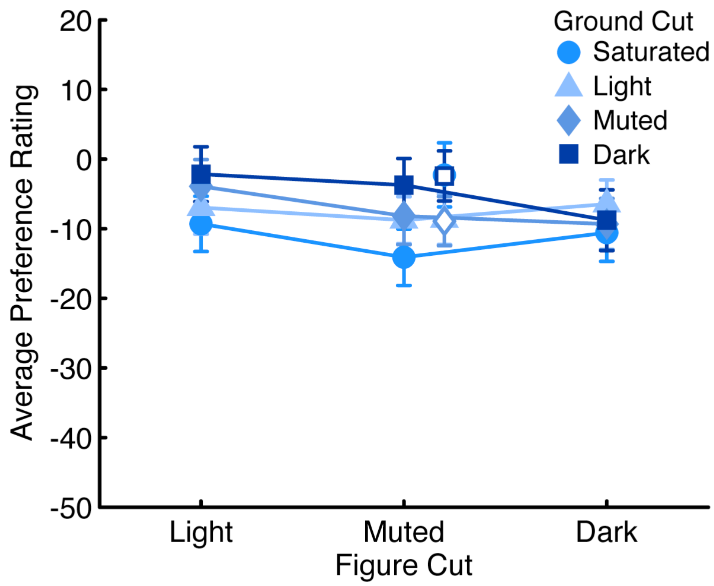

Figure 5. Preference for color pairs for each ground cut (separate lines), as function of figure cut (x-axis). Data points for the saturated (S) figure cuts (open symbols) are plotted separately at an x-axis level similar to the muted (M) colors because they share similar lightness levels, but are slightly offset for clarity. Error bars represent the standard errors of the means (SEM).

The effects of figure and ground cut as a function of hue difference between the figure and ground colors can be found in Figure 6. There were no data points for same-cut pairs with zero hue difference because the component colors were identical, so separate ANOVAs were conducted on two data sets. The first ANOVA included only hue-difference steps of 1 through 4 for all cut comparisons (hereafter "Δ1-4, all cuts"). The second ANOVA included all five hue-difference steps (0 through 4) but only for cut comparisons in which the figure and ground were from different cuts (hereafter "Δ0-4, different cuts"). As shown in Figure 6, pair preferences for all cut combinations decreased monotonically as hue difference between the figure and ground color increased (Δ1-4, all cuts: $F(3,141) = 27.83$, $p < .001$); Δ0-4, different cuts: $F(4,188) = 51.08$, $p < .001$), which is consistent with the inclusion of contrast in Munsell lightness (value) in the previously described best-fitting regression model. There was also a 3-way interaction among figure cut, ground cut, and hue difference (Δ1-4, all cuts: $F(27,1269) = 2.22$, $p < .001$; Δ0-4, different cuts: $F(24, 1128) = 4.31$, $p < .001$), but the size of this interaction is small, its nature is unsystematic, and its interpretation is unclear. The simpler figure cut x hue difference interaction was systematic, however, with preference for pairs with D and M figures decreasing more rapidly than pairs with S and L figures as hue difference increased (Δ1-4, all cuts: $F(9, 423) = 2.25$, $p < .05$; Δ0-4, different cuts: $F(8, 376) = 4.29$, $p < .001$). There was no such interaction between ground cut and hue difference (Δ1-4, all cuts: $F(9, 423) = 1.16$, $p > .05$; Δ0-4, different cuts: $F(12, 564) = 1.61$, $p > .05$).

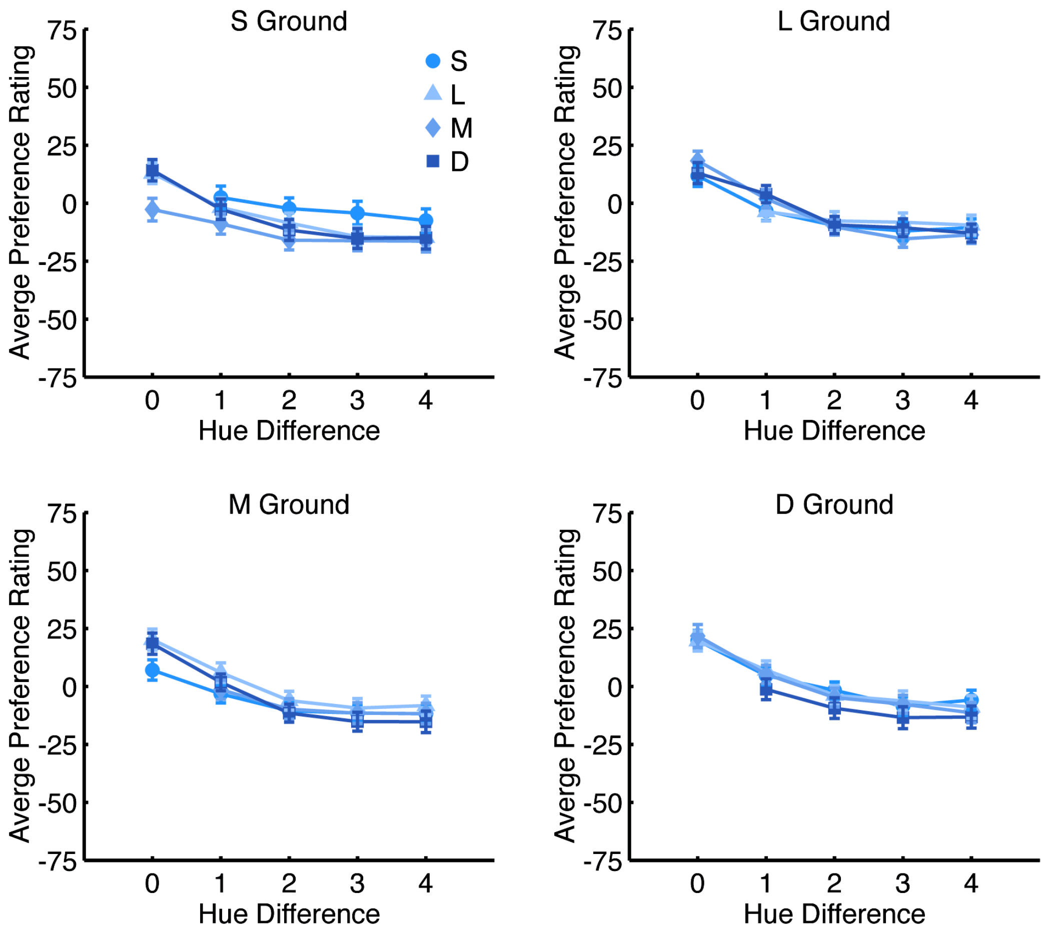

Figure 6. Pair preference ratings for figural color (separate lines) and ground color (separate graphs) of the display for each cut, as a function of the hue difference between the figure and ground colors. Error bars represent the standard errors of the means (SEM).

Figure-ground asymmetries in preference (e.g., warmer-figure/cooler-ground vs. cooler-figure/warmer ground) were also examined to see whether figure-ground status influenced pair preferences by testing the signed difference between the figure and ground color along the Munsell dimensions tested above. Pair preferences were slightly, but significantly, correlated with the differences in coolness, such that pairs with warmer figures on cooler grounds were preferred to the reverse ($r = +.13$, $p < .001$). The same was true of differences in Munsell value: Pairs with lighter figures on darker grounds were preferred to the reverse ($r = +.14$, $p < .001$). Nevertheless, these differences due to spatial figure-ground organization were quite small in comparison with the differences due to different colors. A regression model based on these two spatial predictors explained only 4% of the variance in pair preference, with the value difference accounting for 2% of the variance (lighter figures being preferred) and coolness differences accounting for an additional 2% (warmer figures being preferred). A further investigation of preference asymmetries using a two alternative forced choice task, in which the only difference between the two pairs in the comparison was the figure-ground assignment of the colors, will be presented in Chapter 3. The asymmetries of coolness and lightness noted here are robust in the forced choice task.

2.4. Experiment 2: Color Harmony and its Relation to Preference

Experiment 1 showed that there are clear, systematic patterns in preferences for color pairs that are governed primarily by component color preferences, coolness, hue similarity, and lightness contrast. Experiment 2 investigates what factors influence color harmony ratings and how they relate to pair preference ratings.

Findings previously reported by Ou and colleagues (Chuang & Ou, 2001; Ou et al., 2004; Ou & Luo, 2006) suggest that perceived harmony of color pairs is closely related to pair preference. Chuang and Ou (2001) found that pairs in which both colors were the same in hue were judged as more harmonious than those with different hues, and we found the same to be true for pair preferences in Experiment 1. They also found that pairs that were different in luminance were judged to be more harmonious than those that were similar in luminance, and we found the same to be true for pair preferences in Experiment 1. They further reported that preference for the component colors of a pair influenced harmony judgments: Pairs that included two favorite colors were most harmonious, followed by pairs that included one favorite color and then pairs with no favorite color. Ou and Luo (2006) later reported that pairs were harmonious when colors were similar in hue, different in lightness, had a high combined (summed) lightness, and included light yellow as a component. Unfortunately, many of these conclusions are compromised by Chuang and Ou's definitions of harmony as "that which pleases the viewer" or "that which is harmonious." In the first definition, it is unclear whether "pleasing" refers to how well the colors go together (what we call pair harmony) or how much the observer likes the pair (what we call pair preference). Their second definition of harmony is simply circular and thus meaningless.

The primary goal for Experiment 2 was to obtain harmony ratings that were uncontaminated by confusions with pair preference using the same participants and the same colors as in Experiment 1. These ratings were used to determine how well people's harmony judgments can explain the pattern of variation in their pair preferences (see Experiment 1). In particular, it was predicted that perceived harmony might be the relational variable that would best complement preferences for the component figure and ground colors in explaining people's preference ratings for color pairs. In addition, the findings of Chuang and Ou (2001) were examined using a more refined definition of harmony by including the musical analogy described in the introduction when instructing our observers about the difference between harmony and preference. A secondary goal, was to examine individual differences in "preference-for-harmony" as indexed by the correlation between people's pair preference ratings in Experiment 1 and their harmony ratings in Experiment 2.

2.4.1. Methods

Participants. The participants were the same 48 observers who completed Experiment 1.

Design and Displays. The design and displays were the same as in Experiment 1, except that the left endpoint of the rating-scale line was labeled "dissonant" and the right endpoint was labeled "harmonious."

Procedure. As in Experiment 1, participants were presented with each of 992 chromatic figure-ground combinations, one at a time in a random order. The harmony task was to indicate how "harmonious" the figure-ground color pair was on a scale from "dissonant" to "harmonious." In order to clarify the difference between preference and harmony, participants were told the following: "Your task will be to indicate how "harmonious" you find each combination – how well the colors go together – by clicking a point on a scale like the one below. We are not asking you to rate how much you *like* each pair of colors. Some people like color combinations that are harmonious and others like combinations that are dissonant. For example, in music, some like Mozart and others like Stravinsky, but everyone would agree that Mozart is more harmonious and Stravinsky is more dissonant." The harmony-rating task was completed in a different testing session that took place after the preference-rating task.

2.4.2. Results and Discussion

Because Chuang and Ou (2001) reported that their harmony data were influenced by preferences for the component figure and ground colors, the instructions in this experiment were specifically tailored try to dissociate such effects. To examine the extent to which we succeeded, the influence of figure preference and ground preference on harmony ratings were examined in a two-factor regression analysis. The results show that only 1.4% of the variance in our harmony ratings is due to component color preferences: 1.1% from ground color preference and an additional 0.3% from figural color preference. This amount is an order of magnitude less than the 21.7% of the variance that is due to figure preference and ground preference in the pair preference data of Experiment 1. This striking reduction supports the contention that, with appropriate instructions, observers can make harmony ratings that are essentially unaffected by their single color preferences. This difference between the present results and those of Chuang and Ou (2001) also supports the belief that their observers probably interpreted their instruction to judge how "pleasing" the color pairs were as asking, to some extent, about preference rather than or in addition to harmony (at least as defined it in the instructions).

The pattern of color harmony ratings as a function of figural hue and ground hue is shown in Figure 7A. Notice first that it is strikingly similar to the pattern of results for pair preference ratings but somewhat more exaggerated. Indeed, the correlation between average pair-wise preference ratings and average pair-wise harmony ratings was +0.79, accounting for 62% of the variance. Given this strong positive relation, it is understandable that Chevreul and other color theorists erroneously equated color harmony and color preference: Generally speaking, people do tend to prefer harmonious color combinations. That does not mean that harmony and preference are either conceptually or empirically the same, however. It is also noteworthy that there was greater agreement among participants about their judgments of pair harmony than about their judgments of pair preference. The correlation of each observer's harmony ratings with the group-average harmony ratings (average $r = +.51$) was significantly greater than the corresponding correlation of their preference ratings with the group-average preference ratings (average $r = +.36$) ($t(47) = 5.72$, $p < .001$). This fact indicates that, whatever perceived color harmony might be, people are in better agreement about it than about their preferences for the same colored displays.

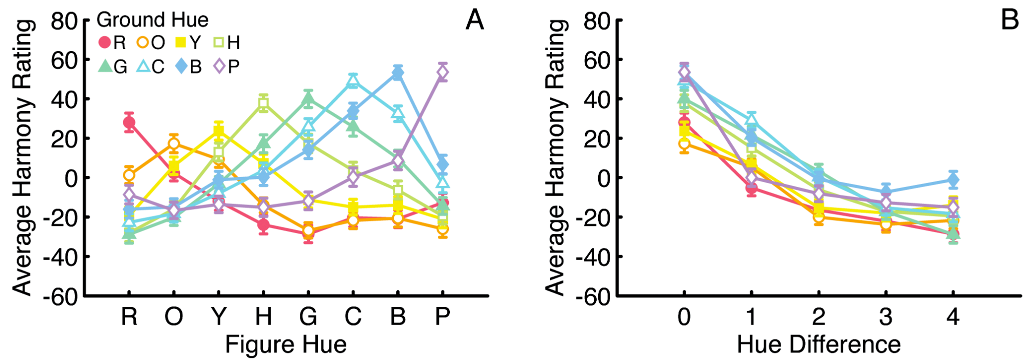

Figure 7. Harmony ratings for color pairs (A) as a function of figural hue (x-axis) and ground hue (separate lines) and (B) as a function of the hue difference (in terms of steps in the present BCP design) between the figure and ground. Error bars represent the standard errors of the means (SEM).

The harmony data in Figure 7A reveal main effects of both figural hue ($F(7,329) = 28.92$, $p < .001$) and ground hue ($F(7,329) = 22.80$, $p < .001$), as well as a strong interaction between them ($F(49,2303) = 64.85$, $p < .001$). Harmony ratings were highest for each pair when the figure and ground hues were the same, and they decreased monotonically as hue difference increased. This result is consistent with Chevreul's (1839) claim that analogous colors are harmonious. It is also consistent with previous empirical studies of color harmony in which harmony was defined as "pleasantness" (e.g., Chuang and Ou, 2001; Ou and Luo, 2006), even though the latter data appear to be contaminated by single color preferences for the reasons outlined above.

As was also true for pair preferences, there is virtually no evidence supporting Chevreul's (1839) claim that contrastive hues are harmonious. If there had been, the harmony curves in Figure 7B, which are plotted as a function of hue difference, would curve upward toward the right end, where the figure and ground hues are maximally contrasting (red-green and blue-yellow). Instead, when these data are averaged over ground hue, there is a reliable *decrease* in harmony ratings for pairs from the hue-step 3 to hue-step 4 conditions ($F(1,47) = 6.11$, $p < .05$). The same is true for hues paired with their paint-complement (blue-orange and yellow-purple): paint-complement pairs were rated as reliably *less* harmonious than the same hues paired with the average of the two hues adjacent to their paint-complement ($F(1,47) = 17.67$, $p < .001$). Thus, the results are not in accord with what Chevreul presumably would have predicted. The only reliable up-turn is for the blue-ground/yellow-figure combination ($F(1,47) = 11.05$, $p < .006$), which may be an artifact arising from the fact that blue and gold (essentially, a shade of yellow) are the official school colors of the University of California, Berkeley, where the experiments were conducted. (See Schloss, Poggesi, and Palmer (in press) for an in-depth study of the influence of school colors on the color preferences of Berkeley and Stanford students.) The reliable increment for blue and yellow combinations over their immediately adjacent neighbors may also be due to the large lightness contrast between them.

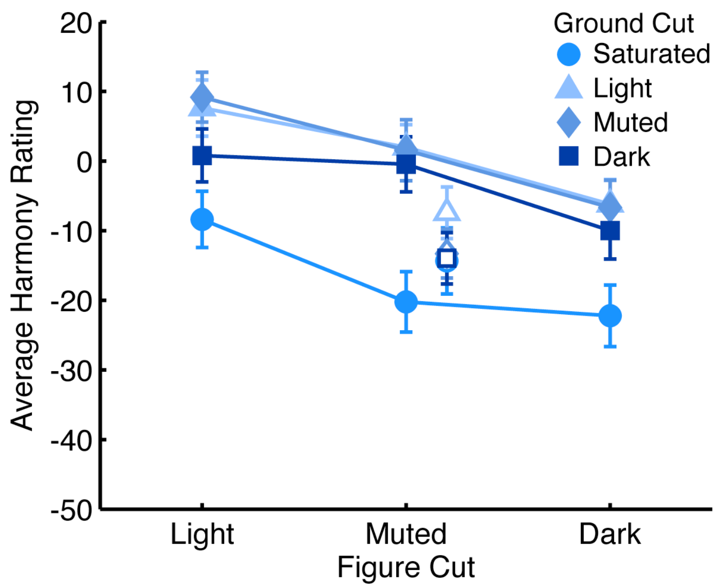

Figure 8. Harmony ratings of color pairs for each ground cut (separate lines), as a function of figure cut (x-axis). Data points for the saturated figure cut (open symbols) are plotted separately at the same x-axis point as the muted colors because they share similar lightness levels, but they are slightly offset for clarity. Error bars represent the standard errors of the means (SEM).

An analysis of the effects of cut (saturation/lightness level) showed main effects of figural cut ($F(3,141) = 28.25$, $p < .001$), ground cut ($F(3,141) = 10.19$, $p < .001$), and their interaction ($F(9,423) = 8.41$, $p < .001$), as shown in Figure 8. Combinations that contained lighter and less saturated colors tended to be rated as more harmonious. The results of paired comparisons between each cut combination can be found in Figure 9. To summarize: The L figures were judged most harmonious against all four ground cuts, and the D and S figures were judged least harmonious against all four ground cuts.

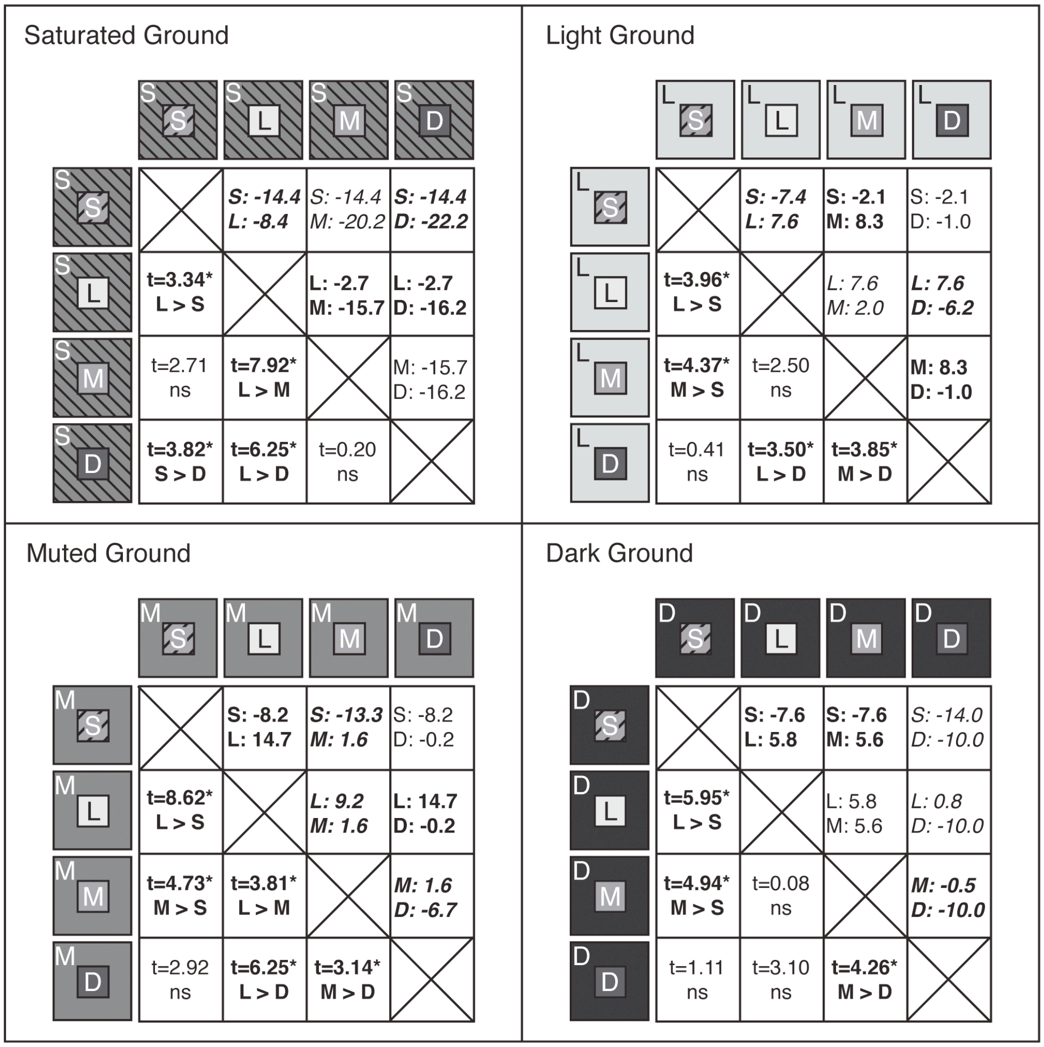

Figure 9. Comparisons between harmony ratings of color combinations with the same ground cut (separate quadrants) and different figure cuts. Icons adjacent to each row column represent the cuts of the figure-ground pairs that were judged. The lower triangle of each ground-cut matrix shows the results of t-tests (df = 47, *p ≤ .002, using the Bonferroni correction) and direction of the difference (e.g., "L > S" in the Saturated Ground quadrant indicates that light figures on saturated grounds were judged more harmonious than saturated figures on saturated grounds). The upper triangle of the matrix shows the means of the pairs that were compared. The means for comparisons with same-cut pairs (italicized text) include only hue-difference steps of 1-4 for both pairs. All other means include all hue-differences steps (0-4). Bold face text indicates differences were significant. (Note: In this diagram, the figure squares.

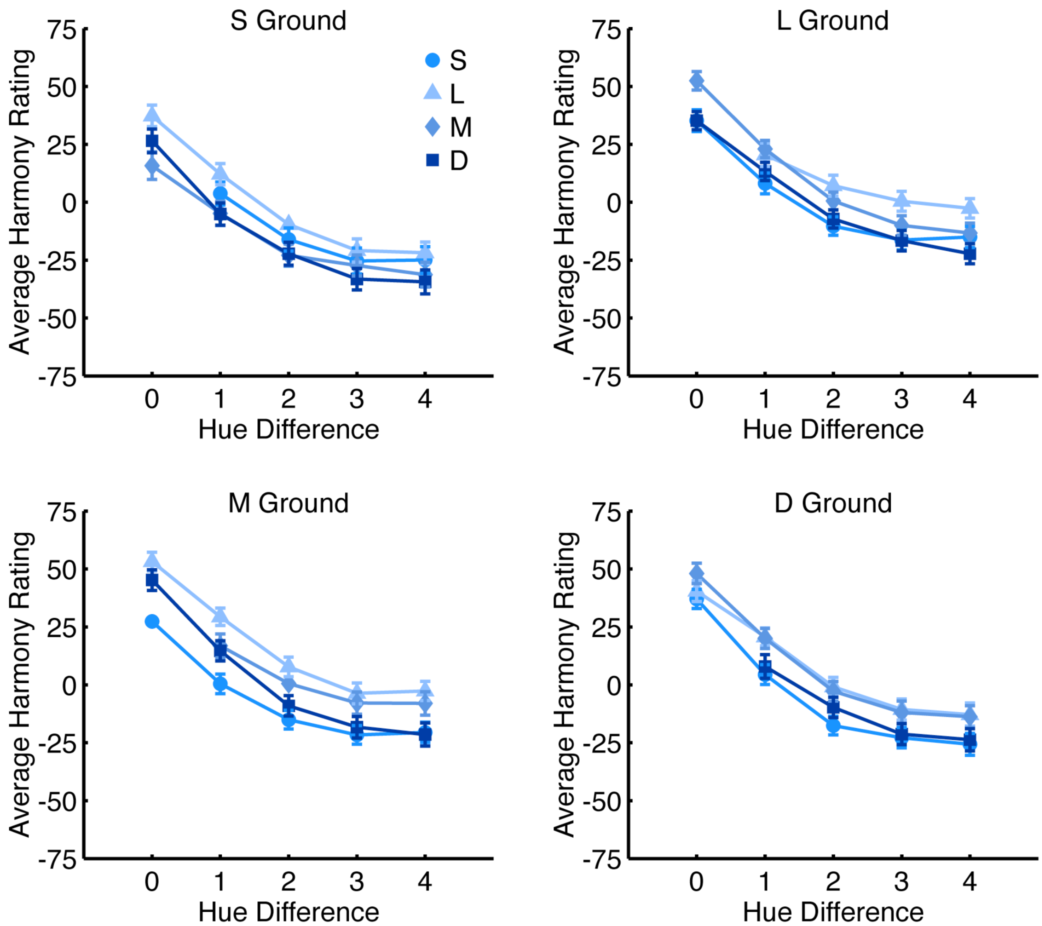

Figure 10. Harmony ratings for figural color (separate lines) and ground color (separate graphs) of the display for each cut, as a function of the hue difference between the figure and ground colors.

Figure 10 shows plots of harmony ratings for figure and ground cuts as a function of hue difference between the figure and ground colors. Similar to pair preference ratings from Experiment 1 (Figure 6), harmony ratings decreased monotonically as the hue difference between the figure and ground colors increased (Δ1-4, all cuts: $F(3,141) = 54.83$, $p < .001$; Δ0-4, different cuts: $F(4,188)= 111.70$, $p < .001$), but the reductions are more pronounced (Δ1-4, all cuts: $F(3,141) = 27.71$, $p < .001$; Δ0-4, different cuts: $F(4,188) = 43.90$, $p < .001$) (compare Figures 6 and 10).

There was a 3-way interaction among figure cut, ground cut, and hue difference (Δ1-4, all cuts: $F(27,1269) = 3.26$, $p < .001$; Δ0-4, different cuts: $F(24,1128) = 6.03$, $p < .001$). Relative to the other figural cuts, saturated figures are less harmonious with muted grounds of similar hues, light figures are more harmonious with light grounds of contrasting hues, light figures are more harmonious with muted grounds of similar hues, and muted figures are more harmonious with light grounds of similar hues. There was also an interaction between figure cut and hue difference (Δ1-4, all cuts: $F(9,423) = 4.19$, $p < .001$; Δ0-4, different cuts: $F(8,376) = 4.73$, $p < .001$) in which harmony ratings for pairs including dark figures decrease more rapidly as hue difference increased, relative to the other cuts. There was no such difference between ground cut and hue difference (Δ1-4, all cuts: $F(9,423) = 1.06$, $p > .05$; Δ0-4, different cuts: $F<1$).

What, then, are the color appearance factors that influence ratings of color harmony? The same 10 Munsell factors tested for pair preference in Experiment 1 were analyzed in regression analyses to predict perceived color harmony. The best fitting model (Figure 4) for the 992 color pairs showed that more harmonious pairs were more similar in hue, cooler, more desaturated, and more similar in coolness (67.3% of the variance explained). When all 10 Munsell factors are included in the full model, 72.6% of the variance could be explained, but there was no clear "best" model with more than four factors.

In the discussion of Experiment 1, we noted that one or more relational variables was required to account for the interaction between figure and ground colors in preference for color pairs. We then identified a set of relational Munsell factors that explained 53.5% of the variance in such preferences. When the inherently relational factor of harmony ratings is also included as a predictor variable, the best linear model accounts for 80.8% of the variance in preference ratings (multiple $r = +.90$). Harmony ratings alone explain 62.3% of the variance (more than all ten Munsell factors combined), preference for the ground color adds another 9.3%, preference for the figure adds a further 4.7%, and the absolute value of the difference in Munsell values (lightnesses) adds a final 4.5% (larger lightness differences being preferred). Although there is a remarkably strong relation between harmony and preference, it falls considerably short of the equivalence that would be required to justify their interchangeable use by Chevreul (1839) and others (e.g., Granger, 1955a-c).

What are the differences between pair preference and harmony? Many differences are found in the effects of cuts (saturation and lightness levels) where preferred pairs contain more dark and saturated colors and harmonious pairs are generally lighter (see Figures 6 and 8). Figure 11 shows a scatter plot of preference ratings (y-axis) versus harmony ratings (x-axis) for each color pair in a way that highlights many of the principal differences. The high correlation between preference and harmony is evident in the strong linear trend of the point-cloud with a slope of somewhat less than unity. Differences between preference and harmony are then evident in systematic deviations from the best-fitting regression line.

First, Figure 11 shows that the color pairs that are more preferred than harmonious (upper left quadrant) are generally high in lightness contrast, whereas those that are more harmonious than preferred (lower right quadrant) are generally low in lightness contrast. Second, it illustrates the dissociation between pair preference and pair harmony in terms of component color preferences. Palmer and Schloss (2010) found that the same participants especially disliked dark yellow and dark orange, and Figure 11 shows that although pairs containing those particular colors were disliked, they were still judged to be harmonious when combined with light colors of similar hues. Figure 11 also highlights some similarities between preference and harmony. First, pairs containing cool colors are generally both more harmonious and more preferred (toward the upper right quadrant) than pairs containing warm colors, which are less harmonious and less preferred (toward the lower left quadrant). Second, saturated red produces particularly disharmonious and disliked combinations (extreme lower left in Figure 11), particularly those pairs containing a saturated red ground.

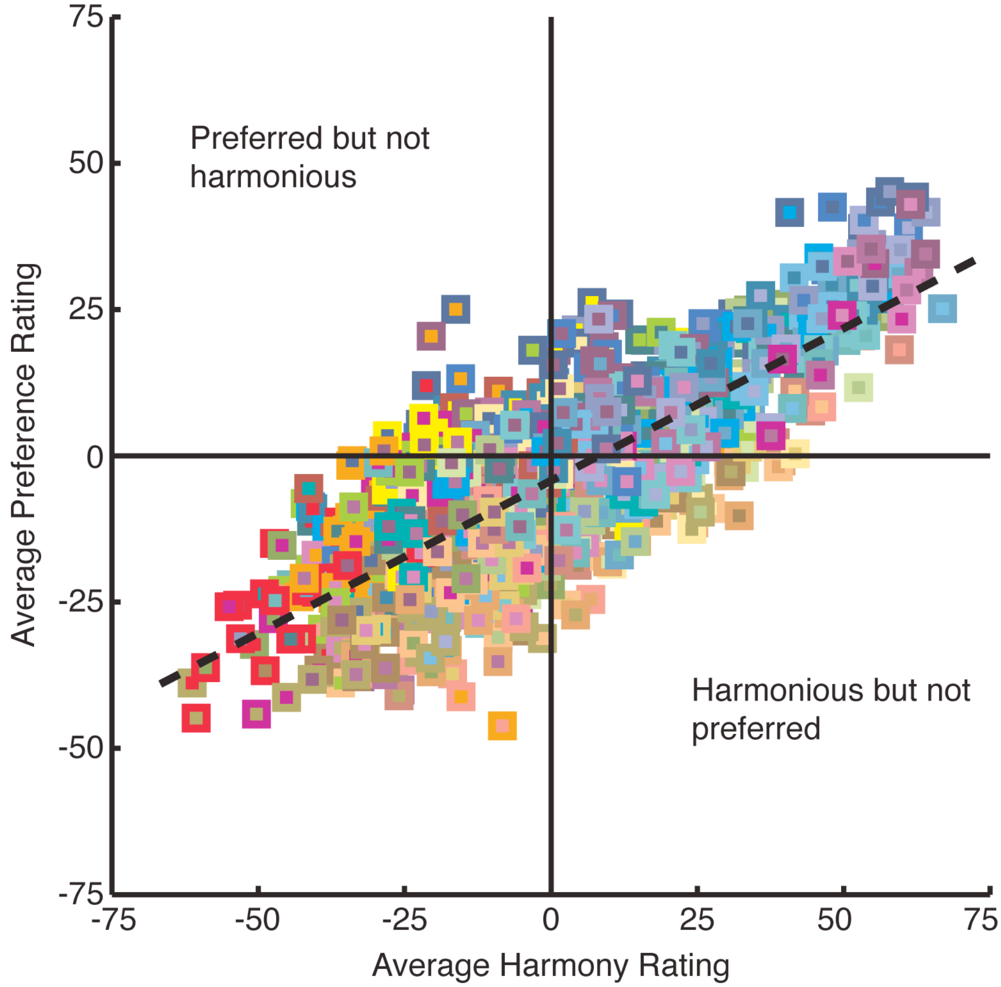

Figure 11. Preference ratings for each color pair plotted as a function of its harmony rating. Each of the 992 data points depicts an approximation of the figural color (small square) and ground color (large square behind the figure). The dashed line shows the best fitting regression line relating preference to harmony (y = -7.93 + 0.52x).

The differences between preference and harmony ratings can be analyzed quantitatively through regression analyses after their mutual variation (62.3%) has been removed. First, as stated above, the residual systematic variance in preference ratings was due to preferable ground colors (9.3%), preferable figural colors (4.7%), and large differences in lightness (4.5%). In contrast, the residual systematic variance in harmony ratings was due to greater hue similarity (i.e., fewer Munsell hue steps apart) (13.7%) and lower overall saturation (i.e., lower sum of the Munsell chroma coordinates) (6.4%). Altogether, pair preference, hue similarity, and low saturation explain 82.4% of the variance in average harmony ratings (multiple-r = +.91). The latter two factors indicate that color pairs that are more harmonious than would be expected from preference ratings were the more similar pairs. Hue difference is clearly a similarity metric, but total saturation is also relevant to color similarity, because pairs of desaturated colors are closer to the center of color space, and all else being equal, closer together in color space than are highly saturated colors of corresponding hues and lightnesses.

Spatial asymmetries in the pair harmony ratings due to figure-ground organization were also examined. The figure-ground asymmetry in lightness found for preference ratings ($r = +.14$) was also present in harmony ratings ($r = +.13$, $p < .001$), in that pairs with lighter figures on darker grounds were rated as more harmonious than pairs with darker figures on lighter grounds. However, the coolness asymmetry that was present in the preference ratings ($r = +.13$) failed to reach statistical significance in the harmony ratings ($r = +.05$, $p > .05$).

Although there is a high correlation between pair preference and pair harmony in the data averaged over all participants ($r = +.79$), the same is not necessarily true for individual participants. Each individual's degree of "preference-for-harmony" was computed as the correlation between his/her preference ratings and his/her harmony ratings over all 992 color pairs. These correlations ranged from a high of 0.75, for the person who most preferred harmonious color combinations, to a low of -0.03, for the person who was most indifferent to harmonious color combinations.[3] A variety of factors that might predict these individual differences in preference-for-harmony were examined. These included Big Five Inventory (or BFI), a 44 item personality inventory that measures five personality factors: Extraversion (talkative, assertive, active, energetic), Agreeableness (sympathetic, kind, appreciative, affectionate), Conscientiousness (organized, thorough, planful, efficient), Neuroticism (tense, anxious, nervous, moody), and Openness (wide interests, imaginative, intelligent, original) (John, Donahue, & Kentle, 1991; see John, Naumann, & Soto, 2008 for more detailed descriptions). Table 2 shows the correlations between participants' scores on each of the BFI factors and their degree of preference for harmony. Although none of the correlations even approached statistical significance, there are trends in which participants who showed a high preference for harmony were less extraverted and more conscientious.

Table 2. Correlations between preference-for-harmony measures and scores on the Big Five Inventory (BFI). Preference-for-harmony is measured for each participant by the correlation between his/her pair preference ratings and pair harmony ratings for all pairwise combinations of the 32 chromatic BCP colors.

BFI factor	Pearson's r	p-value
Extraversion	-0.16	0.27
Agreeableness	-0.01	0.93
Conscientiousness	0.13	0.38
Neuroticism	0.06	0.68
Openness	-0.09	0.55

[3] It may initially seem odd that the highest individual correlation (+.75) is lower than the correlation of the group averages (+.79), but this only indicates that the pattern of deviations across individuals is noisy and tends to cancel out, on average, across individuals.

When all five factors were tested in a linear regression, extraversion (2.7%, less extraverted, higher preference-for-harmony) and conscientiousness (an additional 2.3%, more conscientious, higher preference-for-harmony) were the factors that comprised the best fitting model (multiple-r = .22). Given that our sample size (n=48) is very small for personality research, further data must be collected to determine whether preference-for-harmony is related to these (or other) personality factors.

The only factor that was reliably related to preference-for-harmony was the amount of formal color training that participants reported on a scale in response to the question, "How much formal training have you had in color?" The formal color training scale ranged from 1 (none at all) to 5 (very much), where 3 indicated training to the extent that traditional high school art offers. Figure 12 shows average preference-for-harmony correlations plotted as a function of formal color training.

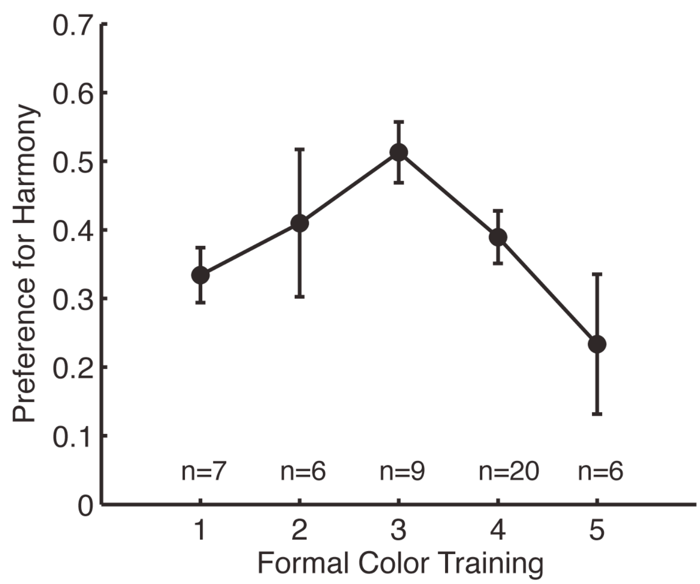

Figure 12. Preference-for-harmony as a function of formal color training. Individual participants' correlations between their own pair preference ratings and pair harmony ratings are plotted as a function of level of formal color training, ranging from 1=low to 5=high. The number of participants in each group is displayed below the corresponding data point. Error bars show the standard errors of the means (SEM).

Somewhat surprisingly, preference-for-harmony was quadratically related to color training ($F(1,47) = 7.58$, $p < .01$). People who reported a moderate amount of formal training in color were most likely to prefer harmonious pairs. It is likely that everyone scoring 3 or above in color training was exposed to the kinds of rules that art theorists have formulated about color harmony and preference (e.g., Chevreul, 1839; Itten, 1961/1973). Thus, they may well have been taught that harmonious combinations are preferable, and this pattern predominates among those with moderate color training. However, our participants who had more formal training, which included professional artists, decorators, and graphic designers, may have discovered through experience how to go beyond those rules in creating effective color combinations even with disharmonious pairs. Finally, those with essentially no formal training may simply have evaluated how much they like the two component colors in the pair, without much regard for the degree of harmony in those combinations.

Figure 13 shows average pair preference ratings for participants with low (n = 7), moderate (n = 9), and advanced (n = 6) degrees of formal color training, respectively. Participants in the low group scored a 1 on the formal training scale, in the moderate group scored a 3 and in the advanced group scored a 5, with those who scored a 2 or 4 being excluded to focus on the clearest cases. To help understand the effects of such training, each group's data was analyzed separately using a regression model that included three factors: that group's average preference rating for the ground color, their average preference for the figural color, and their average harmony ratings (Figure 14). (These three factors were chosen to emphasize the differences in pair preferences between participants with different levels of formal color training, rather than to optimize the total percent of variance explained.)

Preference ratings in the moderate group decreased steadily as the hue similarity between the figure and ground color decreased (Figures 13 C-D). The results of the regression analysis showed that color harmony was indeed the most important factor for this group with only a small amount added by component color preference (see Formal Training group 3).

The preference function for participants with no formal color training appears to be more strongly driven by ground color preference, with pairs containing cool grounds being most preferred, and those containing warm grounds being least preferred (Figures 13 A-B). The same three-factor regression model showed that component color preference component color preferences are relatively more important for those with little or no formal color training than those with moderate color training (compare Formal Training groups 1 and 3 in Figure 14).

Interestingly, participants with advanced color training have generally flatter preference curves (Figure 13E-F), indicating that they are less influenced by hue difference (and harmony) between the figure and ground color than the other two groups. As shown in Figure 14 (Formal Training group 5), only a moderate amount of variance in their preferences can be explained by their harmony ratings, with small amount added from component color preference. Perhaps the group with the most formal color training, who tended to be color professionals of various sorts (painters, designers, decorators, etc.) have had so much experience in working with color combinations that they have become bored with harmonious combinations and have come to appreciate contrastive combinations, in which hue differences are greater. Nevertheless, we see little evidence for Chevreul's (1839) claim that highly contrastive colors are well liked even among this group of people highly trained in color theory.

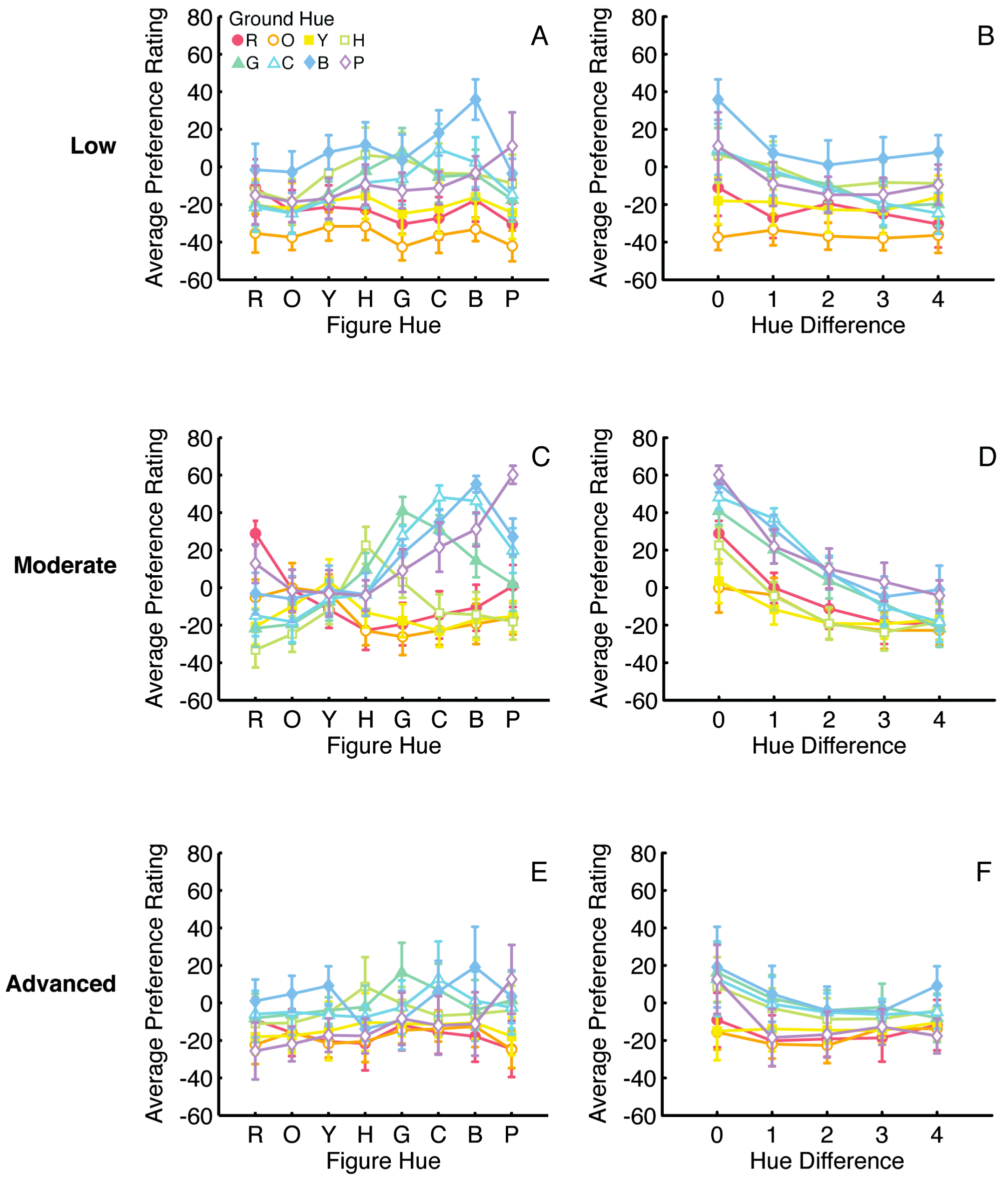

Figure 13. Pair preference for those with (A-B) *low*, (C-D) *moderate,* and (E-F) *advanced* formal color training, as a function of figural hue (x-axis) and ground hue (separate lines) (A, C, E) and as a function of the hue difference (in the present BCP design) between the figure and ground (B, D, F). Error bars represent the standard errors of the means (SEM).

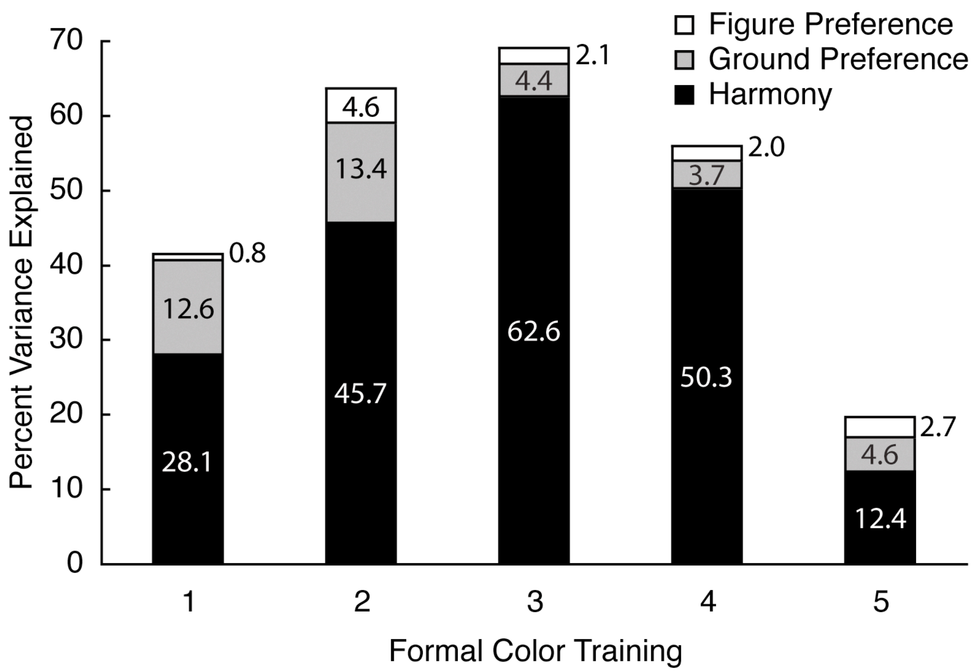

Figure 14. Bars show the percentages of variance explained by the harmony ratings (black stripe), ground color preference ratings (gray stripe) and figural color preference ratings (white stripe) for participants at each level of formal color training (1=none, 5=advanced). The order of the stripes represents the order in which each factor was entered into the regression model (bottom to top).

One question that can be asked about these harmony ratings is whether the instructions we gave produced a "demand characteristic" such that participants inferred that they are "supposed" to give the pattern of data that we observed. There are two noteworthy aspects of our instructions regarding color harmony. One is that they included the musical analogy, which explicitly told participants that their ratings of harmony did not need to conform to their ratings of preference. This analogy certainly does not dictate anything about how an individual "should" rate the harmony of a given color pair because the instructions specifically stated that "some [people] like Mozart and others like Stravinsky," implying that harmony and preference ratings might be either quite similar or quite different. The other noteworthy aspect of the instructions is that they stated that harmonious colors are ones that "go naturally together." Participants might have inferred from this that colors "should be" rated as harmonious to the extent that they are similar. This issue is addressed in Experiment 3, in which we obtain explicit ratings of color similarity and contrast them with ratings of harmony.

2.5. Experiment 3: Color Similarity and its Relation to Preference and Harmony

The results of Experiment 2 provided evidence that color harmony is not only closely related to color preference, but also to color similarity: Harmonious colors are those with smaller hue differences, smaller differences in coolness, and lower total saturation, all of which imply that more harmonious colors are more similar to each other. Two further questions are now addressed. First, how does color harmony differ from color similarity, if at all? Second, which of these two measures of color relations provides better predictions of pair preferences? If color harmony is, in effect, simply another name for color similarity, then similarity ratings should be able to explain as much variance in pair preferences as harmony ratings do. Moreover, there would be no need to consider the somewhat mysterious concept of color harmony if it predicts pair preference no better than the intuitive concept of color similarity. Experiment 3, therefore, measures perceived color similarity of the same color pairs using the same BCP participants who previously made the preference and harmony ratings to examine more closely its relation to pair preference and pair harmony.

2.5.1. Methods

Participants. The participants were the same 48 observers who completed Experiments 1 and 2.

Design and Displays. The design was the same as that of Experiments 1 and 2, but the displays were slightly different. The two colored regions were equal in size (100px x 100px) and positioned side by side, separated by a 20 px gap. We did not use figure-ground displays for the similarity ratings because we wanted our observers to judge how similar the two component colors appeared to them without any spatial asymmetries in the displays (e.g., one color being inside another) or any complications arising from interactions along shared borders. Since all pair-wise combinations of the colors were tested, each pair appeared twice, once when one color appeared on the left and the other on the right, and a second time in the reversed spatial configuration. The left endpoint of the response scale was labeled "different" and the right endpoint was labeled "similar."

Procedure. As in Experiments 1 and 2, participants were presented with each of the 992 chromatic combinations one at a time in a random order. Their task was to rate how similar each pair of colors was on a scale from "different" to "similar." Participants completed this task in a separate session, at least one day after the harmony task had been completed.

2.5.2. Results and Discussion

Average color similarity ratings are plotted in Figure 15A as a function of figural hue and ground hue, averaged over figural cut and ground cut. As is evident by inspection, the hue effects on color similarity ratings are quite similar to the corresponding hue effects on harmony ratings plotted in Figure 7A ($r = +.83$), but even more extreme. They are also somewhat similar to the preference ratings plotted in Figure 2A ($r = +.55$). Color similarity ratings were highly consistent across subjects, with an average correlation of +.75 between each subject's own ratings and the entire group's average ratings. Notice that this consistency measure is substantially greater than the same measure for both the harmony ratings ($r = +.51$, $t(47) = 8.84$, $p < .001$) and the preference ratings ($r = +.36$, $t(47) = 14.39$, $p < .001$).

The similarity data showed main effects of both left color hue ($F(7,329) = 102.58$, $p < .001$) and right color hue ($F(7,329) = 96.22$, $p < .001$), as well as a strong interaction between them ($F(49,2303) = 174.77$, $p < .001$). Like preference and harmony ratings, similarity ratings were highest for each pair when the figure and ground hues were the same and decreased as the hue difference increased. Figure 15B shows the same similarity data re-plotted as a function of the hue difference between the figure and ground colors. As was the case for the preference and harmony ratings in Figures 2B and 7B, perceived similarity decreases monotonically as the hue difference between the two colors increases. The similarity functions do vary systematically over hue, however, with similarity being greater for the cool hues (blues, cyans, and greens) than for the warm colors (yellows, oranges, and reds) ($t(47) = 14.59$, $p < .001$), with purples and chartreuses being generally intermediate.

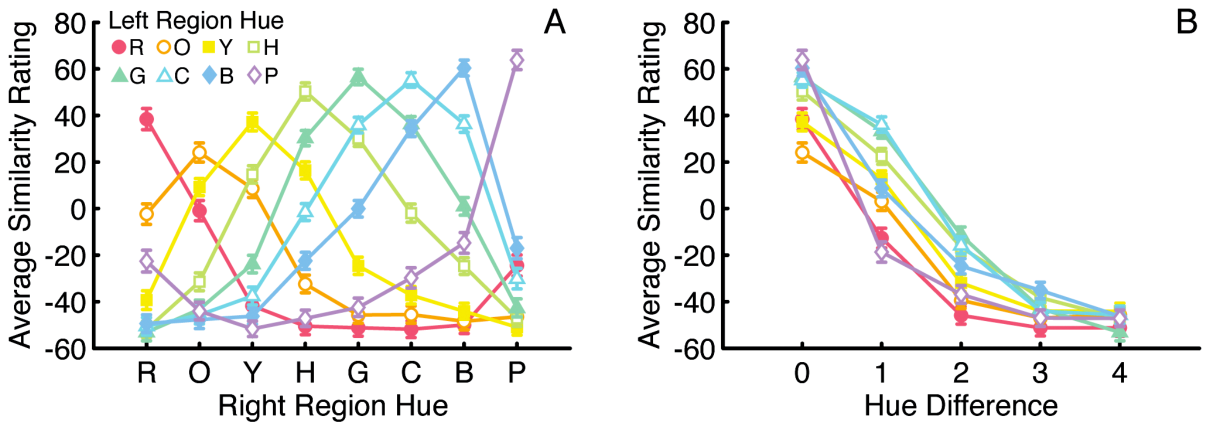

Figure 15. Similarity ratings for color pairs (A) as a function of the hue on the right of the monitor (x-axis) and hue on the left of the monitor (separate lines) and (B) as a function of the hue difference (in terms of steps in the present BCP design) between the right and left colors. Error bars represent the standard errors of the means (SEM).

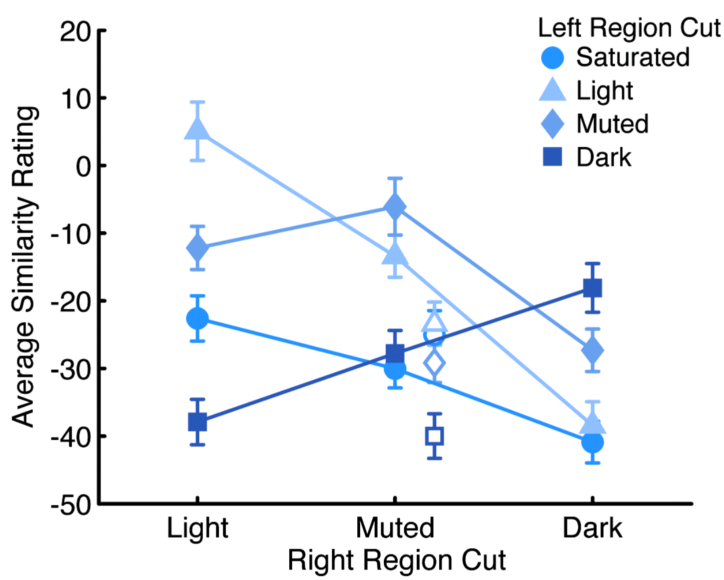

Figure 16. Similarity ratings of color pairs for each left region cut (separate lines), as a function of right region cut (x-axis). Data points for the saturated figure cut (open symbols) are plotted separately at the same x-axis point as the muted colors because they share similar lightness levels but slightly offset for clarity. Error bars represent the standard errors of the means (SEM).

Similarity ratings were also analyzed in terms of cut (saturation/lightness level). As shown in Figure 16, there was a man effect of figure cut ($F(3,141) = 52.13$, $p<.001$) ground cut ($F(3,141) = 66.56$, $p < .001$), and a strong interaction between them ($F(9,423) = 46.40$, $p<.001$). Not surprisingly, pairs containing colors with more similar lightness values were rated as more similar. For example, dark colors were judged more similar to other dark colors than to muted colors ($t(47) = 4.53$, $p < .002$). This pattern of results is different from color harmony ratings (Figure 8), in which colors that generally contained lighter colors were more harmonious (e.g., dark colors were judged more harmonious with muted colors than with other dark colors ($t(47) = 4.26$, $p < .002$).

Figure 17 shows pairwise similarity ratings for left and right region cuts as a function of hue difference between the two colors. Similar to pair preference and harmony ratings, similarity ratings decreased monotonically as the hue difference between the component colors increased (Δ1-4, all cuts: $F(3,141) = 311.08$, $p < .001$; Δ0-4, different cuts: $F(4,188) = 429.59$, $p < .001$), but with even more pronounced reductions than pair preference (Δ1-4, all cuts: $F(3,141) = 175.62$, $p < .001$; Δ0-4, different cuts: $F(4,188) = 199.90$, $p < .001$) and harmony ratings (Δ1-4, all cuts: 56.38, $p < .001$; Δ0-4, different cuts: $F(4,188) = 50.75$ $p < .001$). Further analyses of the interaction between figure and ground cut as a function of hue difference between the two regions can be found in Figure 18.

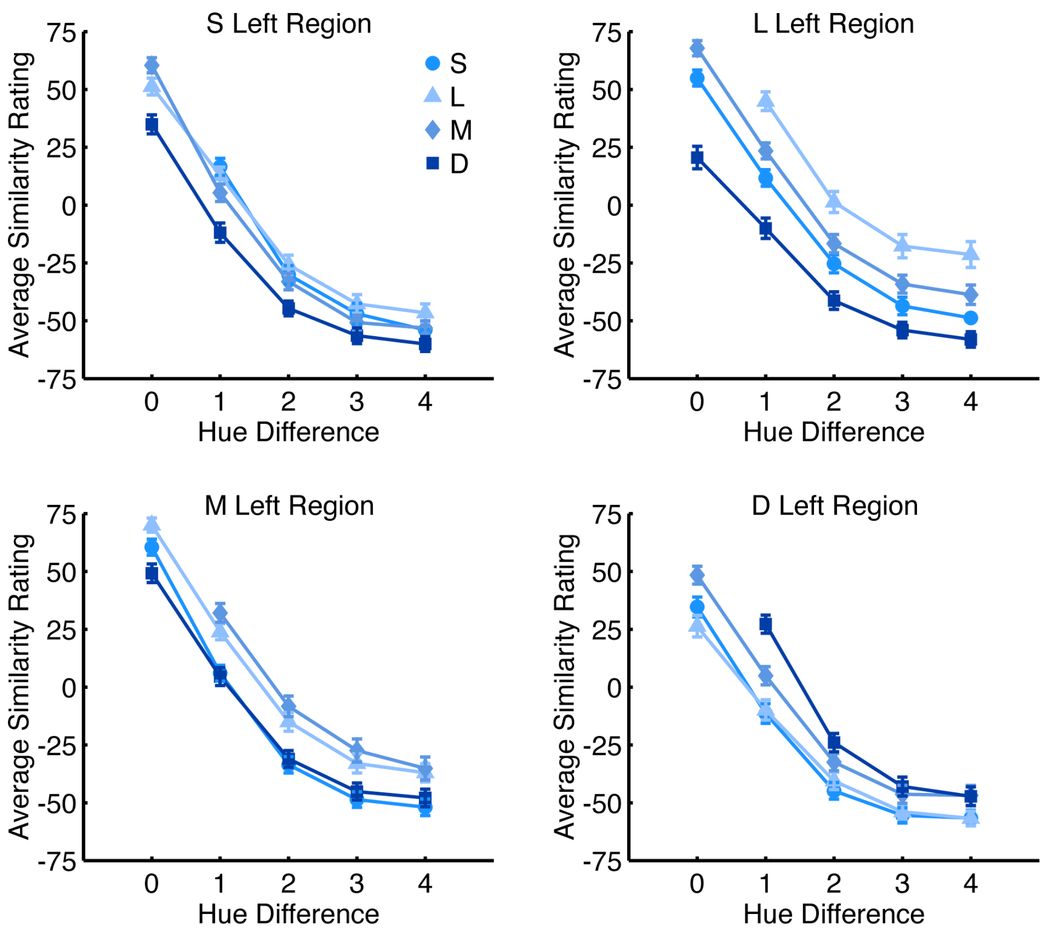

Figure 17. Similarity ratings for the left (separate graphs) and right (separate lines) region cuts, as a function of the hue difference between them.

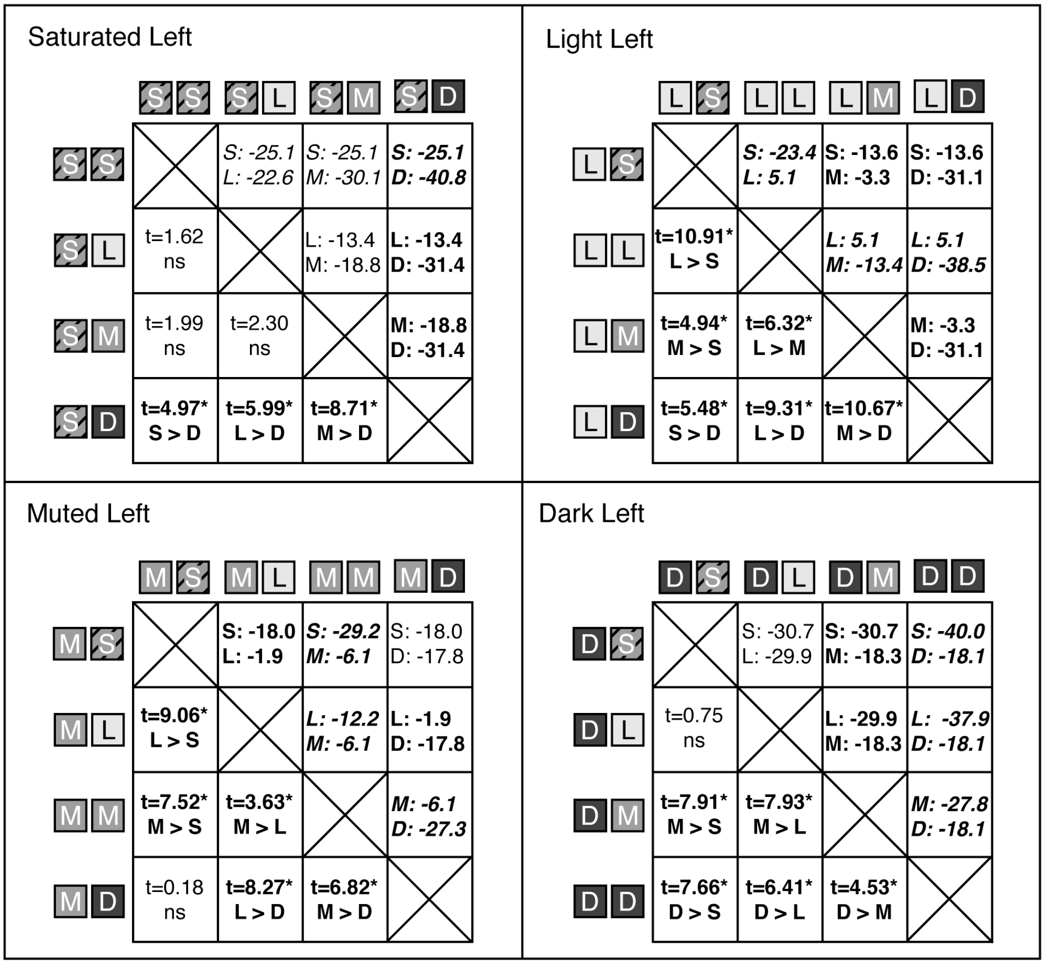

Figure 18. Comparisons between similarity ratings of color combinations with the same left region cut (separate quadrants in the figure) and different right region cuts. Icons adjacent each row and column represent the cuts of the pairs that were judged. The lower triangle of the matrix shows the results of t-tests (df = 47, *p ≤ .002 (using the Bonferroni correction) and direction of the difference (e.g., in the "Saturated Left" quadrant, "S > D" indicates that saturated colors (left) are more similar to saturated colors (right) than to dark colors (right)). The upper triangle of the matrix shows the means of the pairs that were compared. The means for comparisons with same-cut pairs (italicized text) include only hue-difference steps of 1-4 for both pairs. All other means include all hue-differences steps (0-4). Bold face text indicates differences were significant.

When Munsell dimensions were used to predict color similarity ratings for the 992 color pairs, the best model showed that more similar colors were more similar in hue, cooler, more similar in value (lightness), and more similar in coolness, explaining 78% of the variance (see Figure 4). When all 10 factors were included, the full model explained 82.8% of the variance, but there was no clear "best" model among those that included more than four predictors.

As noted previously, color similarity ratings are strongly correlated with harmony ratings (r = +.83). To analyze the differences between them, we looked at the residuals after removing their mutual variation (69.6%) through regression. The only additional predictor entered into the regression equation for harmony was the absolute value of the difference in Munsell value

(+11.0%, with larger lightness differences being more harmonious) for a total of 80.6%, indicating that harmony ratings depended more strongly on lightness contrast (or less strongly on lightness similarity) than did similarity ratings. For similarity ratings, the absolute value of the difference in Munsell value explained an additional 12%, but unlike for harmony, smaller lightness differences were rated as more similar. An additional 7.5% of the variance can be explained by hue difference (the number of Munsell hue steps between the two colors), explaining a total of 89.1% of the variance.

The difference between perceived color similarity and color harmony, therefore, lies primarily in their relation to the lightness contrast of the two colors. Color similarity *decreases* as lightness contrast increases ($r = -.23$, $p < .001$, for the difference between the Munsell values/lightnesses of the two colors), whereas harmony *increases* as lightness contrast increases ($r = +.10$, $p < .01$, for the corresponding difference). This pattern shows that our observers were not judging similarity when making their harmony ratings. If they were, the obtained dissociation between harmony and similarity in the lightness dimension would not be present. It also shows that our observers were not responding to a demand characteristic in which they inferred that harmony was the same as similarity, for their ratings clearly contradict this equivalence in the lightness dimension.

Thus far, it has been established that color similarity is strongly related to, but not the same as, color harmony and that color harmony is strongly related to, but not the same as, preference for color pairs. This raises the important question of whether similarity is more useful in predicting pair preference than pair harmony is. The clear answer is: No. The raw correlation between average pair preference and average pair similarity ($r = +.55$) is substantially lower than the raw correlation between average pair preference and average pair harmony ($r = +.79$). A comparison between these correlations computed separately for each participant shows that the correlations between preference and harmony are reliably higher than those between preference and similarity ($t(47) = 8.24$, $p < .001$). Indeed, if both average harmony ratings and average similarity ratings are included in the predictor variables of a regression analysis, similarity is never entered into the regression equation because it does not explain any additional variance in pair preference. If harmony ratings are not included, the best fitting regression model with similarity ratings accounts for 71.3% of the variance, substantially less than the 80.8% accounted for when harmony ratings are included.

Pair preference, harmony, and similarity are related to each other primarily because all of them increase as the hue similarity between the component colors increases: Color combinations with similar hues are generally more preferred, more harmonious, and more similar to each other. They differ primarily in terms of lightness contrast: Pair preference ratings depend more on lightness contrast than do harmony ratings, and harmony ratings depend more on lightness contrast than do similarity ratings.

2.6. Experiment 4: Preference for Figural Colors on Background Colors

Thus far, the discussion has focused on preference and harmony judgments for color combinations as wholes and have found no evidence favoring art theoretic claims that color combinations with strong hue contrasts are either preferred or harmonious (e.g., Chevreul, 1839). One intriguing possibility is that the art theorists simply confused pair preference and pair harmony with what we are calling figural preference. That is, people may find that figural colors are preferable against contrastingly colored backgrounds even though they do not find such pairs of colors either harmonious or preferred as combinations. This would be consistent with our previous finding that people prefer highly saturated colors to less saturated ones (Palmer & Schloss, 2010), because colors viewed against a background with a strongly contrasting hue are generally perceived as more saturated than when they are viewed against a background with a similar hue (e.g., Lotto & Purves, 2000). Experiment 4 investigates how background color influences observers' preference for the figural color against which it was presented. A rating task, similar to Helson and Lansford's (1970), was used to examine preferences for all 32 figural colors against all 32 background colors in an attempt to determine whether preferences for figural colors seen against different backgrounds vary in systematic ways that might explain art-theoretic claims about the aesthetic virtues of contrastive color combinations (e.g., Chevreul's so-called harmony of contrastive hues).

2.6.1. Methods

Participants. The participants were the same 48 observers who completed Experiments 1-3. They performed the figural color-rating task on a different day that was later than the other three tasks.

Design and Displays. The eight colors from each of the four cuts were placed on each of the 32 background colors to make a total of 128 test displays, each containing all 8 hues from the same cut on a uniform colored background. Each display contained the eight hues arranged to form a square with red in the top left corner, followed by orange, yellow, chartreuse, green, cyan, blue, and purple in a clockwise direction, as illustrated in Figure 1A. Each colored square was 100 x 100 px and was separated from the adjacent squares by 100 px. In displays in which one of the squares was the same color as the background, that square was simply not visible in the display. Below each square was an asterisk, which marked the location of the response text box for each color. When participants typed in a rating, the asterisk below the colored square was replaced by the typed number.

The displays in Experiment 4 (in which all eight colors from a given cut were presented simultaneously on a full-screen background color) were substantially different from the previous three experiments in which pairs were presented one at a time. We chose this configuration because we believed that it helped to emphasize that the task was to judge figural color preference independently of the background color rather than preference for the figure-ground combination as a whole.

Procedure. Each display contained the eight hues from one of the four cuts. Participants were asked to rate how much they liked each figural color on a scale from 1 (lowest) to 9 (highest) using the number keys at the top of the keyboard. They could rate the colors in any order they wished, using the tab key to select which colored square to rate. When a square was selected, the asterisk below it enlarged so that participants knew which square they were currently expected to rate. If they desired, participants could change their ratings by tabbing back to a color square and typing a new rating. In displays that contained a figural color that was identical to the ground there was a zero below the square instead of the asterisk, and that square was skipped when the tab key was pressed.

Participants were told that a given color could look different on different backgrounds, so they need not try to be consistent in their ratings across trials. In addition, they were informed that they could give multiple colors the same rating within a given trial (i.e., if they hated all the colors they could give them all a rating of "1" and if they loved them all they could give them all a rating of "9." Once participants had rated all the colors in a test display, they pressed the "Enter" key to go onto the next display. The 128 displays were presented in a random order and were separated by a 500 ms inter-trial interval.

2.6.2. Results and Discussion

Figure 19A plots the preferences for figural hues on different colored backgrounds as a function of background hue. This pattern is somewhat similar to the pair preferences presented in Figure 2A ($r = +.54$) but is also clearly quite different in that the ground color curves do not peak when the figural color has the same hue, as they do in Figure 2A. When figural preferences for each of the 32 figural colors (averaged over backgrounds) were compared with pair preferences for the same figural colors within figure-ground pairs (also averaged over backgrounds), there was a strong correlation ($r = +.74$), but it was not as strong as preferences for the same 32 figural colors when viewed against a neutral gray background (Palmer & Schloss, 2010) ($r = +.87$). Indeed, when these two correlations are calculated separately for each individual participant and compared statistically, correlations between figural color preference on differently colored backgrounds were reliably more closely related to figural color preferences on a neutral gray background than to pair preferences in which that color is figural ($t(47) = 4.64$, $p < .001$). This finding strongly suggests that the observers in Experiment 4 were indeed rating how much they preferred the figural colors in the present task rather than how much they liked the figure-ground pairs as wholes.

There was a main effect of figural hue ($F(7, 329) = 7.70$, $p < .001$) and ground hue ($F(7,329) = 8.47$), and an interaction between them ($F(49, 2303) = 4.58$, $p < .001$) indicating that figural color preferences are indeed influenced by ground color. As is evident in Figure 19A, figural colors were more preferred on cooler backgrounds ($t(47) = 5.27$, $p < .001$). This was especially true for the warm figural colors (red, orange, and yellow) against the cool backgrounds (blue, cyan, and green) compared with warm figural colors against warm backgrounds ($t(47) = 6.03$, $p < .001$).

A regression model was used to predict preference for figural colors on different colored backgrounds using the same ten Munsell factors as predictors (see Experiments 1-3). The best model (Figure 4) showed that figural colors were more preferred when they contrasted with the background lightness/value, when they and the background were cooler, when they were more saturated and cooler than the background, and when they and the background were more saturated (58.4% of the variance explained). A total of 62.3% was explained when all 10 factors were included in the full model, but there was no clear "best" model containing more than 5 factors.

When single color preferences for the figural color and the ground color (each rated independently by the same observers against a neutral gray background color; see Palmer and Schloss, 2010), pair preferences, pair harmonies, and pair similarities were included in a regression model together with the Munsell factors, a total of 66.0% of the variance in figural preference against colored backgrounds was explained by color preference for the figural color on a gray background (30.3%), pair preference (18.7%), pair similarity (12.0%, larger differences being more preferred), and signed chroma/saturation difference (5.0%, more saturated figures on more desaturated grounds being preferred). The increase in figural color preference as perceived similarity decreases is the first evidence we have obtained that preference of any sort increases as hue contrast increases.

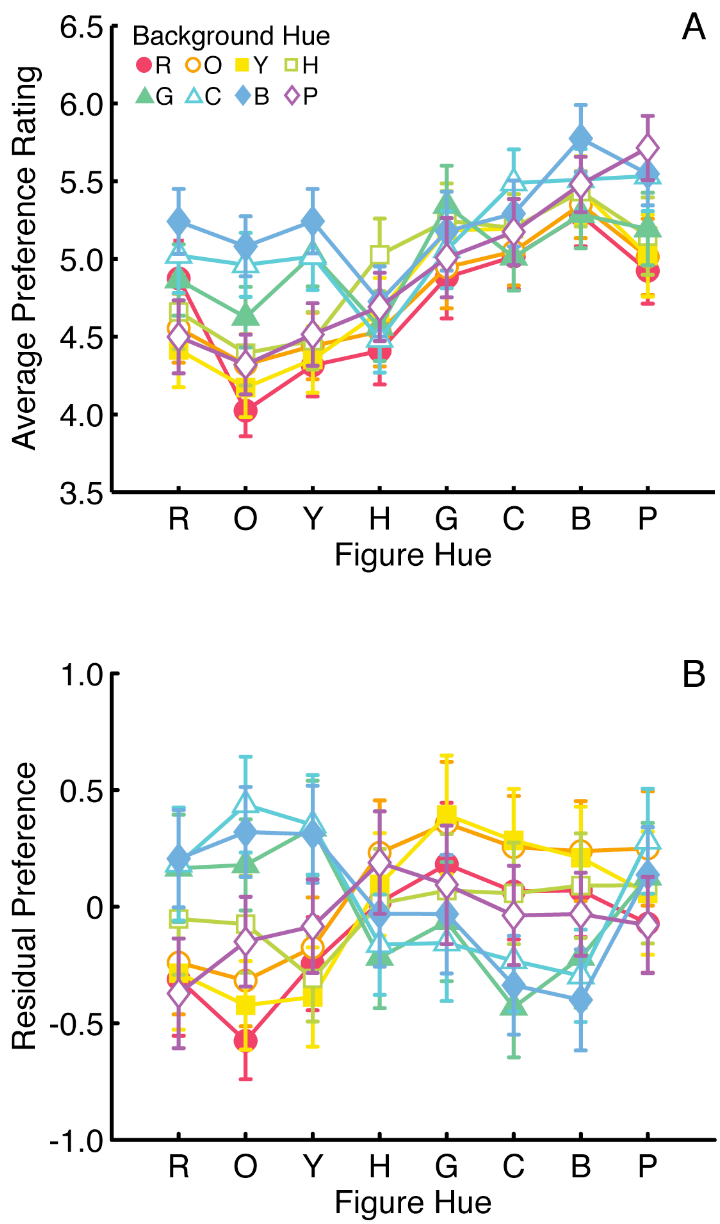

Figure 19. (A) Preference ratings for each figural hue on each of the background hues as a function of figural hue and (B) residual figural color preference after accounting for figural preferences when rated on a neutral gray background (Palmer & Schloss, 2010) and pair preferences plotted as a function of figural hue (B). Error bars represent the standard errors of the means (SEM).

To look more closely at the effects of hue contrast, Figure 19B plots the residual figure preferences after removing the variance due to other sorts of preference: namely, preference for the figural color when viewed against a neutral gray background and preferences for pairs containing the relevant color as figure. There is a clear interaction in the residuals in which warmer hues are preferred on cooler backgrounds and cooler hues are preferred on warmer backgrounds ($F(49,2303) = 7.69$, $p < .001$). This pattern is clearer for the "core" cool hues (green, cyan, and blue) and the "core" warm hues (red, orange, and yellow), than for the "border" hues (chartreuse and purple). Chartreuse followed a similar pattern to the warm hues, but purple peaked over chartreuse, which is the hue that contrasts most with purple.

An analysis of the effects of cuts on figural color preference showed a main effect of figural cut ($F(3, 141) = 8.16$, $p < .001$), ground cut ($F(3, 141) = 8.77$, $p < .001$), and a strong interaction between them ($F(9, 423) = 20.84$, $p < .001$). As shown in Figure 20, saturated figures are generally most preferred, colors on saturated grounds are generally least preferred, light figures are more preferred on dark backgrounds, dark figures are more preferred on light backgrounds, and colors are moderately preferred on muted background.

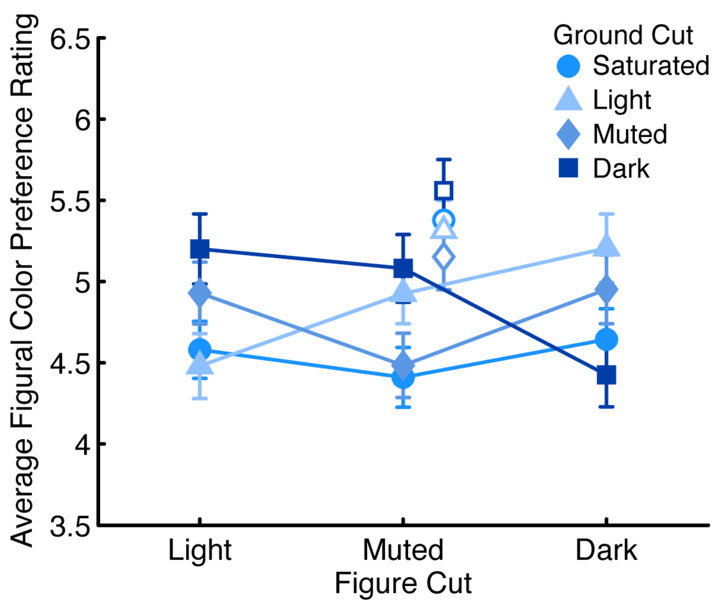

Figure 20. Preference for figural cuts (x-axis) on different background cuts (separate lines). Data points for the saturated figure cut (open symbols) are plotted separately at the same x-axis point as the muted colors because they share similar lightness levels, but they are slightly offset for clarity. Error bars represent the standard errors of the means (SEM).

Supporting statistics of all pairwise comparisons can be found in Figure 21. They provide further evidence for the importance of contrast in figural color preferences. On light (L) grounds, the most contrastive dark (D) and saturated (S) figures are most preferred and the least contrastive L figures are least preferred. Roughly the same is true for the D grounds, for which the most contrastive S and L figures are most preferred and the least contrastive D figures are least preferred. The muted (M) grounds are more contrastive with the S, L, and D figures than with the M figures, and the data follow this pattern as well (although the difference between the

M and D figural colors was not significant after the Bonferroni correction was applied). The only results that appear to contradict this contrast pattern are those for the S grounds, where the S figures are always most preferred even though it seems that these figures should have the lowest contrast with the ground. Notice, however, that highly saturated figures on highly saturated grounds will generally tend to be farther from each other in color space than the corresponding less saturated figures (L, M, and D figures) on highly saturated grounds. Thus, it appears that most of the effects in figural color preferences can be attributed to some form of contrast, which generally enhances preference for the figural color.

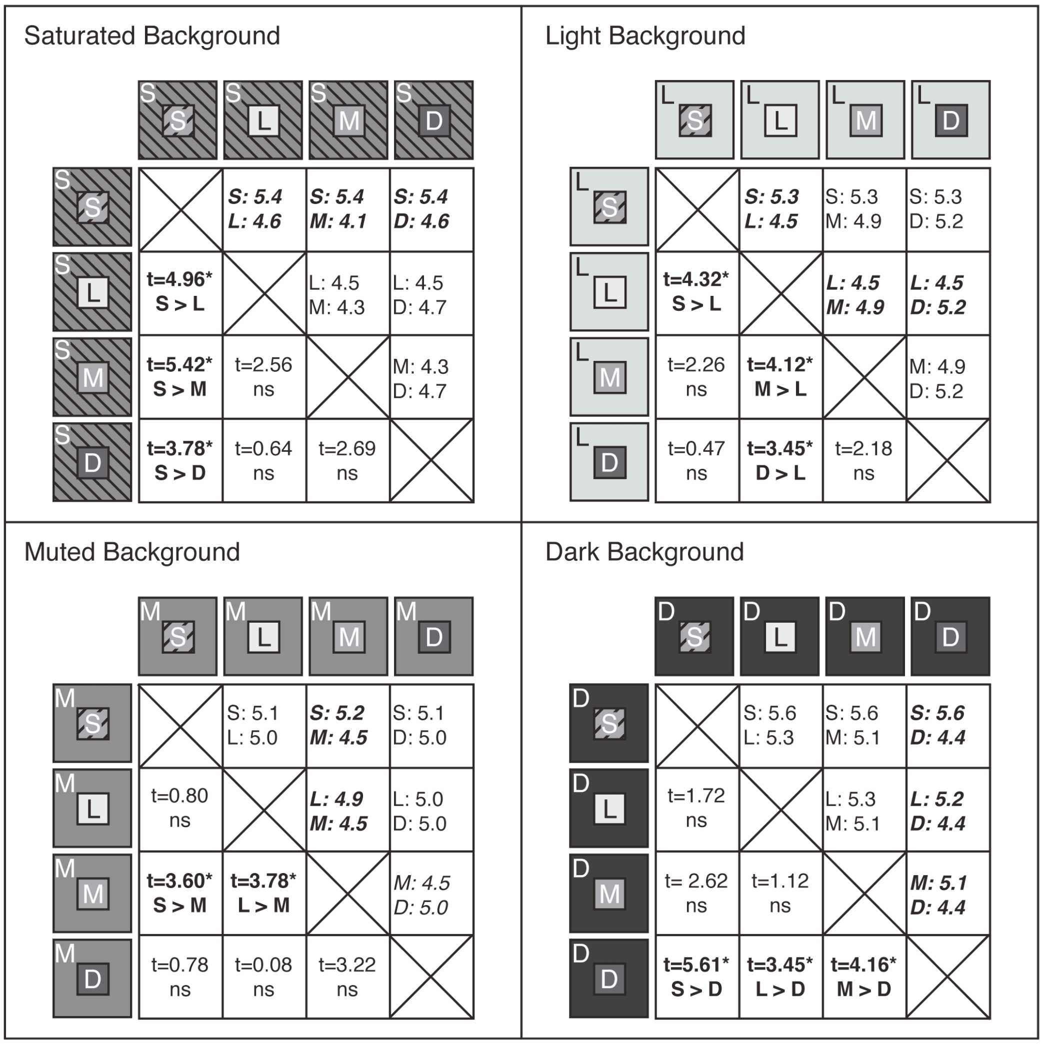

Figure 21. Comparisons between preference ratings for figural colors same background cut (separate quadrants) and different figure cuts. Icons adjacent to each row and column represent the cuts of the colors that were judged. The lower triangle of each background-cut matrix shows the results of t-tests (df = 47, *p ≤ .002 using the Bonferroni correction) and direction of the difference (e.g., "S > L" in the Saturated Ground quadrant indicates that saturated figures on saturated backgrounds were more preferable than light figures on saturated backgrounds). The upper triangle of the matrix shows the means of the pairs that were compared. The means for comparisons with same-cut pairs (italicized text) include only hue-difference steps of 1-4 for both pairs. All other means include all hue-differences steps (0-4). Bold face text indicates differences were significant.

Preference for figural colors, combined across hue and cut, increased as hue difference between the figure and background increased, which is the opposite of the pattern for pair preference, harmony, and similarity (Figure 22). Upon a closer examination, this pattern is primarily limited to color pairs with similar lightness levels, which suggests that hue contrast is more preferable only when there is minimal lightness contrast.

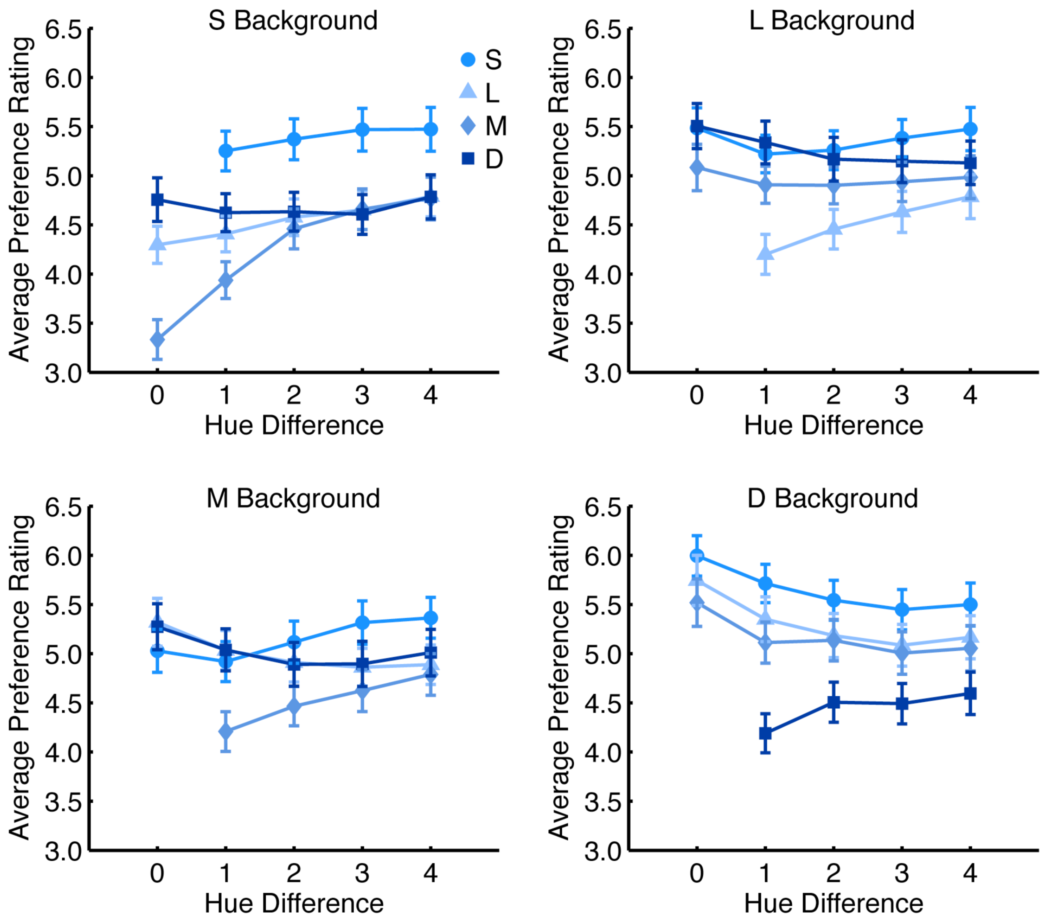

Figure 22. Preference ratings for figural colors (separate lines) on colored backgrounds (separate graphs), as a function of hue difference between the figural and background colors.

There was a 3-way interaction between figural cut, background cut, and hue difference (Δ1-4, all cuts: $F(27, 1269) = 8.68$, $p < .001$; Δ0-4, different cuts: $F(24,1128) = 10.36$, $p < .001$) as well as 2-way interactions between background cut and hue difference (Δ1-4, all cuts: $F(9, 423) = 8.30$, $p < .001$; Δ0-4, different cuts: $F(12,564) = 23.32$, $p < .001$) and figural cut and hue difference (Δ1-4, all cuts: $F(9,423) = 3.75$, $p < .001$; Δ0-4, different cuts: $F(8,376) = 7.41$, $p < .001$). For pairs with similar lightnesses – the four same-cut pairs (S-S, L-L, M-M and D-D) and the two S-M pairs – figural color preferences actually increased as hue difference increased (Δ1-4: $F(3,141) = 22.32$, $p < .001$). Pairs containing S figures on L backgrounds and L figures on S backgrounds also showed this pattern more weakly (Δ1-4: $F(3,141) = 7.20$ $p <.001$), even though L colors were lighter than S colors.

The results of this experiment are roughly consistent with art theorists' claim that hue contrast enhances people's preference for colors in combinations that contain at least certain kinds of hue contrast (e.g., Chevreul, 1839; Munsell, 1921/1969). The main problem with the art theoretic claims is that it is misattributed to increased harmony. In fact, people do not like strong hue contrasts because such combinations are *harmonious*; they like colors against strongly contrastive backgrounds because they make the figural color itself look "better" (more preferred) than it does against a weakly contrastive background. This argument is consistent with the fact that people generally prefer saturated colors over the other three less-saturated cuts when rated on a (zero saturation) neutral gray background (Palmer & Schloss, 2010): Saturated colors are more contrastive than other colors against medium gray. Reasons for why people might prefer colors against strongly contrastive backgrounds will be discussed in the General Discussion section where we address the general question of possible causes of the effects reported in this article.

Thus, it appears that virtually all of the residual effects in these figural color preferences, after variations due to single and pair preferences have been removed, can be attributed to some form of contrast, all of which generally enhance preference for the figural color. In summary, the results show that figural color preference increases as hue similarity decreases, which is opposite the pattern for pair preference ratings, harmony ratings, and similarity ratings obtained in Experiments 1-3, respectively. They also show that pairs are most preferred on backgrounds of contrasting lightness.

Results thus generally support Helson and Lansford's (1970) claim that contrast is a highly influential factor on how much people like figural colors (which they call "object colors") against a background color. They propose that the reason contrast improves figural color preference could be ease of perception on a contrasting background. This fits with the idea that preference in general is related to perceptual "fluency:" the hypothesis that people aesthetically prefer displays that are easier to perceive (e.g., Reber, Schwarz, & Winkielman, 2004).

Finally, we performed a regression analysis to predict pair preference (Experiment 1) from figural color preference (Experiment 4), as well as pair harmony ratings (Experiment 2), similarity ratings (Experiment 3) and Munsell factors. The best-fitting model, which explained 82.6% of the variance in pair preferences, included harmony (62.3%), figural color preference when rated on the correspondingly colored background (12.3%), and ground color preference on a neutral gray background (+8%). This amount is only slightly more than the model from Experiment 2 (80.8%) that included figural color preference on a neutral gray background and lightness contrast, both of which are encapsulated by figural color preference on different colored backgrounds. Nevertheless, this model, which accounts for the most variance with the fewest variables, supports the hypothesis that contextual preference for the figural color (i.e., figural preference on a colored background) has an effect on pair preferences, even though it does not have an effect on pair harmony.

2.7. General Discussion of Experiments 1-4

The results of the preceding four experiments have demonstrated that there are distinct differences among three kinds of perceptual judgments of two-color figure-ground combinations: preference for the color pair, harmony of the color pair, and preference for figural colors against colored backgrounds. Both pair preference and pair harmony vary primarily as a function of hue similarity, such that pairs with similar hues are, on average, both more preferred and more harmonious. Consistent with color theories in art (e.g., Chevreul's (1839) "harmony of analogous colors"), ratings of color preference and harmony were highest for colors most similar in hue. Inconsistent with such theories (e.g., Chevreul's "harmony of contrastive colors"), however, no overall increase was observed in ratings of preference or harmony for complementary hues.

Although preference and harmony are closely related to one another, preferred pairs differ from harmonious pairs in including preference for the component colors and a large lightness contrast component, whereas harmonious pairs are more similar in hue and lower in saturation. Harmony and similarity ratings are also closely related to one another, but harmony ratings do not have the lightness similarity component that similarity ratings have.

Finally, figural color preferences against different background colors are closely related to preference for the same figural colors when rated on a neutral gray background and preference for the combination of the figural color and background color. Once those factors are accounted for, however, clear effects of both hue contrast and lightness contrast are revealed: Warmer figures are preferred on cooler backgrounds, cooler figures are preferred on warmer backgrounds, and figures are generally preferred on backgrounds of contrasting lightness. These results show that Chevreul's so-called "harmony of contrast," at least in the hue dimension, actually applies to preferences for figural colors on different colored backgrounds rather than to pair preferences or pair harmonies.

The present experiments were aimed primarily at establishing the nature of aesthetic preferences for color pairs and their relations to harmony, similarity, and figural preference of color pairs. From these data, we can infer little about the actual causes of pair preferences. Still, we can speculate about causes with varying degrees of confidence for several key aspects of our findings. The primary factors that influence pair preferences appear to be preferences for single colors (of the individual figural color and/or ground color), color harmony of figure and ground, lightness contrast between figure and ground, and figural preference against a colored ground. Before closing, we will consider in turn what factors might underlie each of these factors.

The data from Experiment 1 clearly show that people's preferences for color pairs reliably depend on their preferences for the individual colors of which they are composed (e.g., see Figure 3). Palmer and Schloss (2010) have reported results that strongly support an ecological valence theory (EVT) of single color preferences, positing that people like colors to the degree that they like correspondingly colored objects. For example, people generally like saturated blues and cyans because they like clear sky, clean water, swimming pools, and most other objects that characteristically are these colors. They generally dislike dark oranges (browns) and dark yellows (olive-colors) because they dislike feces, rotting food, vomit, and many other (but not all other – consider chocolate and coffee) objects they associate with these colors. Because one cannot make scientific generalizations about such observations based on just a few examples of desirable and undesirable colored objects, Palmer and Schloss devised a systematic procedure to test their theory.

To obtain comprehensive lists of color-object associations, one group of participants provided verbal descriptions of all the objects they associated with each of the 32 BCP colors in a fixed time period. Another group then rated their affective valence for each verbally described object (i.e., how positive/negative they felt about "clear sky," "feces," etc.). A third group rated how well the colors of each verbally described object matched the BCP color(s) that had elicited it. The affective valence ratings for each described object were weighted (multiplied) by the relevant color-match ratings (higher match ratings produced higher weights) and then averaged for each of the 32 BCP colors to produce the weighted affective valence estimate (WAVE) for each color. The WAVE for a given color, therefore, was calculated as the average weighted valences of *all* objects associated with that color, which could range from very positive to very negative. For example, object associates for brown (BCP dark orange) included "chocolate," which was very positive, "feces," which was very negative, and a large number of other objects with intermediate valences, all of which averaged together gave a net negative WAVE for this color). Using this procedure for all 32 chromatic colors, Palmer and Schloss (2010) found a strong correlation between the WAVEs of the BCP 32 colors and people's average preference ratings for the same 32 colors ($r = +.89$). This result shows that preference for a given color increases as the average weighted valence of all of the objects associated with that color increases.

Because the same single color preferences appear in the regression models for the present pair preferences (see also Figure 3A for pair preference ratings averaged over ground hue and figure hue versus Figure 3B for single color preferences), we assume that this component of the data from Experiment 1 is influenced by the same ecological valences. Moreover, to the extent that certain color combinations are characteristic of entities with strong valences (e.g., red and green with Christmas, blue and yellow with a bright sun against a clear sky, and dark purple and dark green with bruised flesh), the same associative ecological valence principles suggest that color pairs may be more (or less) preferred than would otherwise be expected from the kind of colorimetric relations we have identified in this article (e.g., hue similarity and lightness contrast), depending on the valence of their ecological associations. Of course, one cannot simply point to cherry-picked examples of objects that are associated with color pairs to test for ecological effects on pair preferences. A comprehensive analysis of *all* objects (positive, negative, and everything in between) associated with each color pair would be necessary to test whether the average valence of objects associated with a given color pair is related to preference for that same pair.

The lion's share of the variance in pair preferences, however, is clearly due to abstract color relations: People prefer color pairs that have the same hue but differ in lightness and/or saturation. Our measurements suggest that the best single relational variable in predicting pair preferences is perceived pair harmony, because average pair harmony ratings appear in all of the best-fitting regression models of average pair preference, accounting for 62% of the variance. What, then, might be the cause of the perception of color harmony? Our instructions for rating harmony (aside from the musical analogy) asked observers to report "how well the colors go together," and we presume that this is what they judged, to the best of their ability. Our current conjecture is that color harmony derives from the ecological co-occurrence statistics of color pairs within uniform connected (UC) regions of natural images. Palmer and Rock (1994) defined UC regions as connected areas within an image that are (relatively) homogeneous in terms of many variables, including those related to color. We speculate that the color pairs judged to be most harmonious are those that are most likely to co-occur within UC regions. We are testing this hypothesis by examining ecological statistics in the Berkeley Segmentation Dataset (see Martin et al., 2001; http://www.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/), which contains 200 images that were hand-parsed into regions by human observers. Preliminary results from analyses of the relations among within-region colors suggest that the primary chromatic attribute defining a UC region is hue similarity. This means that pairs of pixels that have the same hue (or very similar hues) are most likely to co-occur within UC regions and that within-region variations in lightness and/or saturation are greater than variations in hue.

A third factor that clearly contributes to pair preference in most of the best-fitting regression models is lightness contrast. Harmony ratings do not depend strongly on lightness contrast, but pair preferences do, with more contrastive pairs being preferred. Why might this occur? One possible explanation comes from the fluency theory of aesthetic preference (e.g., Reber et al., 2004). The basic premise of fluency theory is that people prefer things that are easy to process perceptually. Lightness contrast is one of the primary factors that supports this theory: People prefer images in which the contrast between figure and ground regions is high. Fluency theory frames the relevance of lightness contrast to pair preference in terms of high-contrast figure-ground images being aesthetically pleasing to perceive, but one could also frame the same phenomenon in the opposite terms: Perhaps low-contrast figure-ground images are aesthetically displeasing. This description suggests a possibly different causal account in which isoluminance plays a dominant role: Perhaps people dislike low contrast figure-ground displays as the colors approach isoluminance, making the boundaries between them difficult to discriminate and having perceptually disturbing effects (e.g., Gregory, 1977). We are currently investigating these possibilities, both of which may contain some truth.

Finally, the results of Experiment 4 suggest that hue contrast increases preference for a figural color against a colored background. This effect may be caused by simultaneous color contrast (also known as induced color). The background (or surround) induces a hue shift in the figural color that is complementary to the background color (e.g., da Vinci, 1492; Chevreul, 1839; Helmholtz, 1866/1925; Walraven, 1976; Shevell, 1978). This means that a gray figure on a blue background should appear somewhat yellowish (because yellow is the complement of blue), a yellow figure on a blue background should appear extra yellow (because the yellowness induced by the blue background increases the saturation of the yellow figure), and a blue figure on blue background should appear somewhat grayish (because the yellowness induced by the blue background partly cancels the blueness of the figure). If people generally like more saturated figural colors, as they apparently do (Palmer & Schloss, 2010), and if a contrasting background enhances the saturation of the figural color, then figural colors should be more preferred on backgrounds with strongly contrastive hues. The key question is whether these hue contrast effects will be eliminated if observers first adjust each color on each colored background to look identical to that same color on a neutral gray background. If all of the figural preference effects found in Experiment 4 were to disappear with the appearance-matched figural colors, then simultaneous color contrast is surely their cause. We are currently investigating this possibility.

One concern about the generalizability of the current results is the degree to which preferences for color pairs in concentric-square, figure-ground displays will generalize to preferences for color pairs displayed in other spatial configurations. Preliminary data for color pairs displayed side-by-side with a gap between them suggest that pair preferences still generally increase as hue similarity between the component colors increases. Naturally, certain kinds of spatial factors that produce semantic interpretations of the colored regions could produce fairly pronounced effects on preferences, such as making an orange region carrot-shaped and a green region above it carrot-top-shaped. Future work comparing, some of which is in Chapter 3 below, examines the influence of different geometric arrays is underway to test how the principles established in this chapter apply to color combinations in different spatial arrangements.

The second issue concerns how the present findings from two-color combinations might generalize to combinations of three-colors, four-colors, and beyond. The present data show that preferences for single component colors only weakly predict preference for color pairs, with the lion's share of the variance attributable to pairwise color relations (e.g., harmony). Might the same problem arise when expanding the domain to three-color combinations: i.e., might single and pairwise preferences account for little of the variance, with the lion's share now arising from three-way relations? Preliminary results on preference for color triples, however, suggest that preferences for all possible pairs within triples of colors predict much of the variance within preference for triples as a whole. We speculate, therefore, that once pairwise color preferences are known and understood, enough relational information is available to account for preferences in higher-order combinations.

At the outset of this chapter it was proposed that much of the confusion in the literature on the aesthetics of color combinations was due to confusion among three distinct types of judgments: pair preference, pair harmony, and figural preference on different colored backgrounds. Strong empirical evidence has been shown that these three types of judgments are indeed different, in that they produce systematically different patterns of results. It has also been argued that these results and analyses clarify many of the confusions that have accumulated over the past century. Moreover, it is expected that the new understanding achieved by making clear distinctions among these and related aspects of perceptual response will allow researchers to move beyond the foundational problems of how to define and measure preference and harmony properly to more advanced questions, such as *why* people prefer the combinations they do, both as individuals and as a group, and how color preferences might be influenced by the context and/or intended message of a visual display.

3. The Role of Spatial Organization in Preference for Color Pairs

When combining two or more colors in a visual display there are two primary factors to consider: which colors to use and how to arrange them spatially. In Chapter 2 we analyzed the first of these factors – *which* colors people prefer – and found that, on average, they prefer pairs with cooler colors that are similar in hue, contrasting in lightness, and contain preferred individual colors. In Chapter 3, we investigate how the spatial organization of the two component colors influences people's aesthetic preferences for pairs of colors. Granger (1952) went so far as to suggest that choices of the relative sizes of colored areas may be more important than the choices of colors themselves.

3.1. Introduction to Spatial Aspects of Color Pair Preference

3.1.1. Color-Pair Preference Asymmetries

To address the influence of spatial organization on preferences for color pairs, we tested for the existence of *color-pair preference asymmetries* that would occur if people systematically prefer a given color pair in one spatial configuration over the same colors in the same configuration, but with their spatial roles reversed.

3.1.2. Previous Research on Spatial Aspects of Preference for Color Pairs

In Chapter 2 we presented evidence of preference asymmetries for color pairs organized in a figure-ground arrangement consisting of a small square centered on a larger square. People preferred pairs with warmer figures on cooler grounds and lighter figures on darker grounds. More specifically, preference ratings for each of the 992 pair-wise combinations of the Berkeley Color Project (BCP) 32 chromatic colors were weakly, though significantly, correlated with the signed difference in coolness ($r = .13$) and lightness ($r = -.14$) between the ground and figure colors, with higher preference ratings for warmer, lighter figures on cooler, darker grounds.

The preference asymmetries described in Chapter 2 have left numerous questions unanswered. First, it is unclear whether the correlations were weak because preference asymmetries are simply marginal effects or because the rating task used was not sensitive enough. To address this issue, we measured preferences for the same pairs, in a forced choice task in which the only difference between the two pairs in the comparison was the figure-ground assignment of the colors (Experiment 5). Second, it is unclear which spatial aspects of the figure-ground pairs govern the preference asymmetries. The figural region differs from the ground in multiple ways. Not only is it more 'figural' or object-like, in terms of figure-ground organization (Rubin, 1921/1958), but it is also smaller than the ground and surrounded by the ground, both of which contribute to its figural status. In Experiments 6-7 we therefore isolated which spatial factors are govern these preference asymmetries.

Bullough (1907) tested for such preference asymmetries based on lightness by presenting participants with two color pairs that only differed in their vertical arrangement. He found that participants preferred pairs in which the lower region was darker, which he attributed to color weight, as if people like darker, heavier regions to be lower because they provide more gravitational stability to the image.[4]

Other previous research on spatial aspects of color combinations focused on how balance can be achieved by adjusting the relative area among colored regions. According to Munsell's (1921/1969) principle of inverse ratios of area, color combinations are balanced or harmonious when "stronger" colors occupy less space than "weaker" colors. Accordingly, balance is achieved when the area times value (lightness) times chroma (saturation) of the component

[4] Although Bullough (1907) was adamantly opposed to statistical analysis of aesthetic judgments and reported only qualitative assessments of his data, a quantitative analysis of his data shows that the effects he reported were statistically significant. However, Bullough (1907) had an obvious confound in his test displays, because the color pairs that were consistent with his theory of color weight were always labeled "a" and those that were inconsistent were always labeled "b." A response bias to report "a" would thus have spuriously led to data that he interpreted as supporting his hypothesis.

colors is equivalent. In addition to this relation in area, value, and chroma, Munsell strongly recommend using complementary hues in his space to achieve utmost balance in the sense that they would mix to produce neutral gray on a Maxwell disk. In Birren's contribution to Munsell's (1921/1969) *A Grammar of Color*, he recommended that warmer colors should have higher chroma (and smaller area) than cooler colors within a pair. Birren's proposal is consistent with our previous finding that people preferred warmer figures (small squares) on cooler grounds (large squares) (Schloss & Palmer, 2011).

Moon and Spencer (1944c) proposed a formula similar to Munsell's, but claimed that colors were balanced when the product of the area times distance from the adaptation point products were equal. Their formula accounts for contrast with the background (or adaptation due to the color patches if they are sufficiently large), whereas Munsell's formula is agnostic with respect to the background color. Moon and Spencer's (1944c) and to Munsell's (1921/1969) formulas are equivalent when the Munsell values of the colors are both 5, but otherwise the two formulas produce different area ratios for the same set of colors.

Several studies have investigated which formula, Munsell's or Moon and Spencer's, was more valid empirically. Granger (1953) tested which formula better predicted participants preference for color pairs by having them adjust the relative area of the colored regions to produce the most "pleasing balance." He found that Munsell's rule predicted participants' preferences better than Moon and Spencer's formula. Morriss and colleagues tested the validity of Munsell's and Moon and Spencer's formulas by asking participants to adjust the relative area of two adjacent colored regions until they appeared "balanced" (Morriss, Dunlap, & Hammond, 1982; Morriss & Dunlap, 1987; Morriss & Dunlap, 1988; Linnett, Morriss, Dunlap, & Fritchie, 1991). In all of their studies they asked participants to ignore their color preferences when responding. Consistent with both Munsell's and Moon and Spencer's formulas, participants set the more saturated regions to be smaller when value was held roughly constant (Morriss, Dunlap, & Hammond, 1982; Linnett, Morriss, Dunlap, & Fritchie, 1991). Participants continued to set the higher chroma region as smaller, regardless of the hue difference between the component colors and background lightness (Morriss & Dunlap, 1988). When chroma was held constant and lightness varied, participants set regions with higher lightness contrast with the background to be smaller, which was consistent with Moon and Spencer's (1944c) formula.

In Itten's (1961/1973) discussion of "contrast of extension," defined as the relative area of color patches, he proposed that colors should be combined in a ratio that is reciprocal to their "brilliances" or "intensities." This rule is similar to Munsell's (1921/1969), if one assumes the traditional use of the word "intensity," but Itten was actually referring to Goethe's (1810/2006) order of hue-based "light values," where yellow (9) is most light, followed by orange (8), green (6) and red (6), blue (4), and violet (3). For Itten, if yellow is a "9" in light value and violet is a "3," they should be combined in a ratio of 1:3. Interestingly, Goethe's order of intensities follows the order of Munsell value (lightness) of the most saturated colors in each of the eight named hues in Munsell's color space, and Itten uses highly saturated colors to illustrate his principle. Goethe's order of light values also follows a rough ordering from warmness to coolness.

3.2. Aim of Experiments 5-7: Understanding Color-Pair Preference Asymmetries

In this chapter we analyze preference asymmetries in light of the aforementioned theories. Our analyses do not directly test their formulas because our particular aim is to predict the strength of preference asymmetries in color pairs of fixed ratios rather than the relative area between two colored regions. Still, we use these rules to guide our understanding of the nature of preference asymmetries. We will also measure preference asymmetries in different spatial configurations to determine which spatial factors are most important in contributing to these effects, independent of relative areas.

3.3. Experiment 5: Asymmetries in Preference for Color Pairs

In Experiment 5 we tested for possible asymmetries in preference for figure-ground pairs consisting of a small square centered within a larger square. Preferences would be asymmetric if, for a given pair of colors, observers reliably preferred one color as figure and the other as ground to the reversed figure-ground arrangement.

3.3.1. Methods

Participants. Participants were the same as those from Experiments 1-4.

Design and Displays. As in Experiments 1-4, all pairwise combinations of the 32 chromatic colors from the Berkeley Color Project were used (see Palmer and Schloss, 2010) to generate 992 color pairs.

Each display contained two color pairs, one on the left and one on the right of the monitor. Each pair consisted of a small square figure (100 px x 100 px) centered on a large square ground (300px x 300 px), analogous to the pairs tested Experiments 1 and 2. The two pairs within a trial contained the same two colors but were reversed in figure-round arrangement: i.e., if the left pair contained a figure of Color A and a ground of Color B (denoted pair AB), the right pair had a figure of Color B on a ground of Color A (denoted pair BA). There were 992 trials so that each pair of color-pair displays appeared twice with the spatial positions of the two displays reversed.

Procedure. Participants were instructed to indicate which pair they liked better by pressing the left arrow key if they liked the left pair better, the right arrow key if they liked the right pair better, and the down arrow key if they liked both pairs equally. Displays remained on the screen until participants responded, and trials were separated by a 500ms inter-trial interval.

3.3.2. Results and Discussion

Preferences were considered "asymmetric" if participants preferred color pair AB to color pair BA (or pair BA to pair AB) when the only difference between pair AB and BA was the figure-ground assignment of the component colors. Figure 23 demonstrates the presence of preference asymmetries of hue in figure-ground pairs, averaged over saturation and lightness levels. Each subplot shows the data for all trials containing the hue indicated above it. The square data points represent the proportion of trials on which observers preferred the pairs containing the titled hue as ground color, circular data points represent the proportion of trials on which they preferred the pairs containing the titled hue as figure color, and the gray diamond data points represent the proportion of trials on which observers preferred neither pair. The three points located at each x-axis value within a subplot thus necessarily sum to 1. When the hue of the figure and ground are the same in these plots, the data points are solid to indicate this fact, they differ only in the lightness and/or saturation levels. Because the data in Figure 23 are averaged over lightness and/or saturation levels, there is no distinction between these two cases, making their choice probabilities necessarily the same. However, the proportion of trials on which they indicated lack of preference (gray diamonds) is free to vary. Participants were more likely to choose neither pair when the two colors within the pairs were more similar to one another, as indicated by a reliable positive correlation ($r=+.40$, $p<.001$) between the proportion of times participants chose neither pair on each of the 992 and the same participants' similarity judgments of the two colors within each pair, as reported in Experiment 3. This result suggests that participants simply responded "neither" when the choice was difficult. For this reason, all participants in the subsequent experiments were not given the "neither" option.

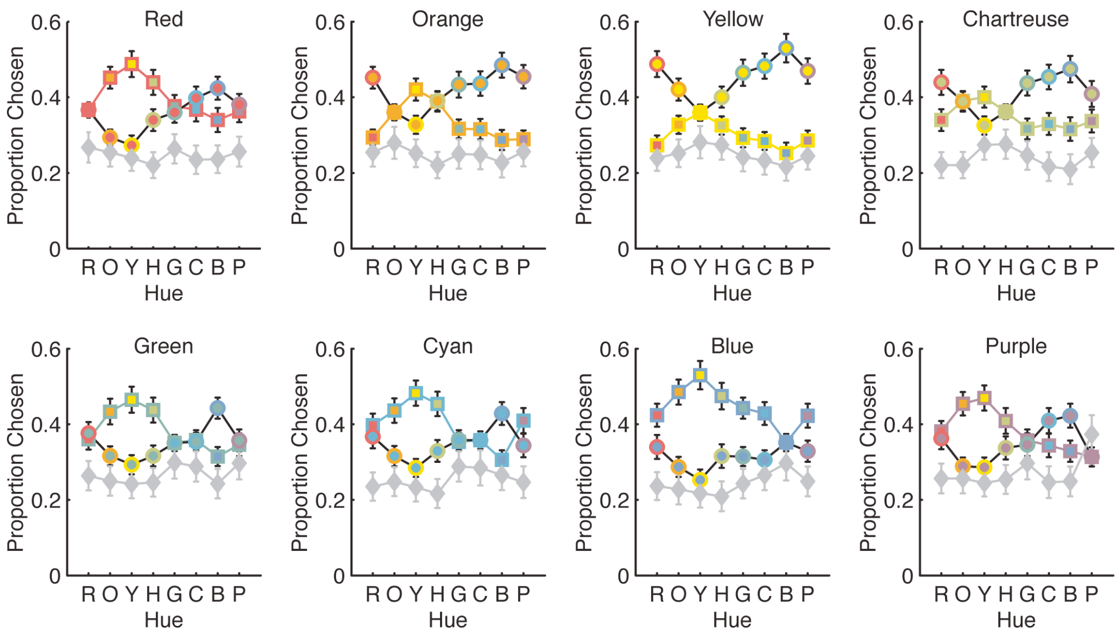

Figure 23. Comparisons between the proportions of times each hue pair was chosen when presented in one figure-ground arrangement relative to the reversed figure-ground arrangement. Each hue sub-plot (indicated by subplot title) compares preference for pairs containing that hue as figure (circles) vs. that hue as ground (squares) when paired with each of the other hues (x-axis). Gray diamonds show the proportion of times neither pair was chosen. Error bars represent the standard errors of the means (SEM).

Close inspection of Figure 23 indicates a striking regularity in the results: participants showed preference asymmetries to the degree that the ground was more blue and the figure was more yellow.[5] This pattern is most apparent in the *Yellow* and *Blue* subplots, where participants always preferred pairs containing yellow to have yellow in the figural region (see *Yellow* subplot) and pairs containing blue to have blue in the ground region (see *Blue* subplot). The same pattern holds for pairs that do not contain yellow or blue specifically, but still differ in the yellowness-blueness dimension, such as orange and cyan (orange figure on cyan ground preferred), and chartreuse and red (chartreuse figure on red ground preferred).

[5]Mexican participants, tested at the University of Guadalajara, show similar yellowness-blueness preference asymmetries, which discounts the possibility that the effects reported here are simply due to Berkeley students' higher preference for Berkeley's primary color pair (gold-on-blue) to its secondary color pair (blue-on-gold) (Schloss, Poggesi, & Palmer, in press)).

To evaluate this blue-yellow asymmetry more quantitatively, Figure 24A plots the proportion of trials on which each of the 992 pairs was chosen as a function of the figure-ground difference in the yellowness-blueness of the colors. The largest values on the x-axis thus corresponding to yellow figures on a blue background and the smallest values to blue figures on a yellow background. The amount of yellowness-blueness, redness-greenness, warmness-coolness, lightness-darkness, and saturation for each color was determined by the same participants' ratings in a previous session (see Palmer & Schloss, 2010), and were normed to range from -1 to 1.

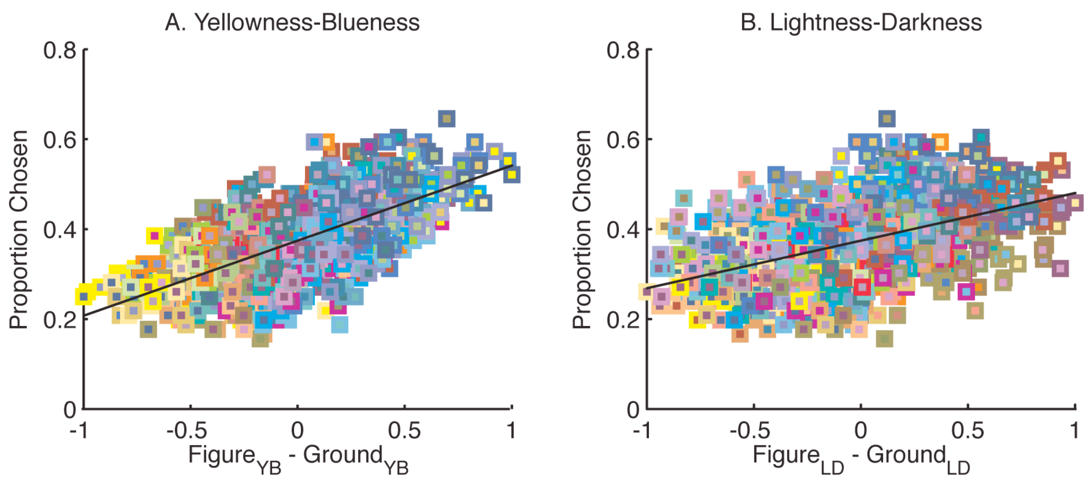

Figure 24. Comparisons between the proportions of times each pair was chosen (i.e., preference) as a function of the difference in (A) yellowness-blueness and (B) lightness-darkness difference between the figure and ground colors in the pair. The colors of the data points represent the colors of the pairs that were judged. The solid black line is the best fitting line as determined by a linear regression equation.

As Figure 24A indicates, there is a clear linear increase in the probability of choosing a given pair to a degree that its figural color is yellower and its ground color is bluer ($r = +.66$, $p < .001$). Figure 24B shows a similar, but weaker, positive correlation between preference and the difference in lightness ratings, in which pairs with lighter figures were more preferred ($r = +.48$, $p < .001$). Corresponding comparisons showed preference for pairs with warmer figures on cooler grounds ($r=+.39$, $p<.001$), a fact that is not surprising given the high correlation between warmness-coolness ratings and yellowness-blueness ratings ($r=+.73$, $p<.001$). There was also a very slight preference for pairs with desaturated figures on saturated grounds ($r = -.09$, $p < .01$), but no relation between preference and figure-ground differences in the redness-greenness of the colors ($r=.01$, $p>.05$).

Palmer and Schloss (2010) found that the single color preferences of the same participants tested here were strongly related to the rated yellowness-blueness of the colors, with a general preference for bluer colors. It is therefore possible that participants like bluer regions to be ground simply because they like the more preferred color in the figure-ground display to be larger. Bullough (1907) also reported that participants often claimed that they chose pairs in which their more preferred color occupied more space. To test this hypothesis we conducted two logistic regression analyses. The first tested how well each participant's component color preference accounted for the participant's corresponding preference asymmetries. The second analysis included that component color preference factor as well as a factor coding the figure-ground yellowness-blueness difference. These two models were compared to see the one containing yellowness-blueness differences explained significantly more variance than the model including only component color preference. The preference asymmetry data (992 trials) were coded as 1 = left display chosen, 0 = right display chosen, and .5 = "neither" response. The component color preferences were obtained from the same participants' single color 2AFC preference data in which they were presented with all pairs of the 32 single BCP chromatic colors and asked to indicate which one they preferred. The 2AFC data were used, rather than the same participants' preference rating data (see Palmer & Schloss, 2010), so that direct comparisons could be drawn between pair preferences and which of the two colors in each pair was more preferred singly. These data were scored so that 1= left display chosen (corresponding to the ground color of the left pair in the pair task) and 0= right display chosen (corresponding to the figure color of the left pair in the pair task). The yellowness-blueness difference was calculated relative to the left pair.

A chi square test was used to compare how well the model including component color preference alone fit the data versus the how well the model with both component color preference and figure-ground blueness difference fit the data. For 33 out of 48 participants ($p < .01$ by a sign test), the model containing both component color preference and figure-ground blueness difference fit their data reliably better than the model with component color preference alone at the .05 level. We conclude that the figure-ground yellowness-blueness effects in the preference asymmetries observed here are not simply a result of people preferring the larger (ground) color to be the more preferred color.

To further understand the preference asymmetries we have observed, we conducted several linear regression analyses on the preference probabilities (averaged across participants) for each of the 992 pairs. We first entered the average 2AFC single color preference described above, which coded the proportion of times the ground color was preferred to the figure color in the pair on the left. This factor reflects the difference in preference between the ground and figure color. This figure-ground difference in color preference explained a total of 29% of the variance, with higher preference for pairs in which the more preferred component color occupied the larger (ground) region. We then added the following five color appearance predictors: the corresponding figure-ground differences in rated yellowness-blueness, redness-greenness, warmness-coolness, lightness-darkness, and saturation in the left pair. Lightness difference explained an additional 22% of the variance (pairs with lighter figures being preferred), followed by yellowness-blueness explaining an additional 9% of the variance (pairs with yellower figures being preferred). The total variance explained was thus 60% (multiple-r = .78.). The other factors did not account for significant amounts of additional variance. It is noteworthy that yellowness-blueness alone explains 40% of the variance, which is a large proportion of the 60% total, but the reason for this dominance appears to be that it conforms with several different biases: (a) the figure-ground difference in the component preferences (because bluish colors are generally preferred to yellowish ones), (b) the figure-ground difference in the lightness of the colors (because yellow is lighter than blue), and (c) the figure-ground difference in yellowness-blueness that is independent of (a).

In the following section we examine preference asymmetries in light of the formulations proposed by Munsell (1921/1969) and Itten (1961/1973). Munsell's rule hypothesizes that people like "balanced" pairs in which the figural color (smaller region) has a higher value x chroma (VxC or lightness x saturation) product. The correlation between average preference asymmetry and the figure-ground difference in the VxC product for the left pairs was +.39, indicating that participants indeed tended to prefer pairs in which the figure had a higher VxC product than the ground. Although this correlation is in accord with Munsell's prediction, the VxC product only explains 15% of the variance in preference asymmetries, which is actually less than a model that includes only figure-ground differences in Munsell value (29%), with lighter figures on darker grounds being preferred. When chroma was added into the model with Munsell value as a separate factor, it explained only 2% more variance than value alone, with more saturated figures on less saturated grounds being preferred. Munsell's multiplicative formulation was thus less accurate than either an alternative based on value alone or an additive model based on the same two factors.

Itten (1961/1973) claimed that the area of a region should be inversely proportional to its "intensity" as given by Goethe (1810/2006), where yellow (9) is most intense, followed by orange (8), green (6) and red (6), blue (4), and violet (3). Because no intensity value was provided for chartreuse or cyan, we interpolated values halfway between yellow and green for chartreuse (7.5) and halfway between green and blue for cyan (5). We then tried to predict preference asymmetries with the figure-ground difference in Goethe's hue-based intensity dimension, where Itten's ratios would predict stronger asymmetries for pairs that have larger differences in intensity. Indeed, 32% of the variance can be explained by this factor, for which people preferred more "intense" figures on less "intense" grounds. When the Munsell dimensions were also entered into this model, value difference explained another 13% (lighter figures on darker grounds being preferred), and chroma difference explained an additional 3% (more saturated figures on less saturated grounds), for a total of 48% of the variance explained.

In Experiment 5 we thus found modest support for both Munsell's (1921/1969) and Itten's (1961/1973) proposals that lighter, yellower regions should occupy the smaller, figural region. However, in further testing Munsell's rule we found that the figure-ground difference in Munsell value explained more variance (29%) than the figure-ground difference in value x chroma (15%). Further, in testing Itten's rule, we found that perceived figure-ground difference in yellowness-blueness explained more variance (40%) than the corresponding difference in Goethe's intensities (32%). We also showed that people tend to like the more preferred color to be the ground, and thus larger, than the less preferred color.

These results indicate that preference asymmetries exist and that the two most potent predictors are figure-ground differences in yellowness-blueness (explaining 40% of the variance) and in perceived lightness-darkness (explaining an additional 9%). We now know a good deal about what the color determinants of figure-ground asymmetries are, but have not yet identified which *spatial* factors are driving the effect. Does the relative area between the two regions modulate preference asymmetries, and if so, is retinal area or perceived area the dominant factor? And do the colors need to be in a figure-ground arrangement to elicit preference asymmetries? We address these questions in Experiments 6-7.

3.4. Experiment 6: Effects of Relative Area on Preference Asymmetries

In this experiment we investigate how varying the area of the figural region relative to a constant ground region modulates preference asymmetries. Such variations necessarily change both the relative *image-based* area and the relative *surface-based* area of the two regions. Image-based area refers to the size of the 2-D regions that are projected onto the retina. Surface-based area refers to the size of the regions that the observer *perceives* after the ground is amodally completed behind the figure (see Palmer, 1999 for a discussion of image-based and surface based representations). The figure's image-based area will always be equivalent to its surface-based area because it is entirely visible. However, ground is partly occluded by the figure, the ground's image-based area and surface-based area will always be different. For example, the ground's surface-based area is constant as the figural area increases, because it is completed behind the figure, but the ground's image-based area decreases as the figural area increases, because the sum of the two is necessarily constant. Thus, when referring to ground area we will always be specific about the type (image-based or surface-based), but we will not have to make that distinction for figural area. To ensure that the figural portion always appeared to be closer (rather than a farther region seen through a thin frame around it), the figural region was divided into numerous texture elements of the same size, all of which appeared to lie in front of the ground region (see Figure 25).

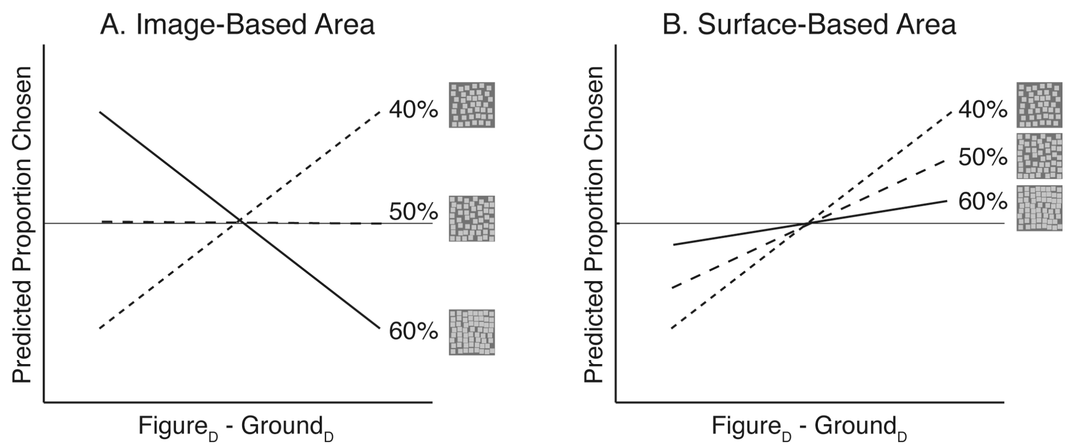

Figure 25. Predicted preferences depending on whether the relative difference in (A) image-based area or (B) surface-based area is dominant in influencing preference asymmetries. The x-axis is the difference along a given dimension (e.g., yellowness-blueness, lightness-darkness) between the ground and figure colors. Separate lines represent the percentages of area the figure occupies relative to the image-based area of the ground.

Different patterns of preference asymmetries should arise as the figural size increases, depending on whether image-based or surface-based area governs figure-ground preference asymmetries. As shown in Figure 25A, if image-based area dominates, there should be a cross-over interaction, in which observers will prefer yellower, lighter figures on bluer, darker grounds when the figural area is small, show no preference asymmetry when the figural area is equated with physical/retinal area of the ground, and will prefer bluer, darker figures on yellower, lighter grounds when the figural area is large. If *surface-based* area dominates, such that the ground is always perceived as fully completed behind the figure, however, observers will always prefer yellower, lighter figures on bluer, darker grounds, because the ground will always be perceived as larger, even though the degree of preference asymmetry will decrease as the figural area becomes larger. The predictions in Figure 25 are drawn as linear functions because the data in Experiment 5 (see Figure 24) were fairly linear.

3.4.1. Methods

Participants. The participants were 32 undergraduates at the University of California, Berkeley, who consented to participate in this study. All had normal color vision (screened using the Dvorine Pseudo-Isochromatic Plates). The Committee for the Protection of Human Subjects at the University of California, Berkeley, approved the experimental protocol.

Design, Displays, and Procedure. All participants completed three tasks in the following order: pair preference, figural size estimation, single color preference. The pair preference task was similar to that described in Experiment 5, but using three possible spatial configurations for the color pairs: one in which there were 40 texture squares (40% figure, 60% image-based ground), one with 50 squares (50% figure, 50% image-based ground) and one with 60 squares (60% figure, 40% image-based ground). Each texture square was 30 px x 30 px and the ground was always 300 px x 300 px.[6] Each trial contained two figure-ground configurations, one on the left and one on the right and participants were instructed to indicate which one they preferred by pressing the left or right arrow keys. (The ”neither” response was not available in this experiment). The pairs within a trial were always spatially identical and only varied in figure-ground assignment of the colors. All pair-wise comparisons of the eight light (L) and eight muted (M) hues were tested to make a total of 240 color pairs. Colors for the saturated and dark cuts were omitted to reduce the number of trials to 720.

In the figural size estimation task, participants were asked to estimate the percent of area occupied by the textural figure relative to the *image-based* area of the ground. Only one figure-ground configuration was presented in each trial, and it was located center of the screen. The four figure-ground configurations included the three just described for the pair preference task plus an additional configuration with even less texture (30% figure, 70% ground). Participants made their ratings along a continuous response line below the configuration that had tick marks delineating 10% intervals ranging from 0% to 100%. They used a mouse to control the position of a vertical line mark on this response scale to “compare the total area of the foreground squares with the total amount of visible area of the background square.” Participants were also told: “Do not consider any background covered by the foreground squares when making your judgment.” Each texture configuration was presented in two color combinations: lighter gray texture squares (63.90 cd/m^2) on a darker gray ground (12.34 cd/m^2) and darker gray texture squares on a lighter gray ground. The background was the same neutral gray background (19.26 cd/ m^2 as is all the other experiments). There were two replications of each condition to make a total of 16 trials.

In the single color preferences task, participants were presented with each of 240 pair-wise combinations of the 16 colors used in the pair preference task (left-right balanced) and indicated which they preferred by pressing a left or right response key. The colors were displayed as two squares (each 100 px x 100 px) on opposite sides of the screen’s vertical midline.

[6] We varied the number of same-sized texture elements rather than varying the size of the same number of texture elements because prior research has shown that when a configuration is composed of many elements, keeping the size of the elements the same and varying their number produces a display that appears more similar perceptually (cf. Kimchi & Palmer, 1982, 1985).

All displays remained on the screen until participants made their response, and trials were separated by a 500ms inter-trial interval. Displays were rendered using Presentation (www.neurobs.com) and were presented on a 20-inch iMac computer (1280 x 768 resolution).

3.4.2. Results and Discussion

In all three texture proportion conditions, participants preferred displays in which the textured figural region was yellower than the ground, but the difference decreased with increasing numbers of texture elements and figural area (Figure 26A-C). There was a strong correlation between the proportion of times pairs were chosen as more preferred and the figure-ground difference in yellowness-blueness for all textural conditions: 40% ($r = .71$, $p < .001$), 50% ($r = .67$, $p < .001$), and 60% ($r = .43$, $p < .001$). There were similar but weaker correlations for the lightness-darkness dimension: 40% ($r = .44$, $p < .001$), 50% ($r = .41$, $p < .001$), and 60% ($r = .06$, $p > .05$). This pattern shows that surface-based area is more important than image-based area in determining which pair is preferred. If image-based area were more important, the data in Figure 26C would be the vertical reflection of the data in Figure 26A because the amount of retinal area subsumed by the texture elements and backgrounds are opposites. Furthermore, the data in Figure 26B would approximate a flat horizontal line because there is no difference in image-based area between the texture and ground regions. Even when the yellower regions covered the majority of the retinal area in a figure-ground display, it was preferred to a figure-ground reversed display as long as the bluer background was perceptually completed behind it.

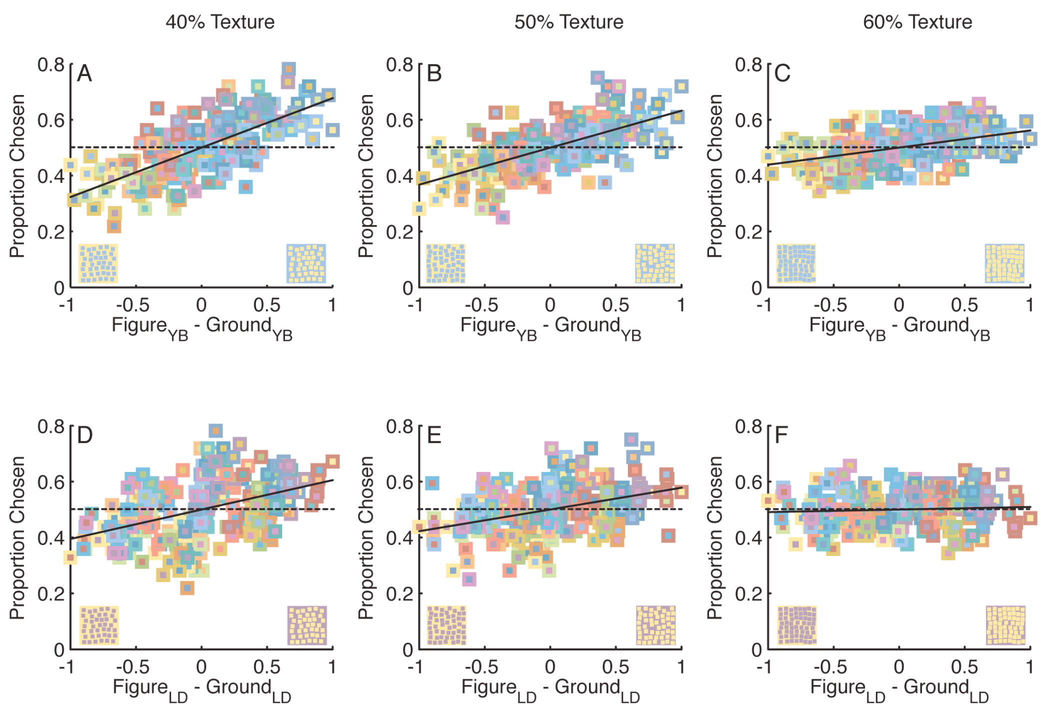

Figure 26. Separate plots show the proportion of times each pair was chosen as function of the yellowness-blueness difference between the figure and ground for the (A) 40%, (B) 50%, and (C) 60% texture amounts and as a function of the lightness difference between the figure and ground for the (D) 40%, (E) 50%, and (F) 60% texture amounts. Data point colors symbolize the color of the pair that was judged. Dashed black lines at a proportion of .5 represent “chance.” Solid black lines represent the best fitting regression line for each texture size condition.

To illustrate and analyze (with inferential statistics) the regularity of the interaction between the figure-ground difference in yellowness-blueness and the relative size of the figural region, we binned the data shown in Figure 26A-C into four groups based on the magnitude of the difference in yellowness-blueness between the figure and ground and averaged over pairs within each bin. The bins were formed by averaging preference for all the pairs that lay between five equally spaced limits from -1 to 1. Figure 27A shows these binned data for the 40%, 50%, and 60% texture conditions for the yellowness-blueness dimension. Figure 27B shows the corresponding binned data for the lightness-darkness dimension.

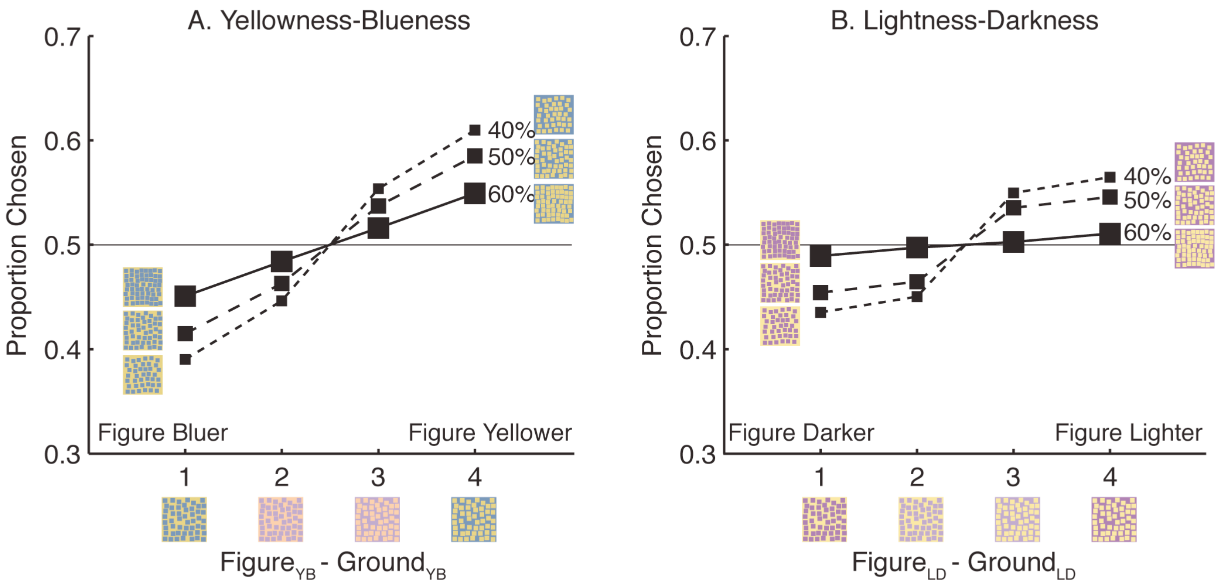

Figure 27. The proportion of trials on which pairs were chosen as a function of the binned difference in rated (A) yellowness-blueness and (B) rated lightness darkness between the ground and the figure.

As shown in Figure 27A, there was a main effect of yellowness-blueness difference ($F(3,93) = 10.19$ $p < .001$), in which preference asymmetry magnitude increased with larger yellowness-blueness differences between the figure and ground. There was also an interaction between yellowness-blueness difference and figural area, in which the preference asymmetries were more extreme when the figural area was small and approached chance (0.5) as figural area approached the surface-based area of the ground ($F(6,186) = 6.33$, $p < .001$). As shown in Figure 27B, there was also a similar relation between pair preference and lightness difference for figural and ground colors ($F(6,186) = 2.87$, $p < .05$), although there was no main effect of lightness difference ($F(3,93) = 1.00$, $p>.05$). Note that these data are more ogival than those found in Experiment 5, and therefore deviate from the predictions shown in Figure 25, which were linear.

One possible explanation for these results is that participants tend to underestimate the amount of image-based area covered by the figural texture relative to that covered by the background. If so, they could always be choosing the pair in which the yellower texture elements *appeared* to occupy less image-based area than the ground. To test this possibility, we asked participants to “compare the total area of the foreground squares with the total amount of visible area of the background square.” They were also told: “Do not consider any background covered by the foreground squares when making your judgment." As shown in Figure 28, participants substantially *over-estimated* the relative amount of area occupied by the figure ($F(1,31) = 114.17$, $p<.001$), which discounts the possibility that participants chose pairs with yellower texture because they thought that the textured region had a smaller retinal area. It should also be noted that the degree to which participants overestimate the figural area increases as figural area increases ($F(3,93) = 12.01$, $p <.001$) and that the overestimation is overall larger for lighter figures on darker grounds than the darker figures on lighter grounds ($F(1,31)= 15.00$, $p < .01$), which is not surprising given previous work on how lightness affects perceived size (i.e., irradiation (Münster, 1941), but see Békésy (1970)).

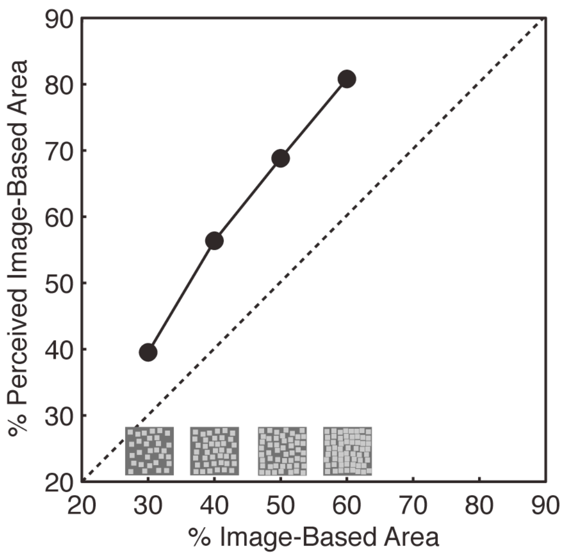

Figure 28. The estimated perceived area (filled circles) occupied by the texture relative to the visible parts of the ground for the 30%, 40%, 50%, and 60% configurations as a function of the image-based area occupied by the texture. The dashed line represents actual percentages.

Based on these data it is clear that participants prefer color combinations to the extent that the figural region is yellower and lighter and the ground is bluer and darker. It is also clear that the perceived surface-based area of the two regions, after amodal completion due to the depth information, is important. Still, it is unclear whether figure-ground organization is required to obtain these preference asymmetries. In Experiment 7 we test whether the same effects exist when the component colors are not nested spatially and are separated by a gap in a side-by-side mosaic configuration that does not produce figure-ground organization.

3.5. Experiment 7: Effects of Area for Separated Regions

Thus far we have only measured preference asymmetries in displays with clear figure-ground organization. In the present experiment we tested displays containing two rectangles separated by a gap so they appeared as a "mosaic" of regions in the same depth plane. If the same pattern of results emerges as reported in Experiments 5 and 6, then figure-ground organization and surroundedness are not prerequisites for preference asymmetries in color combination preferences. It turns out that the answer is different for yellowness-blueness than it is for lightness. We chose to use displays that were divided horizontally to determine whether participants preferred lower regions to be darker, as Bullough (1907) reported.

3.5.1. *Methods*

Participants. The participants were 26 undergraduates at the University of California, Berkeley who consented to participate. All of them had normal color vision (screened using the Dvorine Pseudo-Isochromatic Plates). The Committee for the Protection of Human Subjects at the University of California, Berkeley, approved the experimental protocol.

Design and Displays. There were two types of spatial configurations, one with a short top region (300 x 61 px) and tall bottom region (300 x 239 px) and the other with a short bottom region and a tall top region (see icons below the x-axis in Figure 29 for examples of short-top/tall-bottom displays). The pairs within a trial were always spatially identical and only varied in top-bottom assignment of the colors. All pair-wise comparisons of the eight light (L) and eight muted (M) colors were tested to make a total of 240 color pairs. The spatial arrangements were always the same within a trial. There were a total of 480 trials.

Procedure. The procedure was the same as in Experiment 6, omitting the size estimation task.

3.5.2. *Results and Discussion*

Figure 29 shows the proportion of times each pair was chosen as a function of the difference in yellowness-blueness (Figure 29A) and lightness-darkness (Figure 29B) between the figural and ground colors. These data were averaged over spatial configuration (large-bottom/small-top vs. small-bottom/large-top) because there were no effects of top vs. bottom for the yellowness-blueness difference ($F(3,69) = 2.73$, $p>.05$) or the binned lightness-darkness between the figure and ground colors ($F<1$).

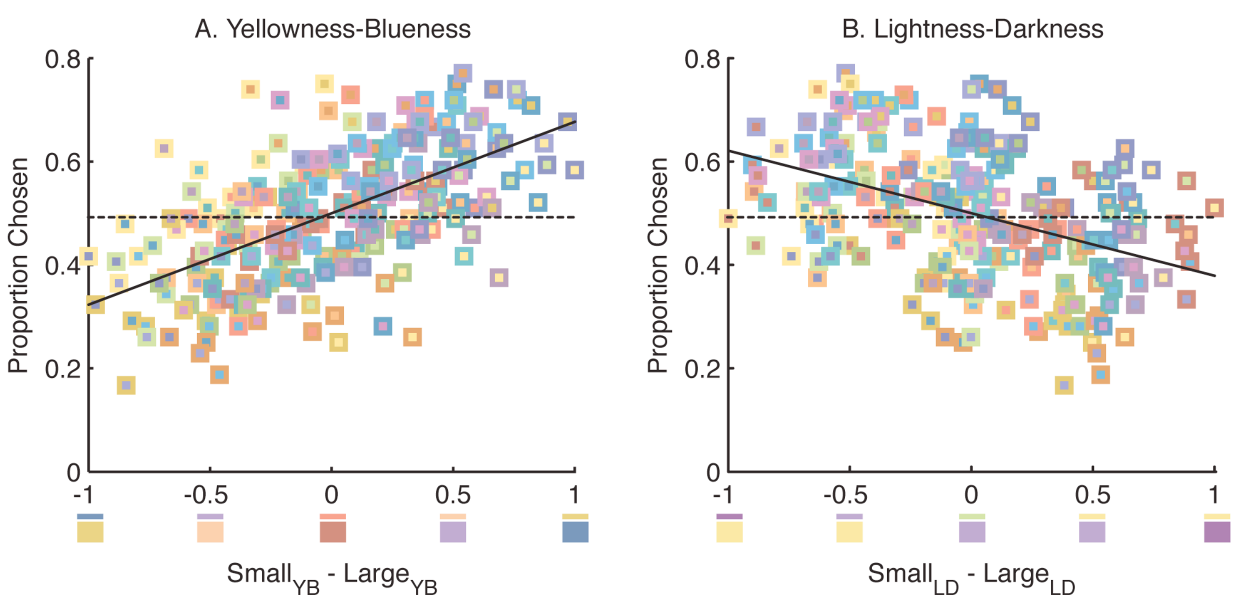

Figure 29. The proportion of times each pair was chosen as a function of the (A) yellow-blueness and (B) lightness-darkness difference between the small and large region. Data point colors represent the colors of the pairs that were judged. Dashed black lines at a proportion of .5 represent "chance." Solid black lines represent the best fitting regression line for each texture size condition.

As in Experiments 5 and 6, there was a positive correlation between pair preference and the difference in yellowness-blueness between the small and large region ($r = .59$ $p < .001$), and there was a main effect of the binned yellowness-blueness difference in an ANOVA $F(3,69) = 6.64$, $p < .01$). Unlike Experiments 5 and 6, there was a negative correlation between pair preference and the difference in lightness between the small and large region ($r = -.42$, $p < .001$), although the binned main effect was not significant ($F(3,69) = 2.02$, $p > .05$). Effects of relative area thus exist even without the figure-ground structure of either perceived depth/occlusion or of image-based surroundedness, but they do not appear to be identical. Yellowness-blueness effects with mosaic displays were comparable with the figure-ground configurations studied in Experiments 5 and 6, but lightness-darkness effects with mosaic displays reversed, such that participants preferred mosaic pairs in which the lighter regions were larger (rather than smaller, as in figure-ground displays) than the darker regions. Although the cause of this reversal is unclear, some participants commented that the mosaic displays made them think of walls and trim and they preferred displays in which the "walls" were lighter than the "trim."

An analysis testing whether participants were more likely to choose the pair that contained the darker region on the bottom, as a theory based on gravitational stability due to "color weight" would predict (Bullough, 1907), showed that no such effect existed in our data ($F < 1$). This null result may have arisen because the two regions in the configurations were separated by a gap, however, as Bullough (1907) argued that that it was crucial that the two colored regions be perceived as part of the same "whole" to produce the color weight effect that he reported.

3.6. General Discussion of Experiments 5-7

The results of three experiments have clearly shown that spatial organization influences people's preference for color combinations in systematic ways. People reliably prefer larger regions of color pairs to be bluer and smaller regions to be yellower, regardless of whether the regions are perceived as a figure in front of a ground (Experiments 5 and 6) or as adjacent figures in the same depth plane (Experiment 7). The effective sizes of the regions in figure-ground displays are determined by their relative perceived areas after the ground has been completed behind the figure rather than by their retinal areas (Experiment 6). Although we found preferences for the component colors have an effect on these preference asymmetries, the yellowness-blueness effects are not solely due to such preference effects because yellowness-blueness accounts for additional variance after the effects of component color preference have been removed (Experiment 5). People also prefer color pairs with smaller regions to be lighter and larger regions to be darker when they are presented in a figure-ground configuration (Experiments 5 and 6), but they prefer smaller regions to be darker and larger regions to be lighter when they are presented in a mosaic configuration of coplanar rectangles (Experiment 7).

Although the pattern of preferences over different spatial organizations is reasonably clear, the reasons for it are not. Here, we will consider two kinds of explanations that attempt to go beyond the colorimetric descriptions given above: namely, accounts based on phenomenology and ecology. Phenomenological explanations appeal to non-colorimetric aspects the observer's subjective experiences in assessing preferences for color pairs. An example of a previous phenomenological hypothesis for pair preferences is Schloss and Palmer's (2011) finding that people's color harmony ratings explain 62% of the variance in their preferences for color pairs ($r = .79$), where color harmony was phenomenologically defined as the degree to which the colors look like they "go together," (using an analogy with the phenomenology of musical harmony). It is not clear how this hypothesis would apply to the kinds of spatial asymmetry effects we report here, however, because "going together" seems primarily to be a symmetric relation. A phenomenological hypothesis that does have an obvious application for the present results is based on Itten's (1961/1973) introspective observation that in order to produce color pairs that are "balanced" in terms of "intensity," high-intensity colors should be confined to a smaller area than lower intensity colors. This is similar to Munsell's (1921/1969) idea that balance is achieved when "stronger" colors occupy less space than "weaker" colors. Munsell's phenomenological "strength" dimension, however, is defined as value x chroma whereas Itten's phenomenological "brilliance" or "intensity" dimension is defined by Goethe's "light values," which are more strongly related to yellowness-blueness (r=.88) than to lightness-darkness (r= .72). Thus, the above results in terms of yellowness-blueness effects in all spatial configurations support for Itten's hypothesis, where colors that feel more "intense" should be smaller in area to achieve balance.

Given that experienced color intensity also varies strongly with the lightness of a colored region, the same limited-intensity hypothesis implies that people should like lighter regions to be smaller and darker regions to be larger. This is true for the figure-ground displays (Experiments 5 and 6), but not for the mosaic displays (Experiment 7). Indeed, the pattern reverses for mosaic displays, with lighter regions being preferred when they are larger rather than smaller. It is not clear why this should be true from a phenomenological standpoint.

A second kind of explanation is that strong ecological associations are responsible for the obtained spatial asymmetries in preference. Palmer and Schloss (2010) found clear support for an ecological account of individual color preferences: people like colors to the degree that they like the things that are those colors. It is therefore reasonable to consider similar kinds of explanations for the present results. Bullough (1907) considered ecological accounts of his result that people prefer darker regions to be lower, but abandoned it when he concluded that there were as many counter-examples as examples in support of his idea. Nevertheless, there are appealing ecological facts about prototypical figure-ground relations that may account for our results with figure-ground displays. The suggestion is that people like to see smaller yellower regions surrounded by larger bluer regions as a generalization of the fact that most people like bright, sunny days, when they see the smaller yellow sun against the larger surrounding blue sky. Because the spatial reversal of this figure-ground combination (blue-on-yellow) has no particular ecological significance, the yellow-on-blue organization would be strongly preferred to its reversal on the basis of ecological factors. The generalization gradient from this prototype might be strong enough that similar, though weaker, asymmetries holds for figures whose color is yellow-ish against grounds whose color is blue-ish. Similar generalizations over spatial factors might also produce preferences for smaller yellower regions next to larger bluer regions in a mosaic configuration. However, there is not yet empirical evidence to support both of these conjectures.

Similarly, people may like to see a smaller brighter region surrounded by a larger darker region because they also like clear, moonlit nights, when they see the small, white moon against the large, dark sky. This preference may be less robust than the yellowness-blueness effects because people probably are not as fond of clear, moonlit nights as they are of bright, sunny days. Another consideration is that the opposite organization (small, dark regions surrounded by large, light regions) also has ecological significance in text and images printed on a white page, although it is not so obvious whether the valences of these images are positive or negative. The further fact that this lighter-on-darker preference does not generalize to smaller light regions next to larger dark regions in a mosaic configuration – indeed, preferences reverse in this spatial structure – may be due to other ecological situations being more relevant to such organizations. As mentioned above, some participants spontaneously volunteered that the mosaic displays reminded them of a wall-and-trim situation, and they preferred the walls to be lighter.

Although the ecological explanations just advanced are clearly ad hoc, they should not be taken lightly. The extremely close link Palmer and Schloss (2010) found between average preferences for single colors and average ratings of the degree to which people like the ecological objects that characteristically have that color – a correlation of .89 using their procedures – obviously raises the possibility that similar effects may underlie preferences for color pairs, and it is quite possible that the spatial structure of such displays plays a significant role. Although Schloss and Palmer (2011) were able to identify strong colorimetric determinants of preference for color pairs – e.g., that people tend to like color pairs to the extent that the two colors are the same or similar in hue but differ in lightness – does not preclude the possibility that ecological effects also influence people's preferences. Indeed, they may be especially relevant to spatial asymmetries in figure-ground displays that may remind observers, consciously or nonconsciously, of particular ecological situations. We are currently pursuing such conjectures in research designed to test them.

4. General Conclusions

The goal of the experiments described here was to understand people's aesthetic responses to color pairs, both in terms of which colors people prefer in combination (Chapter 2) and how the spatial organization of the component colors influences pair preference (Chapter 3).

On average, people prefer pairs with cooler colors that are similar in hue, contrasting in lightness, and contain preferred individual colors. Pair preference is highly related to pair harmony—how well the colors go together—however, pair preference relies more on component color preference and lightness contrast than harmony does. Preference for a color pair as a whole is different from preference for a single figural color, given its background color. Most notably, pair preference relies more on hue similarity, whereas figural colors are preferred on backgrounds that contrast in hue, especially when lightness contrast between the figure and ground color is minimal. Relating these results to color theory, Chevreul's (1839) "harmony of analogous colors" applies to pair preferences and his "harmony of contrast" applies to preference for figural colors on contrasting backgrounds.

Spatial organization plays a strong role in preference for color pairs in that people prefer yellower regions to be smaller in area than bluer regions. This effect it not solely for yellow and blue, but rather follows a generalization gradient away from yellow and blue so that preference asymmetry effects are larger for pairs that differ more on the yellowness-blueness dimension. Surface-based area (i.e., after amodal completion) was more influential than image-based area, such that people preferred figures to be yellower and grounds to be bluer even when the area of the figure was larger than the image-based area of the ground. However, figure-ground organization was not required for such effects, given that the same pattern held when the two component regions were side-by-side and separated by a gap to form a mosaic configuration. For the lightness-darkness dimension, a similar but weaker pattern was present figure-ground preference asymmetries, but the pattern was reversed for mosaic configurations; participants preferred larger regions to be lighter than smaller regions.

Possible explanations for these results are discussed fully in the Results and Discussion sections of Chapters 2 and 3, but will be summarized here. Both phenomenological and ecological explanations have been suggested. For *which* colors people prefer in combination, a phenomenal explanation is that people prefer colors that feel like they "go well" together (are harmonious). But, what determines which colors go well together? One ecological explanation is that colors that "go well" together are colors that literally are found together in the world by co-occurring within uniform connected (UC) regions of natural images (Palmer & Rock, 1994). A phenomenological explanation for color-pair preference asymmetries is that people prefer the more "intense" colors to occupy less space than less intense colors so that the color combination feels balanced (Itten, 1961/1973), and yellower colors feel more intense than bluer colors.

Another possible ecological explanation is that people prefer pairs to the degree that the colors remind them of positive things, in accord with Palmer and Schloss' (2010) ecological valence theory (EVT). Pair preferences are related to single component color preference (Experiment 1), and preference for single colors is highly related to preference for correspondingly colored objects, so it follows that pair preferences might be related to preferences for correspondingly colored objects. This influence could operate on each individual color separately (e.g., yellow figures reminding people of sunflowers and blue grounds reminding them of the ocean) and/or on the color combination as a whole (e.g., yellow figures on blue grounds reminding people of the sun on a clear sky). Of course, according to the EVT, it is not just these specific color-object associations that should be influential, but rather experiences with *all* objects of a given color or color-pair influence preference for a color pair. Following this logic, it is possible that people prefer color pairs that are congruent with ecological scenes, such as a yellow figure on a blue ground associated with the sun against the clear blue sky, rather than the less ecologically valid figure ground color-reversal. Future research will be aimed at testing ecological explanations for color pair preferences.

Before closing, it is important to note that all of the principles presented in this dissertation are not intended to instruct artists and designers on how they *should* construct their work. As both Munsell (1921/1969) and Itten (1961/1973) explain, there are great paintings that do not adhere to such rules of balance because their imbalance has an exciting, provocative effect. Instead, these principles describe average preferences by average viewers of generic displays, and may be used as a default basis that can be manipulated to produce a provocative desired effect.

5. References

Albers, J. (1971). *Interaction of color.* New Haven, CT: Yale University Press.

Allen, E. C., & Guilford, J. P. (1936). Minor studies from the psychological laboratory of the University of Nebraska. X. Determining the affective values of color combinations. *The American Journal of Psychology, 48*, 643-648. doi: 10.2307/1416516

Békésy, G. von (1970) Apparent image rotation in stereoscopic vision: The unbalance of the pupils. *Perception & Psychophysics*, 8, 242-247.

Burchett, K. E. (2002). Color Harmony. *Color Research and Application, 27,* 28-31. doi: 10.1002/col.10004

Bullough, E. (1907) On the apparent heaviness of colours. *British Journal of Psychology,* 2, 111-152.

Camgöz, N., Yerner, C., & Güvenç, D. (2002). Effects of hue, saturation, and brightness on preference. *Color Research and Application, 27,* 199-207. doi: 10.1002/col.10051.

Chevreul, M. E. (1839). *The principles of harmony and contrast of colors.* Birren, F. (ed.) (1967). New York, NY: Van Nostrand Reinhold.

Chuang, M., & Ou, L. (2001). Influence of a holistic color interval on color harmony. *Color Research and Application*, 26, 29 –39.

da Vinci, L. (1492). *Treatise on Painting.* Translated by A. P. McMahon (1956). Princeton, NJ: Princeton University Press.

Goethe, J. W. (1810/2006). *Theory of colours*. Translated by C. L. Eastlake (1840) from German edition "Farbenlehre" of 1810. Cambridge: MIT Press.

Granger, G. W. (1952). Objectivity of colour preferences. *Nature, 170,* 228-780.

Granger, G. W. (1953). Area balance in color harmony: an experimental study. *Science, 117,* 59-61. doi: 10.1126/science.117.3029.59

Granger, G. W. (1955a). An experimental study of colour harmony. *Journal of General Psychology, 52*, 21-35.

Granger, G. W. (1955b). Aesthetic measure applied to color harmony: An experimental test. *Journal of General Psychology, 52*, 205-212.

Granger, G. W. (1955c). The prediction of preference for color combinations. *Journal of General Psychology, 52*, 213-222.

Gregory R. L. (1977). Vision with isoluminant colour contrast: 1. A projection technique and observations, *Perception,* 6, 113-119.

Hård, A., Sivik, L. (1981). NCS - Natural Color System: A Swedish standard for color notation. *Color Research & Application*, 6, 129-138.

Helmholtz, H. von (1866/1925). In J. P. C. Southall (Ed.), *Treatise on physiological optics* (Translated from the 3rd German edition). New York: Dover Publications.

Helson, H., & Lansford, T. (1970). The role of spectral energy of source and background color in the pleasantness of object colors. *Applied Optics, 9*, 1513-1562.

Hurlbert, A. C., & Ling, Y. (2007). Biological components of sex differences in colour preferences. *Current Biology, 17,* 623-625.

Itten J. (1961/1973). *The art of color*. New York, NY: Van Nostrand Reinhold.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm Shift to the Integrative Big-Five Trait Taxonomy: History, Measurement, and Conceptual Issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), Handbook of personality: Theory and research (pp. 114-158). New York, NY: Guilford Press.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). The Big Five Inventory--Versions 4a and 54. Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

Kimchi, R. & Palmer, S. E. (1982). Form and texture in hierarchically constructed patterns. *Journal of Experimental Psychology: Human Perception and Performance,* 8, 521-535.

Kimchi, R. & Palmer, S. E. (1985). Separability and integrality of global and local levels of hierarchical patterns. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 673-688.

Linnett, C. M., Morriss, R. H., Dunlap, W. P., Fritchie, C. J., (1991). Differences in color balance depending upon mode of comparison. *Journal of General Psychology,* 118, 273-283.

Lotto, R. B., & Purves, D. (2000). An empirical explanation of color contrast. *Proceedings of the National Academy of Science, 97,* 12834-12839.

Martin, D., Fowlkes, C. Tal, D., & Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. Paper presented at the 8th International Conference on Computer Vision, July 7-14, Vancouver, British Columbia.

Moon P., & Spencer D. E. (1944a). Geometric formulation of classical color harmony. *Journal of the Optical Society of America, 34*, 46 –59.

Moon, P., Spencer, D. E. (1944b). Aesthetic measure applied to color harmony. *Journal of the Optical Society of America, 34*, 234 –242.

Moon, P., Spencer, D. E. (1944c). Area in color harmony. *Journal of the Optical Society of*

America, 34, 93-103.

Morriss, R. H., Dunlap, W. P., Hammond, S. E. (1982). Influence of chroma on spatial balance of complementary hues. *American Journal of Psychology,* 95, 323-332.

Morriss, R. H., Dunlap, W. P. (1987). Influence of value on spatial balance of color pairs. *Journal of General Psychology,* 114, 353-283.

Morriss, R. H., Dunlap, W. P., (1988). Influence of chroma and hue on spatial balance of color pairs. *Color Research and Application,* 13, 385-388.

Münster, C. (1941). Über den Einflun von Helligkeits Unterschieden in beiden Augen auf die stereoscopische Wabmehmung. *Zeitschrift fur Sinnesphysiologie,* 69, 245-260.

Munsell, A. H. (1921/1969). *A grammar of color*. Ed. F Birren, New York, NY: Van Nostrand Reinhold.

Munsell, A. H. (1966) *The Munsell Book of Color-Glossy Finish Collection*. Baltimore, MD: Munsell Color Company.

Nemcsics, A. (1993). *Colour dynamics: environmental colour design*. New York, NY: Ellis Horwood.

Nemcsics, A. (2007). Experimental determination of laws of color harmony. Part 1: Harmony content of different scales with similar hue. *Color Research and Application, 32,* 477-488.

Nemcsics, A. (2008). Experimental determination of laws of color harmony. Part 2: Harmony content of different monochrome pairs. *Color Research and Application, 33,* 262-270. doi: 10.1002/col.20416

Nemcsics, A. (2009a). Experimental determination of laws of color harmony. Part 3: Harmony content of different hue pairs. *Color Research and Application, 34,* 33-44. doi: 10.1002/col.20457

Nemcsics, A. (2009b). Experimental determination of laws of color harmony. Part 4: Color preference and the color harmony content. *Color Research and Application, 34,* 210-224. doi: 10.1002/col.20489

Ostwald, W. (1932). *Color science*, *Vol. 1*. London: Windsor Newton.

Ou L, Luo M. R., Woodcock A Wright A. (2004a). A study of colour emotion and colour preference, part II: Colour emotions for two-color combinations. *Color Research and Application, 29*, 292–298. doi: 10.1002/col.20024

Ou L, Luo M. R., Woodcock A Wright A. (2004b). A study of colour emotion and colour preference, part III: Colour preference modeling. *Color Research and Application, 29*, 381–389.

Ou L. & Luo, M. R. (2006). A colour harmony model for two-colour combinations. *Color Research and Application, 31*, 3, 191-204. doi: 10.1002/col.20208

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.

Palmer, S.E., and Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin & Review*, *1*. 1, 29-55.

Palmer, S. E., & Schloss, K. B. (2010). An ecological valence theory of human color preference. *Proceedings of the National Academy of Science*, *107,* 19, 8877-8882.

Reber, R., Schwarz, N., Winkielman, P. (2004). Processing Fluency and Aesthetic Pleasure: Is Beauty in the Perceiver's Processing Experience? *Personality and Social Psychology Review, 8*, 364-382.

Rubin, E. (1921/1958) "Figure and ground," in *Readings in perception* Eds. D C Beardslee, M Wertheimer.

Schloss, K. B., Poggesi, R. M., & Palmer S. E. (in press). Effects of university affiliation and "school spirit" on color preferences: Berkeley vs. Stanford. *Psychonomic Bulletin and Review.*

Shevell, S. K. (1978) The dual role of chromatic backgrounds in color perception. *Vision Research, 18*, 1649-166.

Walraven, J. (1976). Discounting the background—the missing link in the explanation of chromatic induction. *Vision Research*, *16*, 289-295.

Westland, S., Laycock, K., Cheung, V., Henry, P., & Mahyar, F. (2007). Colour harmony. *Colour: Design & Creativity, 1,* 1-15.

Wyszecki, G., & Stiles, W. S. (1967). *Color science: Concepts and methods, quantitative data and formulas.* New York, NY: John Wiley.