Differences in Color Naming and Color Salience in Vietnamese and English

Kimberly A. Jameson,[1]* **Nancy Alvarado**[2]

[1] Department of Psychology, University of California, San Diego, CA

[2] Center for Brain and Cognition, University of California, San Diego, CA

Received 21 August 2001; revised 7 March 2002; accepted 19 June 2002

Abstract: The accepted model of color naming postulates that 11 "basic" color terms representing 11 common perceptual experiences show increased processing salience *due to a theorized linkage between perception, visual neurophysiology, and cognition. We tested this theory, originally proposed by Berlin and Kay in 1969. Experiment 1 tested salience by comparing unconstrained color naming across two languages, English and Vietnamese. Results were compared with previous research by Berlin and Kay, Boynton and Olson, and colleagues. Experiment 2 validated our stimuli by comparing OSA, Munsell, and newly rendered "basic" exemplars using colorimetry and behavioral measures. Our results show that the relationship between the visual and verbal domains is more complex than current theory acknowledges. An interpoint distance model of color-naming behavior is proposed as an alternative perspective on color-naming universality and color-category structure.* Col Res Appl, 28, 113–138, 2003; DOI 10.1002/col.10131

Key words: color; color naming; color categorization; color salience

INTRODUCTION

The accepted model of color naming postulates that 11 "basic" color terms representing 11 common perceptual experiences show increased processing *salience* due to a theorized linkage between perception, visual neurophysiology, and cognition. Hardin and Maffi[1] review the many studies demonstrating the cross-cultural robustness of the Berlin and Kay[2] sequence of *basic color terms*. However, the strength of any linkage between basic color terms and salient category focal exemplars remains unclear. Although the literature strongly implies a relationship between basic color terms and perceptually salient color-appearance regions, several recent empirical results suggest that the focal exemplars most frequently labeled by basic color-term glosses are not the same across languages. This implies that there may not be a strong link between basic color terms and specific, perceptually salient focal colors. The goal of our research was to explore the linkage between color terms and color appearances by empirically investigating color naming and cross-cultural salience of best-exemplar color appearances (i.e., the previously identified "focal" or "centroid" colors) using a wide range of both basic and nonbasic color samples presented to three different ethnolinguistic groups. Although this research did not set out to confirm the universality of basic color terms, our results do confirm that basic color terms were widely used to label color samples in all three groups. We observed this despite findings that failed to confirm a linkage between basic color terms and strong perceptual salience.

SALIENCE

The Berlin and Kay theory states that the widespread use of 11 color-category terms and partitions across cultures is attributable to universal panhuman neurophysiological color vision processes.[2–7] Due in part to Heider-Rosch,[8] the concept of "salience" is central to models hypothesizing underlying linkages between visual neurophysiology and universal naming behavior. Although direct physiologic evidence

---

*Correspondence to: Kimberly A. Jameson, Department of Psychology, University of California, San Diego, 9500 Gilman Dr., MC-0109, La Jolla, CA 92093-0109 (e-mail: kjameson@ucsd.edu).

Contract grant sponsor: NSF-9973903, NIMH grant RO3-MH53126-01, Hellman Faculty Fellowship Award (Jameson), UCSD Academic Senate Grant (Jameson), and UCSD Undergraduate Scholastic Research Grant (A. Lewis)

is lacking,[9] theorists propose that a specific visual processing substrate causes certain color appearances to have the behavioral properties of salience.[10–12]

In the domain of color cognition, salience is indicated by "specific and selective nonverbal responses" that are observed for "focal" color-category exemplars.[11] In various models, salient hues have been called "focal," "basic," or "landmark" hues, observed to be more easily located, learned, and remembered than other hues. This model of color salience has been linked to Hering's[13] opponent process model of early visual color sensations, which states that opposing sensations are organized into opponent pairs Red-Green and Yellow-Blue, in conjunction with an achromatic (i.e., light–dark) opponency.[4] These constitute the "fundamental neural response categories" determined by neurophysiology that are thought to result in salience of certain best-exemplar colors within basic color categories and their composites (compounds of two basic colors), as found in the 11 universal color categories identified cross-culturally by Berlin and Kay.[2–4] The initial validation of this concept of focal color salience was provided by Heider-Rosch[8,14] and by the psychophysically rigorous studies of Boynton and Olson.[10,11]

Heider-Rosch[8] found that focal colors were more frequently chosen than nonfocal colors by 3-year-olds in a free-choice situation and were better matched than nonfocal colors by 4-year-olds. For both age groups, focal colors were also found to represent basic color terms more frequently than nonfocal colors. Heider-Rosch concluded that "focal colors are perceptually salient for young children as well as adults, and that color names initially become attached to these most salient areas" (p 454). She then demonstrated empirically that the Dani people of New Guinea form color categories with prototypic exemplars as "foci." In further cross-cultural comparisons, she asserted that these foci were easier to learn, remember, and were most frequently "named" universally. Heider-Rosch[14] stated, "The most saturated colors were best examples of basic color names for both English and for speakers of the other 10 languages represented" (p 13). One criticism of both Heider-Rosch studies is that the stimuli selected were always the maximum saturation available for the Munsell Hue and Value tested. As a result, these studies cannot determine whether the results of focal salience were due to differential perceptual processing or to universal preference for the most highly saturated exemplars. To our knowledge, no study controlling for saturation has confirmed Heider-Rosch's finding of prototypicality for focal color appearances. Thus, despite Heider-Rosch's pioneering work, verification of the Berlin and Kay notion of "focal" color salience remained unconfirmed, as did the relation of perceptually opponent hues to cognitive salience.

Boynton and colleagues expanded upon Heider-Rosch's ideas and gave the first psychophysically rigorous results for "cognitive salience" of color appearances and color naming.[10,11,15,16] Whereas Heider-Rosch used stimulus samples from the Munsell Book of Color, Boynton methodically assessed cognitive salience of the 424 samples from the OSA space, a color-ordered system created by the Optical Society of America.[17] Boynton used several different behavioral measures in his studies, including monolexemic naming consistency, response time, and consensus or majority choice. Note that Boynton and colleagues used a monolexemic naming task in order to meet Berlin and Kay's criteria. For a color term and its associated color-space focus to be considered unequivocally basic, it must be linguistically "monomorphemic," or a single term.[a]

Like Berlin and Kay, Boynton and Olson assessed the salience of color appearances linked to basic color terms. However, Boynton and Olson defined their color appearances differently. They defined a series of salient color category "centroids" derived from an OSA-coordinate average across subject choices, rather than identifying top-ranked best exemplars, as done by Berlin and Kay. One method might identify more salient samples than another, but both approaches identify specific samples, rather than general regions of samples, as salient due to underlying neural response fundamentals. For this reason, we considered it reasonable to test both the Berlin and Kay and the Boynton and Olson definitions of salience in our study of naming behavior. Both theories suggest that certain samples they empirically identify as *focals* or *centroids* have different perceptual processing status and are universally named using basic color terms.

Boynton and colleagues found that Hering's opponent-process colors (red, green, yellow, and blue) were more "salient" than some composite hues (or combinations of basic colors). He termed these more salient hues "landmark hues." However, Boynton and colleagues unexpectedly found that some composite hues demonstrated as much salience as the landmark colors. Such a finding is inconsistent with the privileged processing status believed to be associated with the Hering color-opponent processing.

Boynton's finding of salience for certain non-landmark hues can be related to whether or not a color appearance is named using a *monolexemic* term.[10] For example, in Japanese, the color appearance light blue, considered nonfundamental by opponent-process theory, is named by a monolexemic term, *mizu*. It rivaled the performance of landmark colors on all behavioral measures in the Japanese data, even though the color appearance light blue is generally not considered salient enough to earn rank as a basic color category.[11,15,16] Other composite colors named using monolexemic terms (such as orange and purple) showed the same pattern of results as the landmark colors, or sometimes better results. Despite this, Boynton's psychophysical results have been regarded as evidence supporting a panhuman shared opponent-process neural substrate. To address this strong interpretation of his results, Boynton[9] recently described the limitations of basing color-naming salience on models of lateral geniculate nucleus (LGN) neurophysiologic processing.

---

[a] The Berlin and Kay criteria for basic terms additionally included taxonomic superiority, broad applicability, and salience.

Sturges and Whitfield[18] carried out a large-scale replication of the Boynton and Olson[11] color-naming study using monolexemic naming of Munsell stimuli. Like Boynton and Olson, they observed behavioral differentiation between basic and nonbasic color categories. However, they also found no differentiation between landmark and other basic colors in naming consistency or response time measures. Sturges and Whitfield[18] asked: "Given that the difference between the landmark and other basic colours is small . . . Are the other basic colours sufficiently different from the landmark colours to be classed as less salient?" (p 312). On the basis of their results they concluded that "it would be reasonable to include purple as a landmark colour and to question the very landmark status of red" (p 312). Even so, they suggest, "It would be surprising if . . . results supporting a categorical structure to colour space based on Berlin and Kay's model was not reflected in a neurophysiological correlate" (p 312). Thus, despite ambiguous results, recent color-naming and categorization research continues to suggest opponent-color neural processing as the basis for landmark color salience and focal color universality.[1,4,7,18–20,b]

Roberson and colleagues[21,22] question the validity of the construct of differential focal color salience. They suggest that categorical color perception is based on verbal coding as opposed to visual salience.[22] Moreover, Roberson *et al.*,[21] in a study of color-naming in Papua New Guinea, showed that under a variety of tasks, categorical color perception was in accord with linguistic categories rather than underlying perceptual universals. They also showed that there was no recognition advantage or paired-associate learning advantage for focal stimuli compared to nonfocal stimuli.

Lin *et al.*[23–25] found support for Berlin and Kay's 11 basic color terms but raised new questions about constraints imposed by the empirical practice of monolexemic naming. They compared constrained and unconstrained naming in two linguistic populations, Mandarin Chinese and British English. In an *unconstrained* naming task, they found that whereas monolexemic basic color terms were modal names for roughly half of the samples, all subjects preferred to use modified (not monolexemic) basic names rather than basic names alone.[23] They also questioned Berlin and Kay's definition of basic terms, asserting an additional five Chinese basic terms beyond Berlin and Kay's 11. Under *constrained* (monolexemic) naming, the cross-language similarity in the mapping of basic terms to focal color regions was complicated by conflicting results in two experiments presented.[24]

Questions about the appropriateness of the Hering fundamentals as the basis for color-naming, color categories, and focal color salience have been raised by other investigators as well.[26,28,29] Thus, the results suggest that the linkage between early visual neurophysiology and color cognition may not be as direct as assumed by currently accepted theory. The noted invariance in color naming across cultures is impressive, but the strong model typically suggested for focal color universality and perceptual salience deserves further scrutiny.

EXPERIMENT 1

Experiment 1 used naming behavior to examine the salience of the rigorously defined color category "centroids" identified by Boynton and colleagues.[15,c] Considerable effort was made to reproduce accurately the stimuli used in other work. To assess the impact of empirical naming constraints on color-naming results, the following modifications of previous paradigms were made: (1) Subjects were given unconstrained time to freely name color samples, rather than being provided with terms by the experimenter or encouraged to respond quickly using monolexemic terms; and (2) comparisons were made across two languages in which color categories were expected to vary. We expected that any invariance dependent on underlying neurophysiology should be unaffected by manipulations in task demands.

Participants

Three samples participated: (1) 31 monolingual English speakers, (2) 29 bilingual English and Vietnamese speakers tested in Vietnamese, and (3) 32 monolingual Vietnamese speakers. Bilingual speakers reflect a different access to the lexicon than individuals who are proficient in a single native language.

All monolingual English and some bilingual Vietnamese speakers volunteered through the University of California, San Diego human subject pool and earned partial course credit. Some additional bilingual Vietnamese speakers were paid $8.00 per hour. Monolingual Vietnamese participants were recruited from immigrant Vietnamese communities in the San Diego area and were paid $8.00 per hour. Local Vietnamese communities are sufficiently large to permit individuals to function and work without needing to acquire English. Monolingual and bilingual Vietnamese speakers unable to read and write Vietnamese were excluded. Three subjects with Vietnamese surnames were omitted from the monolingual English sample prior to data analysis. All subjects were screened for normal (corrected) vision and for

---

[b] Kay recently suggested (personal communication, February 2002) that his current theory of color-naming universals does not emphasize a strong linkage between perceptual salience of the Hering primaries and fundamental neural response correlates found in the lateral geniculate nucleus (LGN). A prelude to this new position is seen in Kay and Maffi.[7] Despite this shift in emphasis, much of the current literature continues to rely on the classical linkage between Hering's perceptual primaries and LGN opponent processing mechanisms.

[c] "Centroid" values are a computed color category position derived by averaging the *L,j,g* values of all samples called by a particular name, weighted according to whether the name was used once or twice. As such, it is a focal point, or sample, derived by the aggregate responses of all subjects in a given ethnolinguistic group. It is not unreasonable to view the centroid exemplar as a sort of group aggregate category "focal" in the Berlin and Kay sense, compared to the individual category focus that a given subject may designate, and which may differ from the group centroid. As defined, group aggregate samples do not coincide with an individual's foci presumed to arise from his or her color vision neural processing.

normal color vision with Ishihara's[30] Pseudoisochromatic Test Plates (Concise Edition). Two subjects with anomalous color vision were omitted from the bilingual Vietnamese sample prior to data analysis.

Procedure

Participants in each of the three language groups were provided with a test booklet that comprised 110 individual color samples, one per page (see description of stimuli that follows). For each sample, participants were asked to provide the appropriate name, with no constraints imposed on their choice of semantic label. Subjects also provided a confidence rating (ranging from 1 to 5) for the estimated accuracy of each name. Fifteen booklet variations were generated, representing different random orders of stimuli. Within each language group, the booklet orders were randomly assigned to participants under the constraint that all 15 orders should be assessed before any given order was repeated. The task was self-paced, and was introduced using one practice trial to familiarize participants with the task, followed by the 110 experimental judgments, then color vision screening and debriefing.

For all three participant groups, the task was conducted in a controlled ambient lighting environment. The room was illuminated by an approximated *C* illuminant conforming to spectral power distributions of the CIE daylight model.[31] Ambient illuminant intensity averaged 185.6 cd/m^2—which subjectively approximates indirect daylight illumination—at CIE (1931) chromaticity: $x = .349$, $y = .360$; and CCT = 4856 K.

Stimuli

Color samples were presented in a neutral viewing context, free of potential color contrast and stimulus-set effects that existed in the color grid used by Berlin and Kay.[2,d] The 110 stimulus items included the landmark hue "centroids" identified by Boynton and Olson,[10] plus a random sample drawn from the OSA Uniform Color Scale (UCS) stimulus space.[32,33] Use of the OSA space to characterize stimuli permits direct comparison of this study's findings with results obtained by Boynton and colleagues.[9–11] To compare our results with results for the "focal" colors of Berlin and Kay,[2] a subset of the stimuli were also characterized by reflectances that correspond to Munsell renotated color samples.[34,e] The Munsell best-exemplar samples used by Berlin and Kay, hereafter referred to as "focals," are surface color papers described using hue, value, and chroma parameters under a standard daylight C-illuminant.[31]

The 110-item stimulus sample includes 11 *best exemplar* samples, 8 from typically assessed color categories (i.e., red, green, yellow, blue, orange, brown, purple, pink), plus 3 from the categories of peach, turquoise, and chartreuse (or lime). The initial 8 OSA best exemplars were chosen based upon the mean centroid values for subjects assessed by Boynton and Olson[10] (p 100, Table IV). The additional three categories were included based on prior empirical work suggesting that they are psychologically salient and candidates for new emergent basic color terms.[10,11,35] The best exemplar OSA samples for peach, turquoise, and chartreuse were determined by experimenter consensus (five individuals with normal color vision).

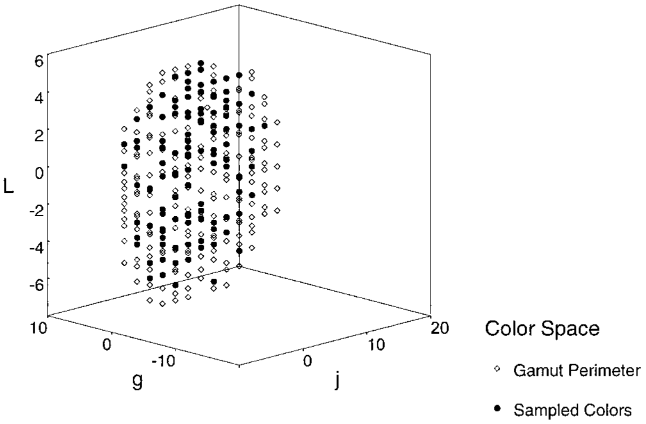

FIG. 1. Distribution of sampled color appearances in OSA space.

The 99 nonfocal stimuli were identified by a heuristic designed to sample the entire OSA stimulus space systematically and isolate a set of items representative of the area of each of the OSA levels. We felt it was important to sample items irrespective of the actual steps or spacing of the OSA within-level steps in order to (1) obtain color-naming results for the full range of variation of color space; (2) avoid any biasing structure that might be inherent in the spacing of the OSA color-space metric; and (3) avoid the selection of a stimulus set that was either uniform or optimized for saturation components across the hue dimension. For the latter, see saturation values for the Berlin and Kay[2] stimuli in Backhaus *et al.*[36] Figure 1 depicts the sampled stimuli scaled by *L,j,g* parameters of the OSA stimulus solid. Note that stimuli are thoroughly and consistently sampled across the entire space. Appendix A describes the selection heuristic.

The resulting stimuli are listed in Appendix B, Table B-1, according to their closest OSA *L,j,g* triples, and according to their measured CIE 1931 chromaticity coordinates.[37] The chromatic properties of stimuli are represented in the (*x,y*) chromaticity plane of the CIE 1931 standard observer.[31] The 110 color stimuli were rendered using an Apple Color StyleWriter 2400 inkjet printer within the most acceptable visual match of the OSA counterparts and subsequently measured with a Pritchard PR704 spectrocolorimeter and determined within an acceptable range of Delta-E($L^*a^*b^*$) tolerance.[31] In the analyses that follow, data for one of the 99 OSA grid-sampled color appearances (item 47) were

---

[d] Unpublished data from the World Color Survey may represent a similar stimulus presentation (see Kay and Berlin[5] for a brief description).

[e] The Newhall *et al.*[35] renotation data are used in this analysis rather than more recent measures of the Munsell stimuli, because they give the closest approximation to the stimuli used by Berlin and Kay.[2]

TABLE I. Between-language comparisons on mean measures of agreement and consensus.

	Measure	Monolingual English vs. monolingual Vietnamese	Monolingual English vs. bilingual Vietnamese	Bilingual Vietnamese vs. monolingual Vietnamese
1	Agreement percentage across all samples	29.1	48.9	41.8
2	Agreement percentage for blue–green samples	2.9	0	26.5
3	Wilcoxon test (two-tailed) for frequency of modal name	$z = 7.67, p = 0.00$	n.s.	$z = 8.25, p = 0.00$
4	Wilcoxon test (two-tailed) for variability	$z = 8.57, p = 0.00$	$z = 6.98, p = 0.00$	$z = 8.99, p = 0.00$
5	Paired t test (two-tailed) for agreement index	$t(108) = 7.78$, $p = 0.00$	$t(108) = 2.37$, $p = 0.02$	$t(108) = 8.57$, $p = 0.00$
6	Spearman correlation (two-tailed) for frequency of modal name	$r = .41, p < 0.01$	$r = .54, p < 0.01$	$r = .43, p < 0.01$
7	Spearman correlation (two-tailed) for variability	$r = .64, p < 0.01$	$r = .59, p < 0.01$	$r = .55, p < 0.01$
8	Pearson correlation (two-tailed) for agreement index	$r = .45, p < 0.01$	$r = .45, p < 0.01$	$r = .55, p < 0.01$
9	Paired t test (two-tailed) for confidence ratings	$t(109) = 18.74$,* $p = 0.00$	$t(109) = 13.32$,* $p = 0.00$	$t(109) = 9.3$,* $p = 0.00$
10	Spearman correlation (two-tailed) for confidence ratings	$r = .65, p < 0.01$	$r = .77, p < 0.01$	$r = .65, p < 0.01$

* Comparison based on 110 samples (item 47 not removed).

eliminated, because subsequent colorimetric measures showed that the sample duplicated (within rendering tolerance) one of the centroids assessed (item 110, turquoise), leaving a total of 109 color appearance samples.

The rendered stimuli measured 1-inch square. Each stimulus was centered on a 3-inch-square neutral gray background (closely approximating Munsell neutral gray 5), leaving on all sides a 1-inch gray border serving as a neutral visual context. The entire configuration was centered on letter-size white paper. The estimated viewing distance was approximately 15 inches, with the stimulus placed flat in the horizontal position, and with the illuminant directly overhead. Specular reflections were minimized by the matte surface of the printed samples and the viewing angle of the stimulus relative to the illuminant position.

An important goal of our study was to compare our results with those obtained by Berlin and Kay,[2] and by Boynton and colleagues.[10,11,15,16] Boynton and Olson[10] centroids for the 11 above-mentioned English categories are listed in Appendix B, Table B-2. The focals for English and Vietnamese empirically identified by Berlin and Kay[2] (Appendix I) are also listed in Table B-2, which provides both the Munsell H V/C notation for the focals and the rendered OSA approximate used in our study. In addition, the CIE 1931 x,y chromaticity coordinate equivalents are given for the Munsell samples (from Wyszecki and Stiles[31]), and our rendered OSA samples and the Delta-xy difference values are presented. Delta-E differences between centroids and focals are perceptually similar enough to permit a comparison of our best exemplar results with those found by the previous researchers mentioned. In the analyses presented below, the term "centroids" refers to the Boynton and Olson[10] sample equivalents, whereas the term "focals" refers to the Berlin and Kay[2] sample equivalents.

Results and Discussion of Experiment 1

Predictions of the Berlin and Kay Model. Two sets of predictions were tested for the Berlin and Kay model. First, if neurophysiology determines perceptual salience, then basic focal colors identified in previous research should be responded to in a consistent manner by individuals across our three language groups. Basic focal colors should produce greater confidence, greater agreement among subjects, and less variability in naming than nonbasic colors, and the same basic focal colors should be identified as in previous cross-cultural surveys. Second, we predicted that altering task demands should not affect such findings because the underlying stimulus-dependent salience should produce consistent response regardless of task. Thus, the Berlin and Kay model predicts (1) similarity of results across languages, and (2) conformance to previous findings.

First, we determined whether the same color samples were assigned names with the same meanings across the language groups. To test this, the modal response to name each sample was identified. The modal response was defined as the single response free listed with the highest frequency to name each color appearance sample. All 109 modal responses given in Vietnamese were translated to English, and the percentage of agreement was calculated for each pair of language groups.[f] Percentage of agreement was defined, across all 109 stimuli, as the number of matches between the names given in two languages, divided by the

---

[f] Translations were initially made by a native speaker of Vietnamese who was fluent in English, then reviewed by a native English speaker to ensure that the same words in Vietnamese were consistently translated to the same words in English. The different word order for modifiers in Vietnamese compared to English was handled consistently and appropriately during translation.

TABLE II. Color terms appearing most frequently as the modal response for multiple samples within each language.

Monolingual English		Bilingual Vietnamese		Monolingual Vietnamese	
Term	No. of Samples	Term	No. of Samples	Term	No. of Samples
Green	11	*Xanh*	16	*Xanh la cay*	8
Orange	11	*Cam*	15	*Cam*	6
Yellow	9	*Xanh la cay*	11	*Hong dam*	6
Pink	7	*Tim*	10	*Tim*	6
Purple	7	*Vang*	10	*Nau*	5
Blue	6	*Hong*	9	*Vang*	5
Brown	6	*Nau*	9	*Vang dam*	4
Red	5	*Do*	7	*Vang lot*	4
Peach	5	*Vang lot*	4	*Xanh nuoc bien*	4

total number of samples named (109), multiplied by 100. In order to be considered a match, the modal response given for a sample must have been the same in each of the languages compared.[g] Percentages for all language pairs are shown in Table I (row 1). The overall percentage of times the modal response matched in all three languages was 25.5%.

Table I shows a low percentage of matches between Vietnamese and English, higher for bilingual speakers than for monolingual Vietnamese speakers. Surprisingly, there is also poor agreement between bilingual and monolingual Vietnamese speakers. We believe there are two sources of disagreement producing these results: (1) differences in categorization of colors, and (2) differences in the use of modifying terms. As noted by Berlin and Kay,[2] the most obvious differences in color naming between the English and Vietnamese languages are the categorization of orange, blue, and green. Blue and green are treated as two different categories in English, but are designated using a single category name (*xanh*) in Vietnamese. Within the larger category of *xanh* (undifferentiated blue or green), distinctions between colors are noted by modifying this basic term (e.g., *xanh la cay*, or leaf green, compared to *xanh nuoc bien*, or ocean blue). Orange is a distinct category in English and rivals the other basic colors (first noted by Chapanis,[38] and recently by Schirillo[28]), but Vietnamese has no basic term for orange.[2] In Vietnamese, orange is usually designated by a modified term for yellow, but less frequently is designated as *cam* (a Vietnamese term that glosses the name of the fruit "orange," as occurs in English).

To confirm this source of disagreement and to evaluate how much of it is due to the blue-green category, Table I (row 2) shows the percentages of agreement between language groups for the color samples designated blue or green in either language. This was calculated by dividing the number of matching terms designating blue or green by the total number of color samples evaluated (109), then multiplying by 100. As shown in Table I (row 2), there is no agreement between bilingual Vietnamese and English speakers on the modal name for samples designated as green or blue after translation. Thus, all of the matches occur in other color categories. In contrast, 63% of the total matches between bilingual and monolingual Vietnamese speakers are for samples whose modal names are within the green-blue category. Thus, the source of disagreement is different when different pairs of language groups are considered. Disagreement occurs between the bilingual and monolingual Vietnamese responses because bilingual Vietnamese (responding in Vietnamese) tend to categorize orange more similarly to English speakers, using the term *cam* more frequently than a modified yellow term (see Table II). Monolingual English and bilingual Vietnamese primarily disagree on blue-green samples. And English and monolingual Vietnamese disagree on a combination of category terms.

The second source of disagreement between English and Vietnamese speakers is the use of modifying terms (e.g., *light* pink, *sky* blue). In general, both bilingual and monolingual Vietnamese speakers use a larger number of modifiers added to monolexemic color terms than English speakers do, resulting in multiple-word combinations (e.g., *xanh la cay* or *xanh nuoc bien*). This result parallels that found by Lin *et al.*[23] in comparing British English with Mandarin Chinese naming. This can be seen in Table II, which lists the frequency of occurrence of specific color terms among the modal responses in each language.

A more detailed way of showing how naming varies across the entire color space is to track measures of naming behavior over some logical partitions of color space. In the group-wise comparisons presented below, we compare naming for subsets of stimuli defined by OSA levels (similar to the response time analysis by Boynton and Olson,[11] Fig. 5). This permits an evaluation of whether the three language groups exhibit similar profiles of naming across the entire color space, or whether their profiles differ in a meaningful fashion across the color space tested. Such an analysis allows the tracking of naming trends across the lightness dimension of color space. In addition, analyses of subsets of stimuli defined by levels permit valid comparisons of similar size sets of noncentroids against centroids (centroid–noncentroid comparisons are presented later). Figure 2 shows the mean number of different monolexemic terms listed in each language, by OSA level (L value). Note

---

[g] Data reduction analyses that relax this strict notion of matching are discussed later in this article.

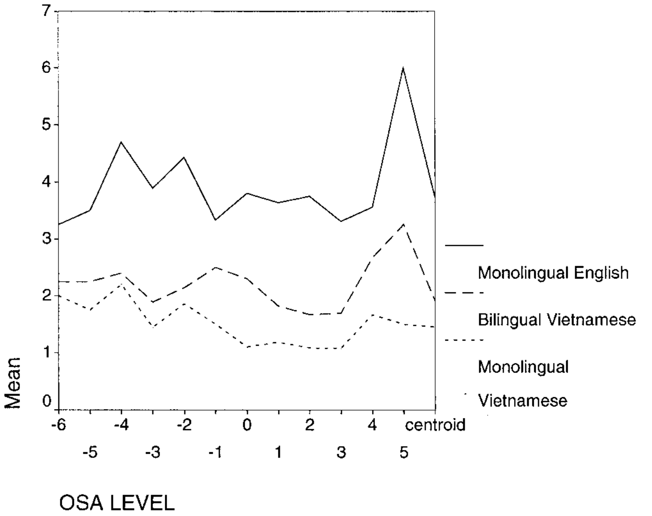

FIG. 2. Mean frequency of different monolexemic terms listed per sample by OSA level and language group.

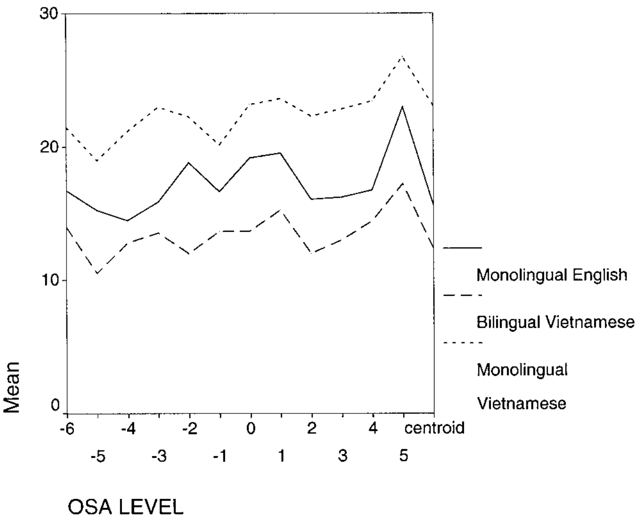

FIG. 4. Mean number of different terms used per sample (variability) by OSA level and language group.

that both Vietnamese groups listed fewer monolexemic terms than English speakers. However, compared to monolingual Vietnamese speakers, bilingual Vietnamese speakers used more monolexemic terms with fewer modifying terms, and their patterns of use by level more closely paralleled those of English speakers (see Figs. 3 through 5).

Accessibility of monolexemic terms is an issue in interpreting previous studies. As shown in Table II, monolingual Vietnamese speakers did not prefer to use monolexemic terms compared to modified terms (5 of the 9 highest frequency names were modified, not monolexemic), whereas English speakers showed higher agreement for use of monolexemic terms and employed a wider variety of such terms in naming (see Fig. 2). Monolingual Vietnamese speakers constrained to use monolexemic terms in a speeded task will be at a disadvantage with respect to accessibility, because such terms are not used with the same frequency as in English and perhaps other languages. How valid is their naming behavior under such a constraint? It may be that those terms found to be "basic" in such tasks are basic by virtue of being monolexemic in the language and thus more readily accessible when performing the task. Indeed, a normative survey of English found that terms for *Red*, *Yellow*, *Green*, *Blue*, *Orange*, *Purple*, *Brown*, and *Pink* (all but the achromatic terms from the Berlin and Kay 11 basic color terms) were the most frequently appearing color terms in the language.[39] If frequency of use of these monolexemic terms in Vietnamese is not on a par with English, are cross-language comparisons of response time and consensus fair to make?

These comparisons show the overall differences between language groups, but what consistency exists for those color appearances identified as most salient by previous researchers? It would be expected that salient color appearances

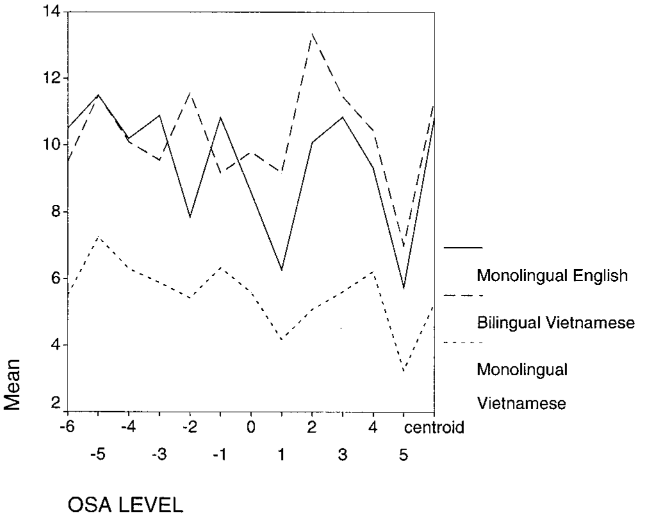

FIG. 3. Mean frequency of the modal term listed per sample by OSA level and language group.

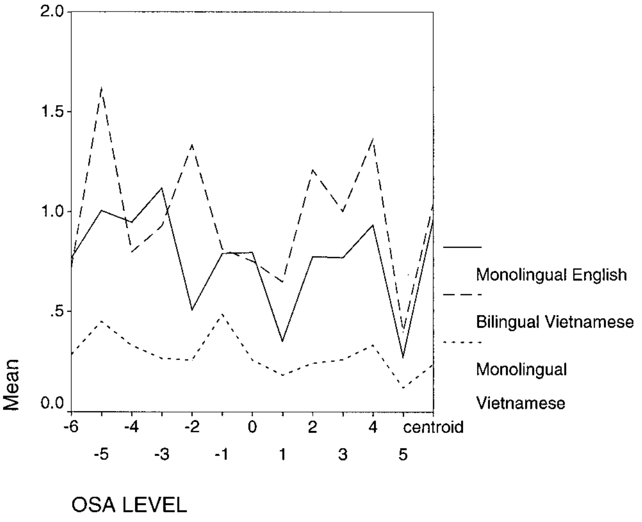

FIG. 5. Mean agreement index per sample (frequency divided by variability) by OSA level and language group.

TABLE III. *L,j,g* parameters of OSA samples[34] rendered as experimental stimuli and empirically ranked by frequency of listing for the modal response within each language.

Monolingual English				Bilingual Vietnamese				Monolingual Vietnamese			
Stimulus ID	Category	OSA *L,j,g*	Frequency	Stimulus ID	Category	OSA *L,j,g*	Frequency	Stimulus ID	Category	OSA *L,j,g*	Frequency
502	orange	0,6,−6	24	506	*tim*	−2,−4,−2	26	11	*vang*	4,8,0	16
89	purple	−3,−3,1	23	11	*vang*	4,8,0	25	503	*xanh la cay*	0,4,4	13
9	yellow	4,6,0	20	38	*cam*	2,8,−6	21	504	*vang*	3,7,−1	11
94	brown	−3,1,−1	19	20	*hong*	3,1,−3	20	506	*tim*	−2,−4,−2	11
78	brown	−4,2,−2	19	27	*vang*	2,4,0	20	92	*tim*	−6,−2,−2	11
92	purple	−6,−2,−2	18	84	*tim*	−5,−3,1	20	9	*vang lot*	4,6,0	10
55	green	−3,3,5	18	19	*vang*	4,10,0	19	41	*xanh la cay*	2,6,4	10
37	pink	2,2,−4	18	34	*hong*	3,−1,−3	19	73	*xanh duong*	−4,−4,2	10
38	orange	2,8,−6	17	77	*nau*	−3,3,−3	19	75	*xanh la cay dam*	−3,3,1	10
66	green	−1,3,3	17	502	*cam*	0,4,−4	19	84	*tim*	−5,−3,1	10
84	purple	−5,−3,1	17	504	*vang*	3,7,−1	19	90	*nau*	−5,−1,1	10
504	yellow	3,7,−1	17	10	*vang*	3,11,−1	18	97	*tim dam*	−4,−2,0	10
508	pink	3,−1,−5	17	29	*cam*	2,4,−4	18	508	*hong*	3,−1,−5	10
20	pink	3,1,−3	16	94	*nau*	−3,1,−1	18	22	*vang dam*	3,7,−3	9
28	yellow	2,8,0	16	30	*cam*	2,6,−6	17	94	*nau*	−3,1,−1	9
509	lime green	1,5,3	16	89	*tim*	−3,−3,1	17	20	*hong*	3,1,−3	8

might be more consistently named in both languages. Our results show that this is not the case. Although the same color terms (e.g., yellow, green, orange) are listed frequently as the modal terms in each language (see Table II), they are not necessarily assigned to the same samples. In each language, there appear certain samples that receive high frequencies of modal response (suggesting high agreement in naming), but these are not the same samples across languages. Table III lists the samples with the highest modal frequencies (the same name listed by the most subjects) within each language. Taking the 16 samples with the highest frequencies (the point where an elbow occurs if modal frequency is plotted), only 5 samples appear on all three lists, and none of these appear among the top 5 samples with the highest frequencies in any language. As can be seen in Table III, different samples produced the highest modal frequencies in each language. Thus, even where considerable agreement exists about the name for a sample, the samples evoking such agreement are different within the three different language groups. Furthermore, only 4 of the color appearances shown in Table III are centroids. Thus, with the task modifications made in this study, the hypothesized salience of the centroids does not appear to result in greater agreement about naming, despite the greater use of basic color terms (see Table II). However, this analysis provides only a rough measure of agreement and does not take into account the impact of the free use of modifiers. A more detailed comparison of centroids and noncentroids is provided below, followed by analyses that use data reduction to remove the impact of differential use of modifying terms across languages.

Descriptive Comparisons of Color Naming. To compare naming of centroids and noncentroids more directly, four quantitative variables describing naming behavior within each language group were created: (1) frequency, (2) variability, (3) monolexemic term use, and (4) agreement index. Means for these variables are presented in Table IV.

Frequency was defined as the frequency with which the modal term for each color sample was listed. Figure 3 compares mean frequency for noncentroids across OSA levels (L values) and centroids by language group. Use of partitions of noncentroids defined by OSA level serves two purposes: (1) tracking naming for a meaningful dimension of color space, and (2) providing partitions of noncentroids that are comparable statistically with the 11-item centroid partition. Bar graphs are generally more appropriate for displaying categorical frequencies, but line graphs are used to present this data, in order to permit easier visual comparison across the three language groups. No continuity between OSA levels, or between such levels and the centroids, is implied.

Variability was defined as the number of different terms listed for each color sample. Figure 4 compares mean variability across OSA levels (L values) by language group. For both frequency and variability, the criteria used for determining similarity of terms were identical to those used to determine matches across languages, as described earlier, except that Vietnamese terms were not translated.

The mean number of different monolexemic terms listed to name color samples was compared across OSA levels by language group, as shown in Fig. 2. We defined an agreement index by dividing the frequency by the variability for each color sample. This produces a more sensitive measure of consensus than either frequency or variability alone, because it describes the degree of concordance between both mode and range of naming. In essence, the agreement index appropriately captures degree of agreement, or deno-

TABLE IV. Mean measures of agreement and confidence by language.

	Monolingual English	Bilingual Vietnamese	Monolingual Vietnamese
Modal frequency	9.60	10.61	5.50
Variability	16.97	13.29	22.61
Agreement index	0.79	0.98	0.28
Monolexemic term use	3.83	2.08	1.46
Confidence rating	3.7	4.1	4.3

tative codability of a given color name, relative to the dispersion of naming choices. Figure 5 compares mean agreement indices across OSA levels (L values) by language group.

Frequency and variability of bilingual Vietnamese were more similar to those of English speakers than to monolingual Vietnamese, as shown in Table IV. The greater variability of the monolingual Vietnamese appears strongly related to the liberal use of stem terms plus modifiers in naming color variations, whereas English speakers appeared to use a wider variety of monolexemic terms when naming such variations (cf. Lin *et al.*[23]). Together, the measures suggest greater cohesion of response among subjects within the bilingual Vietnamese group (higher frequency, lower variability, higher agreement index), especially when compared to monolingual Vietnamese (see Romney *et al.*[40] for a similar bilingual result in a different semantic domain).

We compared means for frequency and variability of naming between languages using paired sample, two-tailed Wilcoxon signed ranks tests, and mean agreement indices using paired sample, two-tailed *t* tests. As shown in Table I, rows 3–5, significant differences were found in pair-wise comparisons of all three language groups for all measures except frequency, where no significant difference was found between monolingual English and bilingual Vietnamese speakers.

Stimulus-related similarities in color naming were assessed by correlation of the frequencies, variability, and agreement indices across language groups, as shown in Table I, rows 6–8. We assumed that if subjects responded similarly to the color samples and used language in a similar manner, these measures should be positively correlated with each other across languages. Such an assumption implies that the same color samples should result in increased frequency or increased variability regardless of which language is assessed. We performed a two-tailed Spearman rank-order correlation for all comparisons except agreement indices, which were compared using a two-tailed Pearson correlation. Correlations are shown in Table I, rows 6–8. Because these correlations are higher than the rough comparison of samples shown in Table III would suggest, it seems likely that the correlations depend as much on disagreement among subjects as they do on agreement. Disagreement would be indicated by low modal frequencies with high variability of naming.

Confidence Judgments for Color Naming. In addition to the free listed names, subjects were asked to rate their confidence in each name given, on a scale from 1 to 5 (with 1 indicating lowest confidence, and 5 highest confidence). Mean confidence ratings for samples grouped by OSA level (L values) are shown in Fig. 6 and listed in Table IV. Paired-sample *t* tests showed significant differences in mean confidence ratings among all three language groups, as described in Table I, row 9. Inspection of responses showed that monolingual Vietnamese underused the lower regions of the rating scale and that many subjects gave a maximum rating of 5 to nearly every sample. This is consistent with cross-cultural differences in rating scale usage noted in other rating contexts by previous researchers.[41] This ceiling effect in the monolingual Vietnamese ratings would tend to restrict range and thereby depress correlations with raters using the whole rating scale.[42] Nevertheless, confidence ratings are highly correlated across languages, as shown in Table I, row 10, suggesting that, in general, samples that elicited lower confidence ratings in one language tended to do so in the other as well. This interpretation is supported by the observation of a stronger correlation between confidence and variability than frequency, as shown in Table V. Note that across language groups, confidence ratings for the monolingual English and bilingual Vietnamese are most highly correlated (Table I, row 10).

TABLE V. Correlations between confidence ratings and frequency, variability and agreement index by language.

	Frequency	Variability	Agreement Index
Monolingual English	.59*	−.77*	.68*
Bilingual Vietnamese	.50*	−.59*	.57*
Monolingual Vietnamese	.60*	−.71*	.66*

* $p < 0.01$, two-tailed

Comparisons of Centroid Naming With Noncentroids. Previous research suggests that differences on the measures described above should be found between the group best-exemplars or centroids identified by Boynton and Olson,[10] and the remaining noncentroid colors sampled in this study. There should be higher frequency, lower variability, higher agreement indices, and higher confidence ratings for the centroids than for the noncentroids. No prediction was made about the use of monolexemic terms. By placing all of the centroids in a single group and comparing them with roughly equal-sized groups of noncentroid samples, segregated by OSA lightness level, the benefits of salience were expected to accumulate and to be more readily visible in the measures. This strategy gives centroids greater opportunity to show a statistically significant difference compared to the remaining noncentroid samples.

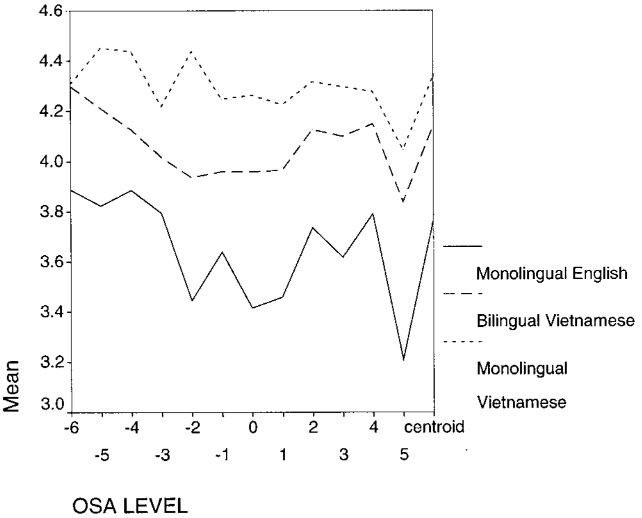

FIG. 6. Mean confidence rating per sample by OSA level and language group.

TABLE VI. Comparison of centroid frequency and variability with noncentroids occurring at the same OSA level and assigned the same color name by subjects.

		Monolingual English			Bilingual Vietnamese			Monolingual Vietnamese		
Stimulus ID	OSA *L,j,g*	Category	Centroid frequency	Non centroid frequency	Category	Centroid frequency	Non centroid frequency	Category	Centroid frequency	Non centroid frequency
Frequency										
504	0,6,−6	yellow	9	20	*vang*	18	25	*vang*	3	16
508	−3,−3,1	pink	6	*	*hong*	7	7	*hong*	5	*
502	4,6,0	orange	24	4	*cam*	16	8	*cam*	4	*
505	−3,1,−1	light blue	15	*	*xanh*	8	6	*xanh bien*	5	*
501	−4,2,−2	hot pink	7	16	*do*	13	14	*hong dam*	5	7
507	−6,−2,−2	light brown	14	19	*nau*	19	13	*nau lot*	7	8
506	−3,3,5	purple	10	*	*tim*	10	*	*tim*	3	*
503	2,2,−4	green	14	6	*xanh la cay*	9	10	*xanh la cay*	13	10
Variability										
504	0,6,−6	yellow	18	6	*vang*	8	4	*vang*	30	13
508	−3,−3,1	pink	15	*	*hong*	13	17	*hong*	21	*
502	4,6,0	orange	6	23	*cam*	11	16	*cam*	23	*
505	−3,1,−1	light blue	13	*	*xanh*	15	15	*xanh bien*	21	*
501	−4,2,−2	hot pink	21	12	*do*	12	12	*hong dam*	22	19
507	−6,−2,−2	light brown	15	9	*nau*	7	11	*nau lot*	22	18
506	−3,3,5	purple	15	*	*tim*	11	*	*tim*	22	*
503	2,2,−4	green	13	10	*xanh la cay*	13	12	*xanh la cay*	20	16

* No similarly named sample was available at the same OSA level.

The eight basic color category centroids identified by Boynton and Olson[10] are listed in Appendix B, Table B-2. For each of the measures analyzed earlier (frequency, variability, agreement index, monolexemic terms, confidence ratings), means for the centroids were compared to means for the remaining samples, classified by OSA level (L values). One-way, two-tailed analysis of variance (ANOVA) was used to test predictions, with a significance level of $p < 0.05$. Although this is categorical data, ANOVA has been shown to produce valid results and to be more informative than chi-square analysis for qualitative data with no extreme dichotomous responses.[45] No significant or near-significant differences, and no extremely nonsignificant differences ($p < 0.90$) were found for any measure except variability in the English language group, $F(12, 108) = 1.91$, $p = 0.04$.[h] As can be seen in Figs. 2 through 6, there is considerable fluctuation in the mean values from level to level on these measures. As shown in Fig. 4, variability of naming is highest for samples at level 5 (the most "white" lightness level) and this is the difference that is statistically significant from the other levels and the centroids. For all variables considered (i.e., modal frequency, variability, agreement indices and confidence), the mean values for the centroids are little different than those for other samples occurring at the same levels, and do not show trends that might be expected if the centroids had greater salience for naming purposes. It might be argued that an ANOVA across all OSA levels introduces too much variability to permit significant differences to be observed. Statistically stronger t tests comparing all measures for centroids with only the samples at level 5 (where no centroids or near-centroids are found) showed no significant or near-significant differences in any language group on any measure except variability.

In a more direct test, Table VI compares each centroid with a noncentroid sample at the same OSA level and assigned the same color name (where available). Note that several noncentroid samples produced higher frequencies, lower variability, and greater confidence than the comparable centroid in each of the languages. A sign test showed that mean frequency and variability of the centroids compared to these selected noncentroids did not vary significantly in the expected direction in any language. Variability tended to move in the opposite direction in both the monolingual English and monolingual Vietnamese groups (i.e., greater variability was found for centroids than for noncentroids).

We performed a similar analysis using an expanded set of centroids, including three additional samples hypothesized to be untested candidate centroid colors for emergent basic color categories: turquoise, peach, and lime green. Using one-way, two-tailed ANOVA to evaluate this expanded set of centroids, we found no significant differences between centroids and noncentroids classified by OSA level in any language group for any measure.

These results seem contrary to findings by Heider-Rosch[14] and others, establishing the importance of focal exemplars for memory and perceptual tasks. We emphasize that we are not questioning the property of salience *per se*,

---

[h] Moreover, no significant differences were observed for any ANOVA measures reported for all three groups when the two centroids with the most rendering deviation—the red and yellow centroids—were eliminated from the analyses (see Table B-2, Delta-E measures).

but rather the grounding of salience in known neural physiology and its linkage to a defined set of color appearances named by using basic terms. As noted earlier, the basis for salience in Heider-Rosch's work is likely to have been saturation of the color samples. Thus, Heider-Rosch's work does not establish that salience results from fundamental neural responses to spectra, defining basic colors, linked to basic color terms.

Reduction to Fewer Categories. As described earlier, a very strict criterion was applied in determining similarity of naming responses. Only spelling errors were regularized. Different word orders were considered different names; thus, yellowish-green was considered a different name than greenish-yellow. Yellow-green was considered a different name than yellowish-green. Light light green was considered a different name than light green.

It might be argued that this strict criterion, coupled with the availability of modified terms as opposed to monolexemic terms, makes it difficult for centroid agreement to exist, within or across languages. On the contrary, we consider it likely that a procedure permitting more precise color naming would enhance the differential effect of salience on naming, not obscure it, if salience does result in greater confidence, greater agreement, and less variability in naming, as widely believed. Because one of the hypothesized properties of salience is the greater tendency for a salient color appearance to be named with the use of basic terms, access to modifiers should not have affected the centroids as much as the noncentroids. Furthermore, our procedures were applied consistently and should have affected all three language groups in the same manner. We found that centroids showed no greater salience than noncentroids on any measure. Thus, we conclude that shared perceptual experiences arising from common neurophysiologic mechanisms do not seem to be the basis for higher agreement about centroids in either language tested, when the same color appearances are named in an unrestricted manner under the same viewing conditions.

Nevertheless, because we used such a strict criterion for evaluating the frequency and variability of naming, we examined the extent to which these null findings might have depended on the method of classification of terms by performing several alternative forms of data reduction. We then retested the predictions made for centroids versus noncentroids in each language group. To more closely approximate the constraints inherent in the response format experienced by Boynton and Olson's subjects, our free listed terms were reduced to a single monolexemic term. In general, this consisted of eliminating modifiers, and little experimenter judgment was required. Reduction of the Vietnamese terms was performed in two ways: (1) before translation (with undifferentiated blue and green considered a single category); and (2) after translation (with terms translating to blue considered one category, those translating to green a different category, and those translating to undifferentiated blue–green a third category). Thus, two different data reductions were performed for each of the two Vietnamese groups. Data reduction of the Vietnamese terms was performed by a native speaker of Vietnamese. All data analyses described above were then performed again.[i]

Data reduction removes much of the difference across language groups—a finding consistent with the results of Boynton and colleagues, and in accord with our suggestion that monolexemic naming shapes salience results. As shown in Fig. 7, means for frequency converge as the number of categories is reduced (cf. Fig. 3). Revised means for frequency and variability are shown in Table VII (Berlin and Kay focal colors are discussed in the next section). Revised Spearman rank order correlations among confidence, frequency, and variability after data reduction are shown in Table VIII. Frequencies, variabilities, and their correlations with confidence ratings appear to improve as the specificity of naming increases. For example, they are higher for Boynton and Olson's less restricted monolexemic categories than for Berlin and Kay's more limited basic categories, and higher when green and blue are considered separate categories rather than a single category. This suggests that the magnitude of the confidence ratings may be related to the specificity of naming, rather than to the characteristics of particular color samples.

One-way, two-tailed ANOVA was used to compare mean frequencies and variability across levels within each language group. Significant differences were found only for the variability measures within both data reduction methods (translated and untranslated), and only within the bilingual Vietnamese language group. As before, the main difference occurred between Level 5 and the lowest levels. Level 5 is the lightest OSA level, or most "white," and subjects offer many modified terms for "white" tinged with some other hue. The same high variability in Level 5 occurred within the monolingual Vietnamese group, but higher variability also existed in the remaining levels, making the difference statistically nonsignificant.

Comparison of Focal Color Naming With Nonfocal Color Naming. In order to compare results of this study with results obtained by Berlin and Kay,[2] we recategorized the free-listed responses based on their 11 hypothesized color categories: red, yellow, blue, green, orange, purple, pink, brown, gray, black, and white.[j] As predicted for the Boynton and Olson centroids, focal colors were expected to show higher frequencies, lower variability, higher agreement indices, and higher confidence ratings compared to nonfocal colors within each language group. Each color name listed by a subject was assigned to one of Berlin and Kay's 11 categories. This required some experimenter judgment but was performed in a consistent manner, with disputes re-

---

[i] Of course, using such reduction methods to translate free-listed naming data into monolexemic naming data does not replicate the task that Boynton and Olson's subjects carried out. Rather, it is used here to approximate their data and thereby make reasonable comparisons between our findings and those of previous researchers. We do not endorse data reduction as the preferred means for obtaining monolexemic naming when such data are desired.

[j] In this study, the achromatic colors (gray, black, and white) were not assessed.

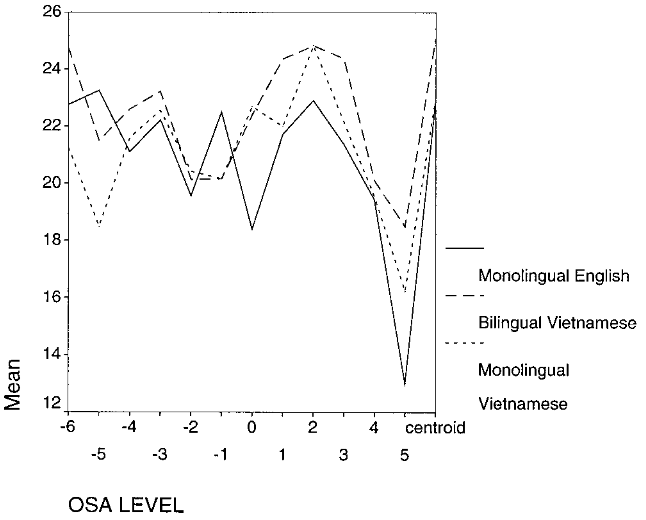

FIG. 7. Mean frequency of the modal term following data reduction to monolexemic stems with blue and green as a single category, by OSA level and language group.

solved by discussion. For example, the difficult-to-classify item "gold" was discussed and then consistently assigned to the category yellow in all three language groups. As described earlier for centroids, Vietnamese terms were classified both before and after translation. Frequency, variability, and agreement indices were then recalculated. Berlin and Kay's identified focal colors are listed in Table B-2 for English and Vietnamese. Note that different focal colors are hypothesized for the two languages. Note also that this analysis makes use of the closest OSA approximate to the Munsell focals of Berlin and Kay. Although the rendering of focals was not perfect (e.g., Delta-$L^*a^*b^*$ values are presented in Appendix B,Table B-2), results of Experiment 2, which directly compared naming for focals and centroids, support our claims that these close approximates were adequate to test focal salience. We performed analyses on English results using the English focal colors, and on bilingual and monolingual Vietnamese results using the Vietnamese focal colors (see Appendix B, Table B-2).

Like the centroid analyses presented above, the focal color analyses compared data for focals in each language group against similar size partitions of nonfocals defined by OSA levels. For each of the measures analyzed earlier (frequency, variability, agreement index, monolexemic terms, confidence ratings), means for the focal colors were compared to means for the remaining samples, classified by OSA level (L values). One-way, two-tailed ANOVA was used, with a significance level of $p < 0.05$. Within the English language group, using English focal colors, we found only one significant difference between nonfocal colors and focal colors and that was for variability, $F(12, 109) = 2.923$, $p = 0.00$. Within the bilingual Vietnamese language group, using Vietnamese focal colors, we found significant differences only for variability and they existed for both methods of data reduction: before translation, $F(12, 109) = 2.961$, $p = 0.00$; after translation, $F(12, 109) = 2.743$, $p = 0.00$. Within the monolingual Vietnamese language group, using Vietnamese focal colors, we found no significant differences. Variability showed no tendency toward near significance for focal colors compared to nonfocal colors classified by OSA level. Thus, in this comparative analysis, salience of focals is not significantly differentiated from nonfocals. This finding is consistent with the free-named centroid analyses presented earlier. Implications of both are discussed below.

In summary, even with reduction to basic terms using a variety of schemes, we found no greater consistency or agreement in naming for centroids or focal colors. Thus, an often assumed neurophysiologically based salience seems to have no effect on naming behavior in a task where subjects are permitted to access the lexicon freely and are given sufficient time to make fine discriminations. The differential impact of use of modifying terms on agreement was eliminated in the data reduction and still no greater agreement or confidence was found for those color appearances hypothesized to have greater perceptual salience. Berlin and Kay's finding that basic terms are used more frequently than nonbasic terms was confirmed for these two languages. When differences in the use of modifiers were eliminated, color appearances were named in a highly similar manner across the three languages. None of the other predictions of the Berlin and Kay model regarding focal salience were supported by our findings.

Mapping of Terms to the OSA Color Space. Reduction to Berlin and Kay categories revealed an uneven distribution of terms across the levels, proportionate to the distribution of hues within OSA color space. Although the terms blue, green, and purple were applied to samples spread fairly evenly from Levels −4 to 3, orange was used only from Levels −1 to 4, red was used from Levels −3 to −1, pink was used from Levels 0 to 4, brown was used from Levels

TABLE VII. Revised measures of agreement by language following data reduction.

Classification Method	Mean frequency	Mean variability
Monolingual English		
Berlin and Kay categories	23.98	3.02
Monolexemic categories	21.11	5.74
Bilingual Vietnamese		
Berlin and Kay categories (before translation)	23.36	2.68
Berlin and Kay categories (after translation)	19.65	3.19
Monolexemic categories (before translation)	22.91	3.05
Monolexemic categories (after translation)	19.40	3.60
Monolingual Vietnamese		
Berlin and Kay categories (before translation)	23.56	3.32
Berlin and Kay categories (after translation)	21.25	3.84
Monolexemic categories (before translation)	21.74	5.56
Monolexemic categories (after translation)	19.50	6.06

TABLE VIII. Revised correlations between confidence ratings and frequency or variability by language, after data reduction.

	Berlin and Kay frequency	Berlin and Kay variability	Monolexemic frequency	Monolexemic variability
Monolingual English confidence	.63*	−.54*	.56*	−.46*
Bilingual Vietnamese confidence				
Before Translation	.36*	−.31*	.34*	−.30*
After Translation	.43*	−.46*	.61*	−.44*
Monolingual Vietnamese confidence				
Before Translation	.47*	−.43*	.44*	−.45*
After Translation	.60*	−.54*	.54*	−.49*

* $p < 0.01$, two-tailed

−5 to −1, and yellow was used primarily from Levels 2 to 5. These patterns are consistent across the three languages. So, in monolingual Vietnamese, yellow is used from Levels 0 to 5, red from −3 to −1, pink from 0 to 4, brown from −5 to −1, whereas blue, green, and purple are distributed across all levels except 4 and 5. The levels where terms occur correspond to the regions identified by Berlin and Kay[2] and higher frequencies tend to occur at the levels of saturation most similar to the Munsell samples they employed. This finding provides indirect confirmation of Berlin and Kay's mapping of terms to the mercator projection of the Munsell stimulus solid they used.

When color samples were selected based on their Berlin and Kay category membership (in English) and mean confidence ratings were plotted by OSA level, an inverted "U" shape emerged for red, pink, yellow, brown, and orange, suggesting that confidence in naming increases until it reaches a point of optimal saturation (as a function of lightness level) and decreases otherwise. For example, confidence peaks at Level 4 for yellow, as shown in Fig. 8. However, a more complex pattern across levels emerges for blue, purple, and green, as shown in Fig. 9 for blue. One interpretation of this jagged pattern is that alternate peaks in confidence occur at the focal points of subcategories within each larger category named by a term such as blue, perhaps corresponding to colors designated by using terms such as turquoise. A similar pattern occurs for purple and green, with confidence peaks where terms such as magenta or chartreuse (lime green) might be used. The larger number of OSA levels encompassed by the Berlin and Kay terms blue, green, and purple may give rise to such subcategories.[28]

Additional support for the idea that confidence is related to language usage rather than color characteristics is provided by the observation that confidence peaks exist at different levels for subjects in different language groups. This may be related to Berlin and Kay's[2] identification of different focal colors in the two languages, as listed in Appendix B, Table B-2. Partial confirmation of this speculation is found by examining further the confidence ratings for the category of yellow. Vietnamese has no consensually applied color term for orange; thus, monolingual speakers tend to apply either the term *vang* (yellow) or *do* (red) to samples called orange in English (and called *cam* by most bilingual Vietnamese subjects). The confidence rating plots

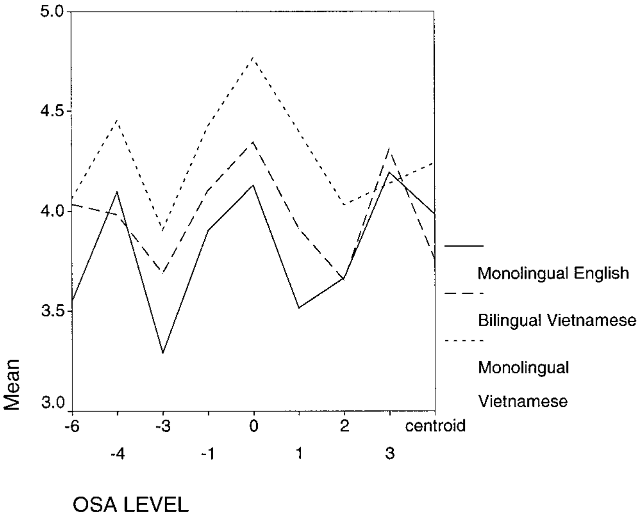

FIG. 9. Mean confidence rating per sample for only those items labeled "blue" by English speaking subjects after reduction to Berlin and Kay basic terms, by OSA level and language group.

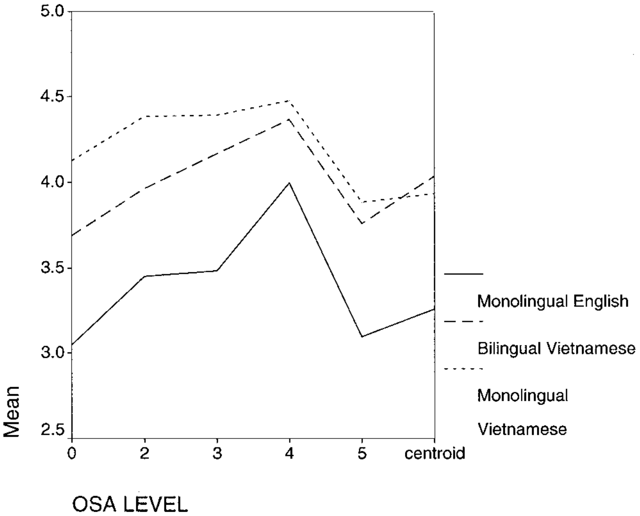

FIG. 8. Mean confidence rating per sample for only those items labeled "yellow" by English speaking subjects after reduction to Berlin and Kay basic terms, by OSA level and language group.

for yellow reflect this wider dispersion of term use by showing no clear peak in the monolingual Vietnamese ratings. Where the term yellow is applied to a smaller set of samples, there exists a clear peak in the English confidence ratings. This pattern is confirmed by inspection of plots for orange and for the samples labeled *vang* by monolingual Vietnamese subjects (omitted due to space constraints). In contrast, when colors are selected based on their English Berlin and Kay categorization in the category blue, the confidence peaks appear in the same places for Vietnamese speakers as for English speakers, even though the Vietnamese language contains no single word for blue. We believe this different pattern of confidence ratings demonstrates that confidence is more dependent on the goodness of fit between names and exemplars, than upon the qualities of the samples being named. In Vietnamese, blue and green are named in a highly consistent manner (see Table II) with the use of multiple word modifications of the term *xanh*. Samples called *orange* by English speakers, or *cam* by bilingual Vietnamese speakers, are called *vang dam* (dark yellow) or *do* (red) by monolingual Vietnamese speakers. Other samples called dark yellow by English speakers are also called *vang dam* by both bilingual and monolingual Vietnamese speakers. Thus, modifiers are not used to distinguish between orange and dark yellow, introducing ambiguity that may relate to the difference in confidence ratings of yellow samples among monolingual Vietnamese.

Experiment 1 seems to suggest that salience for color exemplars may be as closely linked to naming confidence as it is to aspects of perceptual salience inherent in the color sample stimuli. However, direct empirical study of relationships between color naming, salience, and confidence are required before any further discussion of such relationships is warranted.

EXPERIMENT 2

It might be suggested that Experiment 1's failure to find naming differences between best-exemplar stimuli and other stimuli sampled from the OSA space was attributable to a failure to perceptually reproduce the OSA *centroids* or the Munsell *focals* that other researchers have found to be psychologically salient. To rule out this possibility, Experiment 2 makes the following empirical checks: (1) an internal consistency check for centroid naming, and (2) a comparison of task effects for a monolexemic versus an unconstrained naming task. Experiment 2 compares three sets of equivalent stimuli: (1) the rendered centroid stimuli used in Experiment 1, (2) actual Munsell focal chips, and (3) OSA centroid tiles. Experiment 2 compares the same color appearances, rendered in three different ways, in order to determine whether stimulus or task effects might explain the lack of salience observed in Experiment 1. If, in Experiment 2, the naming of rendered centroid stimuli is indistinguishable from the naming of actual Munsell chips and OSA tiles, then it seems unlikely that the Experiment 1 failure to distinguish centroid salience from noncentroid salience is attributable to stimulus properties.

Participants

Two tasks (constrained monolexemic naming and unconstrained free listing) were presented to two groups (monolingual English speakers and bilingual Vietnamese speakers) in a 2 × 2 between-subjects design. The number of subjects ranged from 15 to 17 in each group. All subjects volunteered through the human subject pool and participated for partial course credit, and all were screened for normal (corrected) vision and for normal color vision, as described for Experiment 1.

Stimuli

Subjects were presented with 30 surface color samples, including 10 randomly selected focal hues as identified by Berlin and Kay,[2] 10 corresponding category centroids from Boynton and Olson,[10] and 10 best-exemplar equivalents from the set of rendered stimuli used in Experiment 1. Actual Munsell chips and OSA tiles were used to represent the focals and centroids. For example, if an English focal "red" was randomly selected, then the equivalent OSA tile was also selected (as defined by Boynton and Olson[10]), as was the equivalent rendered stimulus from our set. Stimuli used in Experiment 2 are listed in Appendix B, Table B-3.

Procedure

All aspects of the physical environment (including ambient illumination and viewing distance) were controlled as described in experiment 1. Subjects were shown each sample by an experimenter (the actual Munsell chip, actual OSA tile, rendered centroid stimulus) in random sequence and in isolation. The self-paced task was either to (1) name the stimulus with the use of a monolexemic term, or (2) name the stimulus freely (as described in Experiment 1). In each condition, the subject named all 30 samples twice in different random orders.

Results and Discussion of Experiment 2

Experiment 2 data were examined to address four issues: (1) within-subject naming consistency, (2) frequency of modal naming, (3) cross-language modal term congruence, and (4) denotative equivalence of best exemplars from the three different stimulus sets. Each of these issues is discussed below.

*Individual Naming Consistency.*This analysis evaluated whether individual subjects consistently named color stimulus samples twice. It was expected that a subject would apply the same stem term to the same sample each time it was encountered. The percentage of times this occurred was computed for each subject across the 30 samples. The mean percentages for each task group (monolexemic vs. freelist) were then compared, as shown in Table IX. For the free-list task, where modifiers were permitted, mean modifier use was also calculated. This was defined as the percentage of observed responses in which a stem term was qualified by

TABLE IX. Mean naming consistency, mean modifier use, and mean frequency of modal term use across subjects in two language groups and for two tasks.

	Monolexemic task		Free-list task	
	English (n = 16)	Vietnamese (n = 17)	English (n = 15)	Vietnamese (n = 16)
Naming consistency (%)	96.4	94.7	94.3	95.0
Modifier use (%)	—	—	44.0	36.2
Frequency of modal term use (%)	90.2	87.2	91.3	86.6

one or more modifying terms (e.g., pale, vivid, etc.). On average, subjects were highly consistent when repeatedly naming color samples (see Table IX, row 1). Two-tailed t tests compared (1) English-monolexemic and English-freelist conditions, and (2) Vietnamese-monolexemic and Vietnamese-freelist conditions. No significant differences were found in mean consistency of naming across tasks ($p < 0.05$) within each language. These data lend confidence to the within-subject reliability of naming in both Experiments 1 and 2.

*Modal Name Frequency.*This analysis examined whether the kind of task affected the choice of names assigned to stimuli. It tested whether the modal names given to color stimuli (with or without modifiers) were used with equal frequency across the two naming tasks (monolexemic vs. free listing). High modal frequencies of naming across tasks, if observed, would indicate within-group consensus regarding the modal stem terms used to name color samples, and suggest that across-task modal stem naming is similar for any given color sample. For this comparison, the percentage of subjects using the modal term to name each sample was computed. Table IX, row 3, shows the mean frequency of modal term use for the groups and tasks collapsed across color samples. Two-tailed t tests between the tasks within each language group revealed no significant differences. These data indicate that (1) on average, substantial within-group consensus exists regarding the modal stem term assigned to samples, and (2) even when modifiers are used (e.g., as permitted in the free-list task) the stem term name that is modified is used with the same level of frequency across tasks. This suggests that access to modifiers (constrained in the monolexemic task used here) does not substantially affect the choice of stem term assigned to a sample.

*Cross-Language Modal Term Congruence.*This analysis demonstrated that the design of Experiment 1 did not inherently preclude the possibility of differentiating best-exemplar salience from non-best-exemplar salience. It could be argued that Experiment 1 found no differences in salience for focals and centroids compared to other color samples, because the unconstrained use of modifiers made naming agreement exceedingly unlikely. To address this issue, Experiment 2 tested whether the monolexemic and free-list tasks were capable of yielding agreement in naming across languages. If the type of task (monolexemic vs. free list) had no impact on agreement in Experiment 2, then we can more confidently assume that our use of unconstrained naming did not preclude finding agreement for centroids in Experiment 1 either. Thus, if Experiment 2's two tasks and centroid stimuli allow for similar levels of naming agreement *across languages*, then it is not unreasonable to suggest that a within-language finding of differential salience for centroids relative to noncentroids was an allowable outcome in Experiment 1.

To assess cross-language congruence in naming, we counted the frequency with which an equivalent stem term was used to name the same color sample across the two languages. Names were considered *equivalent* glosses when they corresponded with the Berlin and Kay[2] translations. For example, *red* and *do* were equivalent, *yellow* and *vang* were equivalent, and so on. All such translations were verified by research assistants with native Vietnamese proficiency. Because different focal hues are identified by Berlin and Kay[2] for the two languages, different subsets of stimuli were necessarily used in the two language groups. Thus, only the mean modal-term frequencies for the 12 color stimuli common to both language tests were considered. Across languages, and within task, these mean frequency variables of 12 values were statistically compared. Two-tailed t tests ($p < 0.05$) showed no significant differences in the frequency of equivalent modal gloss use across languages for monolexemic naming or freelisting tasks. Although this test is limited by the modest overlap of the two language's color samples, the results support our belief that congruency for naming individual samples with a translated modal term gloss is possible in the unconstrained naming task presented in Experiment 1.

Comparison of Best-Exemplars From Different Color Order Systems. If the stimuli presented as equivalent are in fact the same, they should be named with the same stem term. This analysis determined (1) whether the equivalent best-exemplar stimuli from three different stimulus sources (Munsell focals, OSA centroids, and our rendered centroids) were all named using the same category stem term, and (2) whether our rendered versions were named with use of the same term as the equivalent Munsell focals and OSA centroids. For example, we reasoned that if subjects assign the same name to all three versions of these samples with high consistency, then our failure to observe salience in measures of agreement in Experiment 1 requires some explanation other than subtle differences between our stimuli and those used in previous studies.

During stimulus creation (described for Experiment 1), colorimetry measures found the rendered centroid stimuli to be close perceptually to measures of OSA centroids and Munsell focals, and deemed perceptually "equivalent." Experiment 2's comparison of rendered stimuli against actual OSA tiles and actual Munsell chips provides a direct empirical test of the colorimetric "equivalents." Across the three stimulus types, we compared (1) frequency of naming with the same modal term, and (2) frequency of modal term use across language groups. We made comparisons within and across both tasks and languages using two-tailed t tests.

Separate analyses of the bilingual Vietnamese group's monolexemic and free-list naming data revealed no significant differences ($p < 0.5$)[k] in mean modal term naming for any of the possible comparisons of Munsell focal chips, OSA tiles, or rendered centroids. Similarly, for our English subject group, separate analyses of monolexemic and free-list naming data revealed no significant differences ($p < 0.5$) in any of the possible comparisons of the three stimulus types, with one exception. Monolexemically named OSA tiles were found to be significantly less likely to evoke the modal stem term compared to monolexemically named rendered centroid stimuli (Table B-3, column 3). This one significant difference most likely occurred because the actual OSA tiles were, on average, less frequently named with a modal stem term (M = .76) and were named with greater variability (SD = 31.2) compared to the overall mean frequency of observed modal naming for the rendered stimuli (M = .98, SD = 7.8). Reasons for this are discussed below. This was the only comparison (out of 12 comparisons made) in which the naming of best-exemplars from one stimulus set was found to be different when compared with other stimulus system equivalents.

These findings, which (1) confirm the equivalence of stimuli, and (2) demonstrate cross-language naming congruence in both tasks, provide indirect evidence that Experiment 1's results most likely reflect real color-naming behaviors and are not simply attributable to aspects of experimental stimuli or design.

These Experiment 2 results for both language groups (31 monolingual English and 33 bilingual Vietnamese speakers) showed an overwhelming tendency for subjects to use the identical color-term stem when naming the Berlin and Kay category focals, our rendered best-exemplar centroid stimuli, and the Boynton and Olson OSA centroids (to a lesser degree). As might be expected from the results of Experiment 1, the same *stem* was used by different subjects to name the three versions of the same item, but the actual naming behavior differed considerably in comparisons across language conditions. The primary characteristics of the observed variation across conditions were (1) extensive use of modifying terms (e.g., dark, strong, pure), and (2) extensive use of objectifying composites (e.g., brick red, sky blue).

In view of this empirical evidence, we believe our stimulus rendering is accurate at a level permitting demonstration of differences between best exemplars and other category members if such differences existed. Beyond these empirical demonstrations, there are additional reasons why criticism of stimulus rendering cannot explain our findings. Imperfect correspondences between best-exemplar stimuli from the Munsell and OSA color order systems are also a product of the different ways the systems structure the color space. For example, the Munsell system includes unitary, well-saturated examples of the category regions for red and brown. The OSA system, on the other hand, does not optimally represent these regions.[44] The comparative stem term usage in Experiment 2 is impacted by presentation of these less-than-optimal OSA tiles. Depressed stem usage for several OSA samples (especially for brown, orange, and red samples) most likely drives the one significant difference observed across the two color order systems. An analogous pattern is seen in the Vietnamese naming of categories *do* (red) and *nau* (brown). This difference is important because it illustrates two problems: (1) the difficulties of finding a color stimulus set with formally modeled isotropic properties that has excellent exemplars for all color categories, and (2) the difficulties of making comparisons between color order systems with different color space properties. Experiment 2 yields satisfactory levels of observed correspondence in stem usage between the Munsell samples and our rendered samples. We are less enthusiastic about the correspondences between the OSA samples and the other two sets, but our data suggest that this exists largely because of deficient best exemplars for red, brown, and orange found in the OSA/UCS tiles.

Previous investigators have made similar observations about the OSA tiles. Boynton and Olson[10] discuss the deficiencies of certain chromatic samples (p 104) and achromatic samples (p 101) in the OSA set. Sturges and Whitfield[44] also discuss the "limited range of OSA space" (p 370) and empirically demonstrate that the sample for red is ranked below purple—a finding clearly at odds with a model that asserts greater salience for all landmark colors (red is landmark, purple is not). Finally, despite these rendering issues, our failure to find centroid salience in Experiment 1 is unlikely to be the result of suboptimal rendering. In Experiment 1, when we omitted from our ANOVA analyses the most poorly rendered samples [based on the Delta-E($L^*a^*b^*$) measures discussed in Appendix A] as listed in Table B-2 (our red and yellow), and recalculated all of our measures, we still found no significant differences in any of the results comparing centroids with noncentroids—no greater salience of centroids by any of our measures.

The rendered centroids used in our study are perceptually and colorimetrically matched (within tolerance) to the stimuli that Boynton and Olson[10] and Berlin and Kay[2] used. Experiment 2 demonstrates that there is little practical difference in the way our best exemplar samples are named compared to the actual centroid samples and focal chips used by the other investigators. Based on these convergent observations, we argue that the failure to demonstrate clear differences in centroid versus noncentroid naming salience cannot be dismissed because of inaccurate rendering of other investigators' best-exemplars. Maintaining the equivalence of stimuli through careful stimulus control is essential to the generalizability of color-naming and psychological salience results to other settings and situations in which color categorization behaviors are assessed. It is important for future color categorization research to provide comparisons of new results with existing results in ways that permit the absolute determination of stimulus differences across

---

[k] No adjustment of the significance level for multiple *t* tests was made, because we wished to allow maximum opportunity for any existing difference to be observed.

research efforts. We have attempted to do that here by using colorimetric measures of color difference and empirically testing the pragmatic aspects of naming best exemplars. It is our belief that this general method of comparison would likely increase the comparability of future research in this area.

GENERAL DISCUSSION

The results presented here demonstrate that real differences are found in the ways that our ethnolinguistic populations name surface color samples—a finding that is not consistent with a strict notion of color-naming universality. Furthermore, little support was found for distinct salience of previously published focal or centroid exemplars. However, it might be argued that salience is a property of regions of the color space, not individual color samples. If so, then the differences observed across ethnolinguistic groups in our study still lead us to believe that salience emerges through the process of naming rather than from a shared neural physiology.

The very purpose of naming seems inconsistent with extreme sensitivity to stimulus characteristics, including the stimulus selection and rendering issues explored in Experiment 2. If we assume that an important purpose of cognitive color naming is to label best exemplars of color categories, and we further assume that naming occurs as a continuous function rather than point-wise (a specific color percept), then any color appearance that is reasonably close to the actual best-exemplar centroid should, with a high probability, be named with the same term as the centroid. This implies that for any given color category, the probability that naming will produce a straightforward mapping between color terms and color appearances will be maximized in a central region of the category, and will continuously decrease as the function extends away from the central region, as suggested by Kay and MacDaniel.[3] Color appearances that are close to the theoretical peak centroid will fall within the envelope of appearances named with the same term as that of the centroid.

As support for a probabilistic description of the operation of naming, consider the differences between the focals identified empirically by Berlin and Kay, and the centroids identified by Boynton. When new samples of each are viewed side by side, these category best exemplars are clearly not perceptually identical. When viewed under the controlled *C*-illuminant conditions or in more naturalistic viewing situations, none of the category best-exemplar samples exactly match. Yet, empirically, these two perceptually different sets of best exemplars both produced results suggesting universality of psychological processing salience (see Experiment 2). Boynton and colleagues[15] suggest equivalence between his centroid findings and those of the focals of Berlin and Kay, but between these two studies, the samples are far from equivalent with respect to any perceptual criterion. A further lack of equivalence across studies arises because individual visual systems will produce idiosyncratic, subtle variation in response to many stimulus centroids, giving rise to individually different metameric equivalence classes. Thus, an important function of naming must be to allow best-exemplar naming tolerances that accommodate slight differences across individuals' color perception, for the sake of communicating with other members of one's linguistic group about color. This same language-based accommodation results in the cultural differences evident in our cross-language comparison. See Jameson[45] for further discussion on the relation between naming and color perception.

Our results do not challenge the existing literature on salience of basic terms or exemplars, but rather question the proposed basis for that salience—a universal underlying opponent-process neural physiology. In the paragraphs that follow, we propose an explanation for the observed salience in previous studies, and the lack of salience in our own research and that of several current researchers. In part, we believe the discrepancy arises from differences in cognitive processes that are made evident with changes in methodology. A key difference between research showing salience and research that fails to find salience is the manner in which the subjects were questioned. In Berlin and Kay's[2] research, subjects were generally asked to provide *color samples for category names specified by the experimenter*, whereas in the research of Boynton and colleagues, subjects were asked to provide *category names for color samples presented by the experimenter*. Similar to Boynton and colleagues, we asked subjects to provide *names for color samples presented by the experimenter*, but unlike the previous studies, we did not constrain subjects to monolexemic naming, but rather permitted them to list freely the best available color name, regardless of naming complexity. The "best" name given was idiosyncratic to each individual subject. Thus, there exist two levels of difference between our experimental task and that of previous researchers: (1) providing *names* versus providing *color samples* and (2) *unconstrained* versus *constrained naming*.

Despite the tendency of color-naming investigations to consider the two modes as directly comparable, different results are obtained when names are assigned to samples than when samples are selected to exemplify provided names. These asymmetries alone may be enough to account for the observation of salience of stimuli in one paradigm and failure to observe it in another paradigm. Moreover, similar task-dependent naming-relation asymmetries have been observed in other psychological domains (such as emotion) naming employing different judgment tasks.[46–48] Thus, it is plausible that the assignment of color terms to color appearances by a cognitive naming function is not symmetrical and does not exhibit the property of reciprocal signification when assessed. Until the consequences of such naming-relation asymmetries are better understood, reasoning from empirical findings toward theories of the psychological salience of color stimuli and semantic labels will be difficult. These asymmetries may be due to selection set effects similar to those found in other psychological domains, such as Kahneman and Tversky's[49] "frame" effects or Braun and Julesz'[50] "stimulus set" effects, or they may be a property of other category-exemplar lexical mappings that have yet to be elucidated. We believe the latter explanation is most likely, for reasons stated below.

As evidence of the impact of task demands, consider the effects of reducing our free-list data to Berlin and Kay's[2] categories. The key result is that the naming data from the three languages converge toward a similar naming pattern, much like that found by Boynton and colleagues and Berlin and Kay and colleagues, using a task that was artificially constrained for experimental simplicity. To see this, compare Figs. 3 and 7. Figure 3 illustrates the frequency differences across languages with the use of our unconstrained color-naming task. Figure 7 represents the same stimuli and responses after applying a data reduction technique that approximates the responses that a subject would provide if given a monolexemic naming task or a Berlin and Kay categorization task.

This is not intended to imply that the Berlin and Kay or the Boynton and Olson results are mere methodologic artifacts—a comment that has been made elsewhere.[51,52] Rather, we interpret the observed discrepancy to imply that *empirical method* greatly influences the structure of resulting data. We believe that uniformity of experimental context has contributed to the consistency of findings in previous research. Berlin and Kay used two different empirical procedures and two different stimulus sets, circumstances that should have yielded two different results, yet they obtained closely similar findings of salience. Before a panhuman universality based on neurophysiology can be assumed, competing explanations must be tested, including other sources of panhuman universality resident in cognition. We propose that their results occurred because the color-naming function is inherently asymmetric, mechanistically, but produces reciprocity when sufficient experimental constraints are imposed (e.g., monolexemic naming). We think the best way to establish a *general* model of color naming and categorization—one that will predict across a variety of stimulus formats, contexts, and tasks—is to identify those theoretical constructs that remain valid across a variety of empirical designs and testing formats. An encompassing set of general principles must reconcile previous findings with results such as those reported here.

It is incorrect to characterize our findings as a mere context effect, just as it would be incorrect to characterize the findings of other notable studies as such. We believe our divergent results emerge from the process of naming itself. Our findings are analogous to an imperfect mapping of color terms across two different languages. For example, if the color lexicon of Language *A* did not translate directly and fully into the color terms available in Language *B*, then this would be most accurately described as an incomplete or nonreciprocal mapping by the naming function of Language *A* to Language *B*. Similarly, in our study, we find that within language and within subject, the mapping of names to colors does not fully predict the mapping of colors to names. We observe this failure of reciprocity as a feature of the color-naming function, rather than a mere context effect. Characterizing this study's findings as a context effect assumes *a priori* that the naming is reciprocal, without evidence to support that assumption. In our opinion, whether the naming function is reciprocal is an empirical question that has not been fully considered in the literature to date.

Evidence from Dichromat subjects suggests that asymmetric color naming is not merely a context effect. For example, an extreme form of such an asymmetry between cognitive organization of color names and color appearances can be seen in the work of Shepard and Cooper.[53] Their results suggest that in subjects with normal color vision, color terms and color appearances are mapped in a symmetric fashion. However, clearly different, asymmetric mappings can also result, depending on whether linguistic or visual criteria are emphasized, as found for color-blind similarity scalings of color names compared with color appearances. Furthermore, one half of the naming function can be formed and maintained in the absence of the other half, as in the case of the color-term similarity scalings of the blind subjects.

The suggestion of a naming function that consists of separate relational mechanisms for terms and appearances is in accord with cases from the neurophysiologic literature. For example, in reviewing neural disorders of internal color space, Roberson *et al.*[54] describe cases in which patients exhibit the ability to name colors and yet cannot perceptually categorize the appearances (pointing to a dissociation in the relational color appearance similarities coexisting with an intact mapping of color names to appearances). Moreover, the opposite preserved ability to categorize colors by name independent of an impaired ability to name colors is discussed (suggesting perhaps a different form of dissociation between naming and perceptual identification). Existing findings also seem independently to localize the ability to discriminate color to right-hemisphere function, compared with a specific deficit of categorization and naming localized to left-hemisphere function (see Davidoff[55]). Although these neurologic data do not prove a mechanistic division between the two mapping directions permitted by the color-naming function, they are consistent with our findings and our assertion of a qualitative difference in underlying cognitive processes that manifests itself differently when making "verbal" versus "visual" responses.

Our findings clearly suggest the need to reevaluate the concept of color salience and its relation to color naming. As can be seen in Table III, considerable agreement exists about the name for a sample, but the samples evoking strongest agreement are different within the three different language groups. This supports the idea of cross-language agreement on semantic labels but does not support the idea that specific *samples* are universally salient exemplars of those labels. We strongly suspect that confidence may be a better indicator of color salience than consensual naming or response time.[56] It may be that the feature that is both salient and universal across our tested groups is the *certainty* with which colors are named. Salience in this free-listing task may be indicated by the degree to which individuals feel they have produced the consensual response. These feelings may be the hallmark of the appearances with which they are associated rather than the names given to the appearances, which vary by language.

What Is the Color-Naming Relation?

The potential for different mappings between the lexicon and color appearances under differing task demands implies existence of a naming relation, a cognitive process that assigns names to objects. We call this process the "naming function." If salience is not a by-product of the stimuli or the terms themselves, as our research clearly suggests, perhaps it is a by-product of this naming relation. We raise the possibility that the similarities among the results of Boynton, Berlin and Kay, and their respective colleagues may be second-order consequences of underlying panhuman properties of the naming function for color appearances. We believe that our results and their incongruence with some previous research can also be well explained by existence of first-order psychologically salient features of color space and color lexicons that are operated upon by the naming function.

We propose that the naming function implements the principles of an Interpoint Distance Model (described by Jameson and colleagues[29]), as an alternative theory of color naming and categorization supported by our results and those of Berlin and Kay,[2] Smallman and Boynton,[57,58] and Roberson and colleagues.[21] The paramount criterion of the naming function in everyday color experience is devising an optimal and meaningful information code for the visible color space given the use of two, three, four, or more basic color terms. When assessed empirically with a color order system such as the OSA/UCS or the Munsell Book of Color, the optimal mapping of color names to color samples may vary to accommodate color space differences. Thus, the actual assignment of terms to color appearances would be expected to vary as a function of the number of terms available, and the number of samples to be denotated, with a consequent impact on which color appearances are considered focal within a given category, and the use of specific terms to identify them.

It seems reasonable to suggest that different naming partitions would result if different stimulus sample sets were initially used. Thus, an extensive sample of color appearances would elicit a greater number of color term names than would a smaller, less-representative, sample. Similarly, it seems likely that placing a naming-task constraint (e.g., monolexemic naming or predefined category term labels) would influence the manner in which color terms are applied to the stimulus space. In that case, confidence might indicate the goodness of the fit between terms and appearances given the options available in the context of a particular task. These aspects of color naming and categorization are addressed by an Interpoint Distance Model (IDM) of color categorization.

For the general case of categorization and naming, the IDM proposes that lexical terms are mapped not onto objects but onto a relational stimulus structure of categories and category members, to form a meaning space structure. This meaning space can be limited to the stimuli presented in the context of an experiment, or it may encompass the entire range of stimuli accessible in memory by a given individual. As the size and content of the meaning space varies, so does assignment of names to objects in that space. Note that this flexible assignment depends on stimulus characteristics and the properties specified by the lexicon, but is mediated by additional considerations of the extent and content of the space to be named. Thus, category structure (e.g., a hierarchy of basic and nonbasic labels for color appearances) exists independent of the mapping onto a stimulus space.[45,48]

Essentially, the IDM asserts that the best exemplars within each category (i.e., the focal or basic color) will be distributed to optimize the codability of color terms through equality of interpoint distances between all best exemplars, while secondarily striving to encompass approximately equal areas across the category partitions (see Jameson[45] for a discussion of spatial and dimensional salience). We also suggest that when resource limitations are imposed on subjects by constraining the available names, by demanding fast response times, or by imposing memory and cognitive processing loads, naming more closely follows category structure. When these constraints are relaxed, subjects are better able to make fine distinctions among colors and color names, and tend to rely less upon basic terms and focal colors.

Predictions from the Interpoint Distance Model (IDM)

The findings in Experiment 1 and Experiment 2 presented earlier support hypotheses following from the IDM just described. For example, removal of the constraints present in previous research (e.g., speeded response, restriction to monolexemic terms) results in emergence of different focal colors than previously identified (as shown in Table III). Consistent with the theory, different focal colors emerge in different languages due apparently to differences in access to terms. The color appearances identified by confidence and agreement were different in the different languages, and different from those found in previous studies, and did not appear to depend upon theoretically assumed category structure or salience of exemplars. This is expected, because the different languages tested possibly contain different numbers of color categories,[2,23] different numbers of monolexemic terms, and tend to make different use of modifiers for fine color distinctions. The IDM predicts, as observed, that specific focal color stimuli identified by confidence and agreement should be different in different languages, different than in previous studies, and should not be strongly correlated with an opponent-process-based category salience.

The existence of universal color terms, without universal application of those terms to the same color appearances, as demonstrated in this study, provides strong support for the IDM. Under that model, the use of color names and labeling of color space seen in previous empirical studies may be largely attributable to (1) the degree to which the sample tested represents all manner of exemplars from visible color space, and (2) the constraints placed on the choice of names by the experimental task used to assess naming, especially monolexemic naming.

The IDM suggests that naming behavior is better predicted by psychophysical judgments and the characteristics of the psychophysical meaning space obtained from such judgments (i.e., Indow's global color metric, 1980) than by early neural coding (see Jameson and D'Andrade[29]). The IDM is in accord with existing findings from the psychological scaling of color space[59–61] and is partially supported by the psychophysics literature, which suggests that it is not a specific set of *colors*, *per se*, that are the crucial markers to the color vision system, but the notion of color opponency —like that explained by Hering.[13] D'Zmura and Knoblauch[62] have reported the existence of what appear to be noncardinal color directions for orange and violet, consistent with the work of Krauskopf *et al.*,[63] who report additional noncardinal axes. Such results suggest a growing number of plausible alternatives to the fundamental colors posited as the salient markers of color phenomenology by Hering.[13]

Two important methodologic consequences emerge from the IDM analysis of color-naming phenomena. First, if the color order sample tested is extensive, representing a wide range of stimuli from all possible dimensions of color space, including hue, brightness, saturation, and so on, then the universal labeling behaviors will most likely reflect names applied to the space in a manner that describes the entire color space. Second, once any given color term exists as an item in a language's lexicon, it will likely have a corresponding cognitive best exemplar that all users of that lexicon can imagine (cf. Hard and Sivik[64]). Despite the existence of such best exemplars, there will be no universal or panhuman agreement about any specific context-free category regions, however large, labeled by terms that gloss the same best-exemplar appearance for a given color category. This latter point is a consequence of the naming relation.

Previous research has shown that terms that gloss the same meaning in different languages map onto different best-exemplar color appearance stimuli cross-culturally. When a given color category's best-exemplar region is extended to encompass the cross-cultural data, it no longer represents a best exemplar but more closely approximates the boundaries defining the actual category. The notions of "best category exemplar" and "precise lexical category descriptors" are countervailing properties of the phenomenon: the former aims for context-invariant specificity, whereas the latter strives for stimulus set generality. Neither serves to optimize psychological salience of specific color percepts. Although the lexicon suggests how one should imagine an appearance tied to a color term, in its purest sense, the fuzzy region defined by a color category only tells one where to find the closest context-invariant neighbor of the imagined appearance. This study makes no claim to have addressed the issue of the universality of color naming and best-exemplar color appearances. A study addressing that issue would need to test a variety of monolingual speakers in their own cultures. We do believe, however, that compared to the standard methodologic approaches, this study presents a stronger test of the notion of best-exemplar salience and color naming by virtue of the pains that have been taken to make valid comparisons between our findings and previous results.

Bilingualism and Color Naming?

Bilingualism (or multilingualism) is a factor that may deserve consideration in its own right in studies of cognitive organization and psychological salience. The greater cohesion of response among bilingual Vietnamese (indicated by higher agreement indices) when compared with the other monolingual groups is similar to a finding of Romney *et al.*,[40] who noted that, compared to Japanese or English monolinguals, Japanese bilinguals also had more cohesion for semantic structures involving emotion terms.

The bilingual subjects we tested may be described as undergoing a process of losing monolingual proficiency of their Vietnamese, while improving English proficiency. Modifier use was an overwhelmingly common behavior in the monolingual Vietnamese subjects we tested, and perhaps this is one way a language with fewer color terms compensates for the reduced number of terms available as descriptors (Vietnamese is "problematic Stage VII," whereas English is "Stage VII"[2]). The assumption that color-naming universalities can be adequately studied by focusing upon monolexemic terms is examined by Alvarado and Jameson.[65] Bilingual individuals we tested demonstrated some hybrid variant of the color-naming function that comingled the naming tendencies exhibited by the two monolingual groups. Bilingual speakers were clearly influenced by the naming practices of both languages, and their choices were identical to neither of the monolingual groups.

CONCLUSIONS

Based on our data, we can conclude that even in a less constrained naming task, there is a high degree of similarity and agreement in naming both within and across languages. Our findings were consistent with those of Berlin and Kay,[2] and other researchers asserting universality of naming with the use of basic color terms. When subjects were not constrained to monolexemic naming or to Berlin and Kay's 11 basic categories, they nevertheless used those terms with high frequency. However, the application of those terms to specific color samples differed across language groups. We found considerable differences between English- and Vietnamese-speaking subjects in their use of modifiers to make fine discriminations among colors appearances (see Alvarado and Jameson[65]). Vietnamese speakers tended to use more modifiers, whereas English speakers tended to use a greater variety of monolexemic terms. Greater consistency of response was found among bilinguals, and some evidence of the effects of acquiring another language was noted in their differing response patterns. We found no confirmation for increased salience of those samples identified as best exemplars of color categories by previous researchers, including Berlin and Kay's focal colors and Boynton's landmark hues, even after data reduction to approximate their response constraints. Because great care was taken to reproduce their stimuli accurately, we believe that this failure to replicate salience demonstrates the effect of the differing judgment task on naming. We propose that previously observed salience may be explained as an emergent property of the application, in part, of a cognitive naming function under conditions in which naming is constrained. We suggest that the need to maximize informational content of the available terms will dictate the region of color space to which they are applied and the focal exemplars within such regions. Our findings also suggest that confidence ratings are more strongly related to the mappings made by the naming function than to inherent properties of the color samples, or their labels.

In view of our findings, the extension of neurophysiologic models to color-naming behavior appears unwarranted. We suggest a reconceptualization of the ideas of color salience and basicness as a flexible phenomenon more closely related to the need to encode perception in language than to color vision. Moreover, salience and basicness may be emergent from task demands and imposed constraints on stimulus sampling and response options, rather than invariant properties of stimuli or their names. Serious consideration of alternative models, such as the IDM proposed by Jameson and D'Andrade[29] seems warranted.

Our data illustrate important complexities in color naming. These complexities make it impossible to ignore the inadequacies of the typically favored neural processing models that greatly oversimplify the cognitive processing of color. Reliance on a view of color categorization as emergent from opponent-process color vision is untenable given that simple manipulations of experimental context and accessibility of terms have a strong impact on observed salience. Universality of naming and salience do exist, but we must consider alternative cognitive processing models in order to fully explain them—in the domain of color naming and in other domains.

ACKNOWLEDGMENTS

The authors are grateful for the assistance provided by A. Lewis, A. Nguyen, A. Garcia, I. Moy, and H. Ngo.

APPENDIX A

Color stimulus specification

The heuristic used to select stimuli from the 424 samples of the OSA/UCS[32] consisted of applying a transparent graph-paper grid overlay to the 13 OSA levels presented by Boynton and Olson (Fig. 1[10]) and defining a midpoint for each *Level* using the graph-paper gridlines, followed by selecting samples along eight possible radial lines at specific grid-intersection intervals (thereby avoiding entirely the use of the OSA spacing metric in the choice of stimuli). Although this method is clearly not a *random* selection technique, its advantages are that it is not explicitly based on any of the dimensions of color space, and it does not rely upon the spacing of the OSA parameters. In addition, it selects a sample that proportionally represents the different lightness levels of the OSA solid, while providing a reasonable representation of all the hues in the space.

*Centroids.*The desired Boynton and Olson[10] *centroid* stimuli were rendered at Delta-E ($L^*a^*b^*$) = 8.5. This is within the acceptable tolerance of Delta-E ($L^*a^*b^*$) < 10.0 found to provide color-difference calculations that model visual color differences (see Pointer and Attridge[66]). Note that much of this variation is not chromaticity (i.e., hue), but is attributable to variation in the lightness parameter (Y) of the CIE (1931) chromaticity measures, as evidenced by the computed difference for the CIE $*ab*$ parameters (ΔE_{ab}) of 6.30 for the *centroids*. Thus, the hue matches are quite accurate, and the rendered stimuli differ primarily in the lightness component.

*Focals.*The closest approximations in our 110-item sample to the Munsell chips identified as *focals* for Vietnamese and English lexicons[2] were found to be rendered at an average $\Delta E_{ab} = 66.54$ for Vietnamese, and $\Delta E_{ab} = 61.52$ for English. Although this color-difference measure is not within the recommended tolerance, it should be noted that the chips visually matched the rendered samples within a reasonable degree of variation. This difference is partly attributable to the high degree of lack of correspondence between the OSA stimulus set and the Munsell Book of Color stimulus set, as discussed by Boynton *et al.*[15] We address this point in Experiment 2 when we argue for the validity of comparisons between our stimuli and the Berlin and Kay[2] results.

*Ninety-nine Additional Samples.*Over the entire set of 110 stimulus items, there was an average ΔE_{ab} difference of 11.88 between the measured OSA tiles and the rendered counterparts used in our study. This is well within an acceptable range of rendering given that the ΔE_{ab} between controlled repeated measures of the same stimulus under the same illuminant can be as great as 2.0. Furthermore, our overall rendering of the 110 OSA stimuli is more precise than the computed differences typically found between printed versions of the OSA tiles and the published colorimetric data for the same tiles.[33] For example, the average ΔE_{ab} between the our measurements of the OSA tiles and the published data (as reported by MacAdam, cited in Wyszecki and Stiles[31]) on the corresponding tiles equaled 15.99. Compare this with the average ΔE_{ab} between the measured OSA tiles and measurements of our rendered approximations of those tiles, which equaled 8.58 for the centroids and 11.88 for all 110 stimuli. These data demonstrate that our renderings are well within the amount of colorimetric difference that can occur when comparing (1) differences in printed versions of the OSA stimulus set, and (2) colorimetric drift in the actual samples as a function of the passage of time. Our computed measure of $\Delta E_{ab} = 15.99$ probably includes the contribution of both sources of variation.

APPENDIX B

TABLE B-1. Measured chromaticities for rendered stimuli.

Rendered OSA stimuli for Boynton and Olson's[10,11] centroids

	OSA triple			Measured CIE 1931 chromaticities			
	L	j	g	Y	x	y	ΔE_{ab}
Red	−4	2	−8	39.06	0.46	0.33	19.59
Green	0	4	4	57.91	0.36	0.45	5.44
Yellow	3	11	−1	99.47	0.44	0.45	21.17
Blue	−1	−3	3	40.69	0.28	0.31	5.12
Brown	−3	3	−3	23.14	0.44	0.38	6.41
Purple	−3	−3	−1	25.66	0.36	0.31	8.07
Pink	1	1	−5	54.76	0.42	0.33	6.85
Orange	0	6	−6	55.47	0.50	0.40	5.04
Chartreuse	0	8	2	70.51	0.41	0.48	8.97
Turquoise	1	−1	5	47.18	0.28	0.37	2.23
Peach	4	4	−4	81.08	0.42	0.39	5.55
						Centroid Mean ΔE:	8.59
						Centroid Mean Δa^*b^*:	6.30

99 additional rendered OSA stimuli

	L	j	g	Y	x	y	ΔE_{ab}
	−6	−2	2	33.15	0.29	0.31	25.07
	−6	0	2	18.02	0.33	0.37	15.28
	−6	2	−4	28.64	0.46	0.36	19.21
	−5	−3	1	21.03	0.34	0.31	9.94
	−5	−1	−1	19.26	0.35	0.34	8.37
	−5	−1	1	20.14	0.33	0.36	10.49
	−5	1	−1	26.54	0.38	0.38	11.94
	−4	−4	2	26.90	0.29	0.31	15.16
	−4	−2	−2	29.60	0.37	0.32	10.89
	−4	−2	0	14.78	0.34	0.33	4.67
	−4	−2	4	29.68	0.30	0.36	16.30
	−4	0	−2	19.28	0.38	0.36	5.20
	−4	0	2	16.64	0.34	0.37	7.42
	−4	2	−4	25.58	0.43	0.36	10.87
	−4	2	−2	22.84	0.44	0.38	4.09
	−4	2	4	27.10	0.33	0.41	11.50
	−4	0	−4	28.09	0.40	0.33	10.63
	−3	−3	1	23.98	0.30	0.30	4.39
	−3	−3	3	26.88	0.31	0.34	12.14
	−3	1	−7	37.76	0.43	0.34	14.89
	−3	1	−5	29.30	0.45	0.36	7.15
	−3	1	−1	26.98	0.39	0.36	4.34
	−3	1	1	20.03	0.34	0.38	4.03
	−3	3	−1	40.42	0.40	0.41	11.12
	−3	3	1	24.05	0.36	0.44	6.48
	−3	3	5	38.79	0.32	0.45	10.38
	−2	−4	−2	38.69	0.34	0.30	8.62
	−2	−2	−4	31.55	0.37	0.29	20.81
	−2	0	−8	43.30	0.46	0.33	8.62
	−2	0	−4	26.06	0.40	0.33	3.59
	−2	2	−8	45.52	0.46	0.35	13.46
	−2	2	−2	48.84	0.42	0.38	9.82
	−2	2	2	47.26	0.33	0.40	11.49
	−1	−3	1	39.49	0.28	0.30	6.20
	−1	1	−1	61.50	0.39	0.37	11.63
	−1	3	−9	38.09	0.50	0.36	6.89
	−1	3	−7	45.92	0.48	0.37	6.48
	−1	3	3	49.38	0.34	0.44	6.04
	0	−4	−2	56.92	0.34	0.32	11.36
	0	−4	4	51.14	0.27	0.31	5.78
	0	−2	−2	52.02	0.35	0.32	4.24
	0	0	−2	68.56	0.37	0.35	10.33
	0	0	2	41.28	0.31	0.36	3.89
	0	2	−8	49.51	0.45	0.35	7.38
	0	4	−4	56.15	0.46	0.38	5.13
	0	4	−2	61.43	0.43	0.42	8.06
	0	4	2	57.46	0.37	0.44	3.60
	0	6	−2	58.82	0.42	0.41	12.26
	1	−3	3	61.89	0.30	0.34	6.83
	1	−3	5	59.68	0.26	0.33	6.82

(*continued*)

TABLE B-1. (*Continued*)

Rendered OSA stimuli for Boynton and Olson's[10,11] centroids						
OSA triple			Measured CIE 1931 chromaticities			
L	j	g	Y	x	y	ΔE_{ab}
99 additional rendered OSA stimuli						
1	−1	−3	51.25	0.37	0.32	6.30
1	−1	1	75.85	0.32	0.35	9.99
1	−1	5	60.13	0.29	0.37	4.45
1	1	−3	54.08	0.40	0.36	0.63
1	1	−1	44.03	0.39	0.36	6.40
1	1	3	60.03	0.31	0.39	5.37
1	3	1	70.84	0.36	0.40	7.21
1	5	−7	49.97	0.50	0.39	3.34
1	5	−3	62.95	0.44	0.41	4.10
1	5	3	75.61	0.38	0.45	8.67
2	0	6	71.15	0.31	0.38	11.42
2	2	−4	68.88	0.42	0.36	1.93
2	4	−4	60.97	0.45	0.38	4.57
2	4	−2	90.72	0.42	0.41	8.29
2	4	0	98.07	0.39	0.42	12.43
2	6	−6	69.10	0.44	0.39	18.24
2	6	−4	77.47	0.45	0.41	6.42
2	6	4	74.23	0.36	0.46	2.21
2	8	−6	62.05	0.49	0.42	9.73
2	8	0	86.60	0.42	0.45	9.68
2	8	2	90.04	0.41	0.47	6.33
2	8	4	64.57	0.40	0.49	5.95
3	−3	−1	93.33	0.34	0.35	12.11
3	−1	−5	78.79	0.39	0.34	5.77
3	−1	−3	79.78	0.37	0.34	1.10
3	−1	−1	10.40	0.36	0.35	41.97
3	−1	3	92.40	0.31	0.35	7.85
3	1	−3	83.79	0.39	0.35	3.81
3	1	1	98.12	0.34	0.37	8.47
3	3	3	84.31	0.36	0.42	5.53
3	5	−3	88.37	0.41	0.39	10.94
3	5	3	84.74	0.37	0.43	4.41
3	7	−3	82.01	0.44	0.42	6.18
3	7	−1	96.52	0.43	0.44	4.36
3	7	1	86.69	0.40	0.45	3.49
4	−2	0	96.56	0.35	0.35	7.20
4	2	−4	93.49	0.38	0.37	10.12
4	2	−2	84.34	0.41	0.38	6.08
4	2	0	12.34	0.36	0.37	47.84
4	4	0	11.10	0.38	0.40	49.63
4	4	2	99.61	0.38	0.42	4.39
4	6	0	11.36	0.41	0.43	52.99
4	8	0	11.01	0.44	0.46	56.00
4	10	0	10.94	0.44	0.46	63.00
5	−1	−1	11.13	0.36	0.36	50.02
5	1	−1	10.35	0.36	0.36	52.20
5	3	−1	10.39	0.39	0.39	53.63
5	5	1	98.52	0.39	0.43	5.15
					Mean ΔE:	11.83
					Mean Δa^*b^*:	6.10

Note. Optical Society of American color space values (L,j,g) for 110 rendered stimuli. CIE (1931) chromaticity coordinates (Y,x,y) of the rendered OSA samples are measurements of the actual stimuli used in the experiments (under illuminant C).

TABLE B-2. Comparison of measured chromaticities for rendered centroids and focals with measured chromaticities for actual OSA centroids and actual munsell focals.

Boynton and Olson[10] CENTROID samples

	OSA	OSA samples measured			Rendered samples measured					
English	L,j,g	Y	x	y	Y	x	y	Δ*x,y*	Δ*a*b**	ΔE_{ab}
Red	−4, 2, −8	16.76	0.53	0.34	39.06	0.46	0.33	0.07	9.82	19.59
Green	0, 4, 4	49.67	0.35	0.46	57.91	0.36	0.45	0.01	3.76	5.44
Yellow	3, 11, −1	82.54	0.48	0.47	99.47	0.44	0.45	0.04	20.39	21.17
Blue	−1, −3, 3	35.05	0.27	0.32	40.69	0.28	0.31	0.01	3.84	5.12
Brown	−3, 3, −3	25.59	0.46	0.39	23.14	0.44	0.38	0.02	6.10	6.41
Purple	−3, −3, −1	21.33	0.33	0.30	25.66	0.36	0.31	0.03	7.23	8.07
Pink	1, 1, −5	52.16	0.42	0.35	54.76	0.42	0.33	0.01	6.74	6.85
Orange	0, 6, −6	46.04	0.51	0.40	55.47	0.50	0.40	0.01	1.90	5.04
Chartreuse	0, 8, 2	52.06	0.43	0.49	70.51	0.41	0.48	0.02	3.90	8.97
Turquoise	1, −1, 5	50.97	0.28	0.37	47.18	0.28	0.37	0.00	1.15	2.23
Peach	4, 4, −4	90.33	0.43	0.39	81.08	0.42	0.39	0.01	4.52	5.55
							Mean Δ values:	0.02	6.30	8.59

Berlin and Kay[2] Vietnamese FOCAL samples

	Munsell	Munsell stimuli in 1931 CIE*			Rendered samples measured					
Vietnamese	H V/C	Y	x	y	Y	x	y	Δ*x,y*	Δ*a*b**	ΔE_{ab}
Do (red)	7.5R 5/14	0.20	0.56	0.34	38.09	0.50	0.36	0.07	33.04	65.52
	7.5R 4/14	0.12	0.60	0.33	39.06	0.46	0.33	0.14	25.79	64.36
	7.5R 3/12	0.07	0.62	0.31	37.76	0.43	0.34	0.19	17.19	62.44
Vang (yellow)	10YR 8/12	0.59	0.48	0.44	82.01	0.44	0.42	0.04	30.19	77.41
	2.5Y 8/14	0.59	0.48	0.47	99.47	0.44	0.45	0.05	39.45	86.66
Xanh (grue)	2.5B 3/6	0.07	0.18	0.25	29.68	0.30	0.36	0.16	8.24	55.38
Nau (brown)	2.5YR 4/10	0.12	0.55	0.39	29.30	0.45	0.36	0.10	18.53	55.83
	5YR 4/10	0.12	0.54	0.41	48.84	0.42	0.38	0.12	13.33	65.67
Tim (purple)	10PB 5/10	0.20	0.25	0.20	38.69	0.34	0.30	0.14	14.54	58.76
Hong (pink)	5RP 7/10	0.43	0.37	0.28	78.79	0.39	0.34	0.06	15.11	73.40
							Mean Δ values:	0.11	21.54	66.54

Berlin and Kay[2] English FOCAL samples

	Munsell	Munsell stimuli in 1931 CIE*			Rendered samples measured					
English	H V/C	Y	x	y	Y	x	y	Δ*x,y*	Δ*a*b**	ΔE_{ab}
Red	7.5R 4/14	0.12	0.60	0.33	39.06	0.46	0.33	0.14	25.79	64.36
	7.5R 3/12	0.07	0.62	0.31	37.76	0.43	0.34	0.19	17.19	62.44
Green	2.5G 5/12	0.20	0.24	0.51	38.79	0.32	0.45	0.10	23.79	61.76
	2.5G 4/10	0.12	0.24	0.50	38.79	0.32	0.45	0.10	25.54	64.12
Yellow	2.5Y 8/14	0.59	0.48	0.47	99.47	0.44	0.45	0.05	39.45	86.66
Blue	7.5B 5/8	0.20	0.20	0.24	59.68	0.26	0.33	0.11	23.75	71.67
	7.5B 4/8	0.12	0.18	0.22	40.69	0.27	0.31	0.13	15.50	61.89
	10B 5/10	0.20	0.19	0.21	40.69	0.27	0.31	0.13	14.10	59.79
	10B 4/10	0.12	0.17	0.20	40.69	0.27	0.31	0.16	14.40	61.62
	2.5PB 4/10	0.12	0.18	0.19	33.15	0.29	0.31	0.16	9.45	56.10
	2.5PB 5/12	0.20	0.18	0.19	51.14	0.27	0.31	0.15	18.20	66.19
Brown	2.5 YR2/4	0.03	0.46	0.35	19.28	0.38	0.36	0.08	4.18	48.30
	5YR 2/4	0.03	0.47	0.37	19.28	0.38	0.36	0.09	4.56	48.34
	7.5YR 2/4	0.03	0.47	0.40	19.28	0.38	0.36	0.10	5.12	48.40
Purple	5P 3/10	0.07	0.28	0.17	22.97	0.35	0.31	0.16	6.35	50.01
	5P 2/8	0.03	0.28	0.17	21.03	0.34	0.31	0.15	6.35	50.13
Pink	5RP 7/10	0.43	0.37	0.28	78.79	0.39	0.34	0.06	15.11	73.40
	5RP 6/12	0.30	0.39	0.26	54.76	0.42	0.33	0.08	22.64	67.51
Orange	10R 5/16	0.20	0.60	0.37	45.92	0.48	0.37	0.12	26.09	66.32
							Mean Δ values:	0.11	16.71	61.73

Note. CIE 1931 chromaticity coordinates for the Munsell notations are from Wyszecki and Stiles[32] [Table I (6.6.1) p 840].

TABLE B-3. Three variants of English and Vietnamese color sample stimuli used in Experiment 2.

English	Munsell H V/C	Rendered stimulus I.D.	OSA sample *L,j,g*
Red	7.5R 4/14	70	−4,−2,−8
	7.5R 3/12	79	−3,1,−7
Green	2.5G 5/12	55	−3,3,5
	2.5G 4/10	55	−3,3,5
		503	0,4,4
Yellow	2.5Y 8/14	10	3,11,−1
Blue	7.5B 5/8	23	1,−3,5
	7.5B 4/8	64	−1,−3,3
	10B 5/10	64	−1,−3,3
	10B 4/10	64	−1,−3,3
	2.5PB 4/10	81	−6,−2,2
	2.5PB 5/12	39	0,−4,4
Brown	2.5YR 2/4	93	−4,0,−2
	5YR 2/4	93	−4,0,−2
	7.5YR 2/4	93	−4,0,−2
		77	−3,3,−3
Purple	5P 3/10	92	−6,−2,−2
	5P 2/8	84	−5,−3,1
		85	−3,−3,−1
Pink	5RP 7/10	508	3,−1,−5
	5RP 6/12	45	1,1,−5
Orange	10R 5/16	63	−1,3,−7
		46	0,6,−6
Chartreuse		25	0,8,2
Turquoise		47	1,−1,5
Peach		21	4,4,−4
Vietnamese	H V/C	I.D.	*L,j,g*
Do (red)	7.5R 5/14	71	−1,3,−9
	7.5R 4/14	70	−4,−2,−8
	7.5R 3/12	79	−3,1,−7
Vang (yellow)	10YR 8/12	22	3,7,−3
	2.5Y 8/14	10	3,11,−1
Xanh (grue)	2.5B 3/6	72	−4,−2,4
Nau (brown)	2.5YR 4/10	88	−3,1,−5
	5YR 4/10	507	−2,2,−2
Tim (purple)	10 PB 5/10	506	−2,−4,−2
Hong (pink)	5RP 7/10	508	3,−1,−5

1. Hardin CL, Maffi L. Color categories in thought and language. Cambridge, UK: Cambridge University Press; 1997.
2. Berlin B, Kay P. Basic color terms: Their universality and evolution. Berkeley: University of California Press; 1969.
3. Kay P, McDaniel C. The linguistic significance of the manings of basic color terms. Language 1978;54:610–646.
4. Kay P, Berlin B, Merrifield W. Biocultural implications of systems of color naming. J Linguist Anthrop 1991;1:12–25.
5. Kay P, Berlin B. Science =/= imperialism: A response to B.A.C. Saunders & J. van Brakel's "Are there non-trivial constraints on colour categorization?" Behav Brain Sci 1997;20:196–201.
6. Kay P, Berlin B, Maffi L, Merrifield W. Color naming across languages. In: Hardin CL, Maffi L, editors. Color categories in thought and language. Cambridge, UK: Cambridge University Press; 1997. p 21–56.
7. Kay P, Maffi L. Color appearance and the emergence and evolution of basic color lexicons. Am Anthrop 2000;10:743–760.
8. Heider-Rosch E. "Focal" color areas and the development of color names. Dev Psychol 1971;4:447–455.
9. Boynton RM. Insights gained from naming the OSA colors. In: Hardin CL, Maffi L, editors. Color categories in thought and language. Cambridge, UK: Cambridge University Press; 1997. p 135–150.
10. Boynton RM, Olson C. Locating basic colors in the OSA space. Color Res Appl 1987;12:94–105.
11. Boynton RM, Olson C. Salience of chromatic basic color terms confirmed by three measures. Vision Res 1990;30:1311–1317.
12. Ratliff F. On the physiological bases of universal color terms. Proc Am Philos Soc 1976;120:311–330.
13. Hering E. Outlines of a theory of the light sense. Hurvich LM, Jameson D, translators. Cambridge, MA: Harvard University Press; 1964.
14. Heider-Rosch E. Universals in color naming and memory. J Exp Psychol 1972;93:10–20.
15. Boynton RM, MacLaury R, Uchikawa K. Centroids of color categories compared by two methods. Color Res Appl 1989;14:6–15.
16. Uchikawa K, Boynton RM. Categorical color perception of Japanese observers: Comparison with that of Americans. Vision Res 1987;27:1825–1833.
17. Nickerson D. OSA color scale samples: A unique set. Color Res Appl 1981;6:7–33.
18. Sturges J, Whitfield TWA. Salient features of Munsell colour space as a function of monolexemic naming and response latencies.Vision Res 1997;37:307–313.
19. Shepard RN. Perceptual-cognitive universals as reflections of the world. Psychonom Bull Rev 1993;2:2–28.
20. Shepard RN. On the physical basis, linguistic representation, and conscious experience of colors. In: Harman G, editor. Conceptions of the human mind: Essays in honor of George Miller. Hillsdale, NJ: Erlbaum; 1997. p 311–356.
21. Roberson D, Davies I, Davidoff J. Color categories are not universal: Replications and new evidence from a Stone-Age culture. J Exp Psychol Gen 2000;129:369–398.
22. Roberson D, Davidoff J. The categorical perception of colors and facial expressions: The effect of verbal interference. Memory Cogn 2000;26:977–986.
23. Lin H, Luo MR, MacDonald LW, Tarrant AWS. A cross-cultural colour-naming study: Part I. Using an unconstrained method. Color Res Appl 2001;26:40–60.
24. Lin H, Luo MR, MacDonald LW, Tarrant AWS. A cross-cultural colour-naming study: Part II. Using a constrained method. Color Res Appl 2001;26:193–208.
25. Lin H, Luo MR, MacDonald LW, Tarrant AWS. A cross-cultural colour-naming study: Part III. A colour naming model. Color Res Appl 2001;26:270–277.
26. Kuehni RG. Focal colors and unique hues. Color Res Appl 2001;26: 171–172.
27. Valberg A. Unique hues: An old problem for a new generation. Vision Res 2001;41:1645–1657.
28. Schirillo JA. Tutorial on the importance of color in language and culture. Color Res Appl 2001;26:179–192.
29. Jameson K, D'Andrade R. It's not really Red, Green, Yellow, Blue: An inquiry into color space. In: Hardin CL, Maffi L, editors. Color categories in thought and language. Cambridge, UK: Cambridge University Press; 1997. p 295–319.
30. Ishihara S. Ishihara's tests for colour-blindness (Concise Edition). Tokyo: Kanehara; 1987.
31. Wyzecki G, Stiles W. (1982). Color science: Concepts and methods, quantitative data and formulas. 2nd Edition. New York: Wiley; 1982.
32. MacAdam DL. Uniform color scales. J Opt Soc Am 1974;64:1691–1702.
33. MacAdam DL. Colorimetric data for samples of OSA uniform color scales. J Opt Soc Am 1978;68:121–130.
34. Newhall S, Nickerson D, Judd D. Final report of the OSA subcommittee on spacing of the Munsell colors. J Opt Soc Am 1943;33:349–385.
35. Miller DL. Beyond the elements: Investigations of hue. In: Hardin CL, Maffi L, editors. Color categories in thought and language. Cambridge, UK: Cambridge University Press; 1997. p 151–162.
36. Backhaus G, Kliegl R, Werner J. Color vision: Perspectives from different disciplines. New York: Walter de Gruyter; 1998.
37. Commission Internationale de l'Eclairage. CIE Proceedings 1931. Paris: Bureau Central de la CIE; 1932.
38. Chapanis A. Color names for color space. Am Sci 1965;53:327–346.
39. Thorndike EL, Lorge I. The teachers' word book of 30,000 words. New York: Teachers College, Columbia University; 1944.
40. Romney AK, Moore CC, Rusch CD. Cultural universals: Measuring

the semantic structure of emotion terms in English and Japanese. Proc Nat Acad Sci USA 1997;94:1–5.

41. Matsumoto D. Cultural influences on research methods and statistics. Belmont, CA: Brooks/Cole; 1994.
42. Hays W. Statistics. New York: Holt, Rinehart & Winston; 1988.
43. Rosenthal R, Rosnow R. Essentials of behavioral research: Methods and data analysis. 2nd Edition. New York: McGraw-Hill; 1991.
44. Sturges J, Whitfield TWA. Locating basic colours in the Munsell space. Color Res Appl 1995;20:364–376.
45. Jameson KA. Culture and cognition: What is universal about the representation of color experience? 2001. Submitted for publication.
46. Alvarado N, Jameson KA. New findings on the contempt expression. Cognition Emotion 1996;10:379–407.
47. Alvarado N, Jameson KA. Varieties of anger: Salience of movement cues in the decoding of meaning in facial expressions. 2001. Motivat Emotion, to appear.
48. Jameson KA, Alvarado N. The relational correspondence between category exemplars and names. Philosoph Psych, to appear.
49. Kahneman D, Tversky A. Choices, values and frames. Am Psych 1984;39:341–350.
50. Braun J, Julesz B. Withdrawing attention at little or no cost: Detection and discrimination tasks. Percept Psychophys 1998;60:1–23.
51. Lucy, J. (1997). The linguistics of color. In: Hardin CL, Maffi L, editors. Color categories in thought and language. Cambridge, UK: Cambridge University Press; 1997. p 320–346.
52. Saunders BAC, van Brakel J. Are there non-trivial constraints on colour categorization? Brain Behav Sci 1997;20:167–228.
53. Shepard R, Cooper L. Representation of colors in the blind, color-blind, and normally sighted. Psych Sci 1992;3:97-103.
54. Roberson D, Davidoff J, Braisby N. Similarity and categorisation: Neuropsychological evidence for a dissociation in explicit categorization tasks. Cognition 1999;71:1–42.
55. Davidoff J. Cognition through color. Cambridge, MA: MIT Press; 1987.
56. Boynton RM, Fargo L, Olson C, Smallman H. Category effects in color memory. Color Res Appl 1989;14:229–234.
57. Smallman HS, Boynton RM. Segregation of basic colors in an information display. J Opt Soc Am A 1990;7:1985–1994.
58. Smallman HS, Boynton RM. On the usefulness of basic colour coding in an information display. Displays 1993;14:158–165.
59. Indow T. Psychologically unique hues in aperture and surface colors. Die Farbe 1987;34:253–260.
60. Indow T. Multidimensional studies of the Munsell solid. Psychol Rev 1988;95:456–470.
61. Indow T, Aoki N. Multidimensional mapping of 178 Munsell colors. Color Res Appl 1983;5:145–152.
62. D'Zmura M, Knoblauch K. Spectral bandwidths for the detection of color. Vision Res 1998;38:3117–3128.
63. Krauskopf J, Williams DR, Heeley DW. Cardinal directions of color space. Vision Res 1982;22:1123–1131.
64. Hard A, Sivik L. NCS—Natural Color System: A Swedish standard for color notation. Color Res Appl 1981;6:129–138.
65. Alvarado N, Jameson KA. The use of modifying terms in the naming and categorization of color appearances in Vietnamese and English. J Cognit Cult 2002;2:53–80.
66. Pointer MR, Altridge GG. Some aspects of the visual scaling of large colour differences. Color Res Appl 1997;22:298–307.