MBS TECHNICAL REPORT 18-05

Further evolution of natural categorization systems: A new approach to evolving color concepts

Maryam Gooyabadi[1,*], Kirbi Joe[1,*], and Louis Narens[1,*]

[1] *Institute for Mathematical Behavioral Sciences, University of California, Irvine*

Abstract

A dynamic model of language evolution is applied to the naturally occurring color categorizations of 110 linguistic communities taken from the *World Color Survey* (Kay, Berlin, Ma, & Merrifeld 1969) in order to assess the stability of their respective color categorizations. The evolutionary dynamics is modeled after human communication and specified by the discrimination-similarity and 2-player teacher games (Komarova, Jameson, & Narens, 2007, "Evolutionary models of color categorization based on discrimination," *Journal of Mathematical Psychology, 51,* 359–382), where color-naming systems are evolved to stable equilibria through agent interactions. This approach remedies the sparseness of empirical, diachronic data – in some cases impossible to attain – that would be ideal for studying natural evolution trends in real human communities by broadly approximating such evolutionary processes. Results suggest that our simulations uphold the original integrity of each community's categorizations – did not impose an external structure – but are still able to evolve each system to a stable equilibrium. Further, our simulation-based approach offers insights in the relative maturity of color concepts within each linguistic population that was previously lacking from analyses of static data. The stability of color categorization can be used as a criterion for grouping different language communities together for the purposes of cross-population analysis. More broadly, we demonstrate that dynamic models initialized with static data can provide valuable insight and have real implications for research across a large variety of fields.

Key words: Simulations, Evolution of meaning, Color Categorization, Evolutionary Dynamics, Naming Systems, World Color Survey, Learning, Agent-based models, Evaluation of stability, Non-Nash equilibrium, Probabilistic models

1 Introduction

1.1 History of Color Naming

A major area of interdisciplinary study concerns how words and signals acquiring meaning. This is a central and heavily investigated issue in the humanities (philosophy and linguistics), the social sciences (psychology and anthropology), and engineering and computer science (robotics and artificial intelligence). This article investigates a special case: the evolution of color concepts in languages of non-industrialized, isolated linguistic communities.

Color naming has a long history in academia, beginning with seminal work on ancient Greek color terminology by Gladstone in 1858 [7] that was extended by other 19th century researchers to additional ancient languages. The subject got a great boost in 1969 by the seminal study of Berlin & Kay's [2], *Basic Color Terms: Their Universality and Evolution,* and later by the the *World Color Survey* (or WCS for short) by Kay, Berlin, Maffi, & Merrifeld [15]. The latter surveyed the color naming and categorization behaviors of over 2500 individuals from 110 non-industrialized and isolated linguistic communities. The WCS presented each individual two nam-

* Corresponding authors: Joe, K. (joek@uci.edu) and Gooyabadi, M. (mgooyaba@uci.edu)

ing tasks on 330 specially selected color chips. This produced a famous data set has been analyzed from a number of perspectives using a variety of methodologies. This article provides a new methodology for analyzing data from the WCS based on evolving naming strategies through game theory.

There is a history of using game theory to understand the evolution of meaning in linguistic and signaling systems. There are many approaches to this subject (e.g., Steels, 2003, 2011; Smith, Kirby, & Brighton, 2003; Hoffman & Singh, 2012; Vylder & Tuyls, 2006; Loreto & Steels, 2007; Centola, Barochelli, 2015; Belpaeme & Bleys, 2005), and several have been used for color naming.

We follow this path and consider color naming to be a convention [23], and use the *2-person Discrimination-Similarity* game developed in Komarova, Jameson, & Narens [16] as a way to evolve population color naming strategies. This game's dynamics is based on a form of reinforcement learning that is used by the individual agents. Unlike some other evolutionary color naming models that emphasize supposed properties of human color vision (e.g., the Hering primaries of red, blue, green, and yellow lights having special perceptual properties and salience), the *2-Person Discrimination-Similarity* Game is based only on primitive ideas about communication and the agents' ability to discriminate one colored stimulus from another.

The WCS revealed across linguistic communities a highly restricted pattern of naming strategies. A number of theories have been developed to explain this. The idea of Berlin & Kay and others is that each community is in a particular stage of color naming evolution. The overall evolutionary process of cultural color naming can be understood by analyzing the increasing complexities of the naming strategies in terms of how they partition the color stimulus space into concepts named a small set of "basic color terms" (BCTs).

Our analysis and goal for the WCS data is different. Like Berlin & Kay, we look at each linguistic community to be a particular stage of color category evolution. But, unlike Berlin & Kay, we don't ask how this stage is related to a next, more complex stage, as displayed, for example, by another unrelated linguistic community having one more BCT. Instead, we ask, how well evolved *as a communication system* is the community's naming strategy, and what would *its* future evolution look like if its population would communicate freely about color by repeated play of the *2-Person Discrimination-Similarity* Game? We use individual data from the WCS to model agents in a society, and have them play the communication game repeatedly until an equilibrium strategy (i.e., a color naming convention) arises and remains stable.

This evolutionary approach is particularly useful for two reasons. First, the WCS assumes that monolingual communities which are in the same "stage of evolution" from across the world have equally stable conventions about color and hence compare data across communities. Instead, by evolving each community's data independently to a stable equilibrium, we can evaluate how far each color naming system was from a stable convention and perhaps group communities accordingly–a different grouping criterion for comparison across communities rather than just number of BCTs. Second, the WCS and similar studies are missing data on how linguistic evolution would have happened naturally. To investigate the evolution of naming systems, it would be ideal to use diachronic data which would allow for a real, evolutionary view of each community. However, data of this type is extremely sparse–and now impossible for the linguistic communities that have gone extinct or are composed of bilingual speakers. Therefore, simulations can generate diachronic data through agent interaction based on human communication. Evolving the initial data in this way provides a unique perspective on possible ways evolution of color naming systems evolve, a better approximation than evaluating static data.

1.2 Theories of Color Categorization

Berlin & Kay modeled color so that each color belonged to a category and had a color name. This is reflected in the data they collected for their 1969 book and in the World Color Survey. Each participant was required assigned a name for each color in the stimulus space. This allowed them to partition perceptual color space into color categories, and then identify which of these were considered "basic". However, some argue and present data, e.g., Levinson 2000 [18], that it is possible to have gaps in color conceptual color space. In other words, there could exist regions in the color space for which there are no color concepts or names. This gives rise to two different theories of color term evolution. In the Berlin & Kay setup, because the whole space is named, the introduction of a new term can only arise by splitting an existing concept. This has arisen in the blue-green category in the history of written languages with a single term covering blue and green categories splitting into two terms one for blue and one for green.

Japanese provides a good example of this with the single term *aoi* for blue and green hues that split in the 14th and 15th century A.D. into *aoi* for blue and *midori* for green. It also happen in Russian with splitting of its blue category–originally called *sinij*–into two categories, *sinij* for darker blues and *goluboy* for lighter blues. This method of evolving color categories is called the *Partition Hypothesis.* Levinson

(2000) proposed a different method. In the earliest stages of color evolution, color terminology did not partition the color space, but rather had gaps in it. He used data he collected on the Yélî Dnye as an example to document this. He then theorized that initially, color terms remain focused around certain substances or objects. Over time, these terms became more generalized, and begin to cover more of the color space. Whenever the need arose for new color terms, they were introduced to fill in the gaps. This theory is called the *Emergence Hypothesis.*

1.3 Basic Color Terms

This idea of "basic color terms" (BCTs) originates from the *universalist* perspective of the *linguistic relativity* debate. Universalism states that color cognition is a universal, physiological phenomenon rather than a cultural one. Berlin and Kay popularized the universalist view by exploring universal features of color categorization through a formalization of basic color terms introduced in their 1969 book, *Basic Color Terms: Their Universality and Evolution* [2]. They defined the set of BCTs to be the smallest set of color terms within a language with which a speaker could name every possible color. For example, Berlin and Kay identify 11 BCTs in English: black, white, red, blue, green, yellow, orange, purple, pink, brown, and gray. Their theories supported a universalist perspective suggesting that for all languages there exists: *(i)* universal constraints on possible number of BCTs within a language, and *(ii)* a universal pattern of evolution for the emergence of new BCTs. Data used to support these claims were collected from "published sources and personal communication with linguists and ethnographers who have specialized knowledge of the languages in question" [2] on 98 language groups and survey data from 20 language groups in the San Francisco Bay Area.

However, the empirical data was met with criticism due to the sparse number of participants per language, bilingual nature of participants who spoke English in addition to the target language, and location of data collection–collected in San Francisco rather than the homeland of the target languages. In response to such concerns, Berlin and Kay initiated a multinational survey referred to as the World Color Survey.

1.4 World Color Survey

The *World Color Survey* (WCS) [15] which was undertaken as an effort to verify the claims made in Berlin and Kay's 1969 book *Basic Color Terms* [2], gathered data on the color categorization of 110 monolingual tribes around the world, each with $\sim$24 participants on average. The data [4] was gathered in person by linguistic missionaries who conducted a mapping task and a naming task. The mapping task asked participants to identify color chips, called *focal chips*, that best represented a specific color category. The naming task asked participants to assign names to 330 standardized Munsell color chips, presented one at a time, in a fixed, random order, on a gray background. The data from these two tasks, taken from 108 of the languages included in this survey[1], are the basis for a new methodology to study the evolution of color categorization.

2 Evolutionary game-theoretic tool – ColorSims

In this framework, naming strategies of simulated populations evolve to stationary equilibria as the agents play a simple communication game. A stationary equilibria is defined as a non-Nash stable point that does not change over a long period of time. The communication game asks agents to assign names to color stimuli and "communicate" repeatedly with members of the population till a naming convention arises. The agents in this evolutionary dynamic are endowed with minimal perceptual and learning abilities. The game is deployed using *ColorSims*, a Python-based program developed by S. Tauber based on the communication game specified by Komorova et al. [16]. This platform has been utilized in several past studies to evaluate the evolution of color naming systems in populations of simulated agents on a one-dimensional color space [16, 17, 11, 12, 20].

The simulation framework used in this paper is an extension of *ColorSims* [27] with *(i)* a higher dimensional color space that more realistically approximate human color perception, *(ii)* real observer population data from the WCS, *(iii)* a measure to assess the stability of a converged solution. This version of the software is referred to as *ColorSims 2.0* [8]. We use Colorsims 2.0 to investigate theories of category evolution, such as Emergence Theory or Berlin & Kay's Theory of category splitting, that has not been done before.

2.1 Games

The communication process features two steps. First, agents discriminate between like and unlike stimuli and assign names accordingly; Second, two agents

[1]Language numbers 62 and 93 were omitted from analysis in this paper because of the poor nature of the data collection at the time of the survey. Any future mentions to "all WCS language" refers to the set of languages that excludes languages 63 and 93.

communicate their naming strategies and update or learn depending on the agreement of their assigned names.

2.1.1 Discrimination-Similarity Game

A round of the *discrimination-similarity game* has two players. The object of the round is for the players to a agree on category names for stimuli from a continuous domain, which for this article will be colors, conceptualized as chips each being perceived as a homogeneous color. Two chips are randomly chosen from a particular set of chips. The players will be both rewarded if they name agree on the naming of the chips subject to the following criterion: If the chips are sufficiently close in color similarity they are to have the same name, and if sufficiently far in color similarity different names. Of course, "sufficiently close" and "sufficiently far" must have objective definitions for this to be a useful criterion, and they will be provided later. The idea is capture this by putting an appropriate perceptual metric on the colors and use a parameter, called *k-sim*. We interpret *k-sim* in a manner similar to Komarova et al. (2007) who write,

> $k_{sim}(a, b)$ is interpreted as being related to the utility of categorizing a and b as the same or different colors. It is defined by the environment and the life-styles of the individual agents. It is used to reflect the notion of the pragmatic color similarity of the patches. For instance, suppose one individual shows another a fruit and asks her to bring another fruit "of the same color." It is a nearly impossible task to bring a fruit of a color perceptually identical to the first, because different lighting, different color background and slight differences in fruits' ripeness contribute to differentiating its perceived color from the comparison fruit. Therefore to satisfy "of the same color" of a fruit's ripeness in practical terms, the individual must be able to ignore such unimportant perceptual differences and bring a fruit that is "of the same color" practically. It may also be as important to be able to distinguish ripe, edible, "red" fruit from the unripe, "green" ones.

Komarova et al. (2007) use the discrimination-similarity game to create for a population of artificial agents a communicable categorization system. This is done through a reinforcement learning algorithm. For the color stimuli, they chose colors from a hue circle of colors. We increase the underlying theory so that it applies to the entire Munsell Color Solid while keeping their learning algorithm. This increase will allow us to apply results of repeated play of the discrimination-similarity game to the World Color Survey that employs a more complicated stimulus set that is surface within the Munsell color solid.

At the beginning of each round of play, two agents, Player 1 and Player 2, are randomly chosen from the total population. Both players are presented with a randomly selected pair of colored stimuli, *chips*, and are individually asked to provide a name for each chip. Players assign names to a color chip based on their probability strategy matrix (see Section 2.3.1). Two chips should be considered of the same category and given the same color name, if the two chips are within 1 *k-sim* of each other and assigned different color names if their distance is greater than 1 *k-sim*. The names each player provide for the two chips are compared to the criteria above and players are awarded either a "personal success" or a "personal failure". The round is a "social success" if and only if both players have personal successes and both assign the same names to the same chips, and a "social failure" otherwise.

2.1.2 2-player Teacher Game

The 2-player teacher game is a particular implementation of the discrimination-similarity that updates the naming strategy of each player through a learning dynamic introduced in Komarova et al. [16]. Depending on the outcomes of the *Discrimination-Similarity game* for each player, the players may then engage in a type of reinforcement learning dynamic referred to as the *2-player teacher game* . There are two cases where the 2-player teacher game is played between the players with at least one player having a personal success:

(i) One Player has personal success and the other has personal failure in which case the successful player is is designated as *teacher* and the other as the learner.

(ii) Both Player 1 and Player 2 have personal successes but they assign different names to the respective chips in which case one player is chosen at random as *teacher*.

2.2 Evolutionary Dynamics

The outcomes of the *Discrimination-Similarity* and *2-player teacher game* are used to perform either an updating or reinforcement leaning dynamic through repeated play until a stable naming system arises in the total population.

Suppose chips i and j were presented to Player 1 and Player 2. Consulting his matrix, Player 1 assigns name α to chip i (with probability $P^1_{i,\alpha}$) and β to chip j (with probability $P^1_{j,\beta}$); Player 2 assigns name μ to chip i (with probability $P^2_{i,\mu}$) and ν to chip j (with probability $P^2_{i,\nu}$).

Updating:

(i) Personal failure for Player 1 and Player 2 – Social failure

Player 1 updates as follows:

$$P^1_{i,\alpha} \to P^1_{i,\alpha} - q, P^1_{i,\beta} \to P^1_{i,\beta} - q$$

$$P^1_{i,\sigma} \to P^1_{i,\sigma} + q, P^1_{i,\tau} \to P^1_{i,\tau} + q$$

where q is an arbitrary probability and σ and τ are two independently, randomly selected chips other than α and β.

Similarly, Player 2 updates as follows:

$$P^2_{i,\mu} \to P^2_{i,\mu} - q, P^2_{i,\nu} \to P^2_{i,\nu} - q$$

$$P^2_{i,\sigma} \to P^2_{i,\sigma} + q, P^2_{i,\tau} \to P^2_{i,\tau} + q$$

(ii) Personal success for Player 1 and Player 2 – Social success

Given that there is social success, it must follow that $\alpha = \mu$ and β=ν. Hence, Player 1 and Player 2 update as follows:

$$P^1_{i,\alpha} \to P^1_{i,\alpha} + q, P^1_{i,\beta} \to P^1_{i,\beta} + q$$

$$P^1_{i,\sigma} \to P^1_{i,\sigma} - q, P^1_{i,\tau} \to P^1_{i,\tau} - q$$

$$P^2_{i,\alpha} \to P^2_{i,\alpha} + q, P^2_{i,\beta} \to P^2_{i,\beta} + q$$

$$P^2_{i,\sigma} \to P^1_{i,\sigma} - q, P^2_{i,\tau} \to P^2_{i,\tau} - q$$

where q is an arbitrary probability and σ and τ are two independently, randomly selected chips other than α and β.

Reinforcement learning:

(iii) Personal success for Player 1, Personal failure for Player 2 – Social failure

Player 1 is the role of the teacher and Player 2 the learner. Player 1 updates as follows[2]:

$$P^1_{i,\alpha} \to P^1_{i,\alpha} + q, P^1_{i,\beta} \to P^1_{i,\beta} + q$$

$$P^1_{i,\sigma} \to P^1_{i,\sigma} - q, P^1_{i,\tau} \to P^1_{i,\tau} - q$$

Player 2 learns from Player 1's naming schema and updates as follows:

$$P^2_{i,\mu} \to P^2_{i,\mu} - q, P^2_{i,\nu} \to P^2_{i,\nu} - q$$

$$P^2_{i,\alpha} \to P^2_{i,\alpha} + q, P^2_{i,\beta} \to P^2_{i,\beta} + q$$

(iv) Personal success for Player 1 and Player 2 - Social failure Then a random player is chosen to be the teacher and the two players update their strategies as in case (iii).

The games are repeated until a stable, population-wide naming convention is reached.

2.2.1 Evaluating Solution Stability

The goal of these simulation investigations is to study how naming systems arise and evolve over time. A naming system emerges when a population converges on a naming system as a whole, or when there exists a high level of agreement in category names and partitions among all individuals in the population.

In this paper, a solution is considered to have reached convergence when agent error is minimized, and the optimal number of categories is reached. Theoretically, these categories are roughly equal in size. Though error is minimized, complete agreement is never reached since errors continue to naturally occur at the color boundaries. This is because at the boundaries there exists two chips, a and b, that are within *k-sim* and should be named α but belong to different categories and are named α and β – leading to personal and social failure. Hence, a convergent solution is a non-Nash equilibrium. Instead, the solution is said to be stochastically stable if the population-wide naming system experiences minimal change over a very long period of time.

There are two measures used to determine the stability of a population's color naming system:

1. The level of lexicon agreement across the whole agent population
2. The amount of change in the global agreement level across various time periods

Global lexicon agreement on round r of the game (A_r) is defined as follows [3]:

[2] Other variants have been tried that yield the same result. For example, Komarova et al. (2007) reports the following with the case of $q = 1$:

> A variant of a reinforcement learner has the following update rules: If a categorization fails, then the value of the category associated with the chip decreases by 1 as before, and all other categories are enhanced by the amount $\frac{1}{m-1}$ In case of a successful game, the successful category is strengthened (unless it is full), and at the same time all other non-zero categories are reduced. It turns out that for our purposes, the two types [the type just described for $q = 1$ and the just described enhancement version with increments $\frac{1}{m-1}$ of reinforcement learners] behave similarly. Most experiments have been performed by using the first type of reinforcement learners. (p. 369)

[3] For robustness, an alternate measure of global measure was defined called C_r. This alternate measure of agreement compares naming strategies between pairs of agents whereas A_r calculates agreement based by chip. These measures have been found to be highly correlated (r = 0.91). Therefore, for simplicity we use A_r exclusively. The definition of C_r can be found in Appendix A.

$$A_r = \sum_{i=1}^{320} \frac{P_{i,\alpha}}{N} \tag{1}$$

where r is the game round number, i is the chip number, α is the most frequently used name for chip i, N is population size, and $P_{i,\alpha}$ is number of agents who assign name α to chip i.

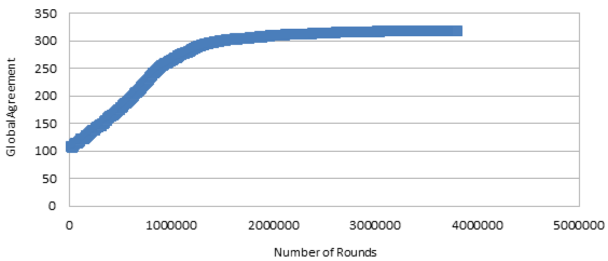

Figure 1: Plot of global agreement agreement measure (A_r) of a population of random agents across the number of rounds played.

The global agreement level is then used to calculate the *Stability Measure*, defined as follows:

$$S_r = \frac{A_r - A_{r-10,000}}{A_r + A_{r-10,000}} \tag{2}$$

Global agreement (A_r) is calculated every 10,000 rounds, so S_r is interpreted as the change in agreement level between rounds r and $(r - 10,000)$, or in other words, the change in agreement level between each "snapshot" of the population.

The simulation marks an r^*, which is the round number at which the stability measure first falls within some predetermined "stability range" (default range is (-0.00175, 0.00175), determined empirically). The simulation will stop if the solution stays stable–S_r stays within the stability range–for another r^* rounds. That is, a naming system is considered to have converged to a stable solution if the system is stable for r^* rounds, resulting in an overall total of $2r^*$ simulation rounds.

Figure 2: Plot of the stability measure (S_r) across numbers of rounds played. We define a "stable" solution to be one which falls within some "stability range" for many rounds. Due to the confusions that agents will have at the boundaries of two color words, the solution will be perpetually subject to minor fluctuations and changes. Thus, we say that a solution which has given a stability measure within some neighborhood of zero for enough rounds is "stable".

We define the S_r in order to understand at which round r the population has converged on a naming system. In the original *ColorSims*, the number of rounds was a parameter given to the simulation. For simulations with "large" parameters, such as populations with many agents or stimulus sets with many chips, determining how many rounds to run the simulation was an educated guess based on trial and error. Naming systems at the end of those predetermined number of rounds were not guaranteed to be stable. However, S_r ensures two things: (i) the system will run for as long as it needs to reach convergence, as determined from S_r falling within a stability range, and (ii) we can check that the solution remains stable over a long period of time, as determined by r^*.

2.3 ColorSims 2.0 Initialization

2.3.1 Agents

The evolution of color categorization is studied by using a learning dynamic on populations of agents. Every agent is endowed with a probability matrix that is updated at the end of each round of interaction. Agents were originally initialized with random probabilities as a basis to test the model and its parameters. Subsequent iterations of the simulation framework initialized agents with real observer data from the World Color Survey.

(i) Initializing Random Agents

Previous work using the ColorSims framework [16, 17, 11, 12, 20] used populations of "random" agents. An agent's probability matrix is organized with colored stimuli, called chips, along the columns and all possible color names along the rows. The matrix is populated with random fractions such that column j of the matrix defines a probability distribution for chip j over all possible color terms. Therefore, the value in cell (i, j) represents the probability that this agent would assign the name i to color chip j. We employ the same process to initialize random agents in our ColorSims 2.0 framework.

When an agent is asked to provide a name for chip j, the name that the agent assigns to chip j is the result of a probabilistic draw over the agent's vocabulary according to the distribution defined in column j of the probability matrix.

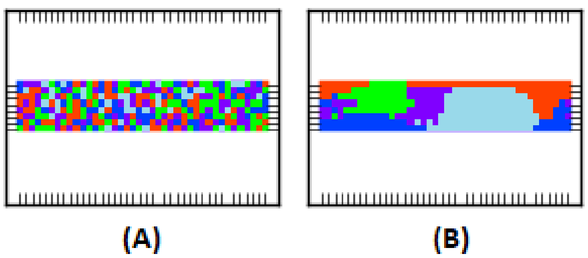

Figure 3: Visual representation of the color naming system of a single agent in a random population at the (A) beginning of the simulation and (B) end of the simulation. The colored rectangle represents the two-dimensional set of stimuli, where each pixel in the rectangle represents a single stimulus chip. The color of each box represents the name which the agent assigns to that given color chip.

Though these random agents are not representative of real observers, they are useful because they allow the evolution of the color naming systems to proceed *de novo*. By initializing random agents, we impose no restrictions on the initial state of the system categorization beyond what is empirically given. Therefore, since agents can reach a shared, stable naming system with minimal assumptions, we can then evaluate how a system that is presumed to be still in the state of evolving, such as those from the World Color Survey, might proceed to stability.

(ii) Initializing Agents with WCS Data

In our simulation-based investigation of color categorization, we initialized some agent populations using all participant data from a single language group. We initialized an individual agent for each participant in the language and populated its probability matrix with the corresponding WCS categorization data from the naming task for a given participant. A cell (i, j) in the agent probability matrix was given an entry of 1 if the participant named chip j with name i or was given an entry of 0 otherwise. These initialized agents represented the data from the WCS for a particular language. The agents' probability matrices were updated with repeated interaction between agents, characterized by the dynamics detailed in Section 2.1.

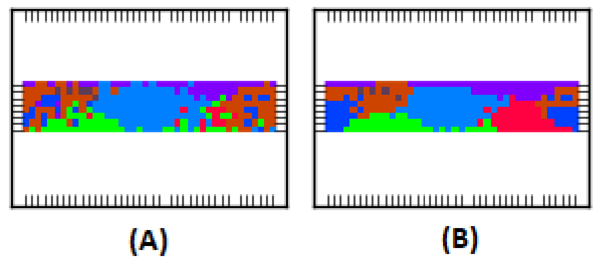

Figure 4: Visual representation of the color naming system of a single participant from WCS Language 16 at the (A) beginning of the simulation and (B) end of the simulation. The naming strategy in (A) represents the actual names that this participant assigned to each of the chips in the grid at the time of the World Color Survey.

2.3.2 Higher-dimensional Color Space

The set of stimuli used in these simulations is a two-dimensional array of colored stimuli. Each individual stimulus in the array is referred to as a color chip. This array is a subset of the standardized set of 330 (320 chromatic, 10 achromatic) Munsell color chips, which is derived from a Mercator projection of a three-dimensional Munsell Book of Color perceptual color space. The set of stimuli used in the simulations is identical to the chromatic component of the stimulus palette used in the World Color Survey [21]. Chips in the grid are organized along the rows according to eight Munsell value (brightness) values and are organized along the columns according to the forty Munsell hue values.

For the purposes of employing a standard color difference metric, each of the chips in this color grid is mapped to its corresponding coordinates in the three-dimensional CIELUV color space, which essentially allows a principled assessment of agents color discrimination judgments based on a standardized uniform perceptual color space. CIELUV was developed by the International Commission on Illumination (CIE) in 1976 with the aim of creating a perceptually uniform space. When presented such stimuli in evolutionary game situations, agents ostensibly "perceive the colors" as 3-tuples (L^*, u^*, v^*), the coordinates of the colored stimuli in CIELUV space. While the set of stimuli is constricted to two-dimensional grid, the underlying space that agents perceive and judge is represented as three-dimensional.

Whereas previous iterations of these simulations utilized a one-dimensional hue space called a hue circle (i.e. a discrete array of color chips arranged according to just-noticeable-differences) [27], the simulation framework used in this article, called ColorSims 2.0 [8], implements three-dimensional perceptual color space in order to create a more realistic representation of how humans perceive color. Most human beings have genes that allow the expression of three retinal photopigment classes–S, M, L. The S-cone peaks in the blue region at 420–440 nm, M in the green region at 534–555 nm, and L in the red region at 564–580 nm. Thus, in principle, all human color sensation can be described as a function of three parameters corresponding to levels of stimulus from these three kinds of cone cells. This biological fact serves as motivation for representing color in three-dimensional models.

The CIELUV color model is appropriate for use here since its approximate perceptual metric permits computation of color difference, or Delta-E, by employing the fact that physical distances between colors map to similar distances in our perception. Similarly, by using a perceptually uniform space, agents can use Euclidean distances between chips in CIELUV space as a basis for judgments of color discrimination when engaged in naming game scenarios used to evolve the population naming conventions of certain chips.

2.3.3 Color Discrimination Measure

The categorizing algorithms in this paper are based on the following idea: colors that are perceptually similar to one another are highly likely to belong to the same category. More specifically, it is based on the following three principles: (i) categorization is important; (ii) to be useful, categorization should attempt to minimize ambiguity, and (iii) when color is a salient or meaningful cue for categorization, two randomly chosen objects that have similar color appearances are more likely to be categorized together than are two objects that have dissimilar colors. These three principles are summarized in the concept of a similarity range, denoted *k-sim* [16].

By definition, *k-sim* is the minimum distance between the color chips for which it becomes important to treat them for pragmatic purposes (and not for perceptual purposes) as different color categories. Since color categorization is important (see principle (i) above), it is beneficial, pragmatically speaking, to assign different category names to colors that are outside the *k-sim* range (principle (ii)), and to assign the same category name to colors that are within the *k-sim* range (principle (iii)).

When the stimulus set is one-dimensional (i.e. a discrete array of chips), distance between color chips i and j is defined as the number of physical chips, or synonymously the number of just-noticeable-differences, that are between i and j. However, when the stimulus set is two-dimensional, we define the distance between color chips to be their perceptual distance. Since the CIELUV space aims to be perceptually uniform, calculating color difference as distances in the physical space will in essence map to similar distances in human perception. Thus, distances between stimuli in the three-dimensional color space are calculated using the standard Euclidean distance metric:

$$d = \sqrt{(L_1^* - L_2^*)^2 + (u_1^* - u_2^*)^2 + (v_1^* - v_2^*)^2} \tag{3}$$

where $d =$ distance between two chips, $(L_i^*, u_i^*, v_i^*) =$ coordinates of chip i in CIELUV space.

For simulations using a one-dimensional color space, Komarova et al. [16] developed a mathematical formulation relating the optimal number of color categories to the total number of chips in the stimulus set and the length of *k-sim* to,

$$C^* = \frac{Q}{\sqrt{2k_{sim}(k_{sim}+1)}} \tag{4}$$

where C^* is the optimal number of terms and Q is the number of chips in the stimulus set. This formula can be used to derive the appropriate *k-sim* to use when initializing *ColorSims 2.0* simulations by setting $Q = 320$ and C^* equal to the number of BCTs identified by Berlin and Kay. Using these values, we can then solve equation 4 accordingly to find *k-sim*. When this formula was developed, the color space was assumed to be one-dimensional and consequently *k-sim* was defined as a distance according to just-noticeable-differences (*jnd*s). Though we are now utilizing multi-dimensional color spaces and measuring *k-sim* according to perceptual distances instead of *jnd*s, we justify using the method to find *k-sim* because the measure of *k-sim* regardless of its underlying metric captures the same thing–how close or far colors are perceptually.

3 Results

To understand the possible theoretical implications of our findings, we first verified that the impact of the simulations were minimal and that it imposed no external structure on the original WCS data. We did so by analyzing categorization change on both a population and participant level (see section 3.1). Given the minimal effects of the simulations, we then explore the possible theoretical implications for the evolution of color concepts and categorizations.

3.1 Simulation Influence

3.1.1 System-wide Change – *global level*

The global agreement level A_r measures the percent correspondence between all individual agents' color naming strategies within a population. A high A_r indicates large population consensus and low A_r indicates a large amount of variability between different agents' individual naming strategies. Due to the nature of the learning and updating dynamics, the global agreement of a population's categorization is expected to increase over the duration of the simulation with language groups that have weak color conventions showing the highest change. To understand this change we define $\%\Delta\ A$ as the amount of change that original naming strategies went through, from the beginning to the end of the simulation, based on A_r:

$$\%\Delta A = \frac{A_{\bar{r}} - A_0}{A_0} * 100 \tag{5}$$

The higher the $\%\Delta\ A$, the higher the number of chips that switched to a different category. The median value for the percent change in agreement level across all WCS languages is 40%, the mean is 44%, and the standard deviation is 23%. The range of percent change extends from 8% to 123%. The level of change for 98.14% of all languages were within one standard deviation from the mean.

The variability in the level of $\%\Delta\ A$ both within each group and between the various language groups of the WCS and is unsurprising. Within-group variation is expected as there is no pragmatic reason for WCS participants to have had high expertise across the entire color space and to place equal importance on all chips–it is unlikely that most individuals would have color names for all 320 chips. Hence, there is a high degree of ambiguity when assigning names to all color chips in the WCS's naming-task.

Even with ambiguity in the color space, it is evident that certain language groups exhibit higher levels of internal consensus than others. The variation between groups is indicative of the stability of color concepts in each group. Starting from the original data, the simulations impose a communication mechanism between agents in such a way that the population converges to a shared, social naming convention. Therefore, languages with very high initial agreement will exhibit very small change in their categorizations as the initial system is "close" to the stable convention to begin with. Conversely, languages with very low initial agreement will exhibit very large percent changes as the color categories are vastly different between agents–a high degree of learning must take place.

While the percent change is useful in quantifying the total amount of change that a language undergoes during the simulation, it provides little information about how these changes impact each agent's underlying categorization. That is, percent change reveals how many chips were assigned to different names in the evolutionary process but provides little intuition about whether these changes are a significant departure from an agent's initial categorization. To gain intuition on this matter, we performed an analysis on the focal chips identified by each participant in the WCS.

3.1.2 Change in Categorization – *local level*

To determine whether the change in naming strategies are significant, we analyze the change to participant's strategy. In the mapping task of the World Color Survey, participants identified focal chip(s) for a set of categories that had previously been elicited by the survey facilitators [15]. A *focal chip*, or *best exemplar*, is a chip (or set of chips) that an individual participant identifies as a best example for a given color word. These chips are those that participants have the highest level of confidence for and find as the most salient examples of a particular color category. The WCS database included the complete set of focal chips that were identified by every participant in every language group surveyed [4].

In our analysis, we investigated the foci chips of each participant as a means to understand the influence of the *ColorSims 2.0* framework. If foci chips remained within their originally assigned category for each participant, over the population, then the influence of the simulation was minimal.

A series of steps were completed: First, focal chips that were located on the achromatic axis of the color grid were excluded from analysis which were not studied in our simulations. Second, only focal chips that were initially identified as focal for a basic color term were considered for analysis. Lastly, chips that were identified as focal for a term x but were assigned a different name y in the naming task were excluded from analysis due to the BCT contradictory nature of these responses.

Using the updated focal chip data, for each participant, we calculated the proportion of focal chips that retained the same category name from the beginning of the simulation to the end. The average of participant focal fractions across the population gave the global measure of focal persistence. Formally, *focal persistence* for language number ℓ's categorization is defined as follows:

$$P_\ell = \frac{\sum_{i=1}^{N} \frac{|S_i|}{|F_i|}}{N} \tag{6}$$

where ℓ is the language number, N is the number of participants in language ℓ, F_i is the set of focal terms for agent i, and $S_i \subseteq F_i$ is the set of focal terms for i which have the same name at the beginning and end of the simulation. The range of P_ℓ is [0,1], where values close to 1 indicate high level of persistence within a language group.

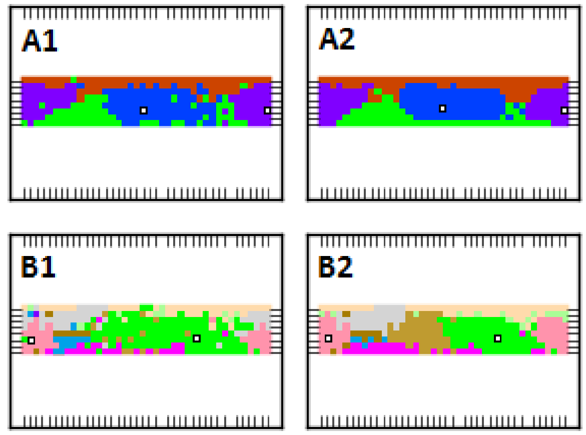

Figure 5: Naming systems for agents in Languages 74 (A*) and 95 (B*) at the beginning of the simulation (*1) and at the end of the simulation (*2) with markers for focal chips identified by those agents in the mapping task of the WCS. Focals are indicated using a white square with a black outline. Language 74 is an example of a language with a small $\%\Delta A$. Language 95 had large $\%\Delta A$. This figure illustrates that despite the fact some languages experience large amounts of change in global agreement as a result of agents converging to a shared solution, the focal chips remain a part of the same category. This suggests that the underlying structure of these languages' naming systems remain the same throughout the simulation.

Based on our analysis, the the median value for *focal persistence* across all of the WCS languages is 0.97, the mean value is 0.94, and the standard deviation is 0.11. This result indicates that, on average, 97% of the focal chips identified for the language's basic color terms retained their original category (name) throughout the simulation – the *important* chips retained their original names while learning took place on the chips that participants were less certain about.

Hence, despite the changes that each language underwent throughout the duration of the simulation (evident from the percent change measures in section 3.1.1), the underlying structure of these languages' naming systems remains the same. In other words, there is support that the communication mechanism in these simulations are not imposing external structure on the original WCS data and are instead improving the categorization across the populations while maintaining important features of agents' categorizations.

This result lends support for the *Emergence Hypothesis*: the focals persist despite large amounts of system change. This may seem contradictory for languages with low initial agreement. Why would foci persist when there seems to be a weak notion of color? In communities where color is not a salient concept, it is reasonable to see highly varied and sometimes inconsistent naming strategies in the WCS color naming-task where participants had to assign names to chips that they had no salient color name for. The mapping-task, on the other hand, allowed participants to choose chips that best represented color categories and they likely chose chips that were highly salient. These salient chips thus served as "anchors" to the category.

In the simulations, as the agents engage in the learning dynamic, they gain expertise in these regions of confusion around the focal chips. Hence, the focal chips for a category persist in the final naming system and learning takes place at the category boundaries. We find that the area of salience and expertise grows outwards from the focal chips to the ambiguous chips.

3.2 Implications for Color Evolution

Emergence theory posits that categories evolve from the center outwards. in support of this theory we found that as a shared naming system emerges in a population, chips that are central to color categories reach full agreement (i.e. all agents in the population agree on the name to assign to that chip) faster than those located on category boundaries. That is, the more central a chip is in a category the more salient it is across the population. Through learning and updating, this salience permeates towards, but also decreases, as we move closer to boundaries. To test this theory, we first develop a methodology to identifying "boundary" chips and track each chip's rate of convergence towards perfect agreement. A high and negative correlation between the likelihood of a chip being on a boundary B_ℓ and how fast a chip moves towards perfect agreement CR_ℓ would provide support for the hypothesis.

3.2.1 Identifying Boundary Chips

The WCS data provides the color categorization of each participant. There is no principled way to aggregating these categorizations into one unified, population-wide naming strategy. In fact, the concept of *boundary* is difficult to define. Hence, we develop a probabilistic approach to identifying category boundaries.

For any chip j in the 320-chip stimulus set, we first calculate the *boundary value* of chip j, which is the proportion of chips within 1 *k-sim* that have the same name as chip j.

$$B_{\ell,i}(c) = \frac{|S_c|}{|K_c|} \tag{7}$$

where ℓ is the language number, i is the numeric ID of the agent, c is the chip number, K_c is the set of chips within 1 *k-sim* of chip c, and $S_c \subseteq K_c$ is the set of chips within 1 *k-sim* of c that have the same name as c according to agent i. This measure increases as a chip becomes more central to a category–the higher the measure, the less likely that chip is a boundary chip.

Let the average *boundary value* for a given chip c across all participants in language ℓ be

$$\bar{B}_\ell(c) = \frac{\sum_{i=1}^{N} B_{\ell,i}(c)}{N}. \tag{8}$$

We then employ the following method to identify the likelihood that chip j is on the "boundary" of a color category. We define $f_\ell : [0,1] \to [0,1]$ to be the probability density function of *boundary values* for language ℓ such that $f_\ell(b) = P(\bar{B}_\ell = b)$, for $b \in [0,1]$. This function is constructed using aggregated frequency data from the *boundary values* of all 320 chips in the stimulus set.

Let $F_\ell : [0,1] \to [0,1]$ be the function that converts *boundary values* to *boundary probabilities* defined by

$$F_\ell(b) = \int_b^1 f_\ell(b)db \tag{9}$$

From this measure, we can then define the probability that a chip is part of a category "boundary" by

$$BP_\ell(c) = F_\ell(\bar{B}_\ell(c)) \tag{10}$$

Therefore, the *border probability* measure, $BP_\ell(c)$, is increasing with the likelihood that a chip is located on the boundary of a category.

3.2.2 Chip Convergence to Full Agreement

We posit that some chips will come to full agreement faster than others. Namely, chips that are central to a category, possibly focal chips, will reach full agreement more quickly than chips that are located on category boundaries due to their ambiguity. Through the measure BP, we can calculate the probability that any given chip is a category boundary. We also track the chip agreement by calculating the proportion of the population that uses the most commonly used name for a given chip. These proportions are calculated for each chip and remeasured at a fixed interval for the duration of the simulation. The proportion change indicates how quickly a chip moves towards full agreement.

The proxy for *convergence rate* of a chip (i.e. the time it takes for a population to reach full agreement on the name for that chip) is the *mean of proportions* as defined below:

$$CR_\ell(c) = \frac{\sum_{i=1}^{\bar{r}} P_{i,\alpha}(c)}{\bar{r}} \tag{11}$$

where $\bar{r}$ is the total number of game rounds, i is a round number, c is a chip number, α is the most frequently used name for chip c, and $P_{i,\alpha}(c)$ is proportion of agents in a population who assign name α to chip c. $P_{i,\alpha}(c)$ equals 1 when the whole population agrees on what name to call c, so the higher value of $CR_\ell(c)$, the sooner in the simulation that chip achieved total agreement.

To evaluate whether there is a relationship between the probability that a chip is a boundary chip and the rate of convergence to full agreement for that chip, we calculate $corr(BP_\ell, CR_\ell)$ across all 320 color chips in the stimulus set. Therefore, for each language we have an aggregated measure of how boundary probabilities relate to convergence time.

The correlations between BP_ℓ and CR_ℓ for all language groups were negative, indicating that chips that

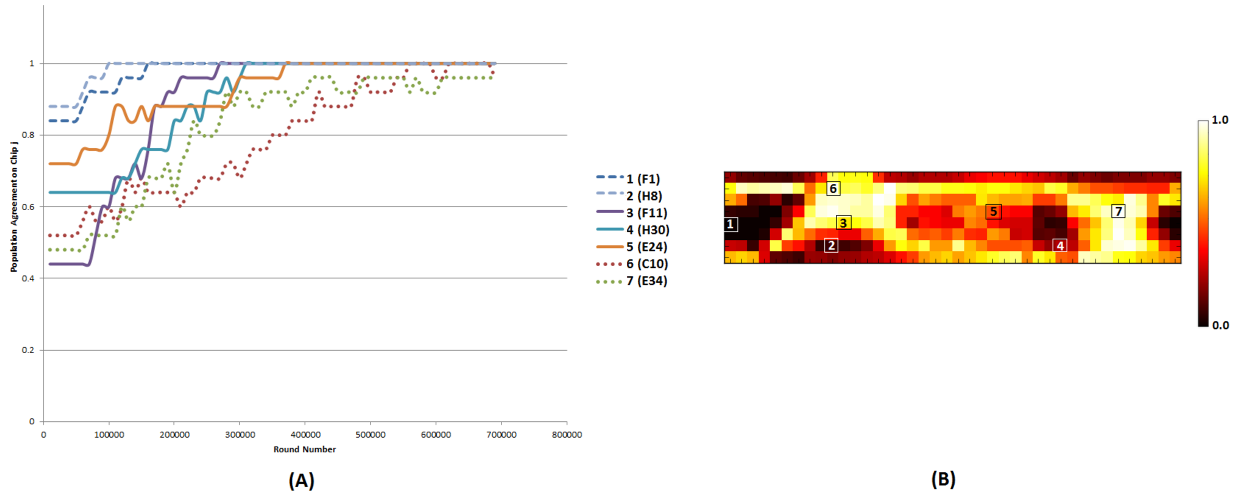

Figure 6: For simulations done on WCS Language 23: (A) Plots of chip agreement ($P_{i,\alpha}(c)$) over time for seven chips. (B) Heat map depicting the boundary probability of each chip, organized on the 320 chip color grid. Boxed numbers on heat map correspond to the chip identified in the plot. From the results depicted in (B), we classify chips 1 and 2 as having *high* border probabilities, chips 3, 4, and 5 as having *moderate* border probabilities, and chips 6 and 7 as having *low* border probabilities. From (A), we observe that *low* border probability chips are the first to reach total agreement ($P_{i,\alpha}(c) = 1$), then the *moderate* border probability chips, and lastly the *high* border probability chips. This suggests that at a social level border chips are the last to be learned, which is consistent with the Emergence Hypothesis

are likely to be on a category boundary are also likely to converge slower – i.e. reach total agreement later in the simulation. Figure 6 presents a specific example of this finding.

Additionally, 88 percent of all languages have a correlation strength of 0.5 and above with close to 50 percent of languages at a strength of 0.7 and above (Figure 7). Therefore, this measure of the relationship between border probability and convergence rate is non-trivial for a majority of the languages tested in this paper.

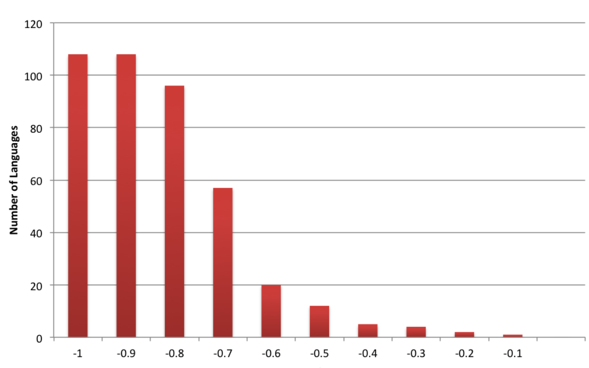

Figure 7: Cumulative histogram of number of languages that have a boundary-convergence correlation measure less than the correlations along the x-axis.

Through the border probability measure, $BP_\ell(c)$, we can obtain an estimate for the geographic location within a color category. The higher the border probability, the more likely that a chip is on a category boundary. The lower the border probability, the more likely that a chip is central to a category. After correlating this measure with the chip convergence rate $CR_\ell(c)$ for all WCS languages, we found that the relationship between these two measures is substantial for all languages and high for most. Therefore, there is evidence that chips located in the center of a category will reach full agreement more quickly than border chips. These results lend support for the *Emergence Hypothesis* of color evolution showing that a society will gain expertise for central chips first and with continuous communication, the expertise will move outwards towards the category boundary.

4 Discussion

This paper was a preliminary exploration of *Color-Sims 2.0* as a tool to analyze the stability of natural color categorizations, using data from the WCS. With minimal perceptual and memory assumptions for agents, we were able to evolve color naming systems, using both randomized and non-randomized agent populations, to a stable solution using the two communication games, *Discrimination-similarity* and *2-player teacher game*. By inducing these systematic interactions, the concept of color was strengthened within a population and was evolved into a more salient, stable concept with high social agreement. As hoped, the impact of these simulations were minimal and imposed very little influence on the original underlying categorizations of the WCS linguistic communities, bolstering our confidence in the findings.

In particular, we observe support for the *Emergence Hypothesis* of color category evolution–namely, foci persistence and the high correlation of chip convergence rates CR_ℓ and boarder probability BP_ℓ. As a central part of their theory about BCTs, Berlin and Kay assumed that the BCTs partitioned color space. This is reflected in WCS's data collection methods by having each participant name each color chip with a BCT from his or her language. This results in each individual assigning a BCT to each chip, and thus provides relevant data for picking a "best BCT" for that chip for the society under investigation. For example, this societal assignment can be done by taking the BCT mode for that chip. There, however, may be little agreement about such "bests," which calls into question the usefulness of such "bests" in actual communicative practice. Similar concerns arise for the selection of a focal chip for a BCTs. [4] This was the main point of Levinson's Emergence Hypothesis.

This Hypothesis starts with color names restricted to colors around certain objects with gaps in the named color space between these colors. Over time, the named colors expand to include nearby colors. If new color terms are needed to described colors occurring in gaps, they get introduced. Because we start of evolutionary modeling with data from WCS, there are no gaps. But some BCTs in some languages have very low agreement, and thus from a social-communicative perspective this can be viewed as a "gap". With this consideration, our evolutionary modeling supports the Emergence Hypothesis to the extent of having foci persistence and the high correlation of chip convergence rates CR_ℓ and broader probability BP_ℓ.

In future research, we will use the stability measure specified and developed in this paper as an effective criterion for the grouping of language communities for cross-population analysis. We also hope to expand the framework to test the robustness of our findings regarding the Emergence Hypothesis and to test other linguistic theories, such as the Partition Hypothesis that is based off of the work of Berlin and Kay.

5 Acknowledgments

This work was in part funded by a University of California Pacific Rim Research Program award, 2010-2015 (Jameson, PI) and a grant for research from the National Science Foundation 2014-2018 (Award: SMA-1416907, K.A. Jameson, PI).

[4] Because of the way BCTs were defined and WCS data was collected, the name assigned to a focal color chip need not be the name of the BCT for that chip. For example, if a society were to assign *crimson* as the name for a chip in the free listing task but *red* emerged from the analysis as the BCT for that chip, and it was determined that *crimson* was a subcategory of *red*, then *crimson* couldn't be a BCT, because, by definition, a BCT cannot be a subcategory of a larger color category. A more general consideration is that given a category C and finding the chip a that is a focal of that category is a different task than given a and asking the participant to assign the best category name for a from a set of categories including C, even if the set only includes BCTs. This kind of asymmetry is discussed in detail in Jameson & Alvarado (2010). In our study it is the likely reason for why across languages a few focals were observed to switch names from free listing to BCT categorization [13].

A Pairwise Agent Comparison Measure

The pairwise agent comparison measure (C_r) is defined to be an alternative method for evaluating the level of agreement of a population's color categorization at a given time. Where as the global agreement measure (A_r) aggregates agreement across the set of color chips (see *Section 2.3*), pairwise agent comparison aggregates agreement across all unique pairs of agents. C_r is formally defined as follows:

$$C_r = \frac{\sum_{i=1}^{N} \sum_{j=i+1}^{N} \sum_{k=1}^{320} [W_{ik} = W_{jk}]}{\frac{(N-1)N}{2}} \tag{12}$$

where i and j are different agents from the population, N is the population size, k is the chip number, r is the number of rounds, W_{ik} is the word that agent i assigned to color chip k, the number of possible pairs of agents is given by $\frac{(N-1)N}{2}$ (see Figure 7), and $[W_{ik} = W_{jk}] = \begin{cases} 1 & \text{if } W_{ik} = W_{jk} \\ 0 & \text{if } W_{ik} \neq W_{jk} \end{cases}$.

References

[1] Belpaeme, T., Bleys, J. Explaining universal colour categories through a constrained acquisition process. *Adapt. Behav.* 13(4), 293–310, 2005.

[2] Berlin, B., Kay, P. Basic color terms: Their university and evolution. *University of California Press*, Berkeley, CA, 1969.

[3] Centola, D., Baronchelli, A. The spontaneous emergence of conventions: an experimental study of cultural evolution. *Proc. Natl. Acad. Sci.* 112(7), 1989–1994, 2015.

[4] Cook, R.S., Kay, P., Regier, T. The world color survey database. In *Handbook of categorization in cognitive science*, pp. 223-241, 2005.

[5] de Vylder, B., Tuyls, K. How to reach linguistic consensus: a proof of convergence for the naming game. *J. Theor. Biol.* 242(4), 818–831, 2006.

[6] Fider, N., Narens, L., Jameson, K.A., Komarova, N.L. Quantitative approach for defining basic color terms and category best exemplars. *JOSA A*, 34(8):1285–1300, 2017.

[7] Gladstone, W. E. Gladstone, W. E. Studies on Homer and the Homeric Age. I. *Oxford University Press*, Oxford, United Kingdom, 1858.

[8] Gooyabadi, M., Joe, K. ColorSims 2.0: An extension to the python package for evolving linguistic color naming conventions applied to a population of agents. *Technical Report Series #MBS18-02.* Institute for Mathematical Behavioral Sciences, University of California, Irvine, Irvine, CA, USA, 2018.

[9] Hoffman, D., Singh, M. Computational evolutionary perception. *Perception* 41, 1073–1091, 2012. (Special issue in honor of David Marr)

[10] Jameson K.A., Komarova N.L. Evolutionary models of color categorization based on realistic observer models and population heterogeneity. *Journal of Vision*, 10(15):26–26, 2010.

[11] Jameson K.A., Komarova N.L. Evolutionary models of color categorization. i. population categorization systems based on normal and dichromat observers. *JOSA A*, 26(6):1414–1423, 2009.

[12] Jameson K.A., Komarova N.L. Evolutionary models of color categorization. ii. realistic observer models and population heterogeneity. *JOSA A*, 26(6):1424–1436, 2009.

[13] Kimberly A. Jameson & Nancy Alvarado The relational correspondence between category exemplars and names. *Philosophical Psychology, 16:1*, 25–49, DOI: 10.1080/0951508032000067752, 2010.

[14] Jameson K.A., Komarova N.L., Tauber S., Narens L. New results on simulated color categorization behaviors using realistic perceptual models, heterogeneous observers and pragmatic communication constraints. *Presentation at The 17th annual meeting of the Cognitive Science Association for Interdisciplinary Learning.* August 2011. Hood River Valley, OR. 2011.

[15] Kay, P., Berlin, B., Maffi, L., Merrifield, W. R., & Cook, R. *The world color survey.* Stanford: CSLI Publications, 2009.

[16] Komarova N.L., Jameson K.A., Narens L. Evolutionary models of color categorization based on discrimination. *Journal of Mathematical Psychology*, 51(6):359–382, 2007.

[17] Komarova N.L., Jameson K.A. Population heterogeneity and color stimulus heterogeneity in agent-based color categorization. *Journal of Theoretical Biology*, 253(4):680–700, 2008.

[18] Levinson, S. C. Yl Dnye and the theory of basic color terms. *Journal of Linguistic Anthropology*, 10, 3–55, 2000.

[19] Loreto, V., Steels, L. Social dynamics: emergence of language. *Nat. Phys.* 3(11), 758–760, 2007.

[20] Narens L., Jameson K.A., Komarova N.L., Tauber S. Language, categorization, and convention. *Advances in Complex Systems*, 15(03n04), 2012.

[21] Regier, T., Kay, P., Cook, R. Focal colors are universal after all. *Proceedings of the National Academy of Sciences of the United States of America*, 102(23), 8386–8391, 2005.

[22] Schwiegerling, J. Field guide to visual and ophthalmic optics. *Spie*, 2004.

[23] Skyrms, B. Evolution of the Social Contract. *Cambridge University Press*, Cambridge, 1996.

[24] Smith, K., Kirby, S., Brighton, H. Iterated learning: a framework for the emergence of language. *Artif. Life* 9(4), 371–386, 2003.

[25] Steels, L. Evolving grounded communication for robots. *Trends Cogn. Sci.* 7(7), 308–312, 2003.

[26] Steels, L. Modeling the cultural evolution of language. *Phys. Life Rev.* 8(4), 339–356, 2011.

[27] Tauber S. ColorSims: A python package for evolving linguistic color naming conventions within a population of simulated agents. *Technical Report Series #MBS15-01.* Institute for Mathematical Behavioral Sciences, University of California, Irvine, Irvine, CA, USA, 2015.

[28] Tauber S., Narens L., Jameson K.A. Evolutionary models of color categorization on networks. *Talk at the annual meeting of the American Society for Mathematical Psychology*, Tufts University, Medford, MA, July 2011.