Evolutionary model of the concepts of 'color chip' and 'color category' in *Evolutionary models of color categorization based on discrimination*

Ragnar Steingrimsson*

Institute for Mathematical Behavioral Sciences, University of California, Irvine, CA 92697-5100

Abstract

Komarova et al. (2007) showed that a color naming system could be evolved by combining perceptual discrimination, basic pragmatic constraints on communication, and simple learning rules. Their system relied on a fixed number of *color chips* and of *color categories*. Here, I construct an evolutionary game theory model where the equivalent concept of the former and the number of the latter are emergent properties of the learning algorithm. The solution involves not a static but a continuously evolving "perceptual space," where the continuous change is arguably a more realistic evolutionary model and echo psychological models in that it is based on a form of memory, possessing a dynamic learning feature. It is shown that such dynamic structure does lead to a stable structure of both color concepts and categories. Simulation results are aimed at replicating the major result of Komarova et al. (2007) both for individual learning and in a population of learners. The concept of *icon chips* is introduced whose number and stability in an *icon structure* are an emergent property of multiple exposures to stimuli (colors) drawn from a continuous color space. The icon chip represents a form of a memory that changes in response to novel stimuli (color) and to the distribution of those stimuli. The results replicate those of Komarova et al. (2007) using constructs they only assumed could be evolved and demonstrate that a stable macro-structures can be maintained with continuously evolving microstructures, arguably a realistic feature evolutionary learning models.

Keywords: icon chip, icon structure, evolutionary game theory, discrimination-similarity game, color, categorization, evolutionary psychology, computational modeling

---

Color categorization and naming phenomena is a widely researched example of the general phenomenon of the cognitive representation of the perceptual

*ragnar@uci.edu

world. Formalizations of the general processes by which humans partition and label their everyday experiences. How such processes vary can give significant scientific insights into individual concept formation and the sharing of cognitive constructs across individuals in groups (e.g., Batchelder and Romney 1988, 1989).[1]

Historically, color categorization and naming research has relied primarily on empirical investigations of color languages and categorization behaviors across a wide range of ethnolinguistic groups (e.g., Kay & Regier 2003) as well as the statistical modeling of such data (e.g., Regier, Kay & Khetarpal, 2007; Lindsey & Brown 2006).

Recent approaches have extended these efforts by adding formal modeling of color categorization behaviors with computational investigations that simulate the categorization phenomena (Belpaeme & Bleys, 2005; Steels & Belpaeme, 2005; Griffin, 2006; Dowman, 2007; Komarova, Jameson & Narens, 2007; Puglisi, Baronchelli & Loreto, 2008; Komarova & Jameson, 2008; Jameson & Komarova, 2009a,b).

A reasonable question to pose is what formal, computational models, that by themselves contain no "psychology" may contribute to the study, here, of color categorization and naming behavior. Obviously, absent a time-machine, our ability to perform a diachronic analysis of color terms is severely limited, in practice, to cultures with written language, of which there are none which are old compared to the emergence of human language. Using evolutionary game theory, it is possible to simulate hypothesized behavior that over time has given rise to observed categorization or naming structures. In this sense, simulations provide proofs-of-concepts. Another use of simulations is that of constructing a minimal set of (ad hoc) assumptions or rules that suffice to mimic observed categorization outcome. In those case, it is possible to identify extant assumptions that may or may not be needed and, in that way, direct empirical research. One last example is that of ethnolinguistic groups that evolve their own categorization systems in isolation but then for reasons of, say, trade must evolve a common communication system; this is another situation to which simulations of the current kind may be applied.

Specifically as relates to color categorization, a benefit of formal modeling and simulation approaches is that they make possible the investigation of time-dependent aspects of color categorization and naming that can not be easily studied with empirical investigations alone. For example, dynamic aspects of the processes involved in color category learning by individual, and color categorization system development and evolution in populations (which requires many years for human populations to develop) can be investigated using formal models of behavior that are implemented in artificial agent populations.

However, there are inherent challenges in color categorization simulation approaches concerning how best to model, for example, individual color category

[1]I am particularly indebted to K. Jameson for giving generously of her knowledge of the color literature for the writing of this introduction.

learning and inter-individual communications about color experiences. One approach to such modeling is to start with a simplest-case scenario, assuming the minimum regarding both features intra-agent learning and color discrimination and features inter-agent communication. This investigative approach to color category simulation was introduced by Komarova et al. (2007).

Komarova et al. (2007) presented evolution-game theoretical simulation techniques to begin exploring the various "...factors that contribute to the universality of color categorization across individuals and cultures" and showed that "the combination of a minimal perceptual psychology of discrimination, simple pragmatic constraints involving communication, and simple learning rules is enough to evolve color-naming systems" (Abstract).

Komarova et al. (2007), as well as in subsequent work (Komarova & Jameson, 2008; Jameson & Komarova, 2009a, 2009b), discuss in detail both the color stimulus space and the form of category learning used. In all cases, a hue circle is used as a natural perceptual subspace of the color appearance solid. They argue the widely held view that a hue circle is a justified subspace since it preserves similarity relations among hue categories regardless of variation in salient hue points across normal individuals. Further, the use of the hue circle capitalizes on a wealth of existing human color perception data for the purpose of modeling artificial agent observer groups. Similar to earlier research, here "color space" simply implies a three dimensional structure that contains perceptual dimensions of hue, saturation, and brightness in their usual configuration (see Komarova et al. 2007, Fig. 1).

As with observed human categorization behaviors, color categories are conceptualized as perceptual partitions of some color subspace that are represented (to a varying degrees) by color names in languages. Typically the perceptual domain from which a color can be sampled is extensive enough to be described as practically infinite. Human perceptual color space typically is defined using descriptions of attributes of perceptions of real colors in the real word. The maximum number of colors in human perceptual color space has been estimated to be about six million (Chapanis, 1965), although the maximum number of colors in human cognitive color space may be no more than thirty (Derefeldt and Swartling, 1995), and often fewer are used in language.

In the algorithm presented by Komarova et al. (2007) and the several subsequent extensions (Komarova & Jameson, 2008; Jameson & Komarova, 2009a, 2009b), practically infinite color space has been abstracted away by assuming a fixed number of color stimuli and a fixed initial (maximal) number of color categories. Specifically, they noted that

> [i]t is reasonable to consider cases where the number of stimuli is so huge and diverse with respect to the number of signals that agents experience only a small fraction of the possible stimuli. Our algorithms, as currently formulated, do not apply to such cases, because they require each chip to be updated, usually a large number of times. One approach to extending the algorithms to cover these cases is to evolve for each name an icon chip. Intuitively an icon

chip approximates one feature of human long-term memory in the naming of a newly presented color chip. (Komarova et al., 2007, p.377)

In short, the present article examines the rationale for these modeling assumptions. In the process, some of the limits and generalizability of the assumptions are explored some new variants of those assumptions are investigated to illustrate novel forms of modeling that are appropriate for these phenomena.

More specifically, to capture this nearly infinite property of perceptual color space and its more restricted cognitive analogue, I develop the idea of *icon chips* (or *icons* for short), which is loosely defined as an internal representation of color (long-term memory) that is the strongest activated memory for a given color exemplar. This concept replaces that of a *color chip,* which in Komarova et al. (2007) is a static pre-defined entity representing color.

A collection of icons is referred to as an *icon structure* and is a dynamic entity (as is memory) in that it is continuously evolving as new stimuli are presented. The structure will serve the purpose that the fixed color chip did in Komarova et al. (2007). Following the evolution of an icon structure, the next task is to examine whether a stable categorization structure can be evolved on top of this icon structure within a single learner, as well as whether it can be extended to the more realistic situation of a population of learners. Both of those categorization tasks must be accomplished by not assuming any existing categories. Thus, the primary aims of the article are the following:

- Develop an algorithm for evolving an icon chip structure.
- Implement categorization algorithms for a single learner as well as a population of learners that are based on an icon chip structure and, for both, remove the procedural need for an initial number of assumed categories.[2]
- Compare the results with those Komarova et al. (2007), a replication of which is equivalent to having demonstrated that the foundation of their main algorithm is sound.

It is important to note that while terms such as colors, memory, and stimuli are used, and at times, inspiration is sought from what is known or hypothesized about human perception, the modeling itself is purely computational and is not as cannot directly inform on psychological realities. In that context, what is meant by more or less realistic modeling, is simply how easy it is to find analogues between the computational choices and some well-known psychological structures.

The article is structured as follows: Section 1 contains three subsections which in turn deal with building an algorithm for an icon chip structure; building

[2]The Komarova et al. (2007) approach started from a specified (maximal) number of possible color categories, categories could fall into disuse. The latter feature is replicated here as well.

an algorithm for color categorization using an icon structure and starting with no categories for a single learner; building an algorithm to extend the single learner situation to a population of learners. The results are summarized in Section 2. Finally, there are several Appendices, each of which presents actual computational implementation of the designed algorithms.

1. Evolving color categories using evolved icon chips

Three separate algorithms and simulations are presented. In Section 1.1, an algorithm for evolving an icon chip structure is developed. Section 1.2 focuses on evolving a categorization algorithm for a single learner that, through the use of an icon structure, can handle an arbitrary stimulus starting with no pre-defined categories. Finally, in Section 1.3, the single learner algorithm is extended to a population of learners. Each section is organized into the format of Introduction, Method, and Results and Discussion.

1.1. Evolving an icon chip structure

With the caveat of the introduction in mind (what is carried out computationally is not a comment on how any human may actually accomplish the analogue task) it may nevertheless be instructive to seek inspiration from what we understand the human reality to be. It is a fact that millions of distinct colors are potentially available to human observers, and human memory does not as practical matter store this effectively infinity of stimuli. Therefore, whether for a human or a machine, computationally keeping track of an unbounded sample of colors is a relatively complicated affair. To deal with this computational complexity Komarova et al. (2007) used the abstraction of a finite number of discrete color chips as stimuli. This is an abstraction that humans must solve differently in that they live in a richer color environment. The concept of a human long-term memory construct of a category exemplar serves as an inspiration for admitting an unbounded sample of color stimuli without a consequent increase in computational complexity. In this framework, a sample of colors, each "sufficiently" similar, are remembered as samples of the same exemplar. This feature allows a color to be remembered without a new memory having to be created for every single stimulus encountered in the world. Here such an exemplar is an icon chip and a collection of those chips is an icon chip structure.

A priori, it is not outlandish to imagine that, e.g., pre-verbal infants have somewhat evolved memory systems of colors, and that later on in the development process they are able to use such memory constructs to form color names. While this view is supported by, e.g., Gwynn & Baker-Ward (2007), I neither wish to, nor can enter into the question of pre-infant development. Hence, the thought is inspirational only. In a manner parallel to this idea, the starting point is the evolution of an icon structure before any category names are attached to any of the icons, but the icon structure itself should never stop evolving as presumably novel stimuli are encountered throughout a life-time.

1.1.1. Icon chip structure: Method

Here, the concepts of an icon chip and an icon structure will be developed. Every color will be associated with one of these icons and, together, these icons will form a structure that represents the organization of (a one-dimensional) color space. The following conventions and definitions are adopted:

- Let the color circle stimuli be represented by the real interval $[0, 1[$ mapped onto a unit circle. (Note: measuring distances between points on this interval as if they were on a circle, requires an extra step. This is ignored in the description).
- Let $c \in [0, 1[$ represent a random color.
- Let i_j be an element of $[0, 1[$, $j = 1, 2, ...n$, called an *icon chip* such that, numerically, $i_{j-1} < i_j$.
- Let k_{mem} be a constant with the characteristic that if $c \in [k_{\text{mem}} - i_j, i_j + k_{\text{mem}}]$, then i_j is considered an icon chip for c.[3]

The conceptual novelty is that of k_{mem}, which represents the longest "distance" between a stimulus and an icon such that it be considered represented by that icon. The inspiration is of long-term memory, where k_{mem} might be thought of as strength of memory activation and the memory the icon; this resembles Hering's (1878/1964) concept of a *memory color*.[4]

Apart from k_{mem}, it is assumed that the learner is able to distinguish any two chips, c and c', from each other—this is akin to human perceptual discrimination ability.[5]

The learning algorithm is divided into three steps, A_1–A_3.

A_1 Pick, uniformly, a $c \in [0, 1[$ (this is the "color" stimulus).

A_2 Ask: is there an i_j such that $c \in [k_{\text{mem}} - i_j, i_j + k_{\text{mem}}]$. If the answer is

No Then, if $j = 0$ (no icon chip exists) make $i_1 = c$, otherwise create a new icon chip $i_t = c$ and insert it into the group of existing icon chips such that $i_{t-1} < i_t < i_{t+1}$.

[3]Here k_{mem} is considered a constant, but it need not be: it is perfectly plausible that memory based discrimination of colors varies across the hue circle if some colors have more perceptual salience or even pragmatic or social utility than others. Such considerations are not addressed here.

[4]Adams (1923) quoting Hering, writes, "The color in which we have most often seen an external object is indelibly impressed upon our memory" and that this color, evoked by an object is the "true" color of the object is what Hering calls a memory color. The idea that frequency of experience is a determinant factor in both the concept of a memory color and and an icon chip. Nothing beyond this similarity in ideas should be read into this history.

[5]It might be noted that this ability is stronger than humans have in that human discrimination is accurate only up to a just noticeable difference (JND). Nothing is lost or gained from this abstraction. Also, in practice, a computer has a finite set of numbers so in practice, a JND may be said to be the numerical precision achived by that computer.

Yes Then pick the i_j with the minimum distance from c and update the value of i_j.

Basic Update: Calculate the value $u = w(|c - i_j|)$, where $w \in]0, 1[$ is a weighting factor.[6] If $c < i_j$, $i'_j = i_j - u$, otherwise $i'_j = i_j + u$.

Under the Basic Update rule, $d = (i_{j-1} - i_{j+1})$ can become arbitrarily (positively) small leading to a large number, n, of icon chips. Hence, the update rule needs to consider how close two icons may become.[7]

Addition A: If $c < i_j$ then calculate the maximum adjustment, $a_{\max} = \max\{i_j - i_{j-1} - k_{\text{mem}}, 0\}$, of i_j and take $u' = \min\{u, a_{\max}\}$, i.e., i_j is never adjusted to be closer than k_{mem} to another icon and if it already is, no adjustment is made—the case of $c > i_j$ is analogous.

If the icons have a perfectly uniform distribution, then the minimal number needed to cover the circle is $I_c = 1/k_{\text{mem}}$ (the uniform distribution would leave no space on the circle into which a new icon could be inserted). As the number of icons approaches I_c it becomes asymptotically more difficult to add the final icon, requiring a flexible addition criterion.

Addition B: Let $\beta \in]0, 1[$.[8] Then two icon chips are adjusted to be no closer than βk_{mem}. Obviously, any $\beta < 1$ is softening of Addition A, but a β close to 1 will make any increase of $n > 1/k_{\text{mem}}$ become exceedingly slow. Here β is a constant, but it could be a function of, e.g., the number of instances of c.

Update Rule: The update rule is the Basic Update along with Additions A and B.

A_3 Repeat steps A_1 and A_2 to produce a continuously evolving icon chip structure.

For simulation purposes, the steps A_1-A_3 must be turned into a computational form; this is done in Appendix Appendix A.

1.1.2. Icon chip structure: Results and discussion

Figure 1 shows the results for evolving one icon chip structure. The simulation ran for a predetermined 10^6 iterations with $k_{\text{mem}} = 1/20$, $w = .25$, and $\beta = .8$. Displayed are seven snapshots along with the final results.

---

[6] Excluding 0 and 1 is a practical consideration in that 0 is equivalent to using none of the new information and 1 to using only the new information. In words, w is an unspecified function of the amount incoming information that should be used.

[7] There are a multitude of ways to deal with this undesirable situation. One is "sudden death" where, when $d \leqslant k_{\text{mem}}$, icon chip i_j is simply deleted, i.e., immediately "forgotten." An alternative is to assign strength values to the icon chips such that a chip is eventually "forgotten." Here, a third way is implemented.

[8] Excluding 0 and 1 is a practical consideration in that 0 or less is equivalent to allowing icon chips to move arbitrarily with regard to one another and 1 or larger does not solve the problem with Addition A.

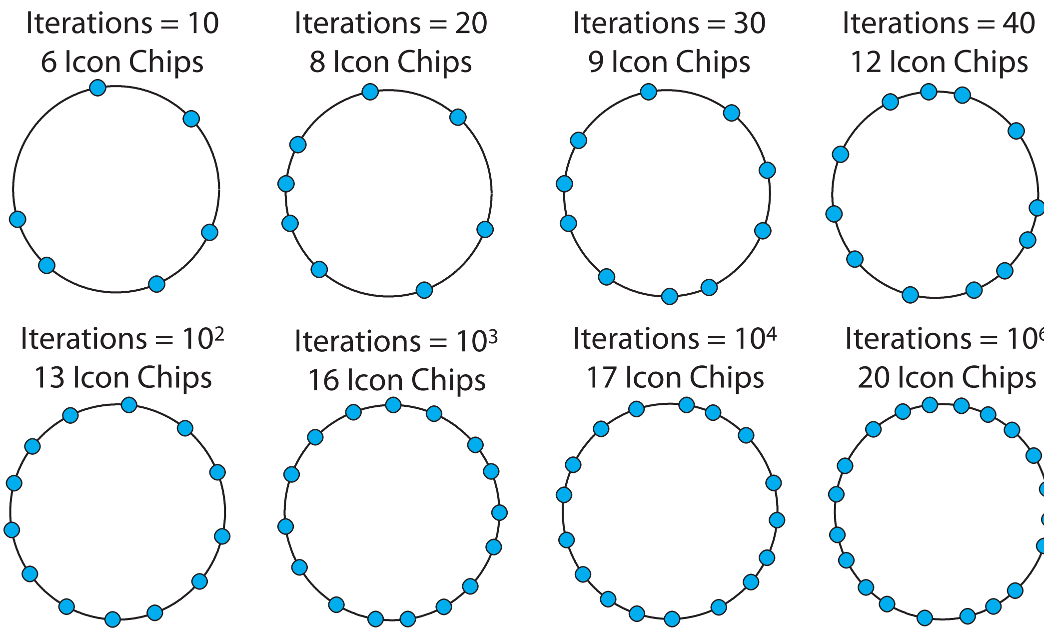

Figure 1: Eight stages of an evolving icon chip structure. Dots correspond to locations of icons on the color circle.

The majority of the icons are generated within the first $\sim 10^2$ iterations, with the final icons taking several orders of magnitude of iterations more to appear. The simulation does arrive at the minimal number of icons, $l_c = 1/k_{\mathrm{mem}} = 20$, needed to cover the space and they are largely uniformly distributed across the representation of a hue circle. The latter feature of the icon chip structure may be considered a desirable one in light of theories relating to the interpoint distance relations among color category best-exemplars (Jameson, 2005; Jameson & D'Andrade, 1997), theories emphasizing the general importance of polar-opposition in category formation (Garner, 1974), findings showing the effects of counter-vailing forces exerted at category boundaries (Jameson & Komarova, 2009a,b), and symmetry breaking results seen when non-uniform sampling of the color space is examined using color space *hot-spots* (Komarova & Jameson, 2008).

Figure 1 shows an icon chip structure formed under uniform sampling of the color space. Real world stimuli may not be so uniform, but a non uniform sampling would not materially alter the results obtained. Were one nevertheless to simulate this scenario, the only alteration needed would be to allow k_{mem} to vary across the hue circle.

However, what is important is that the algorithm does result in an icon chip structure that arrives at relative stability, and yet is continuously evolving.

1.2. Evolving color categories: single learner

The goal is to evolve a stable color categorization for a single learner based on a continuously evolving icon structure within a single learner and to start with no predefined categories. An important consideration is whether the categorization algorithm should start from an empty icon structure or employ a more mature one. Here, again, the inspiration is one current view of human development that pre-verbal infants do evolve a memory system for colors and that they use this system for subsequent naming of colors (Gwynn & Baker-Ward, 2007).

A parallel to this scenario is to assume that the formation of color categories starts from a somewhat stable icon chip structure that continues to develop and evolve in response to the stimuli encountered.

1.2.1. Evolving color categories, single learner: Method

Komarova et al. (2007) investigated a variety of individual learning strategies, of which the most successful was the discrimination-similarity game.

Discrimination-similarity communication game. Komarova et al. (2007) represented colors by a predetermined number of static (i.e., non-evolving) color chips as well a predetermined number of color categories. They introduced the constant k_{sim} which they wrote could be thought of as the "minimum difference between the color chips for which it becomes important to treat them for pragmatic purposes (and not for perceptual purposes) as different color categories." (Komarova et al., p. 363).[9] Obviously, k_{sim} has similarities to k_{mem} from the previous section and it serves some similar purposes for determining the similarity of two color chips as does k_{mem} does in determining similarity of two icons. However, the inspiration for the two clearly differs. In the following, the reader may be helped think loosely of k_{mem} being the distance ideally separates icon chips, whereas as k_{sim} is the distances that ideally separates categories. k_{mem} and k_{sim} are not directly related but somewhat obviously $k_{\text{mem}} \leq k_{\text{sim}}$ for purely logical reason.

The following is the algorithm of Komarova et al. (2007), albeit cast in a different formal language; alterations to the algorithm are given in the following section.

The basic structure for the individual category learning game, two stimuli–color chips–are randomly selected from a stimulus set and are presented to the learner. Based on the learner's existing categorization system, the learner assigns each of the two stimuli a category label, or name. Given two stimuli, c_1 and c_2, the the game can have four outcomes (the words "success" and "failure" mean a desirable and undesirable outcomes, respectively):

1. If $|c_1 - c_2| < k_{\text{sim}}$,

[9] Here k_{sim} is considered a constant, but it need not be: it is perfectly plausible that memory-based discrimination of colors varies across the hue circle if some colors have more perceptual salience or even pragmatic or social utility than others.

   (a) then, if c_1 and c_2 are given the same name, the game is a success and the naming choices are rewarded.
   (b) but if c_1 and c_2 are given different names, the game is a failure and the naming choices are punished while another randomly selected categorization is rewarded.

2. If $|c_1 - c_2| > k_{\text{sim}}$,
   (a) then if c_1 and c_2 are given the same name, the game is a failure and the naming choices are punished while another randomly selected categorization is rewarded.
   (b) but if c_1 and c_2 are given different names, the game is a success and the naming choices are rewarded.

The game is played for a predetermined number of iterations.

Komarova et al. (2007) presented the game in computational terms; here the game is given a generalized description. The following conventions and definitions are adopted:

- Let the color circle be represented by the real interval $[0, 1[$ mapped onto a unit circle. (Note: measuring distances between points on this interval as if they were on a circle requires an extra step. This is ignored in the description.)
- Let $c_i \in [0, 1[$, $i = 1...n$, be equally spaced color chips on the color circle.
- Let κ_j, $j = 1, ..., m$, be a color category which can be thought of as a color word.
- Let $P(\kappa_j|c_i)$ be the probability that the chip c_i be assigned the category κ_j. (How determined, is addressed below).
- Let $k_{\text{sim}} \in [0, 1[$ be a constant with the characteristic that if $|c_a - c_b| \leq k_{\text{sim}}$ then c_i and c_j are considered to belong to the same category κ_k.
- Let $\delta \in]0, 1[$ be the the probability "reward" or "punishment" (as applicable) given to a category.

The algorithm is divided into six steps, B_1–B_6.

B_1 Initialize the game such that $P(\kappa_j|c_i) = 1/m$ for all $i = 1, ..., n$ and $j = 1, ..., m$.

B_2 Pick two color chips, $c_a \neq c_b$ at random where $a, b \in 1, ..., n$. Assign c_a to the category κ_v based on probabilities $P(\kappa_j|c_a)$ and similarly pick a κ_w for c_b.

B_3 If $|c_a - c_b| \leq k_{\text{sim}}$ then the game is a success if $\kappa_v = \kappa_w$ but a failure otherwise. If $|c_a - c_b| > k_{\text{sim}}$ then the game is a failure if $\kappa_v = \kappa_w$ but a success otherwise.

- B_4 If game is a success then if $P(\kappa_v|c_a) = 1$, do nothing. In all other cases "reward" κ_v as follows $P(\kappa_v|c_a) \longrightarrow P(\kappa_v|c_a) + \delta'$ where $\delta' = \min[\delta, 1 - P(\kappa_j|c_a)]$ and $\delta \in]0,1[$ is a constant. Then "punish" a random category $\kappa_t \neq \kappa_v$ as follows: let $\Delta = P(\kappa_t|c_a)$ and if $\Delta \geq \delta' \geq 0$,then $P(\kappa_t|c_a) \longrightarrow P(\kappa_t|c_a) - \delta'$, otherwise $P(\kappa_t|c_a) \longrightarrow P(\kappa_t|c_a) - \Delta$ and distribute the remaining "punishment," $\delta' - \Delta$, in an iterative fashion on available categories. Do the same for κ_w and c_b.
- B_5 If game is a failure then let $\delta' = \min\{P(\kappa_t|c_a), \delta\}$ and $P(\kappa_t|c_a) \longrightarrow P(\kappa_t|c_a) - \delta'$. Then "reward" a random category $\kappa_t \neq \kappa_v$ as follows: if $\Delta = 1 - P(\kappa_t|c_a) \geq \delta'$, then $P(\kappa_t|c_a) \longrightarrow P(\kappa_t|c_a) + \delta'$, otherwise $P(\kappa_t|c_a) \longrightarrow P(\kappa_t|c_a) - \Delta$ and distribute the remaining "reward," $\delta' - \Delta$, in an iterative fashion on available categories. Do the same for κ_w and c_b.
- B_6 Repeat steps B_2-B_6 for the duration of the game.

Algorithm modifications for evolving categories: single learner. Turning to the novel changes: two main changes are necessary. 1. The color chips, c_i, need to be replaced by a continuously evolving icon structure. 2. A mechanism needs to be implemented for creating categories when needed. Item 1 is straight forward: Steps A_1–A_3 (above) involve choosing a c_a, here the change needed is that of choosing an icon, i_i. Item 2 is slightly less clear and the choices made are commented on in the algorithm description.

- C_1 Pick two colors at random and produce the two corresponding icon chips i_a, i_b where $a, b = 1, ..., n$ and n is the current number of icon chips.
- C_2 Assign i_a a category κ_v as follows:
  - (a) If no category is known, then create a category κ_v and set the probability $P(\kappa_v|i_a) = 1$. (Assigning the new category the probability 1 is simply asserting that it is the best knowledge available.)
  - (b) If no category has ever been assigned to i_a, then examine whether any existing category is suitable and pick one of those at random. (Note: if a new category were created each time an icon without a category is encountered, the number of categories created would rise very quickly). A category κ_t is a suitable category for i_a if $|i_a - \bar{\imath}(\kappa_t)| \leq k_{\text{sim}}$ where

    $$\bar{\imath}(\kappa_t) = \frac{\sum_{i=1}^{n} i_i P(\kappa_t|i_i)}{\sum_{i=1}^{n} P(\kappa_t|i_i)},$$

    i.e., it is within k_{sim} of the probability-weighted average $\bar{\imath}$ of the icons that have a non-zero probability of belonging to the category κ_t. If no suitable category is found, create a new category κ_v and set the probability $P(\kappa_v|i_a) = 1$.
  - (c) Otherwise assign to i_a the category κ_v based on the probabilities $P(\kappa_j|c_a)$.

    Similarly pick a κ_w for i_b.

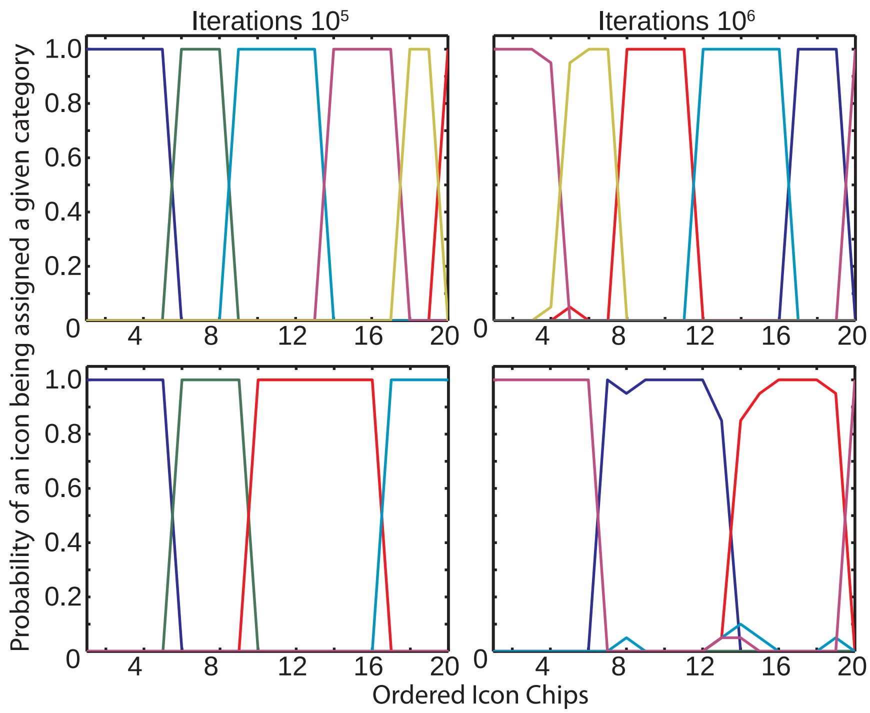

Figure 2: Four separate runs. Starting (and finishing) number of icon chips was $n = 20$, $\delta = 1/20$, and in the first row, $k_{\text{sim}} = 3$ and in the second, $k_{\text{sim}} = 4$.

C_{3-6} Identical to steps B_3-B_6.

Note: each time an icon is picked, the icon chip structure evolves. Hence, both it and the category structure are dynamic at all times.

Appendix Appendix B gives the computational implementation of steps C_1-C_6.

1.2.2. Evolving color categories, single learner: Results and discussion

In the computational implementation of steps C_1-C_3 (Appendix Appendix B), k_{sim} is for practical purposes turned into an integer. This is a computationally simpler implementation and is the choice made by Komarova et al. (2007), the work to which I want to compare.

The simulation was started with an evolved icon-structure, the one shown in Figure 1. This is an icon structure that ran for 10^6 iteration and generated $n = 20$ icons, where $\delta = 1/20$. and simulation was carried out for two instantiations $k_{\text{sim}} = 3$ and $k_{\text{sim}} = 4$ (allowing for direct comparison with the results I seek to replicate). Figure 2 shows four separate simulations.

From Figure 2 it is clear that a fairly regular categorization is achieved in all four cases. For $k_{\text{sim}} = 3$ the number of categories is 5 in both runs and for $k_{\text{sim}} = 4$ the number is 4 and 3, respectively. Komarova et al. (2007, Eq. 8) derived a formula for the optimal number of categories, m_c, namely

$$m_c = n[2k_{\text{sim}}(k_{\text{sim}} + 1)]^{-1/2}. \tag{1}$$

For $k_{\text{sim}} = 3$, $m_c = 4.1$ and for $k_{\text{sim}} = 4$, $m_c = 3.2$. It is not clear whether m_c generally rounds up or down, so the the results are consistent with the calculated expected number of categories and the categorization is achieved. These results compare and replicates those of Komarova et al. (2007), Fig. 12, demonstrating that using an icon structure—and thus an infinity of color stimuli—and starting without any categories, an individual categorization learning is obtained that is in form identical to that obtained by Komarova et al. (2007). The next section explores the same extension for the case of a population of individual category learners.

Additionally, Komarova et al. (2007) used a number of a pre-defined static color chips for what is here a dynamic and evolving set of icon chips. In their algorithm, the placement of a color chip in particular category was reinforced, i.e., the learning involved the placement of a particular chip in a particular category. Here, the reinforcement learning is extended to the color concept itself as a form of a memory construct that is dynamic, evolving in reaction to stimuli. Such learning is arguably a more realistic modeling of this psychological process in that all aspects of the system, from the "remembered" colors, to the number of color categories (increasing and decreasing) are under constant evolutionary pressures, a situation arguably more akin to organic life than are static representations of those entities.

1.3. Evolving color categories: Population of learners

Language is a population code, therefore a natural next extension of the algorithm is to a population of reinforcement learners playing a discrimination-similarity game. The algorithm as presented generalizes the single-learner algorithm to a population of learners; specifically, the discrimination-similarity game (described above) is extended to a game between any two agents at the time, each of which is picked from a larger pool of learners.

1.3.1. Evolving color categories, population of learners: Method

Population of reinforcement learner playing a discrimination–similarity game. In comparison to individual category learners, communication games between agents follows a slightly different scenario. Two agents or learner are chosen randomly from a pool of N agents. The two learners each play the discrimination-similarity game and then communicate to each other their individual outcomes. Each learner can be successful or fail and the naming used by the successful learner is reinforced and individual failures are "punished" further specified below.

In addition to the parameters defined in Section 1.2, the following are also used:

- Let L_k be a learner where $k = 1, 2, ..., N$ and $N \geq 2$ is the number of learners in the population.
- Let $P_{L_i}(\kappa_j|c_i)$ be the probability that the chip c_i be assigned the category κ_j for L_i.

The algorithm is divided into six steps, D_1–D_6.

D_1 For each learner's P_{L_i}, initialize the probability matrix as for a single learner (see B_1).

D_2 Pick two color chips, $c_a \neq c_b$ and two learners, L_c and L_d at random, where $a, b \in 1, ..., n$ and $c, d \in 1, ..., N$. Assign to c_a the category κ_v based on probabilities $P_{L_c}(\kappa_j|c_a)$ and similarly pick a κ_w for c_b. Assign to c_a, c_b, categories κ_s and κ_t, respectively, for L_d in an analogous manner.

D_3 For L_c, if $|c_a - c_b| \leq k_{\text{sim}}$ then the game is a success if $\kappa_v = \kappa_w$ but a failure otherwise. If $|c_a - c_b| > k_{\text{sim}}$ the game is a failure if $\kappa_v = \kappa_w$ but a success otherwise. And similarly for L_d.

D_4 There are four combinations of successes and failures for the learners L_c and L_d, with four associated actions.

   A. L_c fails, L_d is successful: L_c learns from L_d; in that case, L_d is called the teacher.
   B. L_c is successful L_d fails: L_d learns from L_c; in that case, L_c is called the teacher.
   C. L_c is successful, L_d is successful: who serves the role of teacher is picked at random.
   D. L_c fails, L_d fails: each learner does individual update.

D_5 For case A, L_c learns from L_d, L_c performs a successful update for the categories picked by L_d (κ_s and κ_t) in accordance with B_4, and performs an unsuccessful update for its own picked categories (κ_v and κ_w) in accordance with B_5. L_d performs a successful update for its own picked categories in accordance with B_4. The learning situation for cases B and C are analogous.
   For case D, each learner updates according to the individual reinforcement learner algorithm (B_5).

D_6 Repeat steps D_2-D_6 for the duration of the game.

Algorithmic modifications for evolving categories: population learning. In addition to the modifications introduced for evolving categories for a single learner (see steps C_1–C_6), two additional modifications are needed.

The first modification comes from the observation that a learner may encounter categories used by other learners about which the learner has no information. Thus, if a learner encounters an icon chip for which it has no suitable category, the learner may ask if it has previously encountered a category but is not using it. If that is the case, the learner picks at random one of those categories. Only if this search fails does the learner create a new category.

From an algorithmic point of view, this ensures that a category created by one learner can be used by another learner. In motivating the idea, one may imagine an individual learning a language who encounters a color that s/he does not know how to name. One such individual may attempt to name it with a color word that was previously encountered (albeit the meaning of which is still not fully understood), rather than, say, simply make up a new word for the color. This is certainly a practical and perhaps even realistic approach to color category learning. The formalization of this fuzzy vocabulary modification simply entails keeping track of words each individual encounters during the course of the game.

The second modification is that in the course of the game, a learner L_k may be taught a new category κ_t by a observing another learner. In that case the probability for a given icon i_i is updated as $P_{L_k}(\kappa_t|i_i) = 1$. This implements a form of imitation dynamic as discussed in Komarova & Jameson (2008).

For a computational implementation see Appendix Appendix C

1.3.2. Evolving color categories, population of learners: Results and discussion

As for the computational implementation of the single learner, in the computational implementation of steps D_1-D_6 (Appendix Appendix C), k_{sim} is for practical reasons implemented as an integer.

Results for icon chip structure population categorizations are shown in Figure 3. Figure 3 shows two runs where the starting (and finishing) number of icon chips was $n = 20$, $\delta = 1/20$, $k_{\text{sim}} = 4$, and the number of learners $N = 10$.

From Figure 3, in both runs A and B, 6 categories were created in the course of the game but only 3 ended up being used. In all cases, the 10 players achieved a good agreement on the categorization.

From (1), the calculated optimal number of categories for $k_{\text{sim}} = 4$ is $m_c = 3.2$. Thus the results are consistent with the expected number of categories. This result compares to and replicates those of Komarova et al. (2007), Figs. 6 and 7.

2. Discussion and summary

Komarova et al. (2007) showed that "the combination of a minimal perceptual psychology of discrimination, simple pragmatic constraints involving communication, and simple learning rules is enough to evolve color-naming systems" (Komarova et al., 2007, Abstract). They assumed a system with a fixed number of color chips and a fixed number of color categories.

They explicitly realized that they neither addressed how categories could come into being (although they could fall into disuse in their game), nor how a world filled with a practically infinite number of colors could be reduced to a fixed set of color chips. For the latter problem, they discussed the concept of an icon chip as a sort of evolved proxy for a color chip. In general, they assumed that these restrictions were not unreasonable.

Another issue concerns the fact that in Komarova et al. (2007) it was these individual static color chips that were rewarded/punished. From a psychological view point, this is a highly abstract notion and certainly not very realistic. Here this issue has been addressed by implementing a memory-like mechanism in

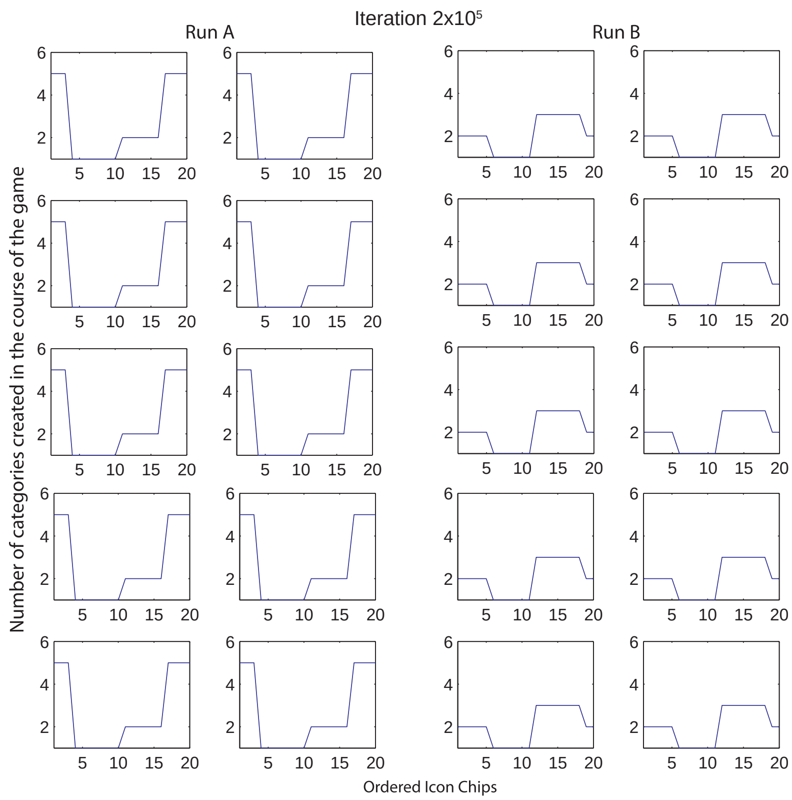

Figure 3: A population of reinforcement learners playing the discrimination–similarity game, with $N = 10$ learners, $n = 20$ icon chips, $\delta = 1/20$ and $k_{\rm sim} = 4$. Each plot corresponds to the categorization of one player.

which experiences are encoded by reinforcing an icon chip that is itself a form of a memory construct that evolves in response to environmental inputs. From a psychological point of view this is a far more realistic modeling approach.

The goal of this paper has been to explore these notions by implementing one possible avenue for evolving icon chips in a fashion that implements a form of a memory construct which serves as the central concept for a continuously evolving category structure. Then this was applied to a single learner and population of learners, as well as to ascertain whether the results of simulations based on these constructs closely replicate those of Komarova et al. (2007). The simple conclusion is that they do; the studies were replicated and thus their assumptions about the basis on which they built their simulation may be assumed to be firm.

In addition to these results, some of the novel elements introduced include:

- The concept of k_{mem} was introduced, which in turn provides a basis for the assumed individual cognitive metric concept inherent in the "distance" between color chips in Komarova et al. (2007).
- To avoid risking the introduction of a new category each time a new icon was encountered, a learner was endowed with the ability to explore its own existing categories to ascertain whether any of them might be suitable for the new icon. This search relied on "memory" access to encountered categories, as well as on having the ability to gauge "perceptual distance" that derived from the concept of k_{mem}.
- To avoid learners in a population from inventing their own categories each time they find none into which to place an icon, the learners were endowed with the ability to "remember" categories that they had encountered while interacting with other learners and to use them even if they had no "understanding" of them.

Acknowledgments

This work was supported in part by National Science Foundation grant NSF#07724228 to the University of California, Irvine. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. I have benefited immensely from discussions and suggestions on earlier versions of this paper by Louis Narens, Kimberly A. Jameson, and Natalia Komarova. I thank K. A. Jameson for relaying to me the relevant historical background of color studies. I also thank Duncan Luce for notational suggestions and comments on a draft and Joetta Gobell for comments on drafts and assistance with several computational issues.

References

Adams, G. K. (1923). An experimental study of memory color and related phenomena. *The American Journal of Psychology, 34*, 359–407.

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

Batchelder, W. H., & Romney A. K. (1989). New results in test theory without an answer key. In E. E. Roskam (Ed.). *Advances in Mathematical Psychology, vol. 2*. Heidelberg: Springer-Verlag. Pp. 229–248.

Belpaeme, T. , & Bleys, J. (2005). Explaining universal color categories through a constrained acquisition process. *Adaptive Behavior 13*, 293–310.

Chapanis, A. (1965). Color names for color space. *American Scientist, 53*, 327–346.

Derefeldt G., & Swartling T. (1995). Colour concept retrieval by free colour naming. Identification of up to 30 colours without training. *Displays, 16*, 69–77.

Dowman, M. (2007) Explaining color term typology with an evolutionary model. *Cognitive Science: A Multidisciplinary Journal, 31*, 99–132.

Garner, W. R. (1974). *The Processing of Information and Structure.* Lawrence Erlbaum Associates.

Griffin, L. D. (2006) The basic colour categories are optimal for classification. *Journal of the Royal Society Interface, 3*, 71–85.

Gwynn, M., & Baker-Ward, L. (2007). Fragile but real: Children's capacity to use newly acquired words to convey preverbal memories. *Child Development, 78*, 448–458.

Hering, E. (1878/1964). *Outlines of a theory of the light sense.* Hurvich, L. M., Jameson, D. (translators); Cambridge, MA: Harvard University Press.

Lindsey, D. T. Brown, A. & M. (2006). Universality of color names. *Proceedings of the National Academy Sciences, 103*, 16608–16613.

Jameson, K. A., & Komarova, N. L. (2009a). Evolutionary models of color categorization. I. Population categorization systems based on normal and dichromat observers. *Journal of the Optical Society of America A, 26*, 1414–1423.

Jameson, K. A., & Komarova, N. L. (2009b). Evolutionary models of categorization. II. Investigations based on realistic observer models and population heterogeneity. *Journal of the Optical Society of America A, 26*, 1424–1436.

Jameson, K. A. (2009). Where in the World Color Survey is the Support for the Hering Primaries as the basis for Color Categorization? In: Cohen, J. & Matthen, M. (eds.), *Color Ontology and Color Science.* Cambridge, Massachusetts: MIT Press.

Jameson, K. A., & R. G. D'Andrade. (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), pp. 295–319.

Jameson, K. A. (2005). Culture and cognition: What is universal about the representation of color experience? *Journal of Cognition and Culture, 5*, 293–347.

Kay, P. , & Regier, T. (2003). Resolving the question of color naming universals. *Proceedings of the National Academy of Sciences, 100*, 9085–9089.

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

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.

Kuehni, R. G. (2001). Determination of unique hues using Munsell color chips. *COLOR Research & Application, 26*, 61–66.

Kuehni, R. G. (2004). Variability in unique hue selection: A surprising phenomenon. *COLOR Research & Application, 29*, 158–162.

Puglisi, A., Baronchelli, A., & Loreto, V. (2008). Cultural route to the emergence of linguistic categories. *Proceedings of the Academy of Sciences ,105*, 7936–7940.

Regier. T., Kay, P. , & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences, 104*, 1436–1441.

Romney, A. K., & Chiao, C-C. (2009), A functional computational model for optimal color coding. *Proceedings of the National Academy of Sciences, 106*, 10376–81.

Steels, L. , & Belpaeme, T. (2005). Coordinating perceptually grounded categories: A case study for colour. *Behavioral and Brain Science, 28*, 469–529.

Appendix A. Computational implementation of the icon structure

The following algorithm is the computational implementation of steps A_1-A_3. The following parameters are defined as:

- $c \in [0,1[$ represents a random color.
- $i_j \in [0,1[$, $j = 1, ..., n$, is an icon chip such that, numerically, $i_{j-1} < i_j$.

- $k_{\text{mem}} \in [0, 1[$ is a constant with the characteristic that if $c \in [k_{\text{mem}} - i_j, i_j + k_{\text{mem}}]$, then i_j is considered an icon chip for c.
- $\delta \in]0, 1[$ is the probability "reward" or "punishment" given to a category.
- $w \in [0, 1[$ is a weighting factor.
- $\beta \in [0, 1]$ is a weighting factor.

The algorithm is as follows:

E_1 Establish an empty vector I and set values for k_{mem}, β, and w.

E_2 Pick a number $c \in]0, 1]$ using a uniform random generator. Search I to determine the existence of icons whose value c is within the distance of k_{mem}, closing the circle at 1—in the following, this closure is assumed. If

No Then, if I is empty $I(1) = c$, otherwise create and insert c into I such that $I(t-1) < c < I(t+1)$.

Yes Then pick the $I(t)$ with the minimum distance from c and update its value as follows. If $c < I(t)$ then calculate the desired adjustment $u = w(I(t) - c)$, the maximum possible adjustment, $a_{\max} = \max\{I(t) - I(t-1) - \beta k_{\text{mem}}, 0\}$, the lesser of the two, $u' = \min\{u, a_{\max}\}$, and set $I(t) \longrightarrow I(t) - u'$. The case for $c > I(t)$ is analogous.

E_3 Repeat steps E_1 and E_2 each time a new color is needed.

Appendix B. Computational implementation of categorization: single learner

The following algorithm is the computational implementation of steps C_1-C_6. It uses as its basis the computational version of Komarova et al. (2007). In addition the parameters defined in Appendix Appendix A, the following parameters are defined as:

- n is the current number of icons.
- κ_k is a positive integer representing the color category k.
- k_{sim} is implemented as an integer value $0 \leq k_{\text{sim}} \leq n - 1$.
- S, δ are integer constants used to represent probability space in integer space. A value S is taken to mean a probability equal 1; δ represents the amount of "reward" or "punishment" given to category naming choices. In probability space, δ/S is a step size.

The game is played as follows:

F_1 Evolve a stable icon chip structure. Establish an empty matrix X and let its rows represent an icon and the columns a category. Pick values for k_{sim}, δ, and S.

F_2 Pick two icons, $i_a \neq i_b$. Assign a category κ_v as follows:

(a) If no category is known, let the row $X[1,:]$ represent the icon i_a and call it by the corresponding row number identified as $i_a^{(r)}$ and the column $X[:,1]$ the category κ_v and set $X[1,1] = S$.

(b) If no category has ever been assigned to i_a examine whether any existing category is suitable and pick one of those at random. A category κ_t is a suitable category for i_a if $|i_a - \bar{\imath}(\kappa_t)| \leq k_{\text{sim}}$ where

$$\bar{\imath}(\kappa_t) = \frac{\sum_{i=1}^{n} I(i)X[i,\kappa_t]}{\sum_{i=1}^{n} X[i,\kappa_t]}.$$

If no suitable category is found, then add a new column to X and assign that column number as the category name κ_v and set $X[I(i_a^{(r)}, \kappa_v] = S$.

(c) Otherwise, if $\varkappa$ is the number of categories in X, assign to i_a the category κ_v by picking a random integer $S \in [1, \varkappa]$ and iteratively sum the cell values of the row $i_a^{(r)}$ until the sum equals or exceeds S. The column number at this point is picked as κ_v.

(d) If **(a)-(c)** do not yield a category, create a new category κ_v by adding a column to X and assign that column number as the name for κ_v.

Similarly pick a κ_w for i_b.

F_3 If $|i_a - i_b| \leq k_{\text{sim}}/n$ then the game is a success if $\kappa_v = \kappa_w$ but a failure otherwise. If $|i_a - i_b| > k_{\text{sim}}/n$ then the game is a failure if $\kappa_v = \kappa_w$ but success otherwise.

F_4 If the game is a success then if $\mathbf{X}[i_a^{(c)}, \kappa_v] = S$, do nothing. Otherwise "reward" κ_v as follows: $\mathbf{X}[i_a^{(c)}, \kappa_v] \longrightarrow \mathbf{X}[i_a^{(c)}, \kappa_v] + \delta$. Then "punish" a random category, $\kappa_t \neq \kappa_v$, such that $\mathbf{X}[i_a^{(c)}, \kappa_t] > 0$, as follows: $\mathbf{X}[i_a^{(c)}, \kappa_v] \longrightarrow \mathbf{X}[i_a^{(c)}, \kappa_v] - \delta$. Do the same for κ_w and c_b.

F_5 If the game is a failure then, $\mathbf{X}[i_a^{(c)}, \kappa_v] \longrightarrow \mathbf{X}[i_a^{(c)}, \kappa_v] - \delta$, and "reward" a random category, $\kappa_t \neq \kappa_v$, by $\mathbf{X}[i_a^{(c)}, \kappa_t] \longrightarrow \mathbf{X}[i_a^{(c)}, \kappa_v] + \delta$. Do the same for κ_w and c_b.

F_6 Repeat steps F_2-F_6 for the duration of the game.

Appendix C. Computational implementation of categorization: population of learners

This algorithm is a generalization of the single-learner algorithm to a population of learners as presented in Appendix Appendix B. It is a computational implementation of steps D_1-D_6. In addition to the parameters defined in Appendix Appendix B, the following parameters are defined as:

- $N \geq 2$ number of learners.

- L_k is the learner $k = 1, ..., N$.

The game is played as follows:

G_1 Do step F_1 (single-learner algorithm) with the addition of picking $N \geq 2$. Initialize a vector $\mathbf{V}$ with with N empty matrices X_{L_k}, $k = 1, ..., N$.

G_2 Pick two learners $L_c \neq L_d$ and two icon chips, $i_a \neq i_b$. For L_c assign to i_a and i_b the categories κ_v and κ_w, respectively, and for L_c the categories κ_s and κ_t as in F_2 **(a)-(d)** with the following addition:

(b) If no suitable category is found, then the learner picks a category the learner has encountered but does not use. If a category κ_q is not used, it is said to have been encountered if the learner previously learned the category κ_r and $\kappa_q < \kappa_r$. If no such category exists, then create a new category according to F_2-**(b)**.

G_3 For L_c, success and failure is as in F_3 and similarly for L_d.

G_4 There are four combinations of successes and failures for the learners L_c and L_d, with four associated actions.

Case	Failure	Success	Consequence
A	L_c	L_d	L_c learns from L_d
B	L_d	L_c	L_d learns from L_c
C		L_c, L_d	Teacher is random
D	L_c,L_d		Individual updates

G_5 For Case A, L_c learns from L_d: L_c performs a successful update for the categories picked by L_d (κ_s and κ_t) in accordance with E_4 with the modification that if L_C does not use κ_s then it sets $\mathbf{X}[i_a^{(c)}, \kappa_s] = S$, and similarly for i_b and κ_t. Then L_c performs an unsuccessful update for its own picked categories (κ_v and κ_w) in accordance with E_5. L_d performs a successful update for its own picked categories (κ_s and κ_t) in accordance with E_4. The learning situation for Cases B and C are analogous.
For Case D, both learners update individually in accordance with F_5.

G_6 Repeat steps G_2-G_6 for the duration of the game.