Semantic and Perceptual Representations of Color: Evidence of a Shared Color-Naming Function

Bilge Sayim*, Kimberly A. Jameson**[1], Nancy Alvarado***, Monika K. Szeszel****

ABSTRACT

Much research on color representation and categorization has assumed that relations among color terms can be proxies for relations among color percepts. We test this assumption by comparing the mapping of color words with color appearances among different observer groups performing cognitive tasks: (1) an invariance of naming task; and (2) triad similarity judgments of color term and color appearance stimuli within and across color categories. Observer subgroups were defined by perceptual phenotype and photopigment opsin genotype analyses. Results suggest that individuals rely on at least two different representational models of color experience: one lexical, conforming to the culture's normative linguistic representation, and another individual perceptual representation organizing each observer's color sensation experiences. Additional observer subgroup analyses suggest that perceptual phenotype variation within a language group may play a greater role in the shared color naming system than previously thought. A re-examination of color naming data in view of these findings may reveal influences on color naming important to current theories.

KEY WORDS

Color categorization, color-naming, cognitive representation, cross-cultural color naming

* Psychology Institute, Christian-Albrechts-Universität, Kiel, Germany

** Center for Research in Language, University of California, San Diego, & The Institute for Mathematical Behavioral Sciences, University of California, Irvine

[1] Author Contact: Kimberly A. Jameson, Institute for Mathematical Behavioral Sciences, University of California, Irvine, Irvine, CA 92697-5100. E-mail: *kjameson@uci.edu*. http://aris.ss.uci.edu/~kjameson/kjameson.html

*** Department of Psychology and Sociology, California State Polytechnic University, Pomona

**** Division of Medical Genetics, Department of Medicine and Cancer Center, University of California, San Diego

Overview

A basic assumption of research on color representation and categorization has been that relations among color terms can be proxies for relations among color percepts because they are isomorphic to each other. Thus, color-naming behavior has been used frequently to define perceptual color category structures. We explore whether this assumption is valid by comparing similarity judgments for color words with judgments of color appearances. Understanding the link between individual cognitive processing of color and a culture's color naming system is important to both psychologists and cross-cultural investigators because it may be generalized to explain the relation between a range of cognitive processes and the cultural systems in which they occur. To this end we examine individual cognitive similarity structures of color to clarify: (1) how individuals partition the color space continuum, and (2) how lexical labels are applied to color space partitions. Cognitive similarity structures of color are thought to directly derive from a single internal cognitive representation of color. Both lexical similarity and perceptual similarity are frequently assumed linked to the same representation. This assumption conflates two cognitive representations that we believe should be modeled as separate. Failure to differentiate perceptual and verbal cognitive representations introduces unnecessary vagueness to related constructs discussed in the literature.

Representations of Color

Shepard and Cooper (1992) asserted that color naming and judgments of color appearances arise from a single internal color representation. They used similarity data to scale the representations of colors and words, and compared the two across different observer groups. Based on this, Moore, Romney and Hsia (2002, p. 3) state ". . . Psychologists have found that in many domains, including color, that 'judgments of similarity among objects are essentially the same whether the objects are presented or only named' (Shepard, 1975, p. 96, Shepard & Cooper, 1992)". Thus, the cognitive representation of color appearances is considered to be structurally similar to its lexical representation. This struc-

tural similarity is sometimes referred to as a *second-order isomorphism*. However the argument for a single representation becomes problematic when results for different observer types ("color-blind" dichromats and blind subjects) are examined. Shepard and Cooper's (1992) results show that for dichromats, color appearance similarity structures differ from color lexicon similarity structures. Specifically, the dichromat's lexical scaling of color terms formed a 'normal' circumplex structure (Newton's color wheel) while their scaling of color appearances did not. This difference between the lexical and perceptual representations of dichromats is consistent with the findings of early psychophysical and developmental psychologists regarding the relationships between naming and perception (Jameson & Hurvich, 1978, Marmor, 1978).

In a cross-cultural study, Moore, Romney & Hsia (2002) confirmed an isomorphism between the color lexicon and color appearance judgments in two ethnolinguistic groups (Chinese and English). Their results supported the view that color term relations can serve as a proxy for color appearance relations in empirical investigations. With respect to identifying similarity relations for color they state: ". . . It does not make a great deal of difference whether the names of the colors or the color samples are used as stimuli . . ." (p. 19).

Like many psychologists studying perception, we believe it is important to distinguish between the perceptual and cognitive processing of color (Jameson & Alvarado, 2001, Jameson & Alvarado, 2003, Jameson, 2005a, 2005b). For example, Derefeldt, Swartling, Berggrund and Bodrogi (2004, p. 8) state, "a distinction between perceptual and cognitive color spaces may be made. A perceptual color space is defined from descriptions of attributes of perceptions of real colors in the real word. The maximum number of colors in a perceptual color space has been estimated to be about 6 million (Chapanis, 1965) . . . [whereas] the maximum number of colors in a cognitive color space may be no more than 30 (Derefeldt and Swartling, 1995)." Similarly, philosophers of color have noted that the epistemology and ontology of color experience suggest that the realm of color naming is rightfully distinct from the realm of color appearance (Dedrick 1997, 1998, Chapter VI).

Thus, it remains an interesting empirical question whether color appearances are cognitively represented independent of the color lexicon.

If they are indeed distinct, there is a gap in the empirical and theoretical literature regarding the processes that might link a perceptual representation of color distinct from a cognitive representation of color language and categories. We have proposed that these spaces must be linked by a cognitive *color-naming function* (Jameson and Alvarado 2001, 2003), and have suggested that such color-naming functions can be expected to vary across individuals from different perceptual observer groups (Jameson, 2005b).

The Question of Individual Variation

Moore, Romney and Hsia (2002, p. 23) recently suggested that "individual differences will become an increasingly important research area" in cross-cultural color naming investigations. Indeed, recent empirical studies (Kuehni, 2001a, 2001b, 2004, 2005a, 2005b) reveal individual differences that challenge the universality of 'unique hue' individual salience – a foundational component of Berlin and Kay and colleague's cross-cultural color naming theory (Kay & Maffi 1999). We too have studied the impact of intracultural variation on color perception behaviors and culturally shared color naming and categorization (Jameson, Highnote & Wasserman, 2001, Jameson & Alvarado, 2001, 2003, Sayim, Jameson & Alvarado, 2003, Jameson, Sayim & Alvarado, 2003, 2005). We have argued that individual variation in perception and naming is an issue central to the study of shared color categorization and lexicalization, and we have aimed to provide an alternative theoretical rationale for such variation (Jameson, 2005a, 2005b).

While a number of investigators suggest that existing color naming systems are culturally shared in spite of intracultural variation because they are either algorithmically ideal (Steels & Belpaeme, 2005), informationally optimal (Griffin, 2004) or environmentally appropriate (Yendrikhovskij, 2001), others now suggest that individual variation is not relevant to the study of cultural color-naming systems (Kay & Webster, In Press, p. 1).

To address this controversy, we examined the extent to which variation in individual perceptual processing produces systematic variation in the way individuals map color names to color appearance. Specifically,

we examined empirically the extent to which individuals share a common color naming function, and the extent to which naming function idiosyncrasies are correlated with perceptual differences likely to cover with color vision phenotype (Jameson, Highnote & Wasserman, 2001). Based on our findings, we argue that shared color naming systems are not necessarily optimal for all individuals in a culture, but are shared because they represent the normative cognitive model for lexicalizing color (Jameson 2005a, 2005b).

In this study, we empirically investigated the nature of the color-naming function: (1) across individuals from a single ethnolinguistic group, and (2) for individuals belonging to specific observer subgroups naturally occurring within that larger ethnolinguistic group. The primary goal of this study was to determine whether significant individual differences in judgments of color similarity and naming existed for these subgroups and how such differences are to be understood in view of socially shared systems of color naming and categorization. The possibility of generalizing these results across linguistic groups is also discussed in the light of known differences in the varying frequencies of observer group phenotypes across different ethnolinguistic societies.

General Method

Two experiments investigated the nature of the color-naming function in native English speakers, across observer subgroups expected to vary in their individual representation of color. General details relevant to both experiments are presented in this section; details specific to each experiment are presented below. Experiment 1 is a *proof of concept* experiment that demonstrates that individual color-naming functions flexibly map color terms to specific color appearances. Experiment 2 presents separate triad tasks using color appearances and their names to directly assess individual color-naming function mappings between color appearances and color terms. An additional color mapping experiment was also presented, but will be reported in a separate article due to space constraints and its bearing on different issues.

Participants

Participants totaled 56 subjects (34 female and 22 male) recruited either through the University of California, San Diego, Department of Psychology human subjects pool, or by posted solicitations. Participants received either cash payment or course extra-credit. Informed consent was obtained by all participants in accordance with UCSD Human Research Protection Program (HRPP) protocol. Four subjects (2 male, 2 female) were omitted from data analyses due to procedural errors during data acquisition. All subjects were native speakers of English.

Participants were also asked to donate a blood specimen for photopigment opsin gene DNA analysis (a protocol approved by the UCSD Medical School branch of HRPP). Three milliliters of venous blood from each subject was collected into EDTA vacutainer tubes by a trained phlebotomist at UCSD's Thornton Hospital for the sole purpose of photopigment opsin genotyping (for procedures see Wasserman, Szeszel & Jameson, 2001, Jameson, Highnote & Wasserman, 2001). Of the 52 participants, 5 (3 male, 2 female) did not volunteer specimens for DNA processing.

General Procedures

Subjects were dark-adapted for at least 10 min. in a dimly lit room illuminated by a diffuse halogen lamp measured at approximately 2 Lux. The experimenter read all task instructions to subjects, and, in computerized tasks, instructions were also visibly displayed. After a series of practice trials, data collection began. At the end of all tasks subjects answered written demographic questions. Finally, each subject's color vision was assessed using the Farnsworth-Munsell 100-hue test (Farnsworth, 1943) and the Ishihara pseudoisochromatic-plate test for color blindness (Ishihara, 1996). Blood specimens for DNA analysis were volunteered within a few days following the psychological component of the experiment.

Experiment 1

A demonstration that subjects flexibly map color names to color stimuli seems an important precondition for arguing that individuals possess: (1)

representations of the similarity relations among color appearances separate from the representation for color words, and (2) a cognitive color naming function that maps the two representations. Experiment 1 demonstrates that individual naming function mappings, at a minimum, depend on configural features of the stimulus and the choice-set in which colors are named.

Experiment 1 was designed to assess color-naming invariance under a single alteration in viewing context. Two specific target color appearances were named under two different viewing conditions to see whether naming was consistent when the stimulus items surrounding the target item were varied. Only the surrounding items (the stimulus choice context) were varied (Braun & Julesz, 1998, Kahneman & Tversky, 1984), not the measured chromaticity (appearance) of the target color samples. Naming for a reddish target and a greenish target were assessed in two contexts, one consisting of surrounding items from widely different regions of color space (e.g., different color categories) and the other consisting of surrounding items from nearby color space (e.g., the same color category). We refer to the widely different items as a "global" viewing context and the nearby items as a "local" viewing context. Figure 1 gives a grayscale illustration of the format and context in which "reddish" target stimuli (indicated by asterisks) were named. Our rationale was that if

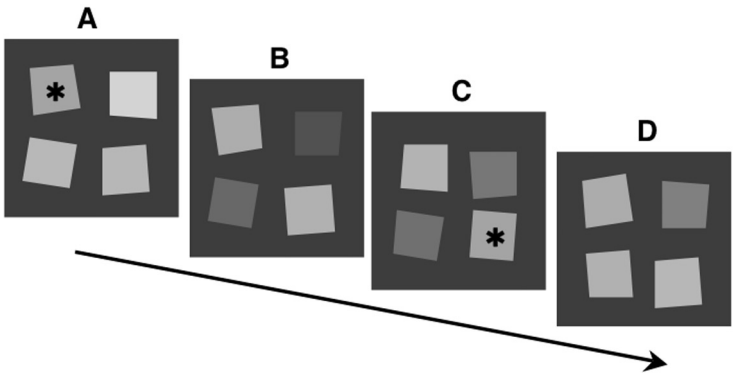

Figure 1. Grayscale schematic of one stimulus order used in Experiment 1's OSA Tile Naming task.

Note. Panels labeled A, B, C, and D correspond to tile groups given in Table 1. Asterisked items in panels A and C illustrate "reddish" tiles that were named across global and local contexts. "Greenish" tiles (not asterisked) were named in Panels B and D.

color names are invariantly mapped to the same color appearances, minimal differences in naming should occur across the local and global viewing contexts. If under these simple viewing circumstances, no such invariance was observed, then it is reasonable to suggest that names may be flexibly mapped to the color appearances depending on the task demands and viewing context.

Subjects

Fifty-two subjects (32 female, 20 male) participated in Experiment 1 as described above.

Apparatus

Experiment 1 stimuli were 2 in. by 2 in. square surface color tiles from the Optical Society of America Uniform Color Scale Atlas (OSA, 1977, MacAdam, 1974); see Table 1 for OSA-UCS specifications. OSA tile stimuli were viewed on a matte background of neutral black poster paper. Subject viewing position was not fixed, but was limited to a range at which stimuli subtended a minimum of 13.42 degrees of visual angle. In addition to the ~2 *Lux* ambient room lighting, OSA tiles were illuminated by a D65 daylight-approximate lamp at an angle that minimized specular highlights on the stimuli. All other nonessential objects were covered with blackout material.

Table 1
OSA-UCS system L, j, g coordinates and stimulus groups used in Experiment 1

Tile Group	OSA-UCS *L,j,g*			
A	3, 9, -1	0, 4, 4	-1, -3, 3	(-2, 2, -6)
B	1, 9, -5	(0, 2, 6)	-3, -3, 3	-3, 3, -3
C	0, 6, -6	0, 0, -2	-4, 2, -8	(-2, 2, -6)
D	(0, 2, 6)	-3, -1, 5	1, -1, 5	-1, 5, 5

Note. OSA notation of four tiles in each tile group as seen in Figure 1.
Parentheses denote *L,j,g* triples of target tiles (reddish: -2,2,-6; greenish: 0,2,6) for which naming invariance was monitored.

Procedure

The procedure used two fixed group sequences (*A,B,C,D* and *B,A,D,C*), and the four tiles comprising each group was fixed (see Table 1), but the physical placement of tiles within a group was random. The selection of which group sequence was presented first was also random with respect to subject, but counterbalanced within experiment. Within each tile group, the four OSA stimuli were hand positioned by the experimenter in a random order on neutral black poster paper (see Figure 1).

The task began with the experimenter presenting the samples for either color group A or B and instructing "Here are four color samples. Please name all these colors. Use whichever terms seem most appropriate to you." The self-paced experiment was typically completed within 10 minutes.

Experiment 1 Results

Data Analyses

Five (4 males and 1 female) of the 52 subjects assessed were excluded from the data analyses due to unexpected stimulus variation, therefore data from 47 subjects (31 females and 16 males) are reported. For each reddish or greenish target tile named, we computed the average frequency with which subjects used the 'same' name, a 'modified' name, or a 'different' name when naming a target across local and global contexts. The same name was coded when targets were identically named across contexts. A modified name was coded when one name given was a linguistically modified form of the other. A different name was coded when two names differed across the two naming contexts.

As shown in Figure 2, on average subjects did not use the same name when naming a target tile, even though the second name was given within 2-3 min. (i.e., one intervening trial) of the first naming. The green target was named the same across local and global conditions by only 23% of the subjects. Similarly, the red target was named the same by only 21% of subjects. For the red target, 11% of subjects provided a modified name across local and global conditions, and 68% gave a

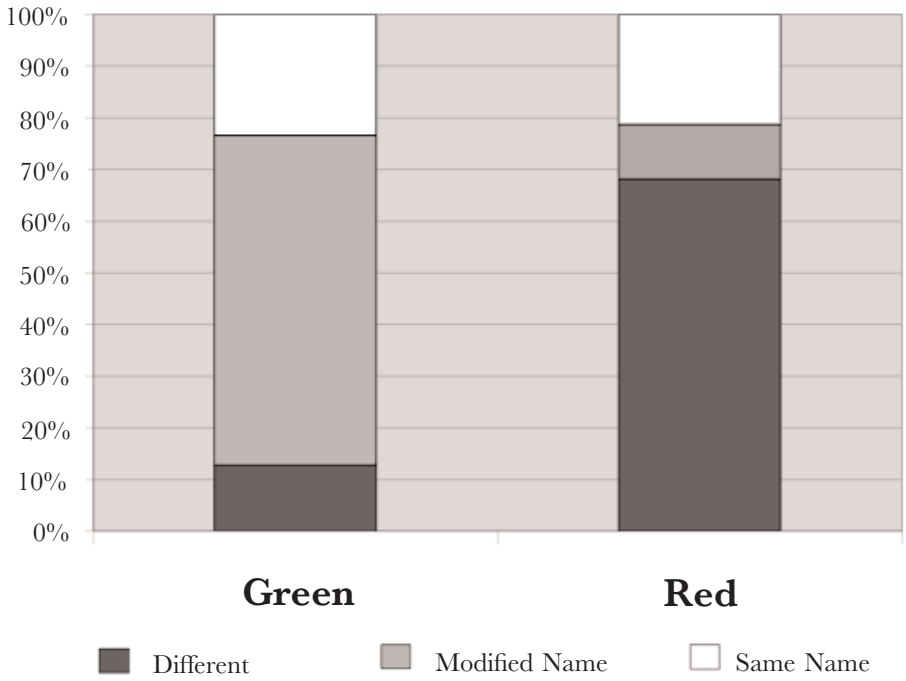

Figure 2. Color-naming variation with stimulus context variation in Experiment 1.

different name. For the green target naming, 64% of subjects provided a modified name, while 13% gave a different name (see Figure 2). Note the interaction between target color category and modified versus different naming. These results suggest that individuals flexibly map color names to color appearances, even when naming a physically identical stimulus within a 3 min. interval, so that names vary when naming occurs in different stimulus contexts. Further interpretation of these findings is presented later.

Experiment 2

Experiment 1 demonstrates that, under very minor changes in context, individual color and word representations are flexibly mapped by a color naming function. Experiment 2 investigates the nature of such a color naming function by comparing triad similarity task judgments for color

appearances and color words for: (1) all subjects and naturally occurring subgroups, (2) subgroups defined by observer's perceptually assessed phenotype, and (3) underlying retinal photopigment genotypes identified by DNA analysis. Whereas Experiment 1 used actual surface color tiles from the OSA Atlas, stimuli presented in the Experiment 2 were OSA surface color approximates presented on a computer controlled CRT display.

Design

All subjects completed the Experiment 2 triad tasks described here and an additional color mapping task, randomly assigned, so that the order of the triad and mapping tasks were counterbalanced. Color mapping task details are not presented in this report. The triad tasks data are described below.

Subjects

The 52 subjects participating in Experiment 1 also participated in Experiment 2.

Apparatus

Stimuli were generated using a PowerPC Macintosh 7200. Stimuli were displayed on a Hitachi RasterOps MC 7515, 21" CRT monitor with a 19" viewable diagonal and EBU monitor phosphors. *CIELUV* 1976 *(u*, v*)* values for phosphors were Red (.4507, .5230); Green (.1206, .5610); Blue (.1754, .1580). The screen resolution was 1024 × 786 at 70 Hz (24 bit color). The experimental display was checked regularly for accurate color rendering using a calibration measurement system dedicated to colorimetry functions (see Appendix A). Stimuli were manipulated on the screen via a trackball mouse and responses were recorded using a *PsyScope* button box (Carnegie Mellon University, Pittsburgh, PA). Experimental procedures were original routines implemented using *PsyScope* 1.2.5 PPC software (Cohen, MacWhinney, Flatt, and Provost, 1993).

Stimuli

Stimulus selection was guided by three requirements. First, basic color terms (Berlin & Kay, 1969) were included to facilitate comparison of results with other studies. Second, we wanted to use a larger number of stimuli than have been presented in previous studies (e.g. Bonnardel, Miller, Wardle, & Drew, 2002, Moore, Romney & Hsia, 2002), to forestall the possibility that consensual naming behavior becomes more likely when using small stimulus sets because the probability of observing large variations in color-to-word assignments is diminished. Third, in view of existing results (Bonnardel, Miller, Wardle, & Drew, 2002, Jameson, Highnote & Wasserman, 2001), stimuli were selected to maximize the potential for different color naming behaviors among different observer groups (e.g., dichromats, trichromats and potential retinal tetrachromats).

Table 2
Global Color Category Stimulus Names and Samples in CIELAB and CIELUV Notation (CIE, 1976)

Number	Color Name	L*	a*	b*	L*	u*	v*
1	red	56.70	67.21	49.42	56.70	149.39	32.90
2	green	84.13	-38.78	41.82	84.13	-29.21	57.33
3	yellow	93.28	-14.31	91.22	93.28	24.82	98.33
4	blue	27.07	10.69	-23.99	27.07	-2.62	-34.88
5	brown	43.76	36.83	52.12	43.76	84.73	36.66
6	purple	46.31	60.81	-22.53	46.31	69.58	-47.27
7	pink	73.53	26.68	-0.42	72.53	42.64	-12.63
8	orange	67.16	39.95	73.75	67.16	107.98	60.15
9	chartreuse	86.67	-23.01	35.76	86.67	-9.40	47.24
10	turquoise	75.88	-55.24	17.80	75.88	-58.97	28.35
11	peach	82.43	15.45	37.78	82.43	51.53	41.23
12	violet	40.72	55.79	-20.74	40.72	61.58	-41.96
13	magenta	53.36	81.16	25.71	53.36	161.38	8.41
14	tan	82.97	1.69	25.26	82.97	22.06	28.64
15	maroon	43.79	60.35	36.81	43.79	123.40	21.45
16	mauve	85.05	13.13	1.44	85.05	24.03	-8.35
17	indigo	29.74	-5.65	-6.83	29.74	-8.52	-10.92
18	teal	58.28	-39.12	2.35	58.28	-44.19	3.70
19	fuschia	59.40	68.71	16.13	59.40	128.79	1.69
20	navy	9.32	27.46	-32.43	9.32	2.13	-26.20
21	aqua	62.57	-22.04	-7.94	62.57	-30.61	-15.08

Table 3
Red Color Category Stimulus Names and Samples in CIELAB and CIELUV Notation (CIE, 1976)

Number	Color Name	L*	a*	b*	L*	u*	v*
1	dark coral	75.65	30.08	18.38	75.65	62.81	13.22
2	dark flamingo pink	66.76	53.05	16.31	66.76	100.62	5.70
3	watermelon red	62.78	60.59	23.11	62.78	119.75	12.00
4	reddish orange	58.40	65.57	68.79	58.40	154.72	44.78
5	dusty red	61.18	41.45	15.54	61.18	78.21	7.48
6	bright rose	64.88	51.39	1.44	64.88	84.06	-13.94
7	bright brick red	64.91	40.42	30.10	64.91	87.56	25.00
8	faded cranberry	69.56	34.43	12.08	69.56	64.78	4.23
9	deep rose	55.44	52.93	16.44	55.44	98.34	5.91
10	brick	58.11	51.47	34.31	58.11	109.38	25.53
11	red	53.20	64.76	50.24	53.20	143.40	32.67
12	burnt orange	57.36	59.97	65.75	57.36	141.44	44.52
13	dark raspberry	56.04	62.38	7.37	56.04	107.75	-7.53
14	light burgundy	54.75	63.63	35.86	54.75	133.38	22.94
15	burgundy red	47.75	63.26	18.91	47.75	117.85	5.83
16	rusty red	57.23	65.25	54.90	57.23	148.18	37.19
17	burgundy	49.56	60.68	32.35	49.56	123.55	19.55
18	wine	45.39	71.88	27.23	45.39	140.98	11.17
19	burnt sienna	51.04	61.02	54.97	51.04	136.76	35.35
20	dark cherry	57.47	59.89	14.77	57.47	110.28	2.25
21	strawberry	58.33	56.28	36.15	58.33	119.94	26.00

Table 4
Blue Color Category Stimulus Names and Samples in CIELAB and CIELUV Notation (CIE, 1976)

Number	Color Name	L*	a*	b*	L*	u*	v*
1	bluish purple	82.20	21.50	-19.05	82.80	20.01	-43.07
2	peacock blue	84.59	-13.34	-9.02	84.59	-21.39	-20.73
3	dark cornflower	85.03	0.79	-8.13	85.03	-1.27	-21.59
4	teal	83.62	-28.30	-5.52	83.62	-38.63	-12.47
5	turquoise	84.30	-20.22	-2.33	84.30	-26.26	-8.53
6	cornflower gray	87.33	-2.97	-3.52	87.33	-3.28	-13.63
7	lake blue	75.05	1.54	-12.67	75.05	-3.82	-27.95
8	blue	80.07	12.42	-15.52	80.07	9.50	-35.19
9	Greek Isle blue	77.32	-22.94	-11.67	77.32	-35.16	-22.87
10	dark teal	80.54	-12.08	-6.63	80.54	-18.06	-16.49
11	colonial blue	79.38	6.02	-9.33	79.38	5.08	-23.75
12	bright blue	68.10	-14.25	-17.05	68.10	-27.25	-31.94
13	Mediterranean blue	72.11	3.72	-14.14	72.11	-2.02	-30.31

Table 4 (*cont.*)

Number	Color Name	L*	a*	b*	L*	u*	v*
14	grayish blue	74.37	-3.64	-8.62	74.37	-8.03	-20.37
15	primary blue	52.71	16.46	-36.22	52.71	-4.11	-64.49
16	royal blue	61.01	24.60	-27.55	61.01	14.75	-53.64
17	sapphire	65.26	-7.75	-15.48	65.26	-18.04	-29.79
18	dark blue	65.99	9.67	-16.36	65.99	4.13	-33.94
19	midnight teal	67.31	-23.65	-9.33	67.31	-33.79	-17.69
20	navy blue	50.00	10.55	-27.95	50.00	-4.66	-49.51
21	midnight blue	43.51	15.05	-26.08	43.51	1.77	-45.11

Under these constraints, three stimulus sets from three specific regions of color space were chosen. Each set consisted of 21 color samples and 21 associated color names. These stimuli include 21 'global' colors, 21 'local red' colors, 21 'local blue' colors, and their *normative* color names (in Tables 2, 3, and 4, respectively). Stimuli are shown in Figures 3-5 and listed in Tables 2-4.

Global stimuli

Global color terms included eight terms empirically identified by Boynton and Olson (1987) to name eight salient color categories (*green, blue, purple, pink, orange, yellow, brown*, and *red*). A yellowish-green category term, *chartreuse,* was included as the ninth lexical term because it is sometimes suggested as a salient color category cross-culturally (MacLaury, 1997, Roberson, 2005). Twelve additional terms represent the most frequently reported color names (other than the nine terms just mentioned) in a free listing task by Furbee et al. (1997). The 21 global terms are shown in Table 2.

Global color appearance stimuli included computer renderings of eight OSA 'centroids' identified by Boynton and Olson (1987) corresponding to the eight salient terms listed above. Colors corresponding to the remaining 13 global terms were computer rendered versions of the OSA tiles named with the highest frequency in a study of unrestricted naming in English of all 424 OSA tiles (D'Andrade, 2003). Table 2 gives the measured *CIE* values for our rendered global color stimuli. Figure 3 shows the measured global stimuli in *CIE L*u*v** space.

Local stimuli

Local stimuli were selected to provide a within-category contrast with global stimuli and to elicit potential differences among observer groups. The rationale was that color similarity judgments should differ for local and global stimulus sets (Indow, 1988) and such differences should be important to the color-naming function. Red and blue categories were selected to explore a potential difference between observers of normal trichromat phenotype and observers with potential phenotype variation due to L-opsin gene dimorphisms (Jameson et al., 2001), particularly when naming reddish colors (Bonnardel, Miller, Wardle, & Drew, 2002). We expected that global colors, sufficiently different from each other, might not be well-suited to reveal observer group differences in our tasks because the social norm for naming has its strongest influence on the frequently used global color names (including 'basic color terms'). We also considered it plausible that greater observer group differences for the local red category region might be found compared to a different set of non-red local stimuli. Thus, local blue colors were included as a control stimulus based on the rationale that in non-deficient subjects red and blue regions should yield similar naming patterns if similarly sampled, whereas such patterns might be quite different if a physiological bias (i.e., L-opsin gene dimorphism) influenced the naming of the red category colors.

The following general selection heuristic was used to select red and blue stimuli. First, the monolexemic naming data of Boynton and Olson (1987) was used to identify 21 OSA tiles, from each category (red and blue), that were reliably named "red" and "blue" by a majority of their subjects. Second, the names given with highest frequency in an *unconstrained naming* study (D'Andrade 2003) were adopted as the corresponding color labels for the 21 red and 21 blue color appearances identified from the data of Boynton & Olson (1987). To constrain the number of stimuli to 21 (as required by the triad design) minor deviations from this strategy were necessary in both categories. Figures 4 and 5 show the 21 red and blue stimuli in *CIE L*u*v** space.

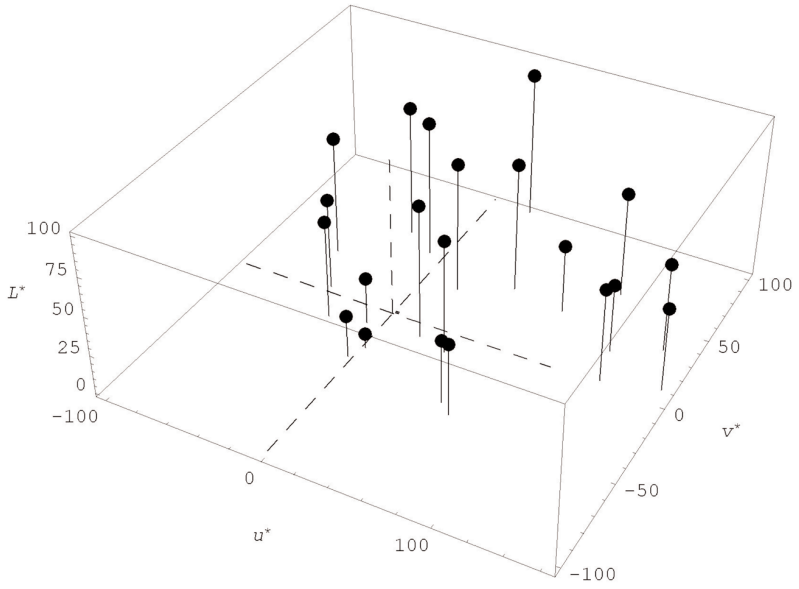

Figure 3. Experiment 2 Global color stimuli plotted in *CIE L*u*v* (1976)* stimulus space

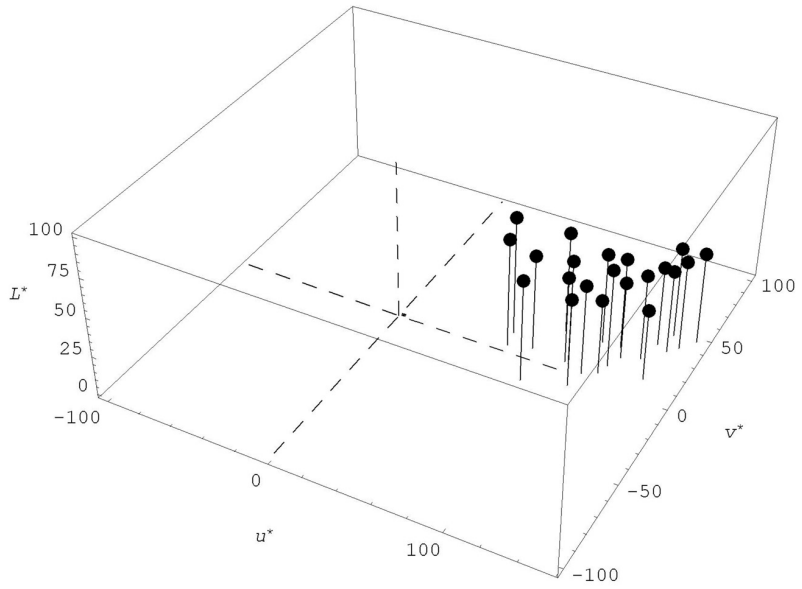

Figure 4. Experiment 2 Local Red color category stimuli plotted in *CIE L*u*v* (1976)* stimulus space

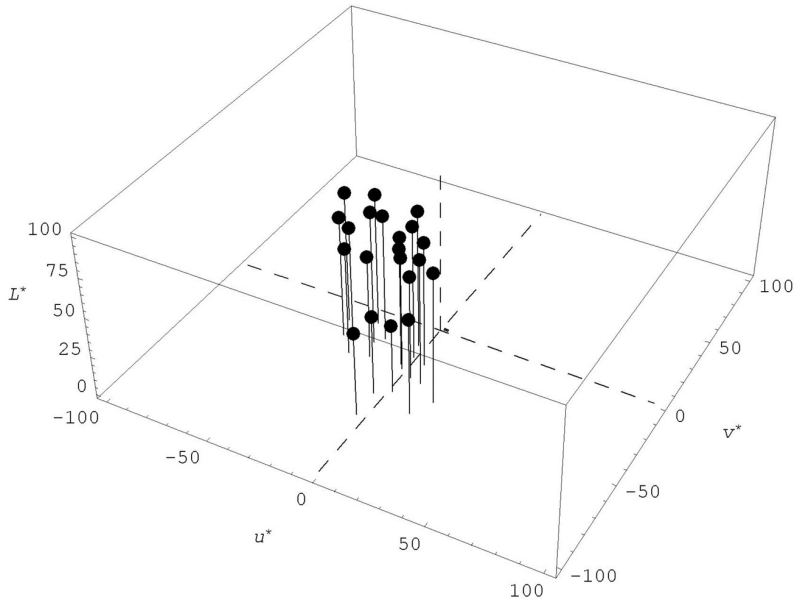

Figure 5. Experiment 2 Local Blue color category stimuli plotted in *CIE L*u*v* (1976)* stimulus space

Procedure

In a triad task, subjects are presented with a series of trials, each consisting of three stimuli. In the variant presented here, the subject must identify which of the three items is most different from the remaining two. Color name triads and color appearance triads were judged separately for global, local red and local blue conditions. This resulted in a total of six triad tasks judged by each subject. A 21 stimulus balanced incomplete block design (BIBD) was used (lambda = 1) for a total of 70 triad judgments per task. Half the subjects judged the three tasks with color names before appearances, the other half judged the three tasks with color appearances before names. One of eight orders of trials within task was randomly assigned to each subject. A sample triad stimulus configuration is shown in Figure 6 (note that subjects saw only 6a or 6b, but not both simultaneously).

Subjects were seated free of viewing restraints at a position ranging 17 in. to 20 in. distant from the computer screen and were instructed on the use of the button-box and trackball. All stimuli were displayed in a

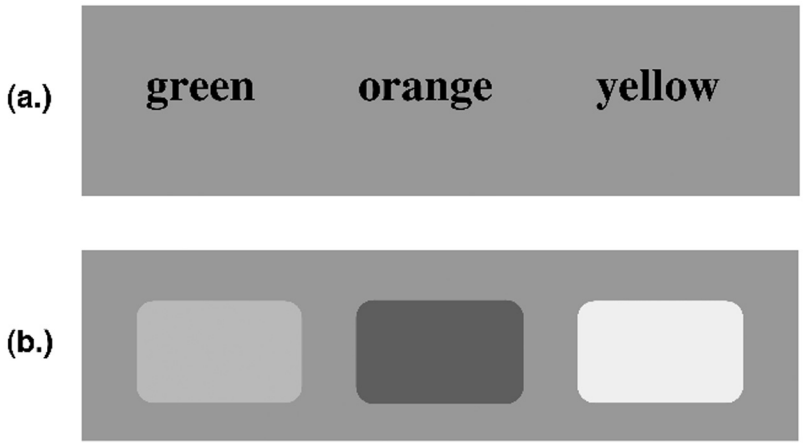

Figure 6. Grayscale approximation of the Experiment 2 Triad Task stimulus for (a) Global word and (b) Global color conditions

uniform midlevel gray background. Each individual color stimulus within a triad subtended a visual angle of 10° or greater (see Appendix A for details). Word triad stimuli appeared in a 24 point computer typeface. Triad judgments were self-paced. At the end of each 70-trial task subjects were permitted to rest before the experimenter initiated the next task.

Experiment 2 Results

Data Analyses

Three analyses were performed: (1) a comparison across tasks for all subjects and subject partitions defined by sex and visual perception assessment; (2) a comparison across tasks for subgroups defined by photopigment opsin genotype analyses; and (3) a comparison of results to a set of theory-based predictions. Analysis 1 examined all 52 subjects' data, whereas a subset of 47 subjects who volunteered blood specimens (30 female, 17 male) comprise the DNA subgroups presented in Analysis 2.

We tested predictions for color name and color appearance triad data compared across groups classified by genotype and perceptual phenotype. We examined individual internal consistency in color similarity

judgments, and its implications for the hypothesized cognitive 'color naming function.' We also examined shared knowledge for a normative color naming model and how it differed across observer groups for the tested color conditions. Finally, we assessed the correspondence between an individual's cognitive representations of color and the normative shared lexical representation, to examine how the individual's cognitive color naming function flexibly maps these two representations.

Our analyses examine two plausible bases for the mapping of names to color appearances performed by the naming function: (1) shared perceptual-physiological processing, and (2) purely cultural shared understanding of color name meanings. To compare these two alternatives we applied two quantitative methods: Consensus Analysis and Matching Analysis.

Consensus Analysis. Consensus analysis (Batchelder and Romney, 1988, 1989) was applied to evaluate the homogeneity of triad response judgments among participants. A detailed description of consensus modeling is provided in Appendix B. In brief, consensus modeling provides information about the pattern of responses within a group and provides a theoretical grounding that links the results for a particular random sample to the population from which it is drawn. Consensus analysis applies a probabilistic model to estimate the likelihood that each particular subject will correctly answer the set of questions presented. It also gives confidence estimates for the correctness of each potential response. The model assumes that if responses across subjects are correlated, it is because the responses are also correlated with latent shared knowledge among subjects. Patterns of correlation can then be used to determine the degree to which individual subjects conform to the consensual shared knowledge, and to estimate the shared correct answers to the triads. An individual's probability of giving a correct answer is called a competence rating. Competence ratings range from -1 to 1, and are normally distributed. An individual's negative rating shows extreme disagreement with the group across the range of questions asked. Thus, idiosyncratic patterns of response are readily apparent. Mean competence is one of several measures used to evaluate whether consensus exists within a group of subjects for a particular set of questions. Criteria for a good consensus model fit to a measured domain are: (1) mean competence above .500; (2) a ratio between the first and second eigenvalues greater than 3:1; (3)

absence or near-absence of negative competence ratings in the group. Data were analyzed using Anthropac (Borgatti, 1990).

Matching Analyses. The 'match' measure is a simple proportion expressing the frequency with which the choice made in the word triad (Figure 6a) exactly matched the choice made in the corresponding color triad (Figure 6b). When corresponding items (as predetermined by the experimental design) are presented in the same order in two triads, this frequency of an individual's matches, expressed as a proportion (or a percentage) across all judged triads, directly measures an individual's naming function, or the mapping across the two domains from which the items were selected (e.g., visual versus verbal domains in our case). This measure is independent of the amount of shared consensual knowledge evaluated using consensus analysis. That is, matching measured in this way does not require an individual to be 'in touch' with the shared color naming normative in the culture because it only measures *consistency of choices* for that individual across the domains assessed in the separate word and color triads. Consistency in matching implies that the representations in the two domains (i.e., words, colors) may be alternatively: (1) the same; (2) distinct representations that maintain an isomorphic within-category structure in relation to each other, resulting in an item-to-item correspondence across domains; or (3) distinct representations tightly mapped by a cognitive assignment process that links one domain to the other. Note that a naming function can, through its mappings, produce an across-domain isomorphism in category structure.

When average match measures for subgroups are compared, we get a comparative index of how reliably linked word representations are to color representations across different perceptual observer groups. For theories hypothesizing similarity or equivalence of such representations across domains, matching should be very high. In subject group comparisons, if word and color representations are the same then people from different observer groups should make the same choices when judging word or color triads.

Within each group, results for consensus analyses and matching are presented separately. All hypothesis tests are two-tailed, $p < .05$. Paired or independent t-tests and repeated measure or factorial ANOVA were used as appropriate to the groups compared.

Analysis 1: All Subjects and Natural Subgroups

Consensus Analyses

In general, consensus measures are based on agreement with the other members of the group submitted for analysis. Thus, consensus scores will vary depending on the members of the group used to calculate consensus. Random sampling assures that the group used to measure consensus will be representative and that the scores, although varying with different samples, provide a valid estimate of the true population mean consensus. Throughout Analysis 1, reported means for subgroups were based on consensus scores for agreement with the entire group, not agreement computed among subgroup members. As such, they are a measure of how much subgroups may differ from the consensual or normative shared knowledge of the larger group.

Global Conditions. Analysis of global task competence showed the expected differences with regard to color perception abilities, and a sex difference in lexical representation. Table 5 presents mean consensus scores for color 'normal' subjects (N = 48) and dichromat (N = 4) subgroups from consensus analyses of global triad tasks.[1] Figure 7A shows global color triad consensus results for groups of male, female and

Table 5
Consensus Analysis for Global Stimulus Triads: Average Consensus Measures for Subgroups

	N	Global Words	sd	Global Colors	sd
All Normal Subjects	48	.643	.144	.700	.099
Females	32	.688	.113	.693	.107
Males	16	.533	.155	.712	.081
Dichromats	4	.500	.146	.370	.179

[1] Here we refer to subjects assessed as color perception impaired as Dichromats, and subjects assessed as having unimpaired as color perception "*normal*," irrespective of their photopigment opsin genotype (discussed further below).

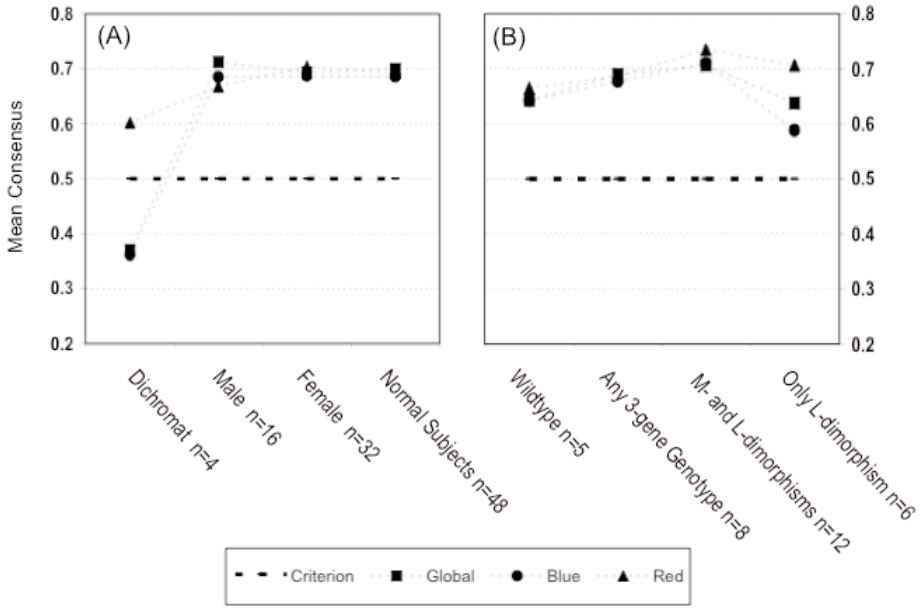

Figure 7. Mean Consensus for Color Triads for Subject Groups

Note. Panel (A) presents consensus scores for groups from Analysis 1. Panel (B) presents consensus scores for genotyped female subgroups from Analysis 2. Measures for global and local conditions are represented by square (global), triangle (local red) and circle (local blue) symbols. A consensus theory criterion is represented as the dashed horizontal line p = .5.

dichromat male subjects. There was no significant difference in global color competence based on sex (Male M = .712, SD = .081; Female M = .693, SD = .107), t(46) = 0.629, p = .533.

By comparison, Figure 8A shows that within the larger group of all subjects (excluding dichromats) global word competence for women (M = .688, SD = .113) was significantly higher than global word competence for men (M = .553, SD = .155), t(46) = -3.420, p = .001.

For all normal subjects (female and male), comparing colors to words, global color competence (M = .700, SD = .099) was significantly higher than global word competence (M = .643, SD = .143), t(47) = -2.595, p = .013. A repeated measures ANOVA with sex as a between-subjects factor was significant for both the word vs. color task and for the task by sex interaction.

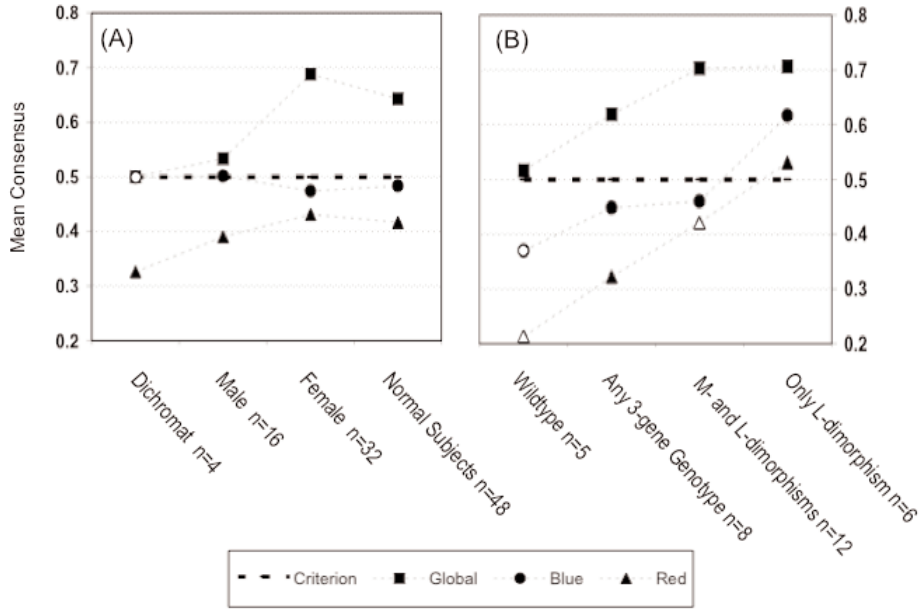

Figure 8. Mean Consensus for Word Triads for Subject Groups

Note. Panel (A) presents consensus scores for groups from Analysis 1. Panel (B) presents consensus scores for genotyped female subgroups from Analysis 2. Measures for global and local conditions are represented by square (global), triangle (local red) and circle (local blue) symbols. A consensus theory criterion is represented as the dashed horizontal line p = .5. Open symbols denote consensus analyses with low mean consensus and eigenvalues failing to show a single dominant factor (see Appendix B for a discussion of these criteria). Such consensus scores are less informative than scores satisfying the discussed criteria.

Excluding dichromats, normal male global color competence (M = .712, SD = .081) was significantly higher than global word competence (M = .553, SD = .155), t(15) = -3.883, p = .001. For females, no significant difference was found between global word competence (M = .688, SD = .113) and global color competence (M = .693, SD = .107), t(31) = -.241, p = .811.

Dichromats (N = 4) were less competent than male and female color normal subjects for global words (Dichromat M = .500, SD = .146; Normal M = .643, SD = .143), t(50) = 1.924, p = .060 and significantly less competent than all normal subjects for global colors (Dichromat M = .370, SD = .179; Normal M = .700, SD = .099), t(50) = 6.027, p = .000.

Table 6
Consensus Analysis for Local Stimulus Triads: Average Consensus Measures for Subgroups

	N	Red Words	sd	Red Colors	sd	Blue Words	sd	Blue Colors	sd
All Normal Subjects	48	.417	.149	.693	.081	.483	.151	.686	.106
Females	32	.431	.163	.704	.082	.474	.162	.687	.117
Males	16	.390	.115	.668	.076	.501	.129	.685	.082
Dichromats	4	.326	.136	.602	.182	.499	.080	.361	.157

As expected, compared to normal male subjects, dichromats were significantly less competent on the global color task (Dichromat M = .370, SD = .179; Normal M = .712, SD =.081), t(18) = 5.874, p = .000, but not significantly less competent for global words (Dichromat M = .500, SD = .146; Normal M = .553, SD = .155), t(18) = 0.619, p = .544.

Local Conditions. Results for the local tasks showed a different pattern than for global tasks. Table 6 presents mean consensus scores for local triad tasks. In addition to global results, Figure 7A and 8A also show local color and local word consensus results for male, female and dichromat male groups.

Contrasting males, females, and dichromats as three separate groups, only the blue color task showed significant differences in competence across groups, F(2, 49) = 16.017, p = .000, with the red color task approaching significance F(2, 49) = 2.742, p = .074. Color perception 'normal' males and females were not significantly different from each other on any of the local tasks (independent sample t-test), thus dichromats account for the differences noted (see below).

Males showed a significant difference in competence when comparing words and color tasks in both the red and blue domains, with word competence consistently lower than color competence. Red word competence (M = .390, SD = .115) was significantly lower than red color competence (M = .668, SD = .076, t(15) = 7.053, p = .000). Blue word competence (M = .501, SD = .129) was significantly lower than blue color competence (M = .685, SD = .082), t(15) = 4.336, p = .001. For males blue word competence (M = .501, SD = .129) was significantly higher than red word competence (M = .390, SD = .115), t(15) = 2.604,

p = .020. There was no significant difference for males in red and blue color competence.

Similarly, females showed a significant difference in competence when comparing words and color tasks in both the red and blue domains, with word competence consistently lower than color competence. Red word competence (M = .431, SD = .163) was significantly lower than red color competence (M = .704, SD = .082, t(31) = 9.197, p = .000). Blue word competence (M = .474, SD = .162) was significantly lower than blue color competence (M = .687, SD = .117), t(31) = 6.420, p = .000. Unlike males, females showed no significant difference between the blue and red word competence. Like males, females showed no significant difference between blue and red color competence. Males and females were not significantly different from each other on any of the local tasks.

Dichromats (M = .514, SD = .261) were significantly different than normal (all other subjects taken as a single group, M = .686, SD = .106) for the blue color task, t(50) = 5.717, p = .000, and nearly significantly different for the red color task, t(50) = 1.921, p = .060. Males (M = .685, SD = .082) and dichromats (M = .514, SD = .261) were significantly different from each other only on the blue color task, t(18) = 5.868, p = .000.

Consensus means were also calculated based on subgroups partitioned from the larger group of all subjects. These subgroups were comprised of males only, females only, and dichromats only, yielding a consensus based on the norms for each subgroup, rather than the norms for the entire group. Means change slightly when consensus is calculated in this manner, but the patterns of significance noted above are the same for this alternative way of calculating consensus.

Match Analyses:

Global Conditions. Table 7 presents mean match proportion measures for color 'normal' subjects (N = 48) and Dichromats (N = 4) from the global triad tasks. Figure 9A shows both global and local match measures for male and female groups. Within the larger group of all subjects (excluding dichromats), global matching for women (M = .614, SD = .089) is

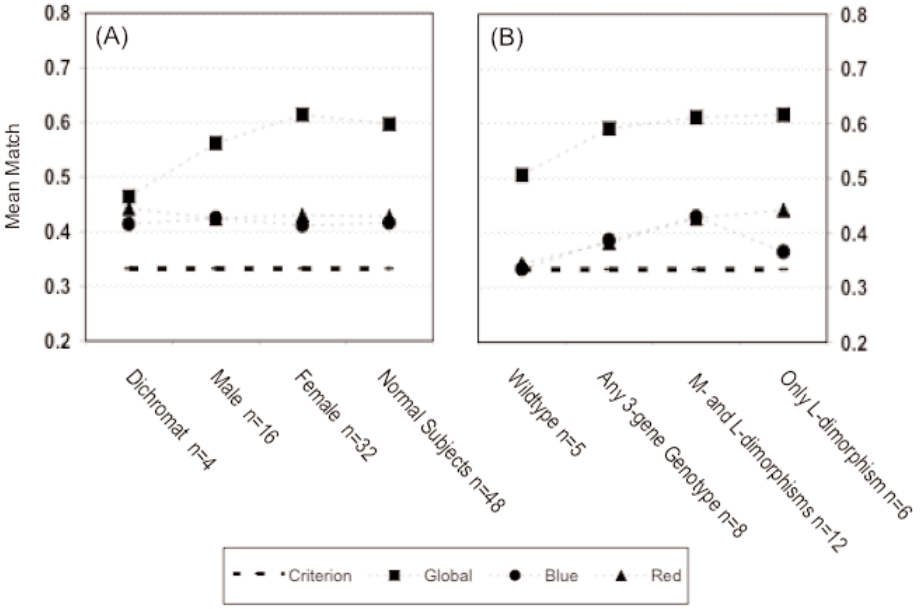

Figure 9. Mean Matches Proportions for Subject Groups

Note. Panel (A) presents consensus scores for groups from Analysis 1. Panel (B) presents match scores for genotyped female subgroups from Analysis 2. Match-proportions for global and local conditions are represented by square (global), triangle (local red) and circle (local blue) symbols. A chance-match criterion is represented as the dashed horizontal line p = .333.

Table 7
Matching Analysis for Global and Local Stimulus Triads: Average Percent Match Measures for Subgroups

	N	Global	sd	Local Red	sd	Local Blue	sd
All Normal Subjects	48	.597	.095	.428	.064	.416	.063
Females	32	.614	.089	.430	.063	.411	.066
Males	16	.562	.099	.425	.067	.425	.056
Dichromats	4	.464	.089	.443	.091	.414	.048

not significantly greater than global matching for men (M = .562, SD = .099), t(46) = -1.857, p = .070.

Dichromat matching was significantly lower than all color normal subjects for global triad tasks (Dichromat M = .464, SD = .089; Normal M = .597, SD = .095, t(50) = 2.687 p = .010. Compared to normal male subjects, dichromats were not significantly different for global triad tasks (Dichromat M = .464, SD = .089; Normal males M = .562, SD = .099), t(18) = 1.783 p = .091. Note that significance levels may have reached the $p < .05$ criterion with a larger sample of dichromats (n = 4).

Local Conditions. Table 7 also presents mean proportion-matches for color 'normal' subjects (N = 48) and Dichromats (N = 4) for the local red and local blue triad tasks. Excluding dichromats, male and female mean proportion-matches were not significantly different for either local condition (see also Figure 9A). Similarly, dichromat mean proportion-matches were not significantly different for either local condition compared to males, or females, or color normal males and females combined.

Matching analyses showed, as expected, that internal consistency of the naming function exists irrespective of color vision ability or gender. These results are interpreted in the discussion section. Note that, based on matching proportions, the local tasks were harder than the global tasks, and results for red tasks were different than for blue. The greater difficulty of the local task is seen in Figures 7A, 8A and 9A.

Analysis 2: Genotype Subgroups

Analysis 1 showed important differences in behavior across groups of subjects varying by demographic features typically used to distinguish population subgroups in the literature. Sex is often used as an attribute to understand how, for example, differences in the socialization of males and females might play distinct roles in color naming and preference. Similarly, scores on tests of color vision ability, also give a natural trait distinguishing color 'normal' observers from observers with mild to seriously impaired color ability. As discussed in Appendix C, recent advances in the genetic basis of color vision present another tool for further specifying observer photopigment opsin gene traits that are known

to affect color perception. Because perception plays a role in cognitive representation, specifying the color vision genotype provides a different route for: (1) understanding differences in color perceptual representation, and their possible relations to color cognitive and linguistic processing, and (2) addressing physiologically based perceptual processing as a plausible contributor to a shared color naming function. Analysis 2 uses color vision genotype information described in Appendix C to define four genotype subgroups that are compared with the naturalistic observer subgroups described earlier in Analysis 1, Tables 5, 6, and 7.

Briefly, the subgroup partitions for Analysis 2 include: (1) a group with a specific form of 3-gene 'wildtype' trichromat genotype (N = 8); (2) a group comprised of four different genotypes each likely to produce a trichromat color vision phenotype (N = 16); (3) a group of individuals with gene dimorphisms for both L- and M-cone opsins (N = 15); and (4) a group genotyped with an L-cone opsin dimorphism and a normal M-cone opsin (N = 7). These correspond to Groups 1-4 in Appendix C (Table C). These genotype subgroups represent the potential for the following phenotypes, respectively: (1) 'normal' trichromacy; (2) trichromacy and anomalous trichromacy; (3) potential retinal tetrachromacy in females through expression of two L-cone classes coupled with the genes

Table 8
Consensus Analysis for Global Stimulus Triads: Average Consensus Measures for Genotype Analysis Subgroups

Genotype Subgroup	N	Global Words	sd	Global Colors	sd
WildType	8	.512	.163	.679	.135
Any 3 gene Genotype	16	.572	.182	.715	.120
L & M dimorphism	15	.658	.124	.711	.072
Females WildType	5	.516	.182	.643	.154
Females Any 3 gene Genotype	8	.619	.220	.690	.145
Females L- & M-dimorphism	12	.703	.069	.707	.078
Females L-dimorphism only	6	.707	.051	.638	.132

Note. Male and female subgroup partitions in rows 1-3, and female only partitions in rows 4-7. All partitions derived from genotype groups described in Table C, Appendix C.

Table 9
Consensus Analysis for Local Stimulus Triads: Average Consensus Measures for Genotype Analysis Subgroups

Genotype Subgroup	N	Red Words	sd	Red Colors	sd	Blue Words	sd	Blue Colors	sd
WildType	8	.292	.197	.686	.074	.413	.171	.679	.160
Any 3 gene Genotype	16	.374	.140	.666	.086	.450	.164	.688	.132
L & M dimorphism	15	.379	.183	.734	.080	.453	.167	.699	.068
Females WildType	5	.213	.442	.665	.097	.370	.371	.642	.168
Females Any 3 gene Genotype	8	.322	.161	.687	.095	.448	.199	.677	.157
Females L- & M-dimorphism	12	.421	.184	.735	.091	.460	.178	.710	.069
Females L-dimorphism only	6	.532	.140	.710	.052	.615	.161	.588	.130

Note. Male and female subgroup partitions in rows 1-3, and female only partitions in rows 4-7. All partitions derived from genotype groups described in Table C, Appendix C.

for two forms of M-cone class[2], and (4) potential retinal tetrachromacy in females expressing two L-cone classes and one M-cone class.

Tables 8 and 9 present consensus analyses of triad data, and Table 10 presents matching analyses of triad data. All three tables present data for subgroup partitions defined by genotyping, and below these are compared with natural subgroups discussed earlier in Analysis 1. Genotype subgroup analyses focus primarily on comparisons among different female subgroups for the purpose of clarifying the role played by sex-linked inherited opsin genes in the assessed behaviors. In addition, we

[2] The expression of the S-cone photopigment (sensitive to short wavelength end of the visible spectrum) is implied throughout the present discussion because it is rarely deleted when autosomally transmitted via Chromosome 7. A retinal tetrachromat is an individual expressing 4 retinal cone classes (short-, medium-, long-, and a second form of long-wavelength sensitive cone type). A trichromat is any genotype leading the possession of 3 cone classes (short-, medium-, and long-wavelength sensitive).

briefly consider whether possessing opsin gene dimorphisms affect observed triad response patterns irrespective of gender. Similar to previous research (Jameson et al., 2001) we compare groups of females with L-cone dimorphisms, females with both M- and L-cone dimorphisms, with females who do not possess opsin gene dimorphisms, with all subjects (excluding dichromats).

Consensus Analyses:

Global Conditions. Table 8 presents average consensus measures for global triads, words and colors, for genotype subgroups discussed in Appendix C. A two-tailed, one-way ANOVA including the four female subgroups shown in the bottom portion of Table 8 approached a significant difference for global words but not for global colors, $F(3,27) = 2.510$, $p = .080$.

By a t-test, however, global word consensus among females possessing both M- and L-dimorphisms ($M = .703$, $SD = .069$) is significantly different from females of wildtype genotype ($M = .516$, $SD = .182$), $t(15) = 3.166$, $p = .006$; and global word consensus comparing females of wildtype genotype is significantly different from females of the L-dimorphism only group ($M = .707$, $SD = .051$), $t(9) = 2.472$, $p = .035$. Furthermore, consensus among females of wildtype genotype is not significantly different from the total group of males and females on any of the consensus measures Table 8 presents. This latter measure gives confidence in the former measured differences between dimorphic and wildtype female groups despite the small size of the female wildtype group ($N = 5$).

Interestingly, global word consensus among females possessing a L-dimorphism is not different from that for females possessing both M- and L-dimorphisms. In sum, consensus differences among subgroups in Table 8 are largest for global words, and the possession of opsin gene dimorphisms contributes to increased consensus. Figures 7B and 8B show the female genotype subgroup trends for color and word tasks respectively.

Local Conditions. Table 9 presents average consensus measures on the local triad tasks for genotype subgroups. In Table 9 the general pattern is similar to that seen for global results (Table 8), with greater consensus for color triads compared to word triads for all groups. But, with the

exception of the L-dimorphism only group (shown in Figure 8B), unlike the global results, local word triads do not meet the .500 criterion for existence of shared cultural knowledge within a domain (see Appendix B for discussion of the criterion).

Despite Figure 7B and 8B's clear trends toward increased consensus among female subgroups possessing opsin gene dimorphisms, only two of the local task differences between subgroups shown in Table 9 were significant. The female L-dimorphism only group (M = .532, SD = .014) and the female *any-3-gene genotype* group (M = .322, SD = .161) were significantly different from each other on the red word task, $t(12) = 2.543$, $p = .026$. And, the female L-dimorphism only group (M = .588, SD = .130) and the female L- & M- dimorphism group (M = .710, SD = .069) were significantly different from each other on the blue color task, $t(12) = 2.543$, $p = .026$. We interpret these differences in the Discussion below as linked to the possession of an L-opsin dimorphism.

To further explore the potential contribution of opsin gene dimorphisms on triad task choice behavior, we used contrast analysis to test a model that females possessing greater opsin gene diversity might demonstrate differences in color judgment and naming compared to female controls with less opsin gene diversity. Comparisons against female controls allows assessment of socialization influences that have been thought to differentiate female color naming behavior from that of males. Thus, in the contrast analysis females with L-cone dimorphisms were assigned a contrast coefficient of "1" while female wildtype and a female 3-gene genotype control group was assigned a contrast coefficient of "-1". The contrast analysis produced a significant result for global words, $t(39) = 3.067$, $p = .004$; and for red words $t(39) = 2.293$, $p = .027$, but no other significant contrasts. Thus for global word and red word triad tasks, the possession of opsin gene dimorphisms were associated with color naming differences (cf., Jameson et al., 2001, Bonnardel et al., 2002).

Matches Analyses

Global Conditions. Table 10 and Figure 9B present mean proportion match measures for some genotype subgroups defined in Appendix C for global and local triad tasks. Overall, Table 10 shows a trend for all groups to have better global matching compared to local matching. Relevant to

Table 10
Matching Analysis for Global and Local Stimulus Triads: Average Proportion of Matches for Genotype Analysis Subgroups

Genotype Subgroup	N	Global	sd	Local Red	sd	Local Blue	sd
WildType	8	.539	.096	.355	.062	.380	.077
Any 3 gene Genotype	16	.585	.116	.405	.067	.401	.072
L & M dimorphism	15	.593	.088	.416	.054	.428	.050
Females WildType	5	.506	.093	.343	.044	.334	.046
Females Any 3 gene Genotype	8	.591	.139	.382	.065	.386	.079
Females L- & M-dimorphism	12	.611	.065	.427	.053	.429	.043
Females L-dimorphism only	6	.614	.094	.441	.059	.367	.059

Note. Male and female subgroup partitions in rows 1-3, and female only partitions in rows 4-7. All partitions derived from genotype groups described in Table C, Appendix C.

Appendix C's subgroup comparisons, global matching was not significantly different for females with L-cone dimorphisms and females with both M- and L-cone dimorphisms. Females of wildtype genotype (M = .506, SD = .093) had significantly lower matching compared to females possessing both M- and L-cone dimorphisms (M = .611, SD = .065), t(15) = 2.683, p = .017. However, female wildtype matching (M = .506, SD = .093) was also significantly lower compared to all male and female subjects matching (excluding dichromats) (M = .597, SD = 095), t(51) = -2.041, p = .046.[3]

Local Conditions. Table 10 also presents average proportion match measures for local red and local blue triad tasks. For local red and blue, the pattern of significant differences parallel those found for global matching. Females of L-dimorphism-only group genotype (M = .367, SD = .059) had significantly lower matching for blue tasks compared to females possessing both M- and L-cone dimorphisms (M = .429, SD = .043), t(16) = 2.55, p = .022. We tentatively attribute this to the additional possession of multiple M-opsin genes in the latter group, although

[3] Although this difference might be explained by the small size of the female wildtype group (N = 5).

given the debate on the utility of M-cone dimorphisms, further data is needed to demonstrate this relationship.

Compared to Figures 7B and 8B, Figure 9B's matching measures also suggest a general trend, albeit non-significant, toward increased matching among female subgroups possessing opsin gene dimorphisms. However, the rather small size of some of our female genotype groups suggest that, similar to the dichromat results presented earlier, the observed trends seen in Figures 7B, 8B and 9B may have reached the $p < .05$ significance level with a larger samples of subjects.

Analysis 3: Tests of Predictions

A series of qualitative predictions were made for the expected results of our global and local naming tasks across observer groups defined by

Table 11
Predictions for Average Consensus and Matching Measures for Subgroups

		Global	Local Red	Local Blue
All Normal Subjects	color triads	Good	Medium	Medium
	word triads	High	Low	Low
	matching	Good	Medium	Medium
Females	color triads	Good	Medium	Medium
	word triads	High	Medium	Medium
	matching	Good	Medium	Medium
Males	color triads	Good	Medium	Medium
	word triads	High	Low	Low
	matching	Good	Medium	Medium
Dichromats	color triads	Poor	Medium	Poor
	word triads	Good	Low	Low
	matching	Poor	Low	Poor
Potential Tetrachromats	color triads	High	Good	Medium
	word triads	High	Good	Medium
	matching	Good	High	Good

Note. Five levels of predictions range Poor, Low, Medium, Good, High. Predictions for triads tasks refer to group average consensus from Consensus Analyses. Predictions for matching refers to group average matching measures. Generally, Low/Poor matching indicates nonisomorphic color and word representations. Good/High consensus for Words indicates a shared lexical representations. Low/Poor consensus for Colors indicates individual variation in perceptual representations.

demographic and genotype information. Table 11 presents initial predictions for the data analyses described earlier.[4] Table 11's predictions are entirely derived from three assumptions about individual perceptual variation and naming patterns, and their relation to shared group naming patterns. Appendix D provides the rationale for the predictions and the principles on which they are based.

For the predictions in Table 11, mean consensus serves as a measure of inter-individual agreement *within each domain* (e.g., word, color, global, and local comparisons). In contrast, mean matching serves as a measure of intra-individual triad judgment consistency *across domains* (i.e., words and colors only), which can be compared across conditions (i.e., global and local). Thus, a high proportion of matches requires that individual subject's have the ability to both differentiate color appearances *and* differentiate color word meanings in a similar manner so that comparative judgments of items within a triad produce the same result in both domains. Individuals who can differentiate items well in both domains (words and colors), will produce a higher proportion of matches. Also, those who have a "tight mapping" of colors and words (i.e., a richly lexicalized color space, as opposed to sparsely lexicalized) should perform better in the match analyses even though they may not be as "in-tune" with the shared knowledge about the modal, or *correct*, names within their linguistic society. Thus, an individual with a tight mapping may have a high proportion of matches while receiving a low 'competence' score from consensus analyses, because the latter depends on agreement with culturally shared knowledge, not individual consistency of mapping. In short, consensus and matching measure two different aspects of naming behavior.

Consensus analysis is ideal for comparing color and word triad performance across domains because the basis for judgment seems likely to be different across the two domains. Lexical triads are likely to be judged using shared cultural knowledge, whereas color triads are likely to be judged using individual perceptual representations. Thus, we predicted in Table 11 that consensus would be more likely for the lexical triads due

[4] Table 11's predictions were made prior to conducting Analysis 1 and 2 and without knowledge of the triad task results.

to reliance on shared knowledge derived from one's culture, with less consensus seen on color triads because of perceptual subgroup differences.

The predictions shown in Table 11 were analyzed for agreement between the nominal predictions coded as ordinal values, and the observed interval-scaled empirical results for each cell-wise relationship presented in Table 11. This ordinal-interval level correspondence was measured using appropriate nonparametric statistics: (1) Goodman and Kruskal's *gamma*, and (2) Kendall's *tau*. Scale differences required that consensus scores and matching proportions be analyzed separately.

Our predictions for consensus corresponded well with the consensus actually observed: The Goodman and Kruskal's *gamma* correlation was $\gamma = .574$, $p = .000$. This can be interpreted as 79% ordinal agreement between the observed scores and Table 11's predicted consensus values. Agreement was also significant using the Kendall's *Tau-b* statistic, $\tau = .506$, $p = .000$ (two-tailed). In contrast, predictions for matching agreed less with the observed proportions: The Goodman and Kruskal's gamma correlation was $\gamma = .266$, $p = .278$. This can be interpreted as 63% ordinal agreement between the observed match measures and Table 11's predicted matching values. This agreement was also nonsignificant using the Kendall's Tau-b statistic, $\tau = .232$, $p = .278$ (two-tailed). The primary reason for this discrepancy was that our initial intuitions were incorrect for dichromats. We expected dichromats to have much less internal consistency in their triad judgments, which would have decreased their match proportions. This was not the case – dichromat matching measures were not much worse than the other groups assessed. However, as shown in Figure 9, panels A and B, the local triad tasks produced above-chance but very low match proportions for nearly all subjects, including those with unimpaired color perception.

Discussion

Experiment 1 demonstrated the existence of a naming function that flexibly maps names to color samples dependent on whether the context in which the items appear is comprised of locally similar colors versus colors from very different regions of color space (global). Note that the target items named in across contexts Experiment 1 never varied in color

appearance (i.e., they were the same physical sample across formats), yet naming varied substantially under simple changes in context.

Experiment 2 demonstrated that different subgroups of observers, including some identified by DNA genotype, produce different color and word triad choice behavior in different global and local contexts. Of particular interest were findings that females possessing greater photopigment opsin gene diversity showed consistent increases in measures compared to other female subgroups, on both the visual and verbal tasks. Experiment 2's results support the idea that individual differences in color representation (presumably linked to individual perceptual variation) may contribute to greater expertise in accessing shared cultural knowledge about naming as well as greater consistency in naming across domains (a richly lexicalized color space, or "tighter mapping" of names to items).

We varied three contributors to behavior in these experiments: (1) context in which the stimulus is judged, including the extent of the stimulus category or color region; (2) individually varying factors such as sex of the observer and color vision ability; and (3) genetic potential to express different cone classes among same-sex observers. Our use of consensus modeling showed that despite these influencing factors, and their variable effect on individuals, subjects exhibited a considerable degree of shared knowledge regarding judgments of global color terms and global color appearances. This suggests that observer perceptual variation need not impede participation in a shared knowledge structure of color semantics, nor do individual differences in perceptual representation necessarily impair an individual's appropriate use of a culture's color naming system (consistent with Jameson, 2005a, 2005b).

While varying from color normal observers, even dichromats showed considerable consensus with respect to the global word task, consistent with existing findings (Jameson & Hurvich, 1978, Shepard & Cooper, 1992). We suggest that a color naming function maps such individual representations of color appearance to a culturally *normative* color-naming system (Jameson, 2005b). As a semantic information code, such a variable mapping is adequate for within-culture communication. Thus, our data support the claim that color language and color perception are independently represented and are linked by a cognitive naming function (Jameson & Alvarado, 2003). The purpose of such a naming function is

to assign names to color referents as needed (or as appropriate) for a specific performance related goal (Jameson & Alvarado, 2003).

One assumption of our theory is that color name assignments should vary under different circumstances and different task demands. This was observed in Experiment 1, where colors were named differently in local contexts compared to global ones, and where features of naming also differed for reddish stimuli (a more circumscribed region of visible color space) compared to greenish stimuli (a broader region of visible color space) for which greater use of modifiers occurs. This observed flexibility in naming challenges the widely-held belief that lexical and perceptual representations are isomorphic. It also raises questions about whether naming robustness depends on invariant stimulus features of certain "basic" or "centroid" colors believed to be perceptually salient, or whether it is instead a measure of the dynamics that guide the application of the naming function.

Explaining observed gender differences

Previous investigations have suggested an absence of gender differences in similarity judgments of color stimuli and words (Moore, Romney & Hsia, 2002). Although our purpose was not to examine gender differences explicitly, we did find significant gender differences in several measures. First, females showed higher global word competence than males. Second, males showed significant differences between global words and global colors and between the red and blue word tasks, whereas females did not. Because consensus modeling estimates latent shared knowledge using the responses of subjects, the more frequently occurring responses will dictate the latent answer key. When males and females are equally represented in a data set (which they were not in this study), or when they share the same knowledge despite their unequal numbers, then this method of modeling consensus should produce a valid result for both individual subjects and the group. If males are a considerable minority within the sample and also draw upon a different shared knowledge, either because of different perceptual experiences or because of differences of participation within the larger culture, then they will produce answers that are less correlated with the latent answer key

and they will be given lower competence scores. Thus, the lower competence scores for males and the significant differences noted in this study suggest that males are drawing upon different shared knowledge. Based on the triad task information alone we cannot identify the source of this difference in shared knowledge. But this difference in competence was found regardless of whether means for male competence was computed based on a total group consensus analysis, or on an exclusively male group consensus analysis. Contrary to Moore et al. (2002) this demonstrates a clear gender difference.

One possible explanation for this observed gender difference is potential behavioral correlates of the possession of L- and M-cone opsin gene dimorphisms among some female subjects. Not all females share this genotype, and not all females with dimorphic genotypes express them phenotypically, but genotype subgroup results show clear trends towards greater consensus scores for those females who do possess dimorphic genotypes (compared to both female wildtype and other 3-gene genotype controls). Consistent with this suggestion, and the idea that specifically L-opsin gene dimorphisms are important, only one group we considered – the L-opsin gene heterozygotes (N = 6) – performed above criterion on *all* consensus and matching measures assessed (see Figures 7B, 8B, and 9B).

Despite a long-standing suggestion that gender differences in color language performance arise primarily due to socialization influences (e.g., Rich, 1977), we believe it is likely that the word triad performance advantages for female subjects found in our data arose, in part, from subtle divergence in perceptual experience and its cognitive correlates (see Jameson, 2005b). Females with the potential to express L-cone dimorphisms were sufficiently frequent (N = 18) here to influence the word triad latent answer key among all subjects. The advantages our female subjects show in Analysis 1 may be attributable to the possibility that females possessing opsin dimorphisms in the sample truly performed differently and as a result elevated the overall female mean consensus score. To evaluate this possibility, we differentiated males in our genotype subgroup analyses. This allowed us to assess whether significant differences existed for female and male subjects based on the presence or absence of opsin gene dimorphisms. As reported above, females possessing dimorphic genotypes were found to differ significantly from females

with non-dimorphic genotypes – a finding that supports our interpretation of genotype contributions to behavior. Moreover, results from non-dimorphic females did not significantly differ from those of the male subgroup. Thus, for some findings, females who do not possess opsin gene dimorphisms exhibit triad task behaviors similar to males. This supports our interpretation that female L-cone dimorphisms contributed to the observed differences in behavior. It does not support the possibility of a simple gender difference based solely on socialization, nor does it support the idea that females in general dictated the latent answer key. This result is highly consistent with genotype-based gender difference found by Jameson et al. (2001).

Note that the gender differences existing in the global data are not present in the *local* triad tasks. There are no analogous differences for either the two local *color* tasks (red or blue) or the local blue *word* task because there is little difference between the responses of males and females in those domains. However, in addition to significant differences among genotyped female groups for red words, blue colors and blue matches, our contrast analysis between females possessing L-dimorphisms and those not possessing dimorphisms confirmed significant contrasts for red word triads, suggesting again that possession of opsin gene dimorphisms in females may be associated with variation in color expertise.

It is unknown what processes are responsible for the trends seen in Figures 7B and 8B suggesting subtle differences in color naming among females with opsin gene dimorphisms. Of particular interest is the finding the only-L-dimorphism group alone performs above criterion on all color, word and matching measures (a finding consistent with Jameson et al., 2001 results). In addition, the finding that this group of L-opsin females shows expertise in word triads (even on our local word triads for which other groups perform very poorly), and show color triad performance differences, suggests that subtle perceptual differences arising from L-opsin gene dimorphisms may underlie the variation in color task consensus scores, and that such perceptual differences do not impair color naming consensus (a finding consistent with Jameson, 2005b). It is possible that subtle differences in perception arising from expressed dimorphisms might bias such females towards developing color expertise by cognitively heightening their color awareness, or by generating a greater interest in color compared to non-dimorphic females. Over a lifetime this

subtle increased interest in color might lead to greater cultivation of color naming expertise, which may result in the more robust lexicalized code evident in word triad tasks. Although the above scenario is speculative, it does accord with our contrast analysis between all females possessing dimorphisms and those not possessing dimorphisms found significant contrasts for red word triads and not blue word triads, because opsin gene dimorphisms mainly impact perceptual processing of longer-wavelength stimuli as opposed to short-wavelength (or bluish) stimuli. Although these findings are in many ways compatible with other results emerging in this research area (Jameson et al., 2001, Jameson, Bimler & Wasserman, 2005a, 2005b, Bonnardel et al., 2002, Jakab & Wenzel, 2004, Bimler, Kirkland & Jameson, 2004), more work is needed to explain these observed opsin gene linked differences in naming behavior.

Compared to the consensus analyses, the matching analyses showed no significant differences between males and females or between color-normal individuals and dichromats, or among genotyped subgroups in general. The global tasks produced higher proportion-matches than the local tasks for all individuals. This suggests that the naming function operates flexibly to produce a consistency of judgment across domains independent of the individual's perceptual experience and their understanding of the culture's shared lexicon for naming colors. Males may sometimes diverge from females in their actual choices, dichromats may diverge from those with normal color vision, yet all these groups showed similar internal consistency of response across domains that suggests that the naming function operates irrespective of the specific representations in either domain. The naming function is not anchored to color appearances or to lexical meaning but operates flexibly to link the two no matter what their representation. Thus, the naming function permits mapping no matter what the individual's color experience or their understanding of the lexicon. The matching proportion essentially measures the tightness of that mapping.

Relevance for cross-cultural investigations

The results presented here are specific to English-language societies and are not intended to suggest patterns of results that should be expected for color naming in other ethnolinguistic groups. Nevertheless, our results

illustrate the importance of individual variation to intra- and inter-cultural investigations of color naming systems. The difference we observed between color word judgments and color appearance judgments may reflect the different influences of conformance to a normative system for naming compared to reliance on a perceptual representation. For our normal color vision subjects, color appearance tasks consistently produced greater consensus compared to word tasks. Observer subgroups identified in our sample showed differences in word task performance that correlated with perceptual phenotype groups (estimated by genotyping). Together these results suggest that: (1) the socially defined lexical domain is more complex and involves factors other than those underlying individual perceptual representations; and (2) individual variation in perceptual experience (even within groups of subjects uniformly diagnosed as color vision normal) can differentially influence observers behavior in color similarity and color naming tasks. Contrary to the view expressed by Webster and Kay (In Press), we believe our findings demonstrate the importance of individual variation within a culture for identifying cross-cultural universals in color naming (consistent with Moore et al., 2002).

Variation in color vision ability is but one of many individually varying factors that shape a subject's perceptual and categorization behaviors. Even so, it seems important to investigate the role that opsin gene frequency has in patterns of cross-cultural color naming. The frequency of opsin gene dimorphisms are know to vary greatly across ethnolinguistic societies (Sharpe, Stockman, Jägle & Nathans, 1999, p. 30), whereas less is known about the frequencies of phenotype expression. Thus another important contributor to cross-cultural color naming may be the extent to which an examined ethnolinguistic group exhibits opsin gene variation within its population. Profound influences on color naming systems attributable largely to color perception phenotype are well known (Sacks, 1997), and are suggested as linked to reorganization of human cortical maps following inherited photoreceptor abnormalities (Baseler, et al., 2002). The challenge for cross-cultural color naming researchers is to identify the contributions of phenotype variation, and to integrate these with cultural factors such as a society's demands of conformance to a social naming norm, and other societal features that make specific forms of naming more salient (as described by Jameson, 2005b). Cognitive processes such as the color naming function described here

mediate such factors in ways that impose a system of shared meaning for effective communication within a culture, despite the range of possible perceptual experiences. We believe that the universalities of cross-cultural color naming do not derive from a uniform physiology, but arise from the similar ways that human cognitive processes solve the problem of how to lexicalize color given different cultural demands and varying perceptual experience.

REFERENCES

Asenjo, A. B., Rim, J. & Oprian, D. D.
1994 Molecular determinants of human red/green color discrimination. *Neuron, 12*, 1131-1138.

Baseler, H. A., Brewer, A. A., Sharpe, L. T., Moreland, A. B., Jägle, H. & Wandell, B. A.
2002 Reorganization of human cortical maps caused by inherited photoreceptor abnormalities. *Nature Neuroscience, 5*, 364-370.

Batchelder, W. H., & Romney, A. K.
1988 Test theory without an answer key. *Psychometrika, 53*, 71-92.

Batchelder, W. H., & Romney, A. K.
1989 New results in test theory without an answer key. In E. Roskam (Ed.), *Mathematical psychology in progress*, (pp. 229-248). Heidelberg, Germany: Springer.

Berlin, B., & Kay, P.
1969/1999 *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.

Bimler, D. L., Kirkland, J., & Jameson, K. A.
2004 Quantifying variations in personal color spaces: Are there sex differences in color perception? *COLOR Research & Application, 29*, 128-134.

Bonnardel, V., Miller, S., Wardle, L., & Drew, E.
2002 Gender differences in colour-naming task. 25th European Conference on Visual Perception. August 27, 2002. Glasgow, Scotland. *Perception, 31, supplement, 71a.*

Borgatti, S. P.
1990 *Anthropac, v. 4.0*. Analytic Technologies, Columbia.

Boynton, R., & Olson, C.
1987 Locating basic colors in the OSA space. *COLOR Research and Application, 12*, 94-105.

Braun, J. & Julesz, B.
1998 Withdrawing attention at little or no cost: Detection and discrimination tasks. *Perception & Psychophysics, 60*, 1-23.

Carrel, L. & Willard, H. F.
2005 X-inactivation profile reveals extensive variability in x-linked gene expression in females. *Nature, 434*, 400-404.

CHAPANIS, A.
1965 Color names for color space. *American Scientist, 53*, 327-346.
COHEN J. D., MACWHINNEY B., FLATT M., AND PROVOST J.
1993 PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*, 257-271. http://psyscope.psy.cmu.edu/.
D'ANDRADE, R. G.
2003 English color naming data for 424 OSA color tiles. University of California, San Diego. Department of Anthropology.
DEDRICK, D.
1998 *Naming the rainbow: Colour language, colour science, and culture*. Kluwer Academic Publishers: Dordrecht.
DEDRICK, D.
1997 Colour categorization and the space between perception and language. *Behavioural and Brain Sciences, 20*, 187-188.
DEREFELDT G., SWARTLING T.
1995 Colour concept retrieval by free colour naming. Identification of up to 30 colours without training. *Displays, 16*, 69-77.
DEREFELDT, G., SWARTLING, T., BERGGRUND, U., BODROGI, P.
2004 Cognitive color. *COLOR Research and Application, 29*, 7-19.
FARNSWORTH, D.
1943 The Farnsworth-Munsell 100 hue and dichotomous tests for color vision. *Journal of the Optical Society of America, 33*, 568-578.
FURBEE, N. L., MAYNARD K., SMITH, J. J., BENFER JR., R. A., QUICK, S., & ROSS, L.
1997 The Emergence of color cognition from color perception. *Journal of Linguistic Anthropology, 6*, 223-240.
GRIFFIN, L. D.
2004 Wherefore the basic colors. *Perception, 33*, 753.
INDOW, T.
1988 Multidimensional studies of Munsell color solid. Psychological Review, 95, 456-470.
ISHIHARA, S.
1996 *Ishihara's tests for colour-deficiency (concise edition)*. Tokyo, Japan: Kanehara & Co., Ltd.
JAKAB, Z., & WENZEL, K.
2004 Detecting tetrachromacy in human subjects. 27th Annual Meeting of the European Conference on Visual Perception (ECVP). *Perception, 33*, Supplement, S64.
JAMESON, D., & HURVICH, L. M.
1978 Dichromatic color language: "reds" and "greens" don't look alike but their colors do. *Sensory Processes 2*, 146-155.
JAMESON, K. A.
2005a Why GRUE? An interpoint-distance model analysis of composite color categories. *Cross-Cultural Research, 39*, 159-194.
JAMESON, K. A.
2005b Culture & cognition: What is universal about the representation of color experience? *Journal of Cognition and Culture, 5*. This Issue.

JAMESON, K. A., BIMLER, D. & WASSERMAN, L.
2005a Re-assessing perceptual diagnostics for observers with diverse retinal photopigment genotypes. In *Progress in Color Studies 2: Cognition*. Pitchford, N. J. & Biggam, C. P. (Ed.s). Amsterdam: John Benjamins Publishing Co. To appear 2006.

JAMESON, K. A., BIMLER, D. & WASSERMAN, L.
2005b Color perception assessment for observers with variable color vision genotypes. *Program for the 17th annual meeting of the American Psychological Association*. Poster Abstract III-077, p. 175.

JAMESON, K. A., HIGHNOTE, S. M., & WASSERMAN, L. M.
2001 Richer color experience in observers with multiple photopigment opsin genes. *Psychonomic Bulletin & Review, 8*, 244-261.

JAMESON, K. A., & ALVARADO, N.
2003 The relational correspondence between category exemplars and names. *Philosophical Psychology 16*, 23-47.

JAMESON, K., SAYIM, B., & ALVARADO, N.
2003 Asymmetries between cognitive and perceptual representations of color: Consequences for theories of color naming. Invited presentation at "Color Perception: Philosophical and Scientific Perspectives." Department of Philosophy, University of British Columbia. October, 2003. Vancouver, Canada.

JAMESON, K., SAYIM, B., & ALVARADO, N.
2005 Semantic and perceptual representations of color and a cognitive color-naming function. *Program for the 17th annual meeting of the American Psychological Association*. Poster Abstract VI-027, p. 237.

JAMESON, K. A., & ALVARADO, N.
2003 Differences in color salience in Vietnamese and English. *COLOR Research & Application, 28*, 113-138.

JORDAN, G. & MOLLON, J. D.
1993 A study of women heterozygous for color deficiencies. *Vision Research, 33*, 1495-1508.

KAHNEMAN, D. & TVERSKY, A.
1984 Choices, values and frames. *American Psychologist, 39*, 341-350.

KAY, P., BERLIN, B., & MERRIFIELD, W. R.
1991 Biocultural implications of systems of color naming. *Journal of Linguistic Anthropology, 1*, 12-25.

KAY, P., & MAFFI, L.
1999 Color appearance and emergence and evolution of basic color lexicons. *American Anthropologist, 101*, 743-760.

KUEHNI, R. G.
2001a Determination of unique hues using Munsell color chips. *COLOR Research and Application, 26*, 61-66.

KUEHNI, R. G.
2001b Focal colors and unique hues. *COLOR Research and Application, 26,* 171-172.

KUEHNI, R. G.
2004 Variability in unique hue selection: A surprising phenomenon. *COLOR Research & Application, 29,* 158-162.

KUEHNI, R. G.
2005a Unique hue stimulus choice: A constraint on hue category formation. *Journal of Cognition and Culture*, This Issue.
KUEHNI, R. G.
2005b Focal variability and unique hue stimulus variability. *Journal of Cognition and Culture*, This Issue.
MACLAURY, R. E.
1997 *Color and Cognition in Mesoamerica: Constructing Categories as Vantages*. Austin, University of Texas Press.
MACADAM, D. L.
1974 Uniform color scales. *Journal of the Optical Society of America, 64*, 1691-1702.
MARMOR, G.
1978 Age at the onset of blindness and the development of the semantics of color names. *Journal of Experimental Child Psychology, 25,* 267-278.
MERBS, S. L. & NATHANS, J.
1992a Absorption spectra of human cone pigments. *Nature, 356,* 433-435.
MERBS, S. L. & NATHANS, J.
1992b Absorption spectra of the hybrid pigments responsible for anomalous color vision. *Science, 258*, 464-466.
MERBS, S. L. & NATHANS, J.
1993 Role of hydroxyl-bearing amino acids in differentially tuning the absorption spectra of the human red and green cone pigments. *Photochemistry and Photobiology, 58*, 706-710.
MOLLON, J. D.
1992 Worlds of difference. *Nature, 356*, 378-379.
MOLLON, J. D.
1995 Seeing color. In Lamb, T., & Bourriau, J. (Eds.), *Color: Art and Science*. Cambridge University Press, Cambridge, United Kingdom (pp. 127-150).
MOORE, C. C., ROMNEY, A. K., & HSIA, T.-L.
2002 Cultural, gender, and individual differences in perceptual and semantic structures of basic colors in Chinese and English. *Journal of Cognition and Culture, 2*, 1-28.
NAGY, A. L., MACLEOD, D. I. A., HEYNEMAN, N. E., & EISNER, A.
1981 Four cone pigments in women heterozygous for color deficiency. *Journal of the Optical Society of America, A, 71*, 719-722.
NATHANS, J., THOMAS, D., & HOGNESS, D. S.
1986 Molecular genetics of human color vision: The genes encoding blue, green, and red pigments. *Science, 232*, 193-232.
NATHANS, J.
1999 The evolution and physiology of human color vision: Insights from molecular genetic studies of visual pigments. *Neuron, 24*, 299-312.
OSA Uniform Color Scales
1977 Washington, Optical Society of America.
POINTER, M. R., & ATTRIDGE, G. G.
1998 The number of discernible colours. *COLOR Research and Application, 23*, 52-54.
RICH, E.
1977 Sex-related differences in color vocabulary. *Language and Speech, 20,* 404-409.

ROBERSON, D., DAVIES, I., & DAVIDOFF, J.
2000 Color categories are not universal: replications and new evidence from a stone age culture. *Journal of Experimental Psychology: General, 129*, 369-398.
ROBERSON, D.
2005 Color categories are culturally diverse in cognition as well as in language. *Cross Cultural Research, 39*, 56-71.
ROMNEY, A. K., WELLER, S., & BATCHELDER, W. H.
1986 Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist, 88*, 313-338.
SACKS, O. W.
1997 *The island of the colorblind and cycad island*. New York : Alfred A. Knopf, Publishers.
SAYIM, B., JAMESON, K. A., ALVARADO, N.
2003 Color naming relations in perceptual color space. *Investigative Ophthalmological and Vision Sciences, 44*, 1911.
SHARPE, L. T., STOCKMAN, A., JÄGLE, H., & NATHANS, J.
1999 Opsin genes, cone photopigments, color vision, and color blindness. In Gegenfurtner, K. R., Sharpe, L. T. (Ed.s), *Color vision: From genes to perception*. Cambridge University Press.
SHEPARD, R., & CHIPMAN, S.
1970 Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology, 1*, 1-17.
SHEPARD R. N., & COOPER, L. A.
1992 Representation of colors in the blind, color-blind, and normally sighted. *Psychological Science, 3*, 97-104.
STEELS, L. & BELPAEME, T.
2005 Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and Brain Sciences, 28*, 469-489.
WASSERMAN, L. M., SZESZEL, M., & JAMESON, K. A.
2001 Long-range polymerase chain reaction method for detection of human red and green opsin gene polymorphisms. University of California, San Diego. http://aris.ss.uci.edu/cogsci/personnel/kjameson/PCR.pdf.
WEBSTER, M. A. & KAY, P.
In Press Individual and population differences in focal colors. *Anthropology of color: Interdisciplinary multilevel modeling*, (MacLaury, R. E., Paramei, G. V., & Dedrick, D., Ed.s). Amsterdam: John Benjamins.
WELLER, S., & ROMNEY, A.
1988 *Systematic data collection*. Thousand Oaks, CA: Sage.
WINDERICKX, J. BATTISTI, L., MOTULSKY, A. G., DEEB, S. S.
1992 Selective expression of human X chromosome-linked green opsin genes. *Proceedings of the National Academy of Science, 89*, 9710-9714.
YAMADE, S., HAYASHI, S., VEYAMA, H., TANABE, S., HUKAMI, K., ICHIKAWA, K. & Tachibana, M.
2001 Red-green pigment gene analysis as a clinical diagnostic tool. *COLOR Research and Application, 26*, S89-S92.
YENDRIKHOVSKIJ, S. N.
2001 A computational model of colour categorization. *Color Research and Application, 26,* S235-S238.

Appendix A

Stimulus Rendering and Color Display Calibration

Stimuli were generated using a PowerPC Macintosh 7200/90 running OS 8.6, 256mb RAM, a single processor with 601 MHz, and 4mb video memory (Apple Computers, Cupertino, CA). Stimuli were displayed on a Hitachi RasterOps MC 7515, 21" CRT monitor with a 19" viewable diagonal and EBU monitor phosphors. *UCS CIELUV* 1976 *(u*, v*)* values for phosphors were Red (.4507, .5230); Green (.1206, .5610); Blue (.1754, .1580). The screen resolution was 1024 x 786, 24-bit color resolution. The Rasterops display was regularly checked for accurate color rendering during the experiment using a calibration measurement system including an LabSphere Integrating Sphere (model #4P-GPS-060-SF) and a collimating lens (Ocean Optics part #74-VIS) via fiber optic to an Ocean Optics spectrometer ISA card (model PC2000) interfaced with a 486 Intel PC computer running *Ocean Optics OOIIrrad* software and dedicated to radiometric and colorimetry functions. The perimeter of experimental display device, and other non-essential equipment surfaces, was obscured by black out material. Each individual color appearance stimulus within a stimulus triad subtended a visual angle of about 13 degrees width and 10 degrees height and the 3 stimuli comprising a color triad subtended approximately 40 degrees of horizontal visual angle. Size of word triad stimuli were also large and were determined by the length of the color labels and the 24 point typeface used.

Color stimulus implementation. The 63 selected OSA color stimuli (21 global and 42 local color appearances) were rendered for computer display. The initial rendering of the 63 appearances was independently carried out by two color-normal observers (the first author and a trained experimenter). Using Adobe Photoshop color mixing software, the color appearance (measured chromaticity) of each color tile was replicated on the screen within rendering tolerance. This was independently achieved by each experimenter through successive adjustment of the color code and measurement of the stimuli using a recently calibrated spectrocolorimeter. This procedure effectively rendered each stimulus to minimize the color difference between the OSA color tile and the rendered monitor version of the stimulus. For each appearance, a second rendering of 63 stimuli was an aggregate of both observers RGB settings for each stimulus. The final acceptance of a rendered stimulus was based on acceptable measurement tolerances for the secondary rendering (including iterative readjustment as needed), the stimulus image files were incorporated into the PsyScope code. While labor intensive, this procedure provided satisfactory renderings of all 63 OSA surface colors as displayable monitor colors. Once implemented as code

the stability of these color renderings were monitored and maintained through regular measurement using the spectroradiometer. Uniform color space measures for the 3 sets of rendered colors are presented in Tables 2, 3, and 4, and depicted in Figures 3, 4, and 5.

Appendix B

Cultural Consensus Theory

A formal description of the consensus model has been provided by Batchelder and Romney (1988, 1989). In addition to evaluating the extent of agreement among participants, the axioms of the formal consensus model provide a theoretical grounding that links the results for a particular random sample to the population from which it was drawn. Consensus analysis is a formal computational model that uses the pattern of responses within a data set to (a) predict the likelihood of correct response for each participant (called the competence rating), (b) provide an estimate of the homogeneity of response among participants (the mean competence), and (c) provide confidence estimates for the correctness of each potential response to a set of questions (the "correct" response is the consensual response). Although this model also makes certain assumptions, it incorporates goodness-of-fit measures that permit an analysis of the extent to which those assumptions have been met. In this research, consensus modeling is used as a heuristic to evaluate the amount of agreement among participants across their responses (as opposed to calculating reliability for separate items). No assumptions are being made about the potential sources of agreement.

The measures used to evaluate results are (1) individual competence scores, (2) mean competence, (3) eigenvalues produced during the minimum residual factor analysis used to estimate the solution to the model's equations, and (4) answer key confidence estimates. Competence scores range from −1.00 to 1.00 and are maximum-likelihood parameter estimates. They are best understood as estimated probabilities of producing the consensual response rather than correlation coefficients. A negative competence score indicates extreme and consistent disagreement with the group across the entire set of responses. Romney et al. (1986), and Weller and Romney (1988) suggest three criteria for judging whether consensus exists in participant responses to questions about a domain: (1) eigenvalues showing a single dominant factor (a ratio greater than 3:1 between the first and second factors), (2) a mean competence greater than .500, and (3) absence or near absence of negative competence scores in the group of participants. Although failure to meet these criteria does not necessarily rule out consensus, it can indicate a poor fit between the data and the model. Establishing these criteria in advance of study and in accordance with other researchers permits a more objective evaluation of whether our results reveal homogeneity of response.

Consensus depends considerably on the shared or dominant response in a group, thus the majority response pattern drives the determination of the consensus answer key. This implies that nonnormative or divergent individuals may not share the consensus solution to the average degree. However, subgroups with coherent alternative knowledge structures may produce their own high consensus measures when partitioned into subgroups and analyzed separately.

Appendix C

Rationale for Photopigment Opsin Genotyping

The rationale for considering that color perception variation arises from different photopigment opsin genotypes is described below. Some discussion of the genetics of color vision is required first.

Research into the molecular genetics of retinal photopigments enables an understanding of photopigment sensitivity as well as the genetic basis for individual differences in color perception (see Nathans, 1999 for a review). Studies show that variation at the level of the photopigment opsin genotype corresponds to response sensitivity shifts in expressed retinal pigments (Asenjo, Rim & Oprian, 1994, Merbs & Nathans, 1992a, 1992b, 1993). The genes for medium-wavelength sensitive (M) and long-wavelength sensitive (L) retinal photopigments are located on the X chromosome, in a head-to-tail array, with the L gene first.

When phenotypically expressed in the human retina, these opsin gene based photopigments define the photosensitive cell classes of M-cones and L-cones. The DNA sequence homology or identity for M and L opsin genes is 98% (Asenjo et al., 1994, Sharpe, et al., 1999). Studies show that photopigment sensitivity to medium (M-cones) or long wave light (L-cones) is determined entirely by substitutions of seven amino acids occurring at codons 116, 180, 230, 233, 277, 285 and 309 of each gene (Asenjo et al., 1994). In particular, individual variability in color perception is associated with genetic variability at one of these critical amino acids, codon 180 in exon 3 of the L- and M-opsin gene. In both genes, the amino acid at codon 180 has been found to be dimorphic (i.e., of two forms). And the frequency of such opsin gene dimorphisms is known to vary across populations of different racial ancestry (Sharpe et al., 1999, p. 30). For example, an estimated 62% of Caucasian males will have the amino acid serine at codon 180 and 38% will have the amino acid alanine at codon 180 in their single X-chromosome inherited L-opsin gene. The average peak sensitivity (λ_{max}) for red light is 557 nm for the 62% segment of males with serine, but in 38% of males with alanine it is 552 nm. Thus in 38% of Caucasian males red-light spectral sensitivity is shifted closer to the λ_{max} for green light, which is 532 nm. This single amino acid substitution in codon 180 gives rise to differences in spectral sensitivity to light and thus, to individual variation in color vision.

Compared to L-opsin genes, substitution in the M-opsin gene – an exchange of serine for alanine at codon 180 – is present in about 9% of Caucasian males, but appears to have a smaller effect on spectral sensitivity to

green light – approximately a 2 *nm* shift in (λ_{max}) peak sensitivity toward the long-wavelength direction (Sharpe et al., 1999). The impact of M-opsin gene substitutions and dimorphisms on spectral sensitivity are the subject of debate, however, due to continuing controversy over the mechanisms for M-opsin gene expression (Winderickx et al., 1992, Nathans, 1999) and recent findings on incomplete X-inactivation effects (Carrel & Willard, 2005), we consider it worthwhile to examine the possible perceptual consequences associated with M-opsin gene dimorphisms, as well as the larger perceptual consequences associated with L-opsin gene substitutions.

Although the expressed retinal phenotype underlying an individual's color sensation is not determined by one's genotype alone (see Yamade et al., 2001), the X-chromosome linked inheritance of M- and L-genes does produce systematic sex differences in retinal phenotypes, and is known to correlate with expressed phenotypic variation. Females, because they have two X chromosomes, have two arrays of M- and L-genes, whereas males, with only one X chromosome, are limited to a single array. As a result, genetic variability in the M- and L-photopigment gene combination is potentially greater for females than for males. As suggested, the actual mechanisms leading to phenotype expression of these genes continues to be studied. However, if a female expressed both of M- and L-opsin dimorphisms, the number of possible M and L genotypes due to codon 180 substitutions would be about double for females compared to that possible for males. For this reason, and the link between substitutions and sensitivity, one might expect greater variability in color perception behavior among females (Mollon, 1992, 1995).

In this study, we address specific hypotheses about the relation between photopigment opsin genotype, perception and naming. To do this we identified within our sample female subjects who were carriers of opsin gene heterozygosities, or multiple forms of genes for L- and M-cone photoreceptors (also called "dimorphisms"). We located these females by advertising for participation by females with male color deficient relatives. The X-chromosome linked inheritance of opsin genes implies an increased probability that females with color deficient male family members possess dimorphisms in their photopigment opsin genotypes (see Sharpe et al., 1999). We refer to females with codon 180 dimorphisms for both M- and L-opsin genes as "potential retinal tetrachromats" due to the fact that they may *express in their phenotype* the usual "wildtype" (or *normal*) forms of photopigments, as well as photopigment variants arising from mutated photopigment genes they possess.

We used modern molecular methods to determine the codon 180 amino acid sequences of each individual's M- and L-photopigment opsin genotype (Wasserman, Szeszel & Jameson, 2001). Briefly, following DNA extraction a

long-range polymerase chain reaction method was used to specify the presence of codon 180 polymorphisms on exon 3 of the red and green genes. The method also makes use of a long-range polymerase chain reaction technique to generate gene-specific PCR products, DNA sequencing to confirm this gene specificity and then PCR and restriction digest to determine M and L codon-180 genotypes. Previous use of such methods isolated genotypes found to be correlated with perceptual behavior (Jameson, Highnote and Wasserman, 2001, Jameson, Bimler & Wasserman, 2005a, 2005b). The method is also used here as a tool to evaluate mechanisms contributing to the variability in color naming and color perception behaviors.

It is important to emphasize that not all females possessing M- and L-opsin gene dimorphisms necessarily express more than three retinal cone classes in their phenotype. Female heterozygotes (possessing dimorphisms both for M and L genes) could be phenotypically color deficient; anomalous trichromat; normal trichromat; or could phenotypically express four, or more controversially five (Nathans, 1999), classes of cones in their retinae. This suggests that even under an assumption of color vision neural trivariance, color perception behaviors might be more variable, or differently distributed, for a group of females possessing both M- and L-opsin gene dimorphisms (5-gene genotypes), or L-opsin dimorphisms (4-gene genotypes), compared to a group of females with 3-gene genotypes. Finally, although the frequency of female heterozygote genotypes has been estimated near 56% in some populations, the actual frequency of female expressors in such a population is yet unknown, and it may be that phenotype expression is much less frequent. Additional molecular genetics research should resolve this issue in the near future.

DNA from 47 subjects (30 females and 17 males) was analyzed from the total group (excluding those who did not give a sample for DNA analysis.) Among these, 4 males were also behaviorally identified as Dichromats and are not included in the group classifications of Table C. Table C classifies the remaining genotyped individuals into 6 different genotypes (numbered (1)-(6) in the right columns of Table C), which are loosely grouped into 4 subgroup partitions differing by potential phenotype and predicted perceptual variation (groups numbered 1 to 4 in the left column of Table C).

The 6 genotypes shown differ with respect to their genetic potential to express different M- and L-cone classes as shown in columns 6 and 7 of Table C. Group 1 includes of subjects with only the genetic wildtype comprising the group of potential "normal" trichromats. The wildtype involves serine at position 180 of the L-cone photopigments (column 6) and alanine at position 180 of the M-cone photopigments (column 7). Group 2 consist of any 3-gene genotype including potential expressors of wildtype and anomalous trichromacy, the latter

Table C
Subjects grouped by Photopigment Opsin Genotype Analyses

Group #	Partitions	N	Female	Male	L-180 genotype	M-180 genotype
1	WildType	8	5	3	(1) Ser	Ala
2	Any 3-gene Genotype	16	8	8	Ser	Ala
					(2) Ser	Ser
					(3) Ala	Ser
					(4) Ala	Ala
3	Two forms of L- & M-cone genes	15	12	3	(5) Ser & Ala	Ser & Ala
4	Two forms of L-cone genes only	7	6	1	(6) Ser & Ala	Ala

Note. Column 6 and 7 give abbreviations for amino-acids identified at codon-180 of the L- and M-opsin genes. 'Ser' denotes serine and 'Ala' denotes alanine amino acids.

possessing a genetic wildtype for one cone class and a mutated form of the other gene, combined with the usual S-cone class. Group 3 of Table C has the greatest genetic potential with two different L-cone class genes (denoted by 'Ser & Ala' amino-acids in column 6) and two different M-cone class genes ('Ser & Ala' in column 7), and is the only genotype group in which all individuals possess genes for five different retinal cone classes (two forms of retinal L-cone, two forms of M-cone, and one S-cone type) – although the expression of more than four retinal cone classes is debated (Nathans, 1999). Group 4 includes individuals with the genetic potential to express two different L-cone classes and one "normal" M-cone class (and the usual S-cone class). We consider both Group 3 and to Group 4 as a potential retinal tetrachromat subgroups with individuals expressing two forms of L-cones – expression of two forms of L-cones is not controversial. However, comparisons between Group 3 and Group 4 is used here to evaluate whether or not behavioral differences are found only for Group 3's double-dimorphism subjects (heterozygous for M- and L-cone opsin genes), or whether Group 4's L-opsin dimorphism females show similar results to Group 3. This comparison aims to clarify whether results are attributed primarily to possession of L-cone dimorphisms, rather than strictly double dimorphism

(see Wasserman, Szeszel, and Jameson, 2001 for a discussion of genotypes included in Groups 3 and 4).

Hypotheses on Table C subgroups

Appendix D presents hypotheses about the behaviors of the subgroups described here. Although some research suggests potential retinal tetrachromats and trichromats differ with respect to some color perception and categorization measures (Nagy et al., 1981, Jordan & Mollon 1993, Jameson, Highnote & Wasserman 1998, 2001, Bonnardel et al., 2002, Jakab & Wenzel 2004, Jameson, Bimler & Wasserman 2005a, 2005b), these differences continue to be debated and it is yet unclear whether and to what extent potential tetrachromats differ in their perceptual representations of color compared to trichromats, and even less known about possible color naming differences between these groups. The present comparisons between Table C groups aims to further clarify how potential retinal tetrachromats differ from trichromats with respect to their perceptual representations. Considering these defined groups, intra-group consensus measures can be used to evaluate shared knowledge consensus among groups defined strictly by genotype, and with those defined in Analysis 1 (i.e., Dichromats). Matching measures can be used to evaluate the consistency between individual's perceptual and linguistic representations, again comparing Table C groups with other naturally occurring groups defined by demographic and color perception assessment information. In particular, trichromats, dichromats, and one genetically identified subgroup of the potential tetrachromats and are analyzed with regard to their intra-group color naming consistency and consensus. Specific predictions are discussed further in Appendix D.

Appendix D

Prediction Matrix Rationale

The prediction matrix presented in Table 11 represents our initial impressions about how the color naming function may vary based on: (1) individual perceptual variation, and (2) color stimulus set variation. The purpose of these predictions is to model how individual color naming relates to the shared cultural knowledge about naming among members belonging to a common linguistic society. If relations among observer groups resemble those in Table 11, we gain further insight into individual color naming functions. Failures in Table 11 predictions may help us refine our color naming modeling and create better predictions in the future. The three principles used in constructing the predictions are stated below.

Consensus Principle 1. Good/High consensus for word triads indicates a strongly shared lexical representation. Low/Poor consensus for words indicates less agreement about lexical representation.

Consensus Principle 2. Good/High consensus for color triads is consistent with uniformity in perceptual representation. Low/Poor consensus for color triads suggests individual variation in the perceptual representation of color.

Matching Principle. High matching measures are consistent with a consistent mapping and thus strong correspondence between color and word representations, and Low/Poor matching is consistent with inconsistent of vague mapping between color and word representations.

The following specific predictions follow from these principles.

Normal Observers

Color vision normal observer group predictions are given in row 1 of Table 11, and separately for females and males in rows 2 and 3. These predictions apply to subjects with an underlying trichromat genotype and a 'normal' trichromat phenotype as shown through color vision assessment using FM 100-hue and Ishihara tests.

Color triad consensus for normals is predicted to be 'Good' (not 'High') due to slight individual differences in representation following perceptual variation among normal subjects. Color triad consensus is predicted as 'Medium' in other cases because of normal individual perceptual variation combined with the increased difficulty of the local task.

Word triad consensus for normals is predicted to be 'High' due to a robustly shared cultural model. Cases where word triad consensus is predicted as 'Medium' is due to good access to a shared cultural model despite the increased difficulty of the local task. Cases where word triad consensus is predicted to be 'Low' follows from a poorly shared cultural model for local stimuli combined with local task difficulty.

Normal females (row 2) are predicted to differ from males (row 3) in local triad task consensus with females showing a slight advantage compared to males ('Medium' for females, 'Low' for males). This prediction follows from evidence that women are more likely to be familiar with the naming of subtle color distinctions.

For 'normal' color ability subjects (females and males aggregated, and separate) mean matching measures between color triads and word triads is expected to be 'Good' for global stimuli, and 'Medium' for local stimuli because we expect that individual color naming functions will closely map global colors with global words, but map local colors with local words less consistently. Again, the assumption is that the cultural model for global stimuli is more widely shared and thus more robustly reinforced in individual color and lexical representations. Matching is a measure of individual subjects' internal consistency on triad choices across the two domains of words and colors, thus there are no predicted differences in female and male performance because equal internal judgment consistency is expected. In general, beyond judgment consistency, greater average matching measures imply tighter and more specific mapping between an individual's color and lexical representations. Lower matching measures imply less consistency and perhaps a diminished similarity between color and lexical representations. Thus, cases of 'Low' matching suggest but don't necessitate distinct perceptual and lexical representations.

Dichromat Observers

Predictions for dichromat observers are given in row 4 of Table 11. These subjects express a color deficient 'Daltonism' phenotype identified during color vision assessment using FM 100-hue and Ishihara tests, irrespective of genotype.

In Dichromat subjects color triad consensus is predicted to be 'Poor' due to individual perceptual deficiency generally producing decreased agreement on triad choices. Local red color triads are predicted at 'Medium' consensus because the red stimuli minimize stimulus contrast on the red-green axis, the source of dichromat color confusions. In comparison, local blue stimuli involve

greater potential difficulty for the dichromat due to chromatic variation in both the red and green color directions.

Dichromat global word triad consensus is predicted to be 'Good' due to access to a robustly shared cultural model despite perceptual deficiencies. Local word triad consensus is predicted to be 'Low' due to a poorly shared cultural model for local stimuli and local task increased difficulty.

Dichromat mean matching measures between color triads and word triads were expected to be 'Poor' for global stimuli, and either 'Low' or 'Poor' for local stimuli. We expected dichromat individual color representations to differ substantially from their lexical representations, so that their color naming functions would not robustly map global color and global word similarity. Also, we predicted that local red color choices – involving stimuli with comparatively less variation along the axis of deficiency – would be matched to local red word choices better than the other stimulus conditions tested.

Diverse Photopigment Opsin Genotype Observers

As described in Appendix C, we expect to see variation in the behavior of observers with diverse photopigment opsin genotypes. Noting, however, that there are no empirical precedents suggesting that triad task similarity is capable of tracking the potential impact of retinal tetrachromacy on cognitive behaviors. Predictions for diverse observers are given in row 5 of Table 11.

For observers with diverse photopigment opsin genotypes color triad consensus is predicted to be 'High' due to a predicted richness in color experience (Jameson et al. 2001) and a related robustness of representation despite perceptual variation among subjects. Cases where color triad consensus are predicted as 'Medium' are due to increased difficulty of the local blue task. Cases where color triad consensus is predicted as 'Good' are due to heightened red category expertise arising from possible expression of two retinal L-cone types. We expect this despite the increased difficulty of the local red task.

For diverse genotypes word triad consensus is predicted to be 'High' due to a robustly shared cultural model. Cases where word triad consensus are predicted as 'Medium' is due to good access to a shared cultural model combined with the increased difficulty of the local task. Cases where word triad consensus is predicted to be 'Good' arise from good access to a shared cultural model, heightened expertise for red category stimuli, combined with local task difficulty.

For observers with diverse genotypes mean matching measures between color triads and word triads are expected to be 'Good' for global stimuli, 'Good' for local blue stimuli, and 'High' for local red stimuli. Again we expect that indi-

viduals with the genetic potential to express greater numbers of retinal cone classes will reflect some form of color expertise. We predict that their color naming functions will closely map global colors with global words, but that they will have greater difficulty mapping local colors with local words less. The red category is an exception. Here their expertise should be optimized because of heightened semantic awareness (partly socialized and partly perceptual based) and a concomitant perceptual expertise in color.

Assigning values to nominal codes

Predictions in Table 11 are expressed as a 5-level nominal code that qualitatively predicts the relationships among conditions (Table D, column 1). With further refinement of the code we can specify the predicted *quantitative ranges* of observed average consensus (Table D, column 2) and average proportion-matches (Table D, column 3) associated with Table 11's nominal code assignments. Table D provides these as μ(consensus) and μ(proportion-matches) denoting the ranges of average measured-consensus and average proportion-matches, respectively.

The numerical ranges above are based on empirical research showing that average consensus less than .5 is a sensible cut off for a tested domain to qualify as a shared knowledge structure. Thus, .5 is adopted as the low value of the scale, and maximum attainable average consensus of 1.0 (although rare in practice) is set as the upper limit of the scale. The range for average proportion-matches is also based on existing results. The low-value cut-off is set near chance performance (p = .333), and the upper limit concurs with match levels observed in previous empirical studies in which the correspondence between tested representations was robust (Alvarado, 1996).

Table D
Quantitative ranges suggested for the nominal code used in the Table 11 prediction matrix

Nominal Code	μ (consensus)	μ (proportion matches)
'High'	.8 – 1.0	.7 – .8
'Good'	.7 – .799	.6 – .699
'Medium'	.6 – .699	.5 – .599
'Low'	.5 – .599	.4 – .499
'Poor'	< .5	.3 – .399

D'Andrade's (2003) English color naming model

Note that all Table 11 predictions assume that subjects share the specific normative English color naming model empirically determined by D'Andrade (2003). Because the model is based on empirical data, and the naming task used did not artificially constrain subject's responses, we believe that the modal names associated with particular samples are correct designations. However, D'Andrade's (2003) sample model may not be representative of the population model. The measures in the Results section can be used to evaluate the degree to which D'Andrade's naming model deviate from the naming model of our subjects. It is possible that the lower consensus observed for word tasks compared to color tasks may be due to discrepancies between D'Andrade's naming model and that of our subjects. The observance of consensus in most tasks encourages us to believe that D'Andrade's model is sufficient to test the observer group differences evaluated in this study. Lower proportions of matches would be the strongest indication that D'Andrade's model is not representative of the model shared by the present sample of subjects. However, again, matches were sufficiently high for us to believe that the model was adequate for our purposes.

Author Notes

Support for this research was provided by a Fulbright Fellowship (Sayim). A preliminary report of portions of this work was presented at the annual meeting of the Association for Research in Vision and Ophthalmology (ARVO), May, 2003, Ft. Lauderdale, Florida (Sayim, Jameson & Alvarado, 2003). Results from a task included in this study, but which are not reported in present article appeared, in *Diplomarbeit* of B. Sayim (Kiel Universität, 2004). The authors acknowledge the support of Linda Wasserman, M.D., PhD. in the supervision of the DNA analyses presented.