Evolutionary Models of Color Categorization: Investigations based on Realistic Population Heterogeneity

Kimberly A. Jameson,[1,*] and Natalia L. Komarova[2]

[1] *Institute for Mathematical Behavioral Sciences, University of California, Irvine Social Science Plaza, Irvine, CA 92697-5100, USA*

[2] *Department of Mathematics, University of California, Irvine Rowland Hall, Irvine, CA 92697-3975, USA*

[*] *Corresponding author: kjameson@uci.edu*

Compiled August 27, 2008

The evolution of population color categorization systems formed on the Farnsworth-Munsell 100 Hue Test continuum is investigated in populations of artificial agents using realistic constraints from human populations: namely, (*i*) varying amounts of normal observer heterogeneity, and (*ii*) varying degrees and forms of observer color deficiency. These constraints are made operational in agent categorization and communication games. They produce a number of interesting consequences for stable, shared categorization solutions that are evolved in agent populations. It is found that the confusion patterns associated with a small fraction of color deficient agents break symmetries in population categorization solutions, and confine the boundaries of color categories to a small subset of available locations. Further, confusion pattern variations across different types of deficient agents lead to changes in category size and number that depend on the form of deficiency represented. In particular, stimulus pairs forming global confusion axes for dichromats tend to attract color boundaries, and local confusion regions characteristic of both dichromats and anomalous trichromats tend to repel color boundaries. Furthermore, the concurrent presence of normal agents and several different kinds of deficient agents produces novel constrained solutions that optimize successful categorization performance in population communication games involving color.

OCIS codes: 330.1690, 330.5020, 330.4060.

KEYWORDS: color categories, color deficiency, Farnsworth–Munsell 100-Hue Test, evolutionary game theory.

1. INTRODUCTION

The empirical literature on human color perception and categorization suggests that there is a good deal of universality in color categorization across different groups of people, although a considerable amount of variation is observed across individual categorizers and ethnolinguistic categorization systems [1]. A long-standing debate in the field is whether specific universal tendencies exist in the ways different human linguistic societies categorize and name perceptual color experiences, and if so, to what factors (e.g., physical environment, human biology, perception, social features) cause them. A major challenge for the area continues to be developing a theory that accounts for universal patterns seen in color categorization data while simultaneously explaining observed differences. This color vision and perceptual processing problem has recently attracted the interest of computational scientists [2 − 7].

This article explores the impact of visual processing constraints on the evolution of shared color category systems. One goal is to systematically examine the ways color category solutions change as a function of simple perceptual processing variations in observer populations, including various types of color discrimination deficiencies found in real observers. Another goal is to isolate the mechanisms that play an instrumental role in simulated category solution variation.

The present modeling framework uses agent based, evolutionary game theory methods previously employed to model human color categorization systems [5, 7].

This earlier research considered only idealized populations of agents with different color perception abilities. By comparison, the present research investigates the impact of more realistic observer variation and population heterogeneity on the formation of stable population category structures. The present article investigates the role of different types of color perception/processing in the formation of shared population color categories. In particular, we examine the following types of observer variation: (1) normal color perception variation, (2) the role of anomalous trichromats, (3) the role of dichromats (both protanopes and deuteranopes), and (4) the role of the color categorization "experts".

In the sections below we first describe the stimulus continuum on which all our agent models are based, then detail the agent models used, then report several investigations that use increasingly realistic population models, and report their results. The last sections of the article discuss the findings of the research.

2. MODELING ARTIFICIAL OBSERVER POPULATIONS

Below we describe an approach used for idealizing some realistic constraints on color categorization and interindividual communication. This approach uses a modeling framework in which individual agents learn to categorize simulated colors through reinforcement learning by playing "discrimination-similarity games," and learn to communicate the meaning of categories to each other [5, 7]. The three main components of this article's evo-

lutionary approach are (*a*) the stimulus space, (*b*) the observer model, and (*c*) the evolutionary game. Component (*c*) is adopted from earlier research [5, 7], whereas components (*a*) and (*b*) are novel and require a detailed description, which is provided below.

The Farnsworth-Munsell 100 Hue Test as a Stimulus Space

Similar to earlier research [5, 7] we use a hue circle as a natural subspace of the color appearance solid. For the task of introducing realistic population heterogeneity into simulated color categorization investigations, the hue circle provides a very useful simplification for several reasons: (*i*) It is a justified subspace according to models of human color perception [8, p. 184]. (*ii*) It is has value as a structure for distinguishing normal observers from color deficient observers based on hue circle similarity relations (e.g., [9]) (*iii*) It is a natural choice as a color subspace, because it preserves similarity relations among hue categories regardless of variation in salient hue points across individuals. That is, the hue circle captures, for most human observers, the appropriate ordinal relations among hues, as well as color-opponent hue structure, even though as a color stimulus it lacks saturation and lightness dimensions generally considered important to individual perceptual color space. And (*iv*) as described in detail below, use of the hue circle capitalizes on existing human color perception data from a widely researched stimulus, for the purpose of modeling artificial agent observer groups.

The hue circle stimulus model employed in the present investigations is the Farnsworth-Munsell 100 hue test, or *FM100* [10]. The FM100 stimulus set forms hue circle that is well understood colorimetrically and is in common use as a diagnostic technique for human color perception deficiencies. This latter feature permits the use of the FM100 in modeling varying discrimination capabilities in our artificial observers.

The FM100 stimulus is a continuous hue circle series, discretized into 85 color "caps" that form a smooth gradient of hue, ostensibly at a fixed level of brightness and a fixed level of saturation. (See Figure 1 for an approximation of the 85 cap hue gradient). The present use of the FM100's 85 stimuli as an approximate hue circle continuum naturally extends beyond the hue continua investigated in our earlier color category studies, and permits more direct generalizations and comparisons with available human color perception data. Also, because the FM100's 85 hues form an approximate continuum of uniform Munsell *Chroma* and comparatively uniform Munsell *Value* levels [11, p. 2239, Figure 1], this provides for the opportunity to relate findings from our simulated population category solutions to existing human color category partitioning results (at a specific Munsell Value) on a widely used stimulus set (i.e., the *World Color Survey* stimulus grid [12, 13]). In the investigations described in this article, populations of simulated agent *observers* engage in communication games (as described in [5, 7] with the stimuli of the FM100 as the domain of color space sampled, instead of a theoretical sample.

Realistic Models of Population Heterogeneity

A major goal of this article is to understand color categorization variation that may be seen under population homogeneity (i.e., populations of artificial agents in which all individual agents possess the same discrimination model) compared with that seen under realistic population heterogeneity (i.e., populations in which individual agent models approximate, to varying degrees, actual dichromat, anomalous trichromat or normal trichromat hue discrimination data). While throughout the present investigations we refer to *normal*, *anomalous trichromat*, and *dichromat* agents, as mentioned above, we consider a subspace of perceptual color space, and therefore the observer models presented reflect only a portion of human color discrimination behaviors found in realistic circumstances.

We now describe the approximations used to define our models for normal, anomalous trichromat and dichromat color observer models. In the present populations, we model only color perception deficiencies linked to human X-chromosome recessive inheritance, or forms of protan and deutan variation. The biological bases and perceptual consequences of these human forms of deficiency are well understood [14]. In protan defects the long-wavelength sensitive photopigment is missing or abnormal. In deutan defects the middle-wavelength sensitive photopigment is missing or abnormal. Under realistic viewing circumstances both protan and deutan defects show varying degrees of color confusions, which can be summarized as confusions among colors within the sets of (1) green-yellows, yellows, and yellow-reds; (2) browns and greens; (3) blue-greens and violets; and (4) purples, grays, and greens. In addition to these four confusion sets, protan deficient individuals may additionally confuse reds with blacks; whereas deutan deficient individuals may confuse greens with blacks [14]. These forms of deficiency are found, for example, in some populations of European descent as shown in Table 1.[1] Note that Table 1 distinguishes between deficient forms that involve a shift of photopigment sensitivity compared with forms caused by a lack of a photopigment class. Table 1's observed human frequencies are ranges, in part, because of measurement and sampling variation across different empirical studies. These data are used primarily as a guide for relative frequencies in our artificial agent populations. They serve the aim of optimizing the influence of population heterogeneity on category solutions (as opposed to aiming to replicate specific human observer population frequencies).

Although the present investigations use human empirical data as the basis for our simulated observers, there are important differences between the color categorization scenarios seen in real human populations and those

[1] Table 1 combines data from several sources [14 – 18]

simulated here. First, we do not model any learned cognitive factors that contribute to color categorization – e.g., factors that include individual color memory, social utility that may be attached to specific colors, or personal color salience. One reason for not including such complex cognitive factors in the present investigations is so that we can begin the inquiry from a simple starting point; another is that we want to clarify the influences of color perception and discrimination differences on categorization behaviors. This initial cognitive minimalist approach permits the gradual increase of cognitive complexity for evaluation – an aim of on-going work [19] – and allows us to systematically study the effects of different factors on color categorization.

Second, we make few assumptions about the salience of color space from the point of view of both (a) variation in environmental color distribution, and (b) variation in the putative importance, or utility, of some subspace of environmental color compared to any other subspace. (See [20] for a discussion.) Thus, in contrast to our earlier investigations [5, 7], this article does not investigate *regions of increased salience* or environmental color *hot spots*, nor are there any investigations at the individual observer level – e.g., specific color appearances that are universally salient, such as Hering's *Urfarben* Unique Hue color appearances – although these are among the aims of our on going studies.

Modeling Different types of Observer Variation within a Population

Both normal and deficient observer group models, based on human data, are used in these simulations. "Normal" agents include those based on normal trichromat performances on the FM100 hue test, with trichromacy confirmed by pseudoisochromatic plate assessment [21]. "Anomalous Trichromat" agents are modeled using individual anomalous trichromat performances on the FM100 as reported by Bimler and colleagues [22, Table 1, p. 166]. And "Deficient" agents are modeled using diagnosed protanope and deuteranope performances found in the FM100 diagnostic database [23]. Qualitatively compared to normal trichromats, protanopes and protanomalous trichromats have reduced sensitivity to long wavelengths, whereas deuteranopes and deuteranomalous trichromats have reduced sensitivity to middle wavelengths [24].

In order to employ data from actual normal and anomalous human individuals, we introduce a probabilistic observer model. The color stimulus (a hue circle) to be categorized by all agents is assumed to be identical (i.e., FM100 stimulus). Colors are chosen randomly from the circle and presented to the observer. We next apply the appropriate color confusion transformation which gives probabilities for the observer to perceive the two chosen color stimuli as the same or different in color. In general the transformation is described as a stochastic $n \times n$ matrix (a matrix whose rows sum up to one), $\{C_{ij}\}$, whose entries C_{ij} give the probability that stimulus i is perceived as stimulus j. The confusion matrix is different for different individuals.

In modeling our agents we distinguish between several types of discrimination based confusions among FM100 stimuli. Here we first describe a form of unsystematic noise in individual agents' categorization performance that resembles the confusion errors often seen when human observers sort the hue continuum of the FM100 stimulus – referred to below as *sorting confusions*. Then we follow with a description of two additional kinds of confusions – referred to below as *local* and *global confusions* – that are modeled in our agents.

Sorting Confusions. Although an individual may have perfectly normal color perception and be expected to show error free performance on FM100 sorting, typically sorting performance shows some random transposition errors, or confusions, that occur between adjacent stimuli. Such confusions tend to be 2-cap exchange errors that are random, seldom reoccur on retest, and are not bunched in any one region of the hue circle [25].

We factor this kind of normal (i.e., unbiased) sorting confusions into our model in the following way. We assume that with probability p, each color cap can be confused with either the cap just to the left or the cap just to the right from it on the FM100 stimulus. (Below we sometimes refer to the p as the "noise" parameter). This leads to the following matrix:

$$C_{ij}^{sorting} = \begin{cases} 1-p, & i=j, \\ p/2, & j=i\pm 1, \\ 0, & \text{otherwise.} \end{cases}$$

Local confusions. Local confusions are confusions among neighboring FM100 stimuli (mostly within, but also spanning, FM100 trays), which go beyond sorting confusions described above. Local confusions for deficient observers are concentrated inside local confusion regions, which are continuous portions of the FM100 continuum where an observer might confuse several adjacent stimuli with one another, perhaps transposing the stimulus order by as much as 10 steps or more in the circle. Segments at which there exists a tendency to transpose the sequential stimulus order can appear at any location in an FM100 tray, including segments of mis-sorting at the end of a given FM100 tray which may be seen to continue when the observer sorts stimuli in the tray most adjacent to the mis-sorted segment. Thus, local confusion regions can span two trays of the FM100, and may be found anywhere along the 85 stimulus continuum dependent on the severity and type of deficient observer tested.

To model local confusions, we assume that there are regions in the color space where the probability of confusion is highly elevated compared to the incidence of sorting confusions, see e.g. Figure 1 for deuteranopes and protanopes. Let us denote the confusion regions as I_1 and

Table 1. Observed Frequency Ranges of Human Color Deficiency used to Model Heterogeneous Agent Populations

Type of color deficiency	Number of photopigments	Percent Frequency Ranges in Populations of European Descent
Protanopia	2: No long-wavelength sensitive photopigment	1.01 - 1.60
Protanomalous Trichromacy	3: Abnormal long-wavelength sensitive photopigment	1.03 - 1.27
Deuteranopia	2: No middle-wavelength sensitive photopigment	1.01 - 1.97
Deuteranomalous Trichromacy	3: Abnormal middle-wavelength sensitive photopigment	4.01 - 5.35

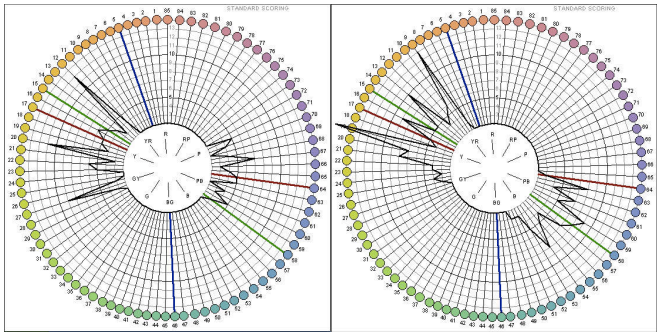

Fig. 1. Hue gradient of the FM100 stimulus model (circumference) shown with confusion data used to model protanopes (left) and deuteranope (right) agents used in simulations. Colored lines show a main axes of confusion for three forms of deficiencies: red=protan, green=deutan, blue=tritan (tritan deficiencies are not investigated here). Deviations from the inner-most circular contour illustrate stimulus sorting errors. An *ideal* normal response, with no confusions among the 85 stimuli, would plot as no deviations from the inner-most circular contour of the figure.

I_2. We have:

$$C_{ij}^{loc} = \begin{cases} exp\{-\frac{(j-i)^2}{w}\}W_i^{-1}, & i,j \in I_1 \text{ or } i,j \in I_2, \\ 1, & i = j \notin I_{1,2}, \\ 0, & \text{otherwise}, \end{cases}$$

where w is related to the characteristic width of the confusion range, and W is the normalization factor for the given confusion region: $W_i = \sum_{j \in I_{1,2}} exp\{-\frac{(j-i)^2}{w}\}$, where the summation is over all the caps inside the confusion region of cap i.

Global confusions. Global confusions are confusions across the hue circle (across FM100 trays). They consist of pairs of FM100 stimuli that form a confusion line pair (see Figure 2) for which color stimuli across the hue circle are linked by a *line of confusion* (similar to an *"axis of confusion"* [7]). Based upon the data shown in Figure 2, global confusions are given by three pairs of caps from opposing sides of the color circle which are commonly confused (or even indistinguishable) by dichromats. Let us denote them x_k and $\bar{x}_k$ with $k \in \{1, 2, 3\}$.

It is convenient to introduce a function g defined on the set $X_{gl} \equiv \{x_1, \bar{x}_1, \ldots, \bar{x}_3\}$ such that $g(x_k) = \bar{x}_k$ and $g(\bar{x}_k) = x_k$. The global confusion matrix is given by the following formula:

$$C_{ij}^{glob} = \begin{cases} 0.5, & i \in X_{gl} \text{ and } j = i, \\ 0.5, & i \in X_{gl} \text{ and } j = g(i), \\ 1, & i = j \notin X_{gl}, \\ 0, & \text{otherwise}. \end{cases}$$

Parameter estimation based on data analysis

We now use individual observer sorting data of the FM100 stimuli to determine the simulation parameter settings for modeling normal, dichromat and anomalous agents.

Local confusion regions. For protan and deutan models, local confusion regions were determined based on the individual results seen in Figure 1. The protanopes have two confusion region ranges: $I_1 = [12, 27]$ and $I_2 = [59, 70]$. For deuteranopes, there are also two regions: caps $I_1 = [9, 25]$ and caps $I_2 = [51, 63]$.

To model the anomalous trichromat local confusion regions, we used data from 10 protanomalous and 17 deuteranomalous individuals diagnosed as severe or extreme anomalous (with diagnoses confirmed by anomalouscope matching and pseudoisochromatic plate assessment) by Bimler, Kirkland & Jacobs ([22], Table 1, p. 166). We computed two average models that combined the FM100 performance vectors across the separate groups of protanomalous and deuteranomalous individuals. To do this we considered individual Farnsworth-Munsell performance as a vector of 85 numbers corresponding to an individual's ordering of the test's 85 color caps, $\{x_1^{(I)}, \ldots, x_{85}^{(I)}\}$, where the superscript (I) is used to distinguish different individuals. From these, we calculated the *averaged* count of Farnsworth-Munsell transpositions in ordering, in the following way. Denote by S_{PN} (S_{DN}) the set of all protanomalous (deuteranomalous) individuals, and by N_{PN} (N_{DN}) the respective numbers of anomalous individuals. Then the average FM100 per-

formance vectors are obtained as follows:

$$\bar{x}_i^{PN} = N_{PN}^{-1} \sum_{I \in S_{PN}} x_i^{(I)}, \quad \bar{x}_i^{DN} = N_{DN}^{-1} \sum_{I \in S_{DN}} x_i^{(I)}.$$

Average FM100 performance vectors for protanomalous and deuteranomalous individuals are shown in Figure 3. Note that 2 has been subtracted from the y-axis values of Figure 3 because usually a FM100 score of 2 is the lowest possible score that any cap can obtain, and represents perfect serial order [25, p. 570]; whereas, for simplicity, here we represent "no error" by the value zero. Two "bumps" in the averaged plots are easily detected; they correspond to the regions of confusion. The boundaries of these local confusion regions are determined as demonstrated in Figure 3: the continuous regions of caps where average cap error scores exceeding 1.5 give the positions of the local confusion regions. For protanomalous individuals, the confusion regions obtained are ranges 14-21 and 57-71; for deuteranomalous individuals they are 10-22 and 51-67.

Sorting errors. The probability of sorting errors, p, is calculated as follows. For normal individuals we used the FM100 data of 10 normal trichromat individuals, and we calculated each individual's total number of cap inversions.[2] The average number of observed inversions across the 10 normal individuals equaled 6.8 inversions. Thus, based on these data, on average a normal individual makes 6.8 randomly distributed sorting errors per 85 caps (1.7 inversions per tray). The probability of inversion *per cap* is given by $6.8/85 = 0.08$, that is, 8%. This value was used to model the unbiased sorting error for normal individual performance across the 85 stimulus set. Consistent with dichromat characteristic patterns of performance, 8% sorting error was also used to model dichromat sorting confusion rates outside local confusion regions and independent of global confusion pairs.[3]

Modeling anomalous trichromats (Figure 3) we noted that this class of observer exhibits patterns of diffuse color confusion, even outside the local confusion regions, where they show considerable sorting difficulties compared to the normal observers. Thus, to calculate the probability of sorting errors for our anomalous agents, we performed an average transposition calculation similar to that for the normals, but we restricted the computation to sequence inversions outside the anomalous local confusion regions.[4] We obtained the following estimates: for protanomalous individuals, the probability of inversion was 26%, and for deuteranomalous individuals it was 24% per cap.

The confusion range. We also defined a width parameter, w, for cap sequence inversions that occur inside the confusion regions. To do this, we find the quantity $q \equiv (1/N) \sum_I (1/n) \sum_i |x_i^I - i|$, where the inner summation is over all the caps inside the confusion regions, n is the number of caps inside the confusion region, and N is the total number of protanomalous or deuteranomalous individuals. Then the value of w is evaluated[5] as $w \approx \pi q^2$. We find that $q = 1.2$ for protanomalous and $q = 1.4$ for deuteranomalous, which roughly gives the value $w = 5$.

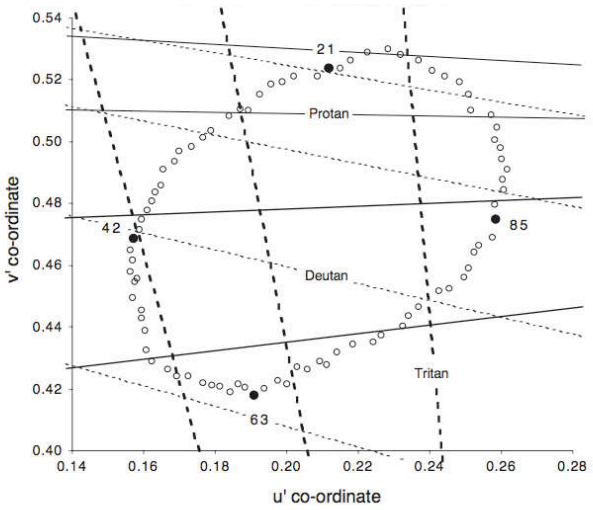

Fig. 2. Dain's depiction of 85 FM100 stimuli displayed in CIE 1976 u', v' coordinate space [26] (reproduced with permission). Inset lines show three global confusion pairs each for protan (solid lines) and deutan (dotted lines) deficient agents as modeled in the present investigations.

Global confusion pairs. Figure 2 data is used to derive values for dichromat global confusion pairs. Protanope global confusion pairs are caps: $x_1 = 9$ & $\bar{x}_1 = 29$, $x_2 = 1$ & $\bar{x}_2 = 40$, and $x_3 = 74$ & $\bar{x}_3 = 53$. Deuteranopes also have three global confusion pairs: $x_1 = 11$ & $\bar{x}_1 = 21$, $x_2 = 2$ & $\bar{x}_2 = 31$, and $x_3 = 79$ & $\bar{x}_3 = 42$.

Summary of the model. Using the considerations described above we incorporated the following agent confusion definitions in our model:

[2]Same as the total number of sequence inversions (inversions of two adjacent caps) needed to recreate the perfect cap ordering of 1 to 85 out of the individual's sorting data.

[3]Making the assumption that dichromats tend to perform on a par with normals away from their axes of confusion, despite Figure 1's suggestion that dichromats can exhibit error-free performance away from their local confusion regions.

[4]Under the observation that protanomalous and deuteranomalous individuals experience two types of confusion errors. One type arising from a systematic bias in color space (consistent with either a shift of the peak sensitivity of the L-photopigment curve towards the M-photopigment (protanomalous), or a shift of the peak sensitivity of the M-photopigment curve towards the L-photopigment (deuteranomalous)) which produces higher confusion in local confusion regions similar to those found for dichromats; and a second, general tendency for diffuse color confusions outside these local confusion regions.

[5]by using $q \approx \frac{\int e^{-x^2/w} |x| \, dx}{\int e^{-x^2/w} \, dx} = \sqrt{w/\pi}$.

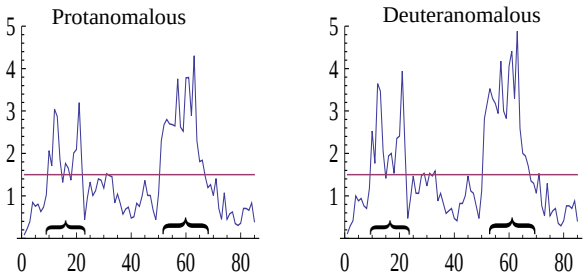

Fig. 3. Average FM100 stimuli sorting score for protanomalous ($n = 10$) and deuteranomalous ($n = 17$) individuals. In both plots, the horizontal axis is the 85 caps of the FM100, and the vertical axis is the observed average error at each cap for the two samples considered. Bracketed regions illustrate continuous regions of confusion, arising from confusion patterns in the data, that obtain the value 1.5 for each anomalous type.

- Normals ($C = C^{sorting}$):
  - Probability of confusion with a neighboring cap is $p = 8\%$.
- Anomalous trichromats ($C = C^{sorting}C^{loc}$):
  - Local confusion regions based on observer data, and supported by the major axes for anomalous subjects reported by Farnsworth [25], are FM100 cap segments 14-22 and 57-71 for protanomalous, 9-22 and 51-67 for deuteranomalous; inside the confusion regions $w = 5$.
  - Probability of confusion with a neighboring cap, both inside and outside the confusion region is $p = 25\%$.
- Dichromats ($C = C^{sorting}C^{glob}C^{loc}$):
  - Local confusion regions based on Figure 1's data range from 12-27 and from 59-70 for protanope agents; and 9-25 and 51-63 for deuteranope agents. Inside the confusion regions $w = 10$.[6]
  - Global confusion pairs based on Figure 2 data: (9,29), (1,40), (74,53) for protanopes, (11,21), (2,31), (79,42) for deuteranopes.
  - Probability of confusion with a neighboring cap, both inside and outside the confusion region is $p = 8\%$.

Other simulation parameters used here are defined in Table 2.

Note that our modeling of global confusion pairs, local confusion regions, sorting confusion rates (p), and the range of confusion within a local confusion region (w) are based on data observed from sampled representative subjects from each observer group modeled. To the degree that the data used to create these models differ from other observer data found in the literature, the models would also similarly vary. In developing these models we tested numerous variants to determine how changes in the underlying data would impact the models, and how, in turn, the categorization solutions would be affected. We found that altering model features like the positions of local confusion ranges, and global confusion pairs, or varying the values of sorting confusion rates, did not fundamentally alter any of the findings presented here, or render less-robust any of the results for symmetry breaking or any of the other category solution effects that we report below.

[6]These dichromat confusion regions closely resemble Mantere and colleagues' [11] simulated dichromat FM100 performance based on the standard Farnsworth method of scoring [10].

3. POPULATION HUE CATEGORIZATION INVESTIGATIONS

The investigations are organized as follows: First we investigate homogeneous populations (Study 1) to compare with results from heterogeneous population investigations (Studies 2, 3, & 4). Study 1 has two emphases: Emphasis 1 examines solution variations obtained from several homogeneous populations. Some of these homogeneous populations are comprised of agents modeled as ideal observers, and other populations consist of ideal observer models made more realistic by allowing them to discriminate the stimuli to be categorized with various degrees of probabilistic confusion. Emphasis 2 evaluates how the introduction of different forms of color discrimination confusions (local confusion regions and global confusion pairs) affect the homogeneous population results, and how they trade off in category solution evolution. Study 2 investigates how heterogeneous populations' categorization solutions are affected by local and global confusion influences. This is done by comparing solutions from populations protan versus deutan biases. In addition, Study 2's normal observer models include models that approximate anomalous trichromat performance. Study 3 presents the first approximately realistic populations of agents. This study is designed to reflect both realistic observer models and approximated population frequencies of common observer types seen in many human populations. Similar to Study 1 and 2, results from Study 3 are examined for the influence of population features on stable categorization solution convergence, levels of population agreement, the presence of symmetry breaking and its causes, the placement of category boundaries seen in solutions, and the number, sizes, and denotative extents of the stable category partitions relative to the FM100 hue circle continuum. Study 4 extends the investigations to examine whether agent models based on FM100 performance data from humans with color perception expertise (i.e., observers empirically shown to have heightened color sensitivity by Jameson and colleagues) impose further constraints on population solutions, and whether such agents can communicate with others in a heterogeneous population

Table 2. Simulation Parameters

Parameter	Definition	Value
N	Total number of agents	100
n	Total number of hue circle stimuli (FM100 samples)	85
k_{sim}	Parameter defining the pragmatic utility of colors	11
w	The *confusion range* operating in a local confusion region	varies
p	The probability of sorting confusion	varies

with no deleterious affects on the generation of a robustly shared population categorization solution.

Discrimination-Similarity Communication Games Interpreted for the FM100 Stimulus

Throughout this article, population naming of colors is achieved through the discrimination-similarity communication game. This game and its rationale as a model for the cultural evolution of color categories is described in detail elsewhere [5, 7]. Here we described only its essentials.

For the purposes of this article, the game is played with the 85 FM100 caps. The goal of the game is for a population of agents to communicate successfully about colors, given certain sensible features involving categorization. At the beginning of the game, a fixed number of names is provided. Agents evolve probabilistic naming strategies for the caps using these and only these names. In a given round of the game, two agents are randomly chosen from the population and two caps are randomly chosen. An agent probabilistic naming strategy is updated by its performance in the round. The basic idea of the updating is that if individuals are successful in communicating about the colors of two caps to each other, that is, agree on the caps' names, then the probabilities that they will give the same names to those caps in the future is increased (i.e., the agents are positively reinforced) and appropriate decreasing adjustments are made in the other caps. To carry out a successful communication in a round, it is also required that the two names meet the following coherence requirement. The degree of similarity of two distinct caps is related to the inverse of the smallest number of caps between them. If the degree of similarity of two caps is large (that is, the two caps are near each other in the circle) and there is not an enormous number of names (which is the case in all of our simulations), then for the naming system to be evolutionarily effective, in general, the two caps should have the same name. Similarly, the names for two caps with a small degree of similarity should, in general, be different. k_{sim} is a parameter that denotes a degree of similarity that determines whether two caps should be given the same name or different names: caps with fewer that k_{sim} caps between them should be given the same name, and those separated by more that k_{sim} caps should be given different names.

The details of the discrimination-similarity game, variants of the game, its relationship to other evolutionary algorithms for population naming, and its rationale as a reasonable approach for understanding central features of the evolution of color categorization are reported in detail elsewhere [5]. The approach extended to populations with ideal heterogeneous color observers with fixed and variable k_{sim} was also recently reported [7]. The present investigations consider situations involving agents made to resemble a variety of real observers with fixed k_{sim}.

In the abovementioned articles as well as this one, the discrimination-similarity game evolves population naming behavior that is near theoretically optimal. (An earlier report [5] details a method for calculating optimal naming behavior.) Optimal naming behavior produces categories that are continuous regions of about equal size. Optimal naming for the discrimination-similarity game cannot be perfect, because when two adjacent chips are at a category boundary such that one belongs to the category and the other does not, a communication failure necessarily results, because of the coherence requirement of k_{sim}.

STUDY 1: Investigating Homogeneous Population Solutions.

Study 1 examines homogeneous populations comprised of individual agents with the same underlying model. The aim of Study 1 is to (*i*) examine categorization solutions under different homogeneous perceptual biases, and (*ii*) provide for comparisons with the heterogeneous population solutions of Studies 2 to 4. Three types of homogeneous populations were separately considered: First, different populations comprised of 100% normal individual agents; second populations of 100% protanopes; and third populations of 100% deuteranopes. The results are described below.

Homogeneous Normal Population. Populations consisting of identical normal observers with $p = 0$ (no sorting errors) or $p > 0$ (a nontrivial probability of sorting errors) exhibit the following behavior.[7] Starting from any initial condition, after a number of iterations of the

[7]The normal model with $p = 0$ is an *ideal* normal observer model that distinguishes all 85 caps with zero sorting error, similar to the ideal observer described in [7], whereas a realistic normal observer model distinguishes all 85 caps with some probabilistic ($p > 0$) amount of sorting error.

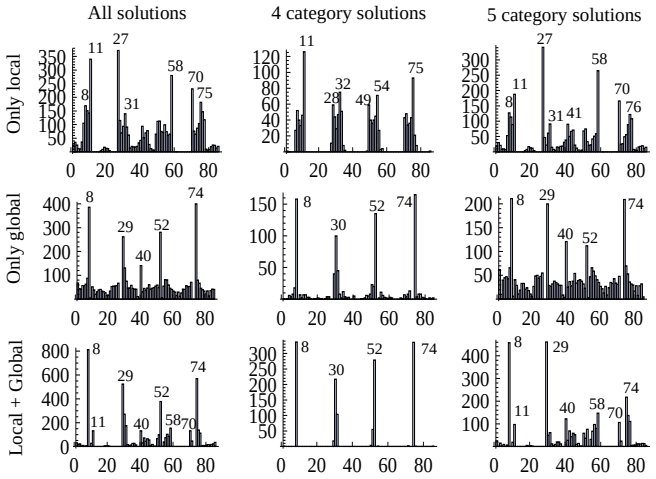

Fig. 4. Population color categorizations from homogeneous populations each consisting of 100% protanope agents. Histograms across columns represent the frequency with which a color boundary position was observed across all solutions (left column), 4-category solutions (middle column) and 5-category solutions (right column). Rows show results found under local confusion regions only (top row), global confusion pairs only (middle row), and both local and global confusion features (right row). The horizontal axis is the 85 caps of the FM100, the vertical axis show how many times out of 1,000 population simulations a color boundary was seen established at a given cap. Other parameters are in Table 2 and as described in Section 2 above.

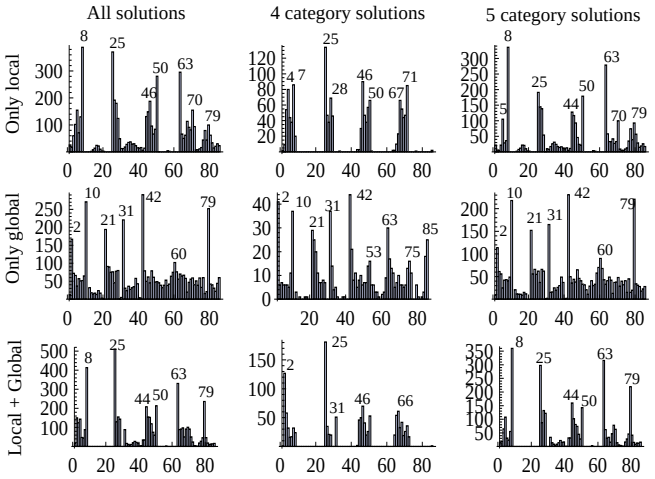

Fig. 5. Population color categorizations from homogeneous populations each consisting of 100% deuteranope agents. Figure layout and parameters are as in Figure 4.

discrimination-communication game, a population converges to an optimal, and shared, categorization solution. With the parameters as in Table 2, such solutions (not shown) normally consist of 4 or 5 distinct color categories of approximately equal size, as predicted in [5]. Raising the values of p does not change the general structure of population solutions, but increases the amount of uncertainty at color boundaries, that is, the degree of population agreement on color terms near color boundaries is lower for higher values of p. With the exception of new affects arising from the introduction of parameter p, Study 1's normal population results confirm earlier category solution findings [5]. That is, for a homogeneous population of observers on FM100 data with fixed k_{sim}, the population naming shows the following form of *symmetry:* Repeated simulations with the same starting parameters produce categorizations with nearly equally spaced boundaries and such that the boundaries varying randomly from one simulation to another. Put differently, if the FM100 caps are viewed with their natural circular ordering, then the categories of one simulation can be rotated into the categories of another simulation, and the simulations can be viewed as random rotations of a given simulation. This kind of situation is predicted by earlier analytical results [5]. A *breaking of symmetry* in this case (illustrated in results of the section below) is a reduction in possible categorizations, all the way to the extreme situation where one observes the same categorization in repeated simulations.

Homogeneous Dichromat Populations. We ran simulations on populations consisting entirely of 100% protanopes (Figure 4) and 100% deuteranopes (Figure 5). As predicted earlier [7], the nontrivial psychophysical transformation characterizing dichromats leads to a symmetry breaking. In populations of dichromats we observed that the structure of solutions was similar to that obtained by normal populations, but the color boundaries had preferred locations. The histograms in the bottom rows of Figure 4 and Figure 5 present frequencies of boundary locations for all stimuli. These distributions are strikingly non-uniform. For example, consider the leftmost bottom histogram in Figure 4, which records the frequencies of boundary locations for all solutions for 1,000 random simulations. There, histogram bar heights show color category boundaries are most frequently found at caps 8, 29, 52 and 74. A much clearer picture emerges when we examine the 4-category solutions separately, shown in the middle histogram in the bottom row of the same figure. Such solutions are almost identical to each other, with boundary locations at caps 8, 30, 52 and 74. By comparison, similar histograms for homogeneous populations of normals resemble essentially uniform (flat) distributions without any clear maxima, except for slight randomness caused by the finiteness of our simulations (these results are not shown).

While homogeneous populations of dichromats do not occur, they are very interesting from a theoretical perspective, as will become apparent later on. Because of this, we have taken additional steps to explain the particular boundary locations observed in these simulations. To do this, the influence of local and global confusions in dichromats are considered separately.

Local confusions. The top rows of Figures 4 and 5 present simulation results in which the color discrimination deficiencies of dichromats are restricted to local confusion regions exclusively (no global confusion pairs). Results show, for example, that the protanope boundaries (Figure 4) are located in well-defined patches along the hue continuum,with few or no boundaries between caps 11 and 27, or between caps 58 and 70 for. This kind of result is depicted in the schematics of Figures 6 and 7 (for protanope and deuteranope solutions, respectively), top rows. There, local confusion regions are marked by thick black circular arcs, and the dashed double-arrows define the allowed locations for color boundaries (these schematics were drawn based on the results in Figures 4 and 5). A general result is that no boundaries occur inside any local confusion regions. For five-category solutions (the schematics at top, right) the two most frequent solutions are drawn as inscribed pentagons. Overall, such results suggest that local confusion regions tend to repel color boundaries. Reasons for this finding are elaborated in the Discussion section below.

Global confusion pairs. Next, we consider the influence of the global confusion pairs in the color categorization solutions, see the middle rows of Figures 4, 5. Color boundaries are very heavily biased to a small subset of stimuli. For example, 4-category solutions for protanopes almost always have boundaries at caps 8, 30, 52, 74. As seen in Figure 6, the left schematic in the middle row, the boundaries of the most frequent 4-category solution align almost perfectly with the positions of four out of 6 caps comprising the global confusion pairs (the pairs are marked by arrows connecting the opposite sides of the circle). 5-category solutions are slightly more complicated because it is hard to find one nearly-regular pentagon with vertices at global confusion stimuli. Therefore, the populations came up with several solutions such that three out of five boundaries aligned with global confusion stimuli. Similar considerations hold for deuteranopes which came up with three most common 4-category and 5-category configurations. Overall, such results suggest that global confusion pairs tend to attract color boundaries. Reasons for this finding are elaborated in the Discussion section below.

Note that technically speaking, the notion of color boundaries is ill-defined in dichromat populations with global confusion pairs. This is because in such populations, optimal categorization solutions are not exactly the same as in populations of normals. Even though they do consist of disjoint regions belonging to one category, there are also "defects" near the location of global confusion pairs. At these locations, two types of solutions are observed: (i) one cap is classified as having a different color category from its neighbors, such as AAAABAAAA (here, cap number 5 has category B and all its neighbors have category A), and (ii) one cap at the boundary between two categories is classified as a third category, such as AAAABCCCC (here, one cap at

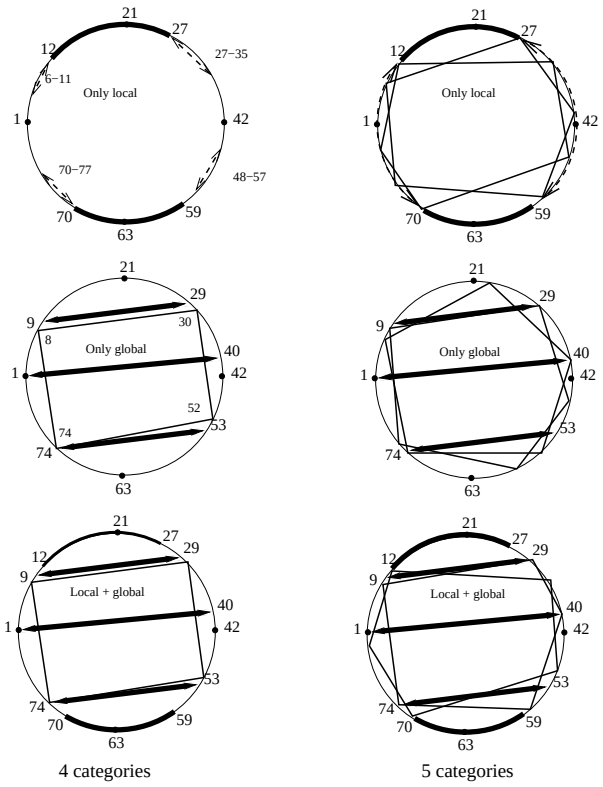

Fig. 6. Population solutions for 100% protanopes. The top row corresponds to local confusions only; the middle row - to global confusions only; and the bottom row - to both local and global confusions. The circle represents the array of 85 caps; here caps 1, 21, 42 and 63 are marked for reference. Local confusion regions are marked by thick black arcs with end cap stimuli indicated. Global confusions are marked by two-point thick black arrows with confusion pair cap numbers indicated. Typical four-category solutions are shown in the left column, and typical 5-category solutions in the right column. In the case of local confusions (top row of figure), the "support" of the solution (the regions where boundaries occur) is marked by dashed double-arrowhead lines.

the boundary between categories A and C is classified as B). In both cases, such defects were smoothed-out for the purposes of Figures 4 and 5. The existence of these defects can be understood as follows: In the presence of global confusions the color space of dichromats has effectively a non-circular geometry, similar to that shown in our previous studies [7], where parts of the circle were collapsed onto each other. In the presence of global confusion pairs, then, the geometry of the optimal solution changes accordingly.

Summary and parameter variation. In the light of the above observations, we can see that the combination of the two constraints described produce a very strong anchoring of color boundaries to a subset of locations along the circle, such that the boundaries generally tend to be attracted by global confusion pairs and repelled by local confusion regions.

The protanopes and deuteranopes in these studies are characterized by parameter values estimated from indi-

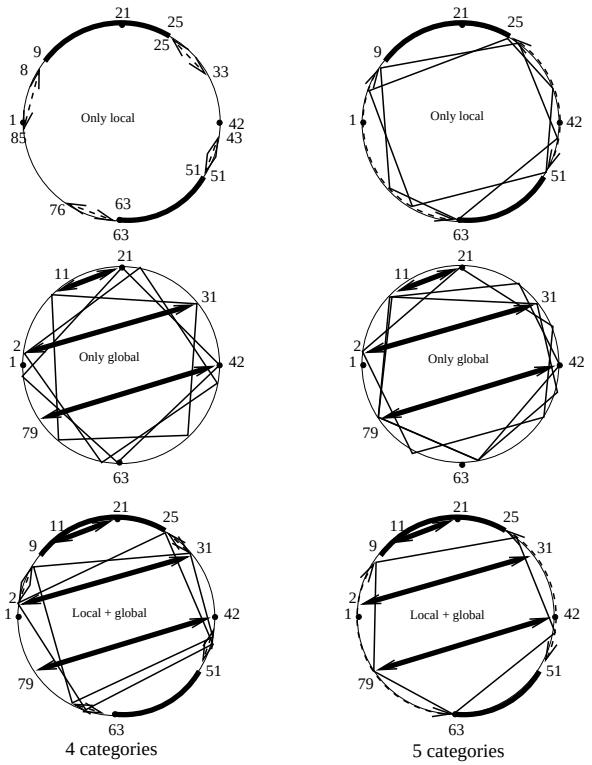

Fig. 7. Same as Figure 6 except the populations consist of 100% deuteranopes.

vidual human data. However, the sample size of data is small, which makes the error in parameter estimation high. Therefore it is important to investigate how particular parameter choices affect the results. We have experimented with various positions of local confusion regions with respect to global confusion pairs, and while the particular boundary locations change, the general conclusions remain robust. We also performed studies with different values of w, the confusion width, and p, the probability of sorting errors. It turned out that the value of p has a minimal effect. On the other hand, decreasing the value of w to low values (say, $w = 1$), changes the picture slightly. In particular, large values of w make local confusion regions virtually impenetrable for boundaries, whereas with $w = 1$, there are some solutions whose boundaries are located inside the local confusion region. Those solutions are still quite infrequent, and disappear as w is increased to more realistic values.

STUDY 2: Population Heterogeneity under varying Protan and Deutan Perceptual Biases

Study 2 investigates the ways categorization solutions vary across populations that are deutanope-biased compared with those that are protanope-biased. Examining (i) whether realistic normal error changes solutions obtained compared to that seen when normals are "ideal" observers, and (ii) any differences found when the realistic comparisons involve only deutan versus protan biases. To address these issues we construct heterogeneous populations of agents, introducing normal, anomalous and deficient observer types at frequencies approximating those found in some real populations, while restricting the forms of bias in the populations to either a protan or deutan type of deficiency. Table 1, compiled from multiple sources [14, p. 98; 15; 16, p. 217; 17, p. 464; 18, p. 30], shows ranges of observed population frequencies approximated in these simulations.

Note that Study 2's deficiency-constrained populations do not resemble human populations, and in general Table 1's ranges of population frequencies are used only as a guide to population frequencies in our simulations, since we do not intend these investigations to serve as replications of any specific real-world population demographics underlying human color category evolution. This is because (a) such frequencies vary across both human populations and different empirical surveys, and (b) the population size used throughout these investigations is necessarily small (for technical reasons) and is held constant at 100 agents, which arguably is too small a population to replicate human groups tasked with developing color categories in real-world settings. Also, due to reason (b), we often round up frequencies in our simulations (e.g., a 1.5% observed frequency to a 2% simulated population frequency) for the sake of permitting evolutionarily important peer-to-peer interactions in our population communication games. Such minor adjustments in population frequencies were well investigated and generally found to have no, or negligible, impact on the population findings reported here.

Two heterogeneous groups investigated consisted of:

1. Population P: 1% protanopes, 2% protanomalous trichromats, and 97% normals, with all parameters estimated as described above.

2. Population D: 2% deuteranopes, 5% deuteranomalous trichromats, and 93% normals, again with all parameters estimated as described above.

The normal and dichromat agents of Populations P and D are given a $p = 8\%$ probabilistic noise level, whereas anomalous trichromat agents are given a p=24% probabilistic noise level. P and D simulation results are shown in Figures 8(a) and 8(b), respectively. Result are summarized as follows:

(1) Categorization solutions found by such populations agree very well with optimal solutions found previously [5] in that they consist of disjoint regions of roughly equal length belonging to different color categories. In other words, the presence of a low percentage of dichromats with global confusion does not lead to a prevalence of defects (as described in Study 1, for homogeneous dichromats with global confusion pairs). Therefore, for this situation, the notion of color boundaries is well-defined.

(2) The symmetry is clearly broken for both types of populations (Figure 8), with some locations being much more preferred by the color boundaries than others. This being said, the peaks in populations of mostly normal

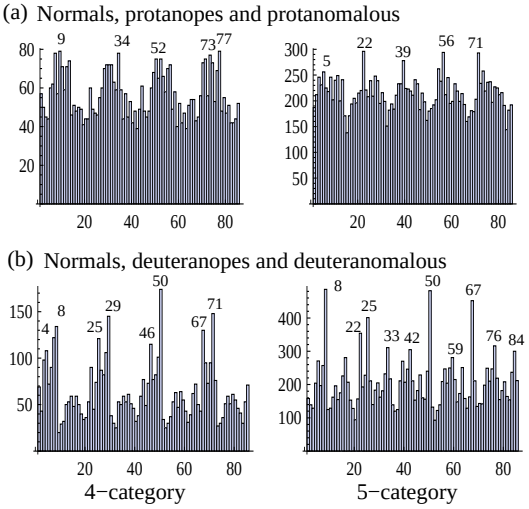

Fig. 8. Symmetry breaking in color categorizations from heterogeneous populations with (a) a protan bias, and (b) a deutan bias. Panel (a) consists of 1% protanopes, 2% protanomalous trichromats, and 97% normals, and (b) consists of 2% deuteranopes, 5% deuteranomalous trichromats, and 93%. Solutions shown in both panels are 4-category solutions (at left) and 5-category solutions (right). The horizontal axis is the 85 caps of the FM100, the vertical axis shows how many times out of 5,000 independent simulations, a color boundary was seen established at a given cap.

observers (like the ones described here) are much less pronounced than the peaks in populations with 100% deficient agents. Because of this, we had to use 5,000 runs to obtain comparable Figure 8 histograms from mostly-normal populations compared to the 1,000 runs that was needed for the Figure 4 and 5 histograms from 100% deficient-only populations.

(3) Figure 8 peaks appear more pronounced for D populations compared with the solutions from P populations. However, a P variant investigated found this result to be due to the total percentage of agents with deficiency in D (at 7%) which is higher than that in populations P (at 3%). That is, P populations with 7% protanope-based observers produced distributions (not shown) as pronounced as that shown for D populations (Figure 8b).
(4) Despite relatively low percentages of individuals with deficiencies, their influence is clearly observed in Figure 8 for both types of population deficiency bias. In fact, we also performed similar runs without anomalous trichromat agents. For example, populations of 1% protanopes and 99% normals, as well as 2% deuteranopes and 98% normals were considered. In these latter cases similarly clear peaks were also observed in the color boundary histograms (not shown).

(5) The locations of the peaks for population P (Figure 8a) are similar to that found in Study 1, where we considered homogeneous populations of protanopes. Similarly, the histograms of population D (Figure 8b) resemble that obtained for Study 1's homogeneous populations of deuteranopes. This underscores the importance of Study 1's separate investigations of unrealistic homogeneous dichromat populations. Through Study 1 we were able to explain the particular positions of peaks based on local and global confusions. The same explanations hold for Study 2's populations where dichromat and anomalous trichromat agents comprise only a small percentage of the populations studied.

STUDY 3: Realistic Population Heterogeneity Comparisons.

Study 2 introduced anomalous trichromat models to increase heterogeneity in populations using *either* a protan *or* a deutan bias. While Study 2's populations P and D do serve to increase heterogeneity of observer types compared to Study 1, they do not achieve the variety of observer types found in realistic populations which contain both classes of deficient biases. Study 3 is designed to increase population heterogeneity with this realism in mind.

Color communications between two individuals with highly similar color perception (say, two siblings with the same photopigment opsin genes yielding similar trichromat phenotypes) are likely to experience fewer perceptually based discordances of color identification and naming (cf., [27]), compared with the interactions between two individuals with different color perception bases, or even compared to two individuals with different forms of normal variation (for example, a normal trichromat compared to an anomalous trichromat). Perceptually based discrepancies in hue naming have been widely examined in the literature and are well understood. For example, as Pokorny and Smith [28] noted:

> Smith, Pokorny and Swartley [29] ... described hue estimations by five deuteranomalous trichromats using restricted response categories "red," "green," "yellow," "blue," and "white." They stated that the response category "blue" was the same for normals and deuteranomals. ... [While] the deuteranomals showed greater intersubject variability than normals in using response categories "red," "yellow," and "green," a phenomenon that does not occur in normal trichromats. ... The deuteranomalous trichromats did not always need three independent color names to describe the spectrum above $510nm$ ([28] p. 1202-1203).

It is plausible, then, that when realistically represented in a population, color perception differences distinguishing normal trichromats from both forms of anomalous trichromats should have an identifiable influence on the shared color categorization system formed by a population. Study 3 examines this possibility, and investigates the ways Study 3's solutions vary from those found with Study 2's partially heterogeneous populations.

Two heterogeneous population types investigated are:

1. 1% protanopes, 2% deuteranopes, and 97% normals.

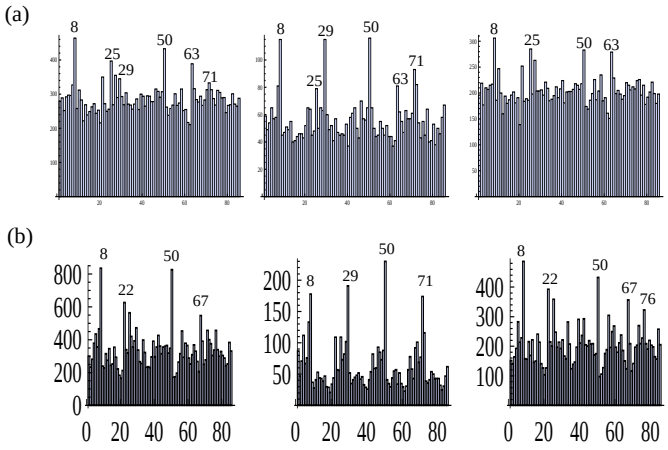

Fig. 9. Population color categorizations from two realistic heterogeneous populations. Panel (a) consists of 1% protanopes, 2% deuteranopes, and 97% normals (noise at the $p = 0\%$ level), and (b) consists of 1% protanopes, 2% protanomalous trichromats, 2% deuteranopes, 5% deuteranomalous dichromats, and 90% probabilistic normals with $p = 8\%$ noise. As in earlier graphs, all solutions (left), 4-category solutions (middle) and 5-category solutions are shown. Each histogram represents the results of 5000 runs.

2. 1% protanopes, 2% protanomalous trichromats, 2% deuteranopes, 5% deuteranomalous dichromats, and 90% probabilistic normals. [8]

Figure 9(a) and (b) present results from the above populations which can be summarized as follows:

(1) Both populations were characterized by close-to-optimal solutions, as in Study 2. Also, both exhibited a tendency for a high degree of symmetry breaking, where certain locations were preferentially chosen as color boundaries.

(2) When compared to Study 2's Figure 8, the histograms in Figure 9 more closely resemble those obtained for Study 2's D-population results, than those obtained for P-populations. Thus, not surprisingly perhaps, we find that a higher total percentage of deuteranope-like types (7%) in a population, in the company of fewer protanope-like types (3%), exert a stronger influence on the boundary locations of Study 3's solutions (similar to Study 2's result item (3)).

(3) Figure 9(a) was created from runs where the probability of sorting errors p was zero. To further evaluate the role of sorting confusions we performed other Population 1 simulations by increasing the value of p to higher and more realistic values. Such increases did not make any tangible difference in the outcomes observed (not shown).

(4) We have found that adding realistic percentages of anomalous trichromat agents to a population (Figure 9(b)) produced no differences in categorization solutions compared with the solutions obtained from populations that differed only by the absence of anomalous trichromats (Figure 9(a)). From a comparative color vision standpoint, and from what is known of human color naming behaviors (e.g., [28]), this result seems counter intuitive, and is elaborated on in the Discussion below.

(5) Comparing the color boundary positions in populations with and without anomalous trichromats (Figures 9(a) and (b)), we can see that the presence of anomalous trichromat agents does make a quantitative difference. In particular, it makes the symmetry breaking effect stronger due to increased total percentages of color deficiency in the population, 3% of the population without anomalous types vs 10% of the population with anomalous types included.

(6) We have checked our results for robustness by changing the positions of local confusion regions for anomalous types slightly. In particular, in one set of simulations, we made the confusion regions for all protan-types identical, and similarly, the confusion regions for all deuteranope-types identical. We observed that such slight modifications did not noticeably change the resulting histograms for obtained categorization solutions.

STUDY 4: Modeling "Expert" Agents using data from Observers with Diverse Photopigment Opsin Genotypes.

Studies 1, 2 and 3 investigated perceptual modeling features related to X–chromosome linked color vision deficiencies. In Study 4 we investigate consequences of non-deficient color processing variation that may arise due to the presence of individuals with X–linked photopigment opsin gene heterozygosities in a population (cf. [30, 31]).

Because the suggestion of perceptual variation associated with opsin gene heterozygosities is a somewhat novel proposition, we first provide below additional explanation of this observer type, then clarify the rationale for investigating the possible impact of such observers on population color categorization solutions, and explain the basis for observations of non-normative FM100 performance variation in such individuals.

Observer groups defined by photopigment opsin gene variation

The protan and deutan observer group models investigated above are analogs of human observer types due to congenital inheritance based variations of the long- (LWS) and medium-wavelength sensitive (MWS) photopigment opsin genes. The genes for the LWS and MWS photopigments are located in a tandem (head-to-tail) array on the X–chromosome, and have 96% of the exon coding regions of their nucleotide sequences in common

[8] As mentioned earlier, here populations' observer-type frequencies are in some cases selected for algorithmic appropriateness and to optimize communication game interactions in a population of size 100 agents – e.g., rounding Table 1's frequency ranges up to 2% (as opposed to down to 1%) to achieve communication games between two same-kind observers while secondarily approximating Table 1's relative frequencies. This is appropriate given the aims and goals of the present research.

[32, 33]. The LWS and MWS gene nucleotide sequences differ at only 15 sites of these amino acid sequences, and single nucleotide substitutions at three particular sites (codons 180, 277 & 285, Exon 3) have been found to produce the most dramatic shifts in spectral sensitivity [18], and shifts in spectral sensitivity are known to increase monotonically as the number of substitution sites increase [34, 35].

Heterozygous females who, for example, carry different forms of the LWS opsin gene on each X–chromosome (e.g., both allelic variants at codon 180) have the genetic potential to express more than the usual three classes of retinal photopigments, and are referred to as putative retinal tetrachromats [30]. It has been suggested that individuals with four retinal photopigment classes might enjoy a heterozygous advantage and experience a dimension of perceptual experience denied to trichromat individuals, possibly gaining tetrachromatic vision [36]. For many years vision scientists have known that some humans inherit the genetic potential and actually express four photopigment variants as distinct retinal cone classes with different spectral sensitivity distributions. Frequency estimates of females of European descent who are potential tetrachromats vary between 15% and 47% depending on the heterozygote genotypes considered. Less is know about the actual frequency of expressing four retinal cone classes.

Perceptual differences associated with retinal tetrachromacy

While the potential for human perceptual tetrachromacy exists, whether it actually exists is the subject of debate, and accepted theory suggests that humans process no better than a trivariant color signal. Thus, four retinal cone classes are a necessary (but not a sufficient) condition for tetrachromatic color perception, since, for true tetrachromacy tetra–variant cortical color signal processing is also needed. A long-held view has been that individuals with four photopigment opsin gene variants (i.e., carriers of color deficiency) themselves exhibit impaired chromatic discrimination. However, some research suggests that carriers of protan deficiencies are not chromatically impaired compared to male or female trichromat controls ([37], p. 2898), finding that some heterozygote carriers need significantly less red/green contrast to detect the chromatic modulation in a grating. The latter result suggests that carriers of protan deficiencies have lower red/green chromatic discrimination thresholds compared to normal trichromats.

Jameson and colleagues [31] present a detailed investigation of FM100 performance of females with LWS (S-180-A) heterozygous opsin genotypes. A portion of Jameson et al.'s heterozygotes were found to be free of any form of color vision impairment or any complaint of color vision weakness, and showed increases in sensitivity for detecting chromatic bands in a diffracted spectrum task [30]. Jameson and colleagues' [30, 31] interpretation of these results is that some female carriers of protan deficiencies differ from female trichromat controls (by standardized color assessment and empirical demonstrations) but, in general, their variation does imply they generally differ in a deficient way. Rather, they differ from normal controls in that they detect significantly more chromatic contrast, which produced increases in categorical color differences, in the chromatic banding task used [30] . Consistent with this, Birch [24] indicated that females who are compound mixed heterozygotes for protan and deutan color deficiency (as were some of those studied by Jameson et al. [30, 31]) are usually reported to have normal, not deficient, color vision.

Research on similarity judgments of color circle stimuli from individuals with varying photopigment opsin genotypes underscores the importance of considering opsin genotype observer models in the present hue circle categorization studies. In a twin-study of perceptual color space, Paramei, Bimler & Mislavskaia [27] showed that projected color space similarity mappings of color circle stimuli are nearly identical across individual same-sex monozygotic twins possessing the same underlying photopigment opsin genotypes, but that such color circle mapping spatial relations can vary substantially across twin pairs due, by implication, to differing opsin genotypes, even though these individuals are similarly diagnosed as normal trichromat observers [27, p. 308, Figure 2]. The implication for the present research is that observers with differing photopigment opsin genotypes who generally possess normal color vision, may experience different color similarity relations among hue circle stimuli. Although such differences may have little impact in some perceptual categorization studies, when expressed in FM100 data used as the basis for observer models, they have the potential to greatly impact the categorization solutions acheived under evolutionary dynamics.

Study 4 uses FM100 data from opsin gene heterozygotes as the basis for an "expert" observer agent model. This study supposes that high degrees of FM100 sorting errors in otherwise normal color vision individuals may represent useful information that contributes to evolutionary game theory dynamics of color categorization development, in contrast to being uninterpretable patterns of confusion errors having no potential to impact inter-observer color communications and population color categorization solutions.

Modeling Expert Observers. Modeling experts on the basis of FM100 performance presents some interesting theoretical predictions and possibilities. The first is that experts with regard to the FM100 test perform as ideal normals, making no transpositions on sorting the 85 cap color continuum, and show no probabilistic sorting error. Experts modeled this way would produce personal stimulus color continua that replicate the "correct" sequential ordering, from 1 to 85, of the FM100 caps. In this sense, the "correct" ordering prescribed by diagnostic would replicate the continuous color sequence of an expert observer.

A second way to conceptualize expert performance on the FM100 is that, for whatever reasons, experts have a different, personal, "correct" ordering of the FM100 continuum which does not exactly follow the FM100 sequential 1 to 85 order. In this case, experts would exhibit transposition errors in the sorting task when scored according to the diagnostic's standard. Where these transpositions might occur and why this is a topic of interest is discussed further below. Also, if such transpositions were sufficiently structured, they could present a new potential for symmetry breaking in addition to the kinds already observed in Studies 1–3 above.

Another possibility is that cap transpositions in the FM100 performance of these experts could be modeled as stemming from non-normative *jnd* structure variation on one or more color space dimensions (as opposed to simply reflecting poor color sensitivity). Such a modeling feature could impact categorization performance by varying the *jnd* mapping of the FM100 and changes the stimulus relative to the parameter k_{sim}. Other specific perceptual and decisional features of possible mechanisms underlying FM100 performance variation of these expert individuals are being investigated, but their description is beyond the scope of the present article.

We model our "expert" agents based on existing observer data [31]. FM100 data from five female individuals with diverse photopigment opsin genotypes were used for modeling agents with non-normative color vision expertise. All five individuals were identified as heterozygous for both the L-opsin (S-180-A) gene and the M-opsin (A-180-S) gene [30, 31].[9]

As detailed by Jameson and colleagues [31], compression parameter analyses of the five subjects' FM100 data indicates that their individual patterns of FM100 confusion all followed a similar trend, displaced in a specific direction corresponding to an axis of compression of approximately 15° in a polar coordinate compression parameter space ([31], Figure 2a, p.12). Their performance differs from normative age-adjusted FM100 performance by an average Z-value of 1.31 standard deviations (indicating individually they either failed or performed very poorly on the FM100 test), but they were described as heterozygotes which the FM100 can diagnose as "... false-positive deficient when their color perception is otherwise unimpaired and their color sense is generally regarded as excellent." [31, p. 15].

Based on these individuals' data, Figure 10 shows an average FM100 performance vector for the experts, with local confusion regions computed similar to that described for anomalous trichromats earlier. For the experts they are 21–28 and 43–47. The confusion range, operating within the expert local confusion regions, was calculated as $q = 0.92$ which corresponds to the value $w = 3$.

Similar to what was found for anomalous trichromats

[9]Subjects 27, 52, 58, 61, and 85 as reported in [31]: age ranged 18-21 years, all performed with above average chromatic banding behaviors ($\mu(medianbands) = 10$ vs. 7.9 for control subjects).

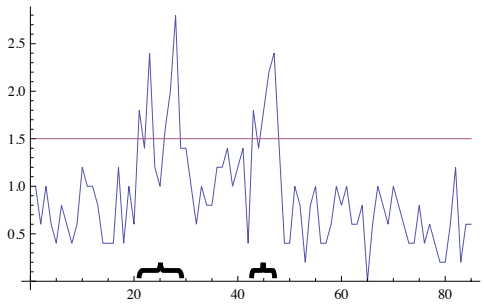

Fig. 10. Confusion patterns found in the FM100 performance of five expert observers from Jameson, Bimler & Wasserman [31].

(Figure 3), expert observers exhibited patterns of diffuse color confusion, showing a large amount of sorting confusion compared to the normal observers. Thus, similar to anomalous model, to find the probability of sorting errors for our experts we performed an average transposition calculation. We obtained the following estimates: For experts the probability of inversion outside the expert local confusion regions is 22% per cap.

Below we present results from simulations that include expert observers at frequencies thought likely to reveal any potential influence on obtained categorization solutions. In populations like those summarized earlier in Table 1 at least 15%, of women of European ancestry are carriers of X–chromosome linked deficiencies of color vision (cf., [37]). The primary question Study 4 addresses is: How do solutions from realistically heterogeneous populations that include expert agents compare with those from populations that are realistically heterogeneous but do not include "experts"?

One possibility is that the presence of a sufficient number of "expert" agents in a population disrupts the population's ability to stabilize a shared categorization system. Generalized beyond this study, such a finding would imply either (*a*) the existence of some human or cultural ability to reconcile color category communication difficulties potentially introduced by a group of individuals sharing a similar, but non-normative, color expertise; or (*b*) would rule out the real world feasibility of our hypothesized "expert" influence, especially in light of the successful human color categorization systems that have been evolved by populations containing such observers.

To illustrate a possible influence of such individuals on population solutions we present the following two populations:

1. A heterogeneous population of 1% protanopes, 2% protanomalous trichromats, 2% deuteranopes, 5% deuteranomalous dichromats, 75% normals, and 15% experts.

2. A homogeneous 100% expert population.

Results for populations including expert observers. The simulation results for populations includ-

ing expert observers are presented in Figure 11(a,b) and are in part predictable from some of the patterns found in studies 1-3.

(1) Categorization solutions for realistic heterogeneous populations including expert observers agree with the usual near optimal solutions found previously [5] . The same holds for populations consisting of 100% experts because of the absence of global confusion pairs for such observers.

(2) The presence of 15% experts (Figure 11(a)) does not change the categorization histogram dramatically, compared with populations without experts, Figure 9(b). Beside a slight shift in boundary locations for 5-category solutions, the biggest difference manifests itself in the existence of some newly apparent "gaps" in color boundary distributions, the most significant being between caps 42 and 47, and another one between caps 20 and 28. This is consistent with our conjecture that local confusion regions repel color boundaries. For expert observers, the local confusion regions lie between caps 21–28 and 43–47. It is exactly where color boundaries occur less often. A more pronounced representation of this effect is seen in Figure 11(b) where we investigate homogeneous expert populations. There, no boundaries occur in these regions of local confusions. In addition, 4-category solutions reflect a third gap between chips 62 and 70 as a direct consequence of the absence of color boundaries inside the local confusion regions (see the middle histogram in Figure 11(b)). The same tendency is also observed in realistic heterogeneous populations in the presence of experts, see the middle histogram in Figure 11(a).

(3) Increasing the percentage of expert observers to 25% of the total population does not change the resulting qualitative picture (result not shown).

In general, Study 4's results suggest that non-deficient color perception observers (who may exhibit a form of genetically based color expertise but perform poorly by FM100 test standards), can be a considerable proportion of a heterogeneous population without substantially disturbing the stablization and form of the populations' shared category solution of the FM100 stimulus. The more detailed differences observed in Study 4's investigations are elaborated in the context of the Discussion below.

4. DISCUSSION

The sections above considered some realistic approximations of population heterogeneity and investigated their impact on simulated color category evolution for a continuous hue circle stimulus. Our population color categorization simulations show that trade-offs exist between the features of various observer discrimination models and the frequency of different observer models in heterogeneous populations. These factors balance in a dynamic fashion during the formation and stabilization of a near optimal color categorization solution by a population of agents. Even in the face of diverse population heterogeneity, observed solutions reach stabilization by maximizing successful color communications among agents via the discrimination-similarity game. This is achieved in ways compatible with an optimal partitioning theory of color space that uses an Interpoint Distance heuristic [38, 39, 40]. Moreover, detailed analyses of observer discrimination features suggest specific mechanisms underlie the formation of category boundaries in simulated population solutions. These findings provide insights for analogous factors thought to underlie the formation of human color categorization systems.

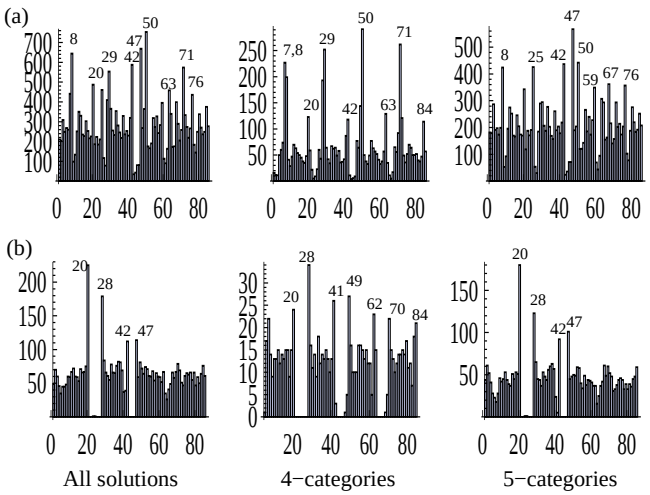

Fig. 11. The role of experts in population color categorizations. The solutions show results from two populations. (a) The population consists of 1% protanopes, 2% protanomalous trichromats, 2% deuteranopes, 5% deuteranomalous dichromats, 75% normals, and 15% experts. (b) A homogeneous population of 100% experts. As in earlier graphs, all solutions (left), 4-category solutions (middle) and 5-category solutions are shown. The parameters are as in the other studies described. Each histogram represents the results of 5000 runs for (a) and 1,000 runs for (b).

Summary of findings

(a) In both populations with and without realistic heterogeneity, observer discrimination variations involving a pair of caps on a confusion line exert significant symmetry breaking effects on categorization solutions; compared to unsystematic error variations or random sorting confusion variations, which have little impact on category solution stabilization.

(b) Random sorting–confusion variations only seem to decrease the degree of agreement at color category boundaries.

(c) Global confusion pairs break symmetry in categorization solutions by anchoring category boundaries.

(d) Local confusion regions break symmetry in categorization solutions, and also alter category–size uniformity, by introducing ambiguity regarding the lo-

cations of a category's center-of-gravity. This results in the repelling of category boundaries.

(e) Increasing the number deficient dichromat agents in a heterogeneous population results in greater symmetry breaking affects on population solutions.

(f) The addition of anomalous trichromat agents to a population has a minor impact (beyond that imposed by dichromats) on the location of category boundaries in the observed category solutions. Although, as discussed in detail below, this is explained by similarities between the FM100's assessment of anomalous trichromats and dichromats.

(g) Individual variation in observer features within agent populations impacts the obtained category solutions. This underscores the importance of tracking heterogeneity in observer modeling for color category learning [42].

(h) Realistic approximations of observer types and population frequencies confirm our earlier symmetry breaking and confusion axis findings from idealized populations. To our knowledge the modeling of color categorization dynamics based on realistic data has not been previously attempted. The present findings underscore the importance of tracking heterogeneity at both the individual observer level and population level when modeling color category systems.

(i) Category solutions are constrained by the observer features employed. The solutions obtained under such constraints are easily modified by introducing to the simulations external factors, such as variable color utility (cf., [7]). We consider this to be a first step in modeling realistic constraints thought to contribute to human color category systems.

(j) Introducing further population heterogeneity by incorporating the modeling of expert agents did not diminish the stability of shared category solutions; instead, it produced unambiguous categorizations of new connected portions of stimuli. A full explanation of this result, which appears to require an alternative form of investigation, is outside the scope of this article.

Understanding how imperfect observer performance impacts categorization solutions

In Study 1 we considered populations consisting of identical normal observers with $p = 0$ (no sorting errors) and several variations of $p > 0$ (nontrivial probability of sorting errors). Across several investigations in which p was varied, in general, the primary consequence of increasing probabilistic sorting errors in the observer models was decreases in the degree of population solution agreement **at color boundaries**. Otherwise, such increases of probabilistic sorting errors produced no other effects on the observed category solutions. This result is interesting from the perspective of human color categorization theory, because, if generalized, it suggests limited varying of precision in individual category assignments during communication games does not hinder the formation of a shared population color categorization system; rather such varying of precision simply adds to the degree of classification uncertainty – or fuzziness – observed at a population's color category boundaries. These findings linked to sorting–error increases resemble the null effects seen following random observer variation in other color category simulation research (see [3], p. 517).

Counter-vailing influences from local confusion regions and global confusion pairs help establish category boundaries

To understand how local and global confusion features trade off in our simulation studies, it is useful at this juncture to consider some implications arising from human color categorization tendencies. For example, realistic populations contain much observer heterogeneity. That is, although a population may have almost no incidence of deficiency and consist largely of "normal" observer types,[10] one finds that normal individuals' color category denotata differ. Such normal observer variation creates reliable differences in individuals' choices of the category focus of, for example, *green*. Suppose the reasonable position that, in such cases, everyone in the society believed that the center of gravity of the population's green category region should be the location of the best–exemplar of green. When a population of otherwise similar individuals (who, for the present discussion, are assumed to have the same *jnd* structure on the hue circle), chose different category best exemplars, then the locations of both population category boundaries and population category best-exemplars are impacted. A likely consequences is that both foci and boundaries become fuzzy – beyond the probabilistic fuzziness inherent in individuals repeated measure variation – in terms of what they denote for the population.

Local confusion regions. Similar to realistic phenomena, in the present simulations, agents have slightly different best-exemplar points for the same k_{sim} partitions for which they've been tasked with learning through individual interaction. This sort of best-exemplar variability interacts with local confusion influences, as follows: When all members of a population of varying individuals need to learn and evolve the same category structure, then the greater the amount of local confusion present (and the greater the ranges of nearest neighbor confusion) the more pressure exerted on category boundaries to form at greater distances from a category center of gravity. Given the normal individual variation assumed in this example, such dynamics will naturally emerge during the process of constructing a shared, unambiguous, best-exemplar region for a *population's* categories.

This counter-vailing interaction (a pragmatically fixed

[10] Or if one considers only the "normal" observers of a heterogeneous population with some color deficiency.

k_{sim} structure trying to accommodate an ambiguous central exemplar region) seems also to play a role in the unequal category-size observed using the present approach (cf., [7]). That is, more category area is needed, when best-exemplar ambiguity exists, to meet a socially prescribed k_{sim} requirement. This analysis also applies when local confusions occur in other places besides a category center, but the case of best-exemplar variation is the most straightforward example of the potential impact of local confusion regions on category boundaries.

Global confusion pairs. By comparison, how do global confusion pairs impact category boundaries? From the present studies, and results from earlier research [5, 7], it appears that global confusion pairs attract boundaries though symmetry breaking, by providing necessary perceptual anchors (for some observers) in the categorization system of a population. Anchoring is the process of creating and maintaining the correspondence between percepts that refer to the same physical objects, or in this case, two objects (or 2 object labels) that share the same percept, or are indistinguishable [41].

How specifically do two terminals of a global axis of confusion provide anchors, and thereby serve as boundary attractors, in such a hue circle situation? One possibility is that during a system's evolution a dichromat segment of the population may tend to label one terminal of their global confusion pair 50% of the time with label "A" and the other 50% of time they label the same terminal "B."[11] In this case, qualitatively there is a point on the hue circle that is functioning very much like a boundary (a cusp between actual categories) even though agents may have yet to master (*i*) the shared meaning of "A" or "B" categories, or (*ii*) the relational structure of the two categories relative to other yet-formed categories and their meanings. For these confusion-pair dichromat agents (and, by invasion, some other agents without global confusions), then, a natural candidate position for a boundary (in a near–optimal population solution) is at a terminal of a confusion pair (or even near them). This, in brief, is the way global confusion pairs attract boundaries in the present simulation studies. Study 2's Figures 6 & 7 illustrate how this occurs, where perceptually indistinguishable terminals of a confusion axis form the boundaries for categories that probabilistically contain both terminals.

In summary, global confusion pairs anchor not so much because labels "A" and "B" serve as "dichromat synonyms" for a given point on the circle, but because the dichromat tendency for confusing the terminals of a confusion pair makes the outcomes of that pairs' games functionally behave like the other (actual) category boundaries, and this creates circumstances where a boundary becomes established as a solution stabilizes.

[11] This can intially happen due to chance variation. But may persist due to the influence of communication games with "normal" agents who are observed labeling one terminal "A" and the second terminal "B" because a normal agent model allows the confusion pair terminals to be distinguished, while the dichromat agent model does not.

Effects of systematic individual variation on population color naming

Jameson [42] suggested that adding non-random individual variation to a population – for example, introducing individuals with some systematic X-linked photopigment deficiency – should influence the color categorization solution the population achieves. This is based, in part, on the substantial variation seen when comparing normal trichromats to X-linked anomalous trichromats on psychophysical tasks that in essence provide category best-exemplar settings, such as "unique yellow" (e.g., [43, 44]). Webster and Kay suggest that to the extent that sources of individual variation (that include physiologically-based perceptual differences of the type simulated here) within and across populations "... affect differences in color appearance, they are also important for understanding the similarities or universal tendencies in color naming." ([20], p. 34).

The results of Study 1, 2, and 3 support the suggestion that individual variation influences population color categorization solutions. For example, it was shown that homogeneous populations of simulated normal observers produce category solutions that differ in important ways from those produced by heterogeneous populations (that only differ by the addition of realistic proportions of dichromat observers). Influences were also seen when the simulation parameters reflecting individual differences were made to approximate realistic observer variation. Second, results from local and global confusion analyses suggest that a possible universal tendency in color naming is to evolve a system that minimizes the likelihood that colors that are perceptually confusable by some individuals in a population tend to be classified by the entire population into different color categories.[12]

We also observed cases where similar color naming solutions across populations occur even when the underlying population response models had different physiological biases (Studies 2, 3 & 4). This rules out the possibility that uniformities in color naming necessarily emerge when the color relevant aspects of physiology are uniform (see [20], p. 34).

Effects arising from anomalous trichromat variation

Study 3 examined the possibility that color perception differences that distinguish normal trichromats from anomalous trichromats could impact the shared color categorization system formed by a population containing both observer types. In Study 3 we reported a result suggesting that anomalous trichromat agents did not have any unique, differentiable impact on population color categorization solutions. In view of Pokorny & Smith's

[12] Such a minimization is more likely when confusion pairs or categories are few).

[28] and others' observations on anomalous trichromat behaviors, this result appears counter-intuitive. However the noted absence of an anomalous trichromat impact is primarily due to observing little difference between solutions from populations comprised of only normal and deficient observers, compared with solutions from populations comprised of normal, anomalous and deficient observers. This apparent absence of a specific anomalous trichromat influence depends on the fact that performance on the FM100 is not particularly useful for distinguishing between dichromats and anomalous trichromats [45]. The reason for this is that anomalous trichromats perform as poorly on FM100 sorting as do dichromats (due to the fact that the test was designed primarily to detect dichromacy). The net result for our simulations is that the presence of anomalous agents only makes dichromat symmetry breaking effects stronger, due to the anomalous agents' FM100 response patterns reinforcing the dichromat biases that are the basis for boundary locations.

In our initial decisions to model population color categorization evolution on a hue circle subspace, the FM100 hue test was chosen because it is routinely used to evaluate individual variation seen in the perceptually based ordering of a standardized stimulus set forming a continuous gradient of hue. The present modeling of observer types is necessarily limited by the FM100's measures of the different observer types assessed. Thus, to the degree that the FM100 test and analysis does not unequivocally differentiate, for example, an anomalous trichromat from a dichromat observer, this ambiguity is inherent in the data we used to model those agent populations. It seems likely that if our observer models were constructed on data that more decisively distinguished anomalous trichromats from dichromats, we would observe a more specific set of anomalous trichromat influences on categorization solutions. Thus, anomalous trichromats may be more influential to such category solutions than those revealed in FM100–based investigations.

This modeling limitation is similarly inherent in the modeling of "expert" color discriminators of Study 4. Since the FM100 was not designed to assess individuals expressing more than three photopigment classes, the test is arguably not an appropriate method for evaluating the perceptual differences that might follow from individuals with weak (retinal) tetrachromacy. If additional forms of hue circle data (in addition to the FM100 data) were available to help model the different observer types seen in human populations, clearly those data would provide a better and more realistic basis for carrying out these simulations, and the resulting solutions may reflect more clearly any unique variation present in the "expert" data. Despite the strong ties of our simulated category systems to the data used for observer modeling, the results for symmetry breaking and local and global confusion influences on boundary settings in stable solutions are unavoidable consequences arising from the observer features and the hue circle communication game scenario we used. Whether such results are also found in categorization solutions on three-dimensional color spaces is of additional interest, especially in view of expected observer model complexities, such as variation in saturation discrimination across observer types [46, 14].

Understanding "expert" color discriminators' poor FM100 performance, and its impact on population color categorization solutions

It is of interest to note that, generally, Study 4's expert subjects show FM100 performance similar to that of our Anomalous Trichromats, in that they have a large degree of diffuse color sorting error around the FM100 hue circle. Our "expert" subjects differ, however, in important ways from anomalous trichromat subjects.

First, as seen in Figure 10 the FM100 confusion regions shared in "expert" average performance plots do not coincide with the confusion regions for protan and deutan defects, as was seen in the anomalous trichromat data considered. Second, the expert group's confusion regions are most pronounced in Tray 2 of the FM100 test, and in the Greenish-yellow region (Tray 2 includes caps 22 to 42) of the test. These two regions may have significance beyond a simple model of deficient sorting.

For example, in eigenvector analysis of cone responses to reflectance spectra measures of the 85 caps of the FM100, Mantere and colleagues showed that when the stimuli are scaled in a two-dimensional opponent color space representation,[13] departures from a smooth color circle contour can be clearly seen in the range from approximately cap 22 through cap 40 ([11], Figure 2, p. 2239).[14]

Thus, the possibility exists that Tray 2 produces subtle disruptions in the smoothness of perceptual spacing of color circle, which were caused when the stimuli were sampled from 100 to 85 caps during the the FM100's development. Farnsworth noted:

> "... The Munsell 100-hue series of colored papers ... was fairly well spaced perceptually. Gross inequalities in discriminative differences were eliminated by dropping from the series one of each pair of papers which was most often confused by normals as rated on standard tests for color vision. Eighty-five color papers were found to be about equal in chroma and value for normals and each was perceptably different in hue from its neighbors. Some of the steps were more easily noticeable than others so that the whole circuit of colors was estimated to embrace over 90 J.E.N.D. (Just Easily Noticeable

[13] With axes corresponding to the second and third vectors of the Eigen solution – namely, a Y-B axis and a R-G+B axis, respectively.

[14] Disruptions in this region of the color circle are similarly seen when spectral reflectance measures of the FM100 stimuli are plotted in a Munsell color metric space using methods described by A. K. Romney [47].

Differences) ..." ([25], p. 568).

If perceptual nonuniformities exist in some regions of the FM100 sequence they may be ignored by color "normal" observers, but noticed by observers of heightened color experience, and result in non-normative sorting in some sections of the FM100, and normative sorting in others. It is also possible that despite aims to maintain iso-brightness of the test, lightness differences in the FM100 caps (see [26], p. 283) might provide sorting advantages independent of color vision capabilities, especially in individuals with L-opsin gene (protan) polymorphisms (such as the five experts used in Study 4's modeling). Also, some combination of both subtle hue and lightness may contribute to the non-normative FM100 sorting of Study 4's expert color discriminators.

Consistent with these ideas, Jameson and colleagues suggested heterozygotes' regions of FM100 confusions could be

> "... construed as reduced color discrimination in those zones; or, more directly, as evidence that a different sorting sequence for the stimuli is more natural for some heterozygotes. In the former case, there is a possibility that the reduced discrimination is compensated for by heightened discrimination in other zones (not picked up by the F-M 100, because of floor effects). Such heightened discrimination might dictate a subjective ordering of the F-M 100 samples that is at odds with the 'correct' ordering recommended by the scoring manual. In any case, this kind of discrimination difference does not necessarily translate into perceived dissimilarities when the color differences are larger" ([31], p. 17).

Thus, observing that while the FM100 seemed able to detect the color perception variation associated with these excellent color discriminators, in some cases the FM100 scoring procedure "... does not distinguish between non-normative deficient and non-normative good color perception" ([31], p. 15).

Initially, Figure 10's expert confusion regions may appear to be simply box-end sorting artifacts (cf., [48]). However, the idea that some perceptual nonuniformities in the FM100 might be accentuated for some observers, but not others, is consistent with other research [49, 50]. The cited studies used roughly circular hue continua comprised of 32 stimuli from the Farnsworth-Munsell D15 and and D15-DS tests, and found that the color similarity space parameters of males and females show differences in the salience of color-space axes, with males attending more to a lightness axis and less to a red-green axis compared to females [50, 49]. Similar to previously mentioned research [31], the cited works of Bimler and colleagues similarly consider the existence of photopigment heterozygosity among females as a possible explanation for the gender linked differences in color similarity relations observed in those studies.

Other research suggests various forms of non-uniformity in the FM100 stimuli and in *normal* performance on the FM100. Craven [48] describes a sorting artifact in the FM100 arising from the way the cap stimuli are grouped into four boxes, resulting in the caps near the ends of each box being scored less compared to caps near the center of box. Others report the converse, namely, increases in scores of caps at the end of boxes [51].

Despite these modeling obstacles, Study 4's results seem especially informative. The suggestion that our "expert" color observers may frequently occur in a heterogeneous population without disrupting the stablization and form of a populations' shared category solution, and without driving it to an indeterminant oscillating solution, or a non-optimal solution, is a very informative result. It provides a proof of concept for human population heterogeneity by implying that violations of a trichomatic constraint would not destroy a population's ability to communicate optimally about color.

Indentifying universal tendencies in color categorization system evolution

Frequently empirical comparisons of human color categorization systems emphasize the similarities and differences across languages concerning the placement of central category members, called color *foci* or best-exemplars (e.g., [1, 12, 13], and others). Of human color categorization systems, Webster and Kay generally noted "the concordance of basic color terms across languages provides strong evidence for universal tendencies in color naming, but does not preclude all cultural or linguistic influence" ([20], p. 37). In their study of 110 color categorization systems from the World Color Survey (WCS) they observe *constellations of similar color categories* across languages, but conclude that "...the WCS data suggest that different language groups do vary in the average focal choices for nearest-neighbor terms ... [and] ... these group differences place some limits on universal tendencies in color naming," and surmise "...whatever factors give rise to large individual differences in focal colors, they may themselves have strong universal tendencies" ([20], p. 50).

Consistent with Webster and Kay's [20] observations, the present simulations reveal universal tendencies in the evolution of color category solutions by artificial agent populations. Following from the present focus on endogenous observer model variation, the universal tendencies found here include:

(i) a minimal impact of unsystematic perceptual variation (i.e., noisy perception or random sorting error) on categorization solutions;

(ii) substantial constraints imposed on categorization solutions by systematic perceptual variation;

(iii) counter-vailing mechanisms (arising from perceptual variation) that trade-off in the process of arriving at a stable categorization solutions;

(iv) increases in trade-off demands towards solution stabilization as perceptual constraints (*e.g.*, color confusion pairs and confusion regions) become either more varied or more frequent in a population;

(v) increases in trade-off demands towards solution stabilization as perceptual constraints (*e.g.*, color confusion pairs and confusion regions) exploit more extensive areas of the color space;

(vi) a tendency to evolve categorization solutions that optimize communication among all population members regardless of observer-type diversity, as opposed to evolving solutions based on majority-rule or a marked population specialization.

While we emphasize that the results of the present simulations do not aim to replicate human categorization results, we believe it likely that some of the categorization solution tendencies observed here suggest mechanisms that may similarly underlie universal tendencies in color categorization systems of human populations.

Regarding Webster and Kay's [20] expected cultural or linguistic influences mentioned earlier, some of our previous research has considered exogenous environmental and socio–cultural influences, and we observed strong symmetry breaking affects in simulations by incorporating color *regions of increased salience* and environmental color *hot spots* [5, 7]. Such influences generalized from pragmatic constraints found in human color naming situations are strong candidates for modeling cultural or linguistic influences that universally exist in real human color categorization and communication circumstances.

Thus, the present research supports the suggestion that various constraints and sources of influence give rise to universal tendencies in color categorization scenarios,[15] and that such tendencies may in turn underlie color naming similarities across human languages.

ACKNOWLEDGEMENTS

Author Contacts: Jameson (kjameson@uci.edu) and Komarova (komarova@math.uci.edu). The authors thank Louis Naren and Ragnar Steingrimsson for helpful suggestions, and David Bimler for providing the Bimler, Jacobs & Kirkland [22] data. Partial support was provided by NSF#07724228 grant from the SES-Methodology, Measurement, and Statistics board. N.K. gratefully acknowledges support of a Sloan Fellowship.

References

1. Kay, P. & Regier, T. (2003). Resolving the question of color naming universals. *Proceedings of the National Academy of Science, 100*, 9085–9089.
2. Belpaeme, T. & J. Bleys (2005). Explaining universal color categories through a constrained acquisition process. *Adaptive Behavior, 13*, 293–310.
3. Steels, L. & Belpaeme, T. (2005). Coordinating Perceptually Grounded Categories: A Case Study for Colour. *Behavioral and Brain Sciences, 28*, 469–529.
4. Dowman, M. (2007). Explaining color term typology with an evolutionary model. *Cognitive Science, 31,* 99–132.
5. Komarova, N. L., Jameson, K. A. & Narens, L. (2007). Evolutionary models of color categorization based on discrimination. *Journal of Mathematical Psychology Psychology, 51*, 359–382.
6. Puglisi, A., Baronchelli, A. & Loreto, V. (2008). Cultural route to the emergence of linguistic categories. *Proceedings of the National Academy of Science, 105,* 7936–7940.
7. Komarova, N. L. & Jameson, K. A. (2008). Population Heterogeneity and Color Stimulus Heterogeneity in Agent–based Color Categorization. *Journal of Theoretical Biology, 253*. 680–700.
8. Cornsweet, T. N. (1970). *Visual Perception*. Academic Press, Inc. N.Y., New York.
9. Shepard, R. N., & Cooper, L. A. (1992). Representation of colors in the blind, color blind, and normally sighted. Psychological Science, 3, 97-104.
10. Farnsworth D. (1949/1957). The Farnsworth-Munsell 100 Hue Test for the examination of color vision. Baltimore: Munsell Color Company.
11. Mantere, K., Parkkinen, J., Mäntyjärvi, & Jaaskelainen, T. (1995). Eigenvector interpretation of the Farnsworth-Munsell 100-hue test. *Journal of the Optical Society of America, 12*, 2237–2243.
12. Cook, R. S., Kay, P. & Regier, T. (2005) The World Color Survey Database: History and Use. In Cohen, Henri and Claire Lefebvre (eds.) *Handbook of Categorisation in the Cognitive Sciences*. Amsterdam and London: Elsevier.
13. Regier, T., Kay, P. & Cook, R. S. (2005). Focal colors are universal after all. *Proceedings of the National Academy of Science, 102*, 8386–8391.
14. Birch, J. (2001). *Diagnosis of Defective Colour Vision*. Butterworth–Heinemann, 2nd Edition.
15. Nelson, J. H. (1938). Anomalous trichromatism and its relation to normal trichromatism. *Proceedings Physiological Society, 50* 661–702.
16. Pokorny, J., Smith, V. C., Verriest, G., and Pinckers, A. J. L. G. (1979). *Congenital and Acquired Color Vision Defects*. Grune & Stratton, New York.
17. Wyszecki, G. & Stiles, W. (1982). *Color Science: Concepts and Methods, Quantitative Data and Formulas (Second Edition)*. New York: Wiley.
18. Sharpe, L. T., Stockman, A., Jägle, H., & Nathans, J. (1999). Opsin genes, cone photopigments, color vision, and color blindness. In K. R. Gegenfurtner & L. T. Sharpe (Eds.), Color vision: From genes to perception (pp. 3–51). Cambridge: Cambridge University Press.
19. Narens, L., Jameson, K. A., Steingrimsson, R. & Komarova, N. L. (In–Preparation). Language, Categorization and Convention.
20. Webster, M. A. & Paul Kay, P. (2007). Individual and Population Differences in Focal Colors. In *Anthropology of Color*, R. E. MacLaury, G. V. Paramei and D. Dedrick (Ed.s). Amsterdam: John Benjamins. 29-53.
21. Sayim, B., Jameson, K. A., Alvarado, N. & Szeszel, M.

[15] Such as categorization symmetry breaking caused by regions of increased salience or observers' common confusion regions.

K. (2005). Semantic and Perceptual Representations of Color: Evidence of a Shared Color–Naming Function. *The Journal of Cognition & Culture, 5*, (3–4), 427–486.
22. Bimler, D. Kirkland, J. & Jacobs, R. (2000). Colour-vision tests considered as a special case of multidimensional scaling. *Color Research and Applications, 25*, 160–169).
23. Farnsworth-Munsell Scaling Software (1997). Version 2.1. Copyright 1997. MacBeth Division of Kolmorgen Corporation.
24. Birch, J. (2003). Extreme Anomalous Trichromatism. In *Normal & Defective Colour Vision.* J.D. Mollon, J. Pokorny, & K. Knoblauch (Ed.s). Pp. 364–369.
25. Farnsworth D. (1943). The Farnsworth-Munsell 100-Hue and Dichotomous Tests for color vision. *Journal of the Optical Society of America, 33*, 568–578.
26. Dain, S. J. (2004). Clinical colour vision tests. *Clinical and Experimental Optometry, 87(4-5)*, 276–293.
27. Paramei, G. V., Bimler, D. & Mislavskaia, N. O. (2004). Colour perception in twins: Individual variation beyond common genetic inheritance. *Clinical and Experimental Optometry, 87*, 305–312.
28. Pokorny, J., & Smith, V. C. (1977). Evaluation of single-pigment shift model of anomalous trichromacy. *Journal of the Optical Society of America, 67*, 1196–1209.
29. Smith, V. C., Pokorny, J. & Swartley, R. (1973). Continuous hue estimation of brief flashes by deuteranomalous observers. *American Journal of Psychology, 86*, 115-131.
30. 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.
31. Jameson, K.A., Bimler, D. & Wasserman, L. M. (2006). Re-assessing Perceptual Diagnostics for Observers with Diverse Retinal Photopigment Genotypes. In *Progress in Colour Studies 2: Cognition.* Pitchford, N.J. & Biggam, C.P. (Ed.s),(Pp. 13–33). Amsterdam: John Benjamins Publishing Co.
32. Nathans, J., Thomas, D. & Hogness, D. S. (1986). Molecular genetics of human color vision: the genes encoding blue, green and red pigments. *Science, 232*, 193–202.
33. Nathans, J., Piantanida, T. P., Eddy, R. L., Shows, T. B. & Hogness, D. S. (1986). Molecular genetics of inherited variation in human color vision. *Science 232*, 203–210.
34. Asenjo, A. B., J. Rim & Oprian, D. D. (1994). Molecular determinants of human red/green color discrimination. *Neuron, 12*, 1131–1138.
35. Yokoyama, S. & Radlwimmer, F. B. (1998). The "Five-Sites" Rule and the Evolution of Red and Green Color Vision in Mammals. *Molecular Biology and Evolution, 15*, 560–567.
36. Jordan, G. & Mollon, J. D. (1993) A study of women hererozygous for colour deficiencies. *Vision Research, 33*, 1495–1508.
37. Hood, S.M., Mollon, J.D., Purves, L. & Jordan, G. (2006). Color discrimination in carriers of color deficiency. *Vision Research, 46*, 2894–2900.
38. Jameson, K. & D'Andrade, R. G. (1997). It's not really Red, Green, Yellow, Blue: An Inquiry into cognitive color space. In *Color Categories in Thought and Language.* C.L. Hardin and L. Maffi (Eds.). Cambridge University Press: England. 295–319.
39. Jameson, K. A. (2005). Culture and Cognition: What is Universal about the Representation of Color Experience? *The Journal of Cognition & Culture, 5*, 293–347.
40. Regier, T., Kay, P. & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Science, 104*, 1436–1441.
41. Garner, W. R. (1974). *The Processing of Information and Structure.* Lawrence Erlbaum Associates: Hillsdale, New Jersey.
42. Jameson, K. A. (2005). Sharing Perceptually Grounded Categories in Uniform and Nonuniform Populations. Commentary on Steels, L. & Belpaeme, T. (Target Article). Coordinating Perceptually Grounded Categories through Language. A Case Study for Colour. *Behavioral and Brain Sciences, 28*, 501–502.
43. Shevell, S. K. & He, J. C. (1997). The visual photopigments of simple deuteranomalous trichromats inferred from color matching. *Vision Research, 37*, 1115-1127.
44. Rushton, W. A. H., Powell, D. S., White, K. D. (1973). Pigments in anomalous trichromats. *Vision Research, 13*, 2017–2031.
45. Birch, J. (1989). Use of the Farnsworth-Munsell 100-hue test in the examination of congenital colour vision defects. *Ophthalmological Physiological Optics, 9*. 156–162.
46. Paramei, G. V. (1996). Color space of normally sighted and color–deficient observers reconstructed from color naming. *Psychological Science, 7*, 311–317.
47. Romney, A. K. (2008). Relating reflectance spectra space to Munsell color appearance space. *Journal of the Optical Society of America A, 25*, 658–666.
48. Craven, B. J. (1993). A Second Box-End Scoring Artifact in the Farnsworth-Munsell 100-Hue Test. *Investigative Ophthalmology & Visual Science, 34*, 503–506.
49. Bimler, D. Kirkland, J. and Jameson, K. A. (2004). Quantifying Variations in Personal Color Spaces: Are there Sex Differences in Color Vision? *COLOR Research & Application*, 29(2), 128–134.
50. Bimler, D., & Kirkland, J. (2002). Sex differences in color vision and the salience of color-space axes [Abstract]. *Journal of Vision, 2*, 28a, http://journalofvision.org/2/10/28/, doi:10.1167/2.10.28.
51. Victor, J. D. (1988). Evaluation of Poor Performance and Asymmetry in the Farnsworth-Munsell 100-Hue Test. *Ophthalmologica, 168*, 128–131.