Advances in Complex Systems
Vol. 15, No. 3 (2012) 1150022 (21 pages)

DOI: 10.1142/S0219525912003536

LANGUAGE, CATEGORIZATION, AND CONVENTION

LOUIS NARENS

Department of Cognitive Sciences,
University of California, Irvine,
Irvine, CA 92697-5100, USA
lnarens@uci.edu

KIMBERLY A. JAMESON

Institute for Mathematical Behavioral Sciences,
University of California, Irvine,
Irvine, CA 92697-5100, USA
kjameson@uci.edu

NATALIA L. KOMAROVA

Department of Mathematics,
University of California, Irvine,
Irvine, CA 92697, USA
komarova@uci.edu

SEAN TAUBER

Cognitive Sciences,
University of California, Irvine,
Irvine, CA 92697-5100, USA
stauber@uci.edu

Received 17 May 2011
Revised 31 October 2011
Accepted 10 November 2011

Linguistic meaning is a convention. This article investigates how such conventions can arise for color categories in populations of simulated "agents". The method uses concepts from evolutionary game theory: A language game where agents assign names to color patches and is played repeatedly by members of a population. The evolutionary dynamics employed make minimal assumptions about agents' perceptions and learning processes. Through various simulations it is shown that under different kinds of reasonable conditions involving outcomes of individual games, the evolutionary dynamics push populations to stationary equilibria, which can be interpreted as achieving shared population meaning systems. Optimal population agreement for meaning is characterized through a mathematical formula, and the simulations presented reveal that for a wide variety of situations, optimality is achieved.

Keywords: Convention; evolution of meaning; evolutionary game theory; population learning algorithms; color category learning.

1. Introduction

It has been recognized for some time that the relationship between a word and its meaning is conventional. The philosopher W. V. Quine and others have argued that the nature of this kind of convention must be of a different sort than those that arise through negotiation or other processes dependent on language (see [18] for a review). The philosopher D. Lewis, in his seminal book on convention [18], writes:

> "Quine [22–25] and White [34, 35] argue that the supposed conventions of language cannot be very much like the central, well-understood cases of convention. Conventions are agreements — but did we ever agree with one another to abide by the stipulated rules in our use of language? We did not. If our ancestors did, how should that concern us, who have forgotten? In any case, the conventions of language could not possibly have originated by agreement, since some of them would have been needed to provide the rudimentary language in which the first agreement was made. We cannot even say what our conventions are, except by long trial and error. Did we know them better when we first adopted them? We have no concept of convention which permits language to be conventional, but we cannot say why. We may indulge this inclination — Quine himself does [*Lewis' footnote:* At the end of 'Carnap and Logical Truth' where he says: 'The lore of our fathers . . . is a pale grey lore, black with fact and white with convention.'] — but we do not understand language any better for doing it. Conclusion: the conventions of language are a myth. The sober truth is that our use of language conforms to regularities — and that is all". (See p. 2 of [18].)

Lewis' reply to Quine and White consisted of providing a game-theoretic based concept of "convention" and applying it to a very narrow and streamlined linguistic situation involving simple semantics and pragmatics. Others in philosophy continue this tradition of applying game theory to simple signaling systems for investigating the evolution of meaning (e.g. [2, 4, 9, 10, 28–30]).

The use of game theory for evolving semantics for simplified linguistic and signaling systems has also spawned a literature outside of philosophy, particularly in biology, economics, artificial intelligence, and psychology. Philosophical and other game-theoretic evolutionary modeling of linguistic and signaling systems have focused almost entirely on situations in which there are enough names or signals so that each object can have at least one unique name or signal. In such situations, the necessity for categorization disappears, that is, grouping objects together so that they each can be denoted by a single name may not be needed. Color naming on the other hand considers situations where there are many more distinguishable colors (e.g. an estimated 10 million [20]) than there are names for them.

This article uses evolutionary game-theoretic approaches to color naming to explore foundational issues involving language, convention, and categorization. For example, the nature of color categories allows some evolutionary naming processes to achieve equilibria that deviates in fundamental ways from Lewis' definition of "convention" and its use in developing language semantics. Also, in idealized color naming, there only exist a small finite number of names to categorize an infinity of continuously distributed colors in perceptual color space. This presents conceptual challenges not generally encountered in the evolutionary modeling of simple signaling games; for example, not all colors will be encountered during evolution, because it follows from the infinite, continuous nature of colors that a given color stimulus, if encountered, will almost certainly never be encountered again. Also, situations naturally occur in which the semantics of color naming systems drift over time, with its pragmatics and logical structure remaining intact. And, of course, there is the problem of how to introduce psychology into the evolution of perceptual categories as well as the nature and role of natural kinds in the evolution of color categorizations.

2. Discrimination–Similarity Games

The central focus of this article is a color naming game — called here the *2-player teacher game* — that was introduced by Komarova *et al.* in [13]. This game is repeatedly played by a population of agents or players, where, in each round of the game, two randomly chosen players interact. In this article, we extend this game to situations on a lattice where players only interact with their neighbors. The aim of both kinds of games is for the population to create a simple naming semantics for a set of colored patches (sometimes referred to as "color chips"), where there are many more patches than names. Successful strategies in the game require each patch to receive a name. Because of this, some names must apply to more than one patch. As a consequence, naming in this game can be viewed as a form of categorization, that is, the denotation of a name may be viewed as a category — the set of colored patches denoted by the name.

The colored patches are assumed to be endowed with a simple perceptual structure derived from elementary, non-controversial psychological considerations involving discrimination. In particular, it is assumed that there are continuous circular arrays of colors — called *hue circles* — that contain all hues. Using English color terms, this means that colors vary continuously in hue from blue to green to yellow to orange to red to purple back again to blue without ever repeating a hue except for the starting and ending hue, and without ever becoming black, white, or gray.[a] The discrimination–similarity game is played on a finite approximation of the hue

[a]From various mathematical modeling perspectives, such hue circles are the most natural one-dimensional reduction of the color space. For example, in the analyses of the dimensional information of the human visual system extracts from spectral stimuli, Kuehni [15] describes how the reflectance functions of twenty Munsell color chips that form a hue circle are reconstructed in the correct perceptual order (i.e. the Munsell order) in a Principle Components space. He suggests this implies "that the hue circle is the natural outcome of dimension reduction of spectral functions having any possible spectral form to three dimensions," and "a circular contour is generated as

circle that is designed to match a psychological hue naming experiment. To this end, a finite number of colored patches are selected and are arranged in the similar pattern as their hues on the hue circle. It is assumed that these patches are of equal brightness, equal saturation, and that they are equally spaced in hue in terms of *just-noticeable differences* (*jnds*). The latter means the following:

Different color patches a and b of equal brightness and saturation are said to be *just-noticeably different in hue*, or more briefly, 1 *jnd in hue*, if and only if the typical subject is 75% correct in saying whether or not a and b are the same patch. Colors c and d on a hue circle of equally bright, equally saturated colors are said to be m jnds apart if and only if m is the smallest positive integer such that there is a sequence $c = c_1, c_2, \ldots, c_m, c_{m+1} = d$ such that c_i and c_{i+1} are 1 jnd apart for $i = 1, \ldots, m$. Color patches of equal brightness and saturation on a circular arrangement are said to be *equally spaced in hue* if and only if every pair of adjacent patches are the same number of jnds apart.

In our simulations, we often use a hue circle of 20 equally spaced colored patches. This is in line with the number of hues employed in many color naming experiments. The World Color Survey [7] uses color samples of 40 equally spaced hues of equal brightness with varying saturations for cross-cultural color naming tasks.

The evolutionary link we use between perception and naming is based on the following idea: Because there are considerably fewer color names than colored patches, two colors within a relatively small number of jnds will have a very strong tendency to be given the same name. This idea is an obvious consequence of the following three principles: (i) categorization is important; (ii) to be useful, categorization should attempt to minimize ambiguity, and (iii) objects of a kind with highly perceptually similar colors tend to similar properties within the kind. The concept of *similarity range*, k_{sim}, formalizes these three principles for a circle of color patches.

By definition, for color patches a and b, $k_{\mathrm{sim}}(a, b)$ is the minimum number of colored patches between a and b for which it becomes important to treat a and b for *pragmatic purposes* (and not for *perceptual purposes*) as belonging to different color categories. It is assumed that k_{sim} is less than the number of color patches. Pragmatically speaking [see principle (i) above], it is beneficial to assign colors

long as at least two dimension reduction functions are themselves curved and overlapping. It is the result of mathematics alone" [16, 17].

A hue circle (as opposed to a hue "line", for example, a line of saturation or a line of brightness) also captures features of color space relational structure considered to be characteristic of color (light mixture) stimulus spaces, for example, classical additive light mixture and cancellation relations (hues that when mixed yield an hue-less appearance) and complementary hue relations (such as complementary negative afterimage pairs — well described since [36]. In this sense, the hue circle we use is relationally similar to Newton's color circle — which was one of the earliest, and most widespread, ordered system of colors — and reflects hue similarity as it is typically presented in the literature. Shepard and Cooper [27] found that this structural ordering was present in the cognitive representation of English color categories, even in individuals who were unable to perceive color. For these reasons, and a long-standing history of its use as a representational structure in the cognitive color processing literature, we chose a hue circle as our color stimulus domain.

outside their k_{sim}-range to different color categories [principle (ii)], and colors within their k_{sim}-range to the same color category [principle (iii)].

$k_{\text{sim}}(a, b)$ 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 its perceived color. 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.

In many of this article's evolutionary scenarios, $k_{\text{sim}}(a, b)$ will have constant value for all a and b on circle of patches. In this case we write k_{sim} for $k_{\text{sim}}(a, b)$. In these scenarios, k_{sim} is intended to set a scale at which color differences become important in the everyday world. It tells us that most of the time, certain objects with colors of a kind within the k_{sim} range will have similar pragmatic properties, whereas other objects of the kind with larger color differences need not.

It is important to emphasize that k_{sim} is not another perceptual version of "just noticeable difference," because, in general, there will be many colors within k_{sim} that are easy to discriminate from one another.

2.1. *Optimal categorization for the discrimination–similarity game*

The colored patches named by the name α defines a *category* (named by α).

Let $\mathbf{N}$ be a set of names. A *naming strategy* Δ for $\mathbf{N}$ and a discrimination–similarity game with Q colored patches and constant k_{sim} is a strategy that probabilistically assigns names from $\mathbf{N}$ to patches. Δ is said to be a *success* for patches a and b (with the possibility that b is a duplicate of a) if and only if the same name is assigned to both patches if they are within k_{sim} and different names if they are outside k_{sim}. Δ is said to be an *optimal naming strategy* if and only if its probability of success for a randomly selected pair of patches (with the possibility of duplicates) is at least as high as that for any set of names and any naming strategy based on those names for a discrimination–similarity game with Q patches and k_{sim}.

Results of [13] show that the number of names C^* from $\mathbf{N}$ that is used in an optimal naming strategy is a natural number closest to

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

Furthermore, each category of an optimal naming strategy is an arc on the hue circle, and the numbers of patches in any two such categories differ at most by 1.

The facts about optimality and Eq. (1), which was formulated for individual agents, extend in obvious ways to naming agreement of a population of agents playing a repeated discrimination–similarity game.

As an example of this result, the evolutionary simulation in Sec. 3.1 starts with $Q = 20$ patches and $k_{\rm sim} = 4$. By Eq. (1), $C^* = \sqrt{10} \approx 3.16$, and $C = 3$. Thus the optimal categorization of this situation will have three categories, each being an arc on the circle of hues with each such arc containing either 6, 7 or 8 patches — which, incidentally, is what is observed in the simulation in Sec. 3.1. Of particular interest about this simulation is that, as described below, although it begins with a specific number of color categories, it is found to drop categories as required by Eq. (1), to achieve an optimal solution.

3. Population Discrimination–Similarity Games

The population version of the discrimination–similarity game is a repeated game. Before the game begins, a finite set of names containing at least two names and a value for $k_{\rm sim}$ are specified. (Because of the nature of the game, the case of a set containing only one name produces a degenerate situation that is of no interest for this article.) In a given round of the game, a subset of the population consisting of at least two members play. In each round, the players are presented two colored patches randomly drawn from the hue circle. These patches can be duplicates, so it is possible that they are identical in color appearance to the players. The players must independently assign names for the two patches from a set of names provided at the first round. It is assumed that if the patches are duplicates, then each player will use a single name to name them. For each player, the round is a *personal success* if and only if the player gives the same name to both patches if they are within $k_{\rm sim}$ and different names if they are outside $k_{\rm sim}$. Otherwise the round is a *personal failure* for that player. The round is said to be a *social success* if and only if it is a personal success for both players and they name the color patches in exactly the same manner.

The game becomes an evolutionary game by (i) specifying an initial naming strategy for each player in the population, and (ii) providing an updating rule for each player in each round based on the following: (a) her current naming strategy, (b) her personal success or failure, and (c) the round being a social success or a social failure. Komarova *et al.* [13] considered various kinds of updating, satisfying the above. This article extends one of these updating rules to a new situation involving players on a lattice, where each round consists of a player playing with a neighbor determined by the lattice. Due to space limitations, extensions of the game to rounds involving more than two players are not considered in this article.

Evolutionary signaling games typically encountered in the literature usually converge in late rounds to completely successful strategies. This cannot happen in a discrimination–similarity game, because, for each player's naming strategy in each round, there will always exist two names α and β and two colors a and b within

k_{sim} such that α names a and β names b, and thus, because a and b are within k_{sim} but have different names, this produces a personal failure and therefore also a social failure. As discussed in Sec. 4, this feature of the discrimination–similarity game produces semantics that yield conventions having very different features from the conventions discussed by Lewis in [18].

3.1. *Evolving categorization through social reinforcement*

A *social reinforcement discrimination–similarity game* proceeds by having a population of players play a discrimination–similarity game. At the beginning of the game, each player has the probabilistic strategy of naming any given colored patch by randomly selecting a name from a uniform distribution of names. In each round of the game, two players are selected at random from the population. The selected players use their current strategies. (If it is a first round for a player, that player uses a uniform distribution strategy.) At the end of the round, the players in the round update according to a reinforcement rule that depends on their personal success or failure in the round and their social success or failure in the round. The other players in the population, who had no interactions with other players during the round, perform a *null update*, that is, they update by leaving their strategy fixed.

The following notation is used for describing the evolutionary behavior of this game:

- Q is the number of colored patches.
- $\mathbf{N}$ is the set of names.
- $p_{c,\alpha}$ stands for the probability of having the patch c named by the name α.
- For each colored patch c and each name α, $S_{c,\alpha}$ stands for a number, called the *reinforcement strength,* for c having name α.
- Reinforcement strengths are non-negative and vary from 0 to a maximum M.

Reinforcement strength and probability are related by the following equation: The probability of the name α in $\mathbf{N}$ naming the colored patch c is

$$\frac{S_{c,\alpha}}{\sum_{\beta\in\mathbf{N}} S_{c,\beta}}.$$

At the beginning of the first round, $S_{c,\alpha} = \frac{M}{Q}$ for each patch c and each name. For simplicity, we will restrict our consideration to the case where $\frac{M}{Q}$ is a positive integer. To keep matters brief, details of updating involving the minimum 0 and the maximum M reinforcement strengths will sometimes be omitted. The following provides some basic ideas for updating:

When conditions suggest that $S_{c,\alpha}$ should be increased by A, $S_{c,\alpha}$ is updated by (i) increasing it by A, if $S_{c,\alpha} + A \leq M$, and by (ii) increasing it to M if $S_{c,\alpha} + A > M$. When (ii) applies, the excess, $S_{c,\alpha} + A - M$ is distributed among the other reinforcement strengths in a manner so that none exceed M. We have experimented with various methods of distributing the excess, e.g. adding it to a $S_{c,\beta}$ that was selected by a specific process involving random selection, dividing it among the $S_{c,\gamma}$ that were not used in the round, etc., and reasonable methods did not have any effect on the end results of our simulations. Similar comments hold for decrements of reinforcement strength and the minimum reinforcement strength, 0.

With the above caveat about the maximum and minimum reinforcement strengths, we now proceed to continue the updating algorithm for a special kind of social reinforcement discrimination–similarity game, called *the 2-player teacher game.*

3.1.1. *2-player teacher game*

In a round, suppose the chips i and j are presented to two players, say Player 1 and Player 2, and Player 1 gives chip i the name α and the chip j the name β and Player 2 gives the chip i the name μ and the chip j the name ν. The following four cases explain the rules for updating the players reinforcement strengths at the end of the round, where for each Player Q in the population, $\mathcal{S}^{Q}_{k,\gamma}$ is Q's reinforcement strength for giving color patch k the name γ.

(i) *Suppose Players 1 and 2 have personal failures* (*and therefore social failures*). Player 1 updates as follows:

$$S^1_{i,\alpha} \to S^1_{i,\alpha} - 1, \quad S^1_{j,\beta} \to S^1_{j,\beta} - 1.$$

Player 2 updates in a similar manner, using μ in place of α and ν in place of β in the above expression. Two names σ and τ different from α and β are selected randomly (without replacement) and Player 1 updates as follows:

$$S^1_{i,\sigma} \to S^1_{i,\sigma} + 1, \quad S^1_{j,\tau} \to S^1_{j,\tau} + 1.$$

Player 2 updates in a similar manner, using μ in place of α and ν in place of β and performing independent random selections. For all the other players Q different from Players 1 and 2 and all patches k and all names γ, Q performs the null update

$$S^Q_{k,\gamma} \to S^Q_{k\gamma}.$$

(ii) *Suppose both players have social successes* (*and therefore personal successes*). Then it follows from the social successes that $\alpha = \mu$ and $\beta = \nu$. Player 1 updates as follows:

$$S^1_{i,\alpha} \to S^1_{i,\alpha} + 1, \quad S^1_{j,\beta} \to S^1_{j,\beta} + 1,$$

and Player 2 updates in the same manner. Two names σ and τ different from α and β are selected randomly (without replacement) and Player 1 updates as follows:

$$S^1_{i,\sigma} \to S^1_{i,\sigma} - 1, \quad S^1_{j,\tau} \to S^1_{j,\tau} - 1.$$

Player 2 updates in a similar manner, using an independent random selection of names. For all the other players Q different from Players 1 and 2 and all patches k and all names γ, Q performs the null update

$$S^Q_{k,\gamma} \to S^Q_{k\gamma}.$$

(iii) *Suppose one player has a personal success, say Player 1, and the other,* (*Player 2*) *has a personal failure.* Then in this situation, Player 1 is called *the teacher,* and her updating for i and j is as follows:

$$S^1_{i,\alpha} \to S^1_{i,\alpha} + 1, \quad S^1_{j,\beta} \to S^1_{j,\beta} + 1.$$

Two names σ and τ different from α and β are selected randomly (without replacement) and Player 1 updates them as follows:

$$S^1_{i,\sigma} \to S^1_{i,\sigma} - 1, \quad S^1_{j,\tau} \to S^1_{j,\tau} - 1.$$

Player 2, called *the learner,* updates by learning from the teacher as follows:

$$S^2_{i,\alpha} \to S^2_{i,\alpha} + 1, \quad S^2_{j,\beta} \to S^2_{j,\beta} + 1, \quad S^2_{i,\mu} \to S^2_{i,\mu} - 1, \quad S^2_{j,\nu} \to S^1_{j,\nu} - 1.$$

For all the other players Q different from Players 1 and 2 and all patches k and all names γ, Q performs the null update

$$S^Q_{k,\gamma} \to S^Q_{k\gamma}.$$

(iv) *Suppose no social success but both players have personal successes.* Then one player is chosen randomly to be the teacher and the other the learner and they update as in case (iii) above.

Komarova *et al.* [13] carried out many simulations involving the 2-player teacher game with variations in the parameter settings. The following example is typical. The simulation had 10 players, 20 color chips, 4 color names, $M = 12$ as the maximum reinforcement strength, and $k_{\text{sim}} = 4$. At 10,000 rounds the beginning of the formation of categories was discernible, and at 80,000 rounds the color chips were categorized by three categories and remained classified by those categories up to 1,000,000 rounds when the simulation was stopped. Note that although the simulation started with four names, the final categorization solution involved only three.

The simulation yielded an optimal solution according to Eq. (1). Other simulations of the 2-player teacher game yielded similar results [13]. They also investigated the case of variable k_{sim}. They showed that the 2-player teacher game using variable k_{sim} and an optimality formula related to Eq. (1) yielded simulations with optimal number of solutions. Jameson and Komarova also investigated 2-player teacher games involving players with heterogeneous perceptions of the color chips, e.g. normal color observers and color deficient ones, and they used human data to determine variable k_{sim} for their players [11, 12, 14].

As a whole, existing results in [11–14] and additional simulations given later in this article show that the 2-player teacher game is robust in producing near optimal solutions in terms of maximizing success for a variety of parameter settings for populations varying from 9 to 100, including situations where the color choices had different impacts on success, and players were heterogeneous in their color perceptions. For populations greater than 100, only a few simulations have been conducted, and they are reported in Sec. 5. They indicate the emergence of *stationary equilibria,* that is, they produce categorizations that remain near one another for a very long time. However, these equilibria were not near optimal in terms of total number of successes. As a general principle, as the population increases in size, with the other parameters of the 2-player teacher game held fixed, more rounds are needed to achieve a stationary equilibrium, e.g. 10^8 for a population of size 121 agents.

3.1.2. *Implications for Lewis' theory of convention*

The simulations for 2-player teacher did produce categorizations that were near optimal in terms of total successes for populations ≤ 100. This is in accord with the use of "convention" in [18]: His final definition of "convention", which Lewis used 73 pages of text to arrive at, allows for almost all of the population to be in a (coordinated) equilibrium-like state instead of the entire population being in a (coordinated) equilibrium. In the following quotation from [18], P stands for a population of agents and S for a convention:

> "There is no harm in allowing a few abnormal instances of S which violate some or all of the clauses [of Lewis's definition of convention as a Nash equilibrium of a coordination game]. So we replace 'in any instance of S of members of P' by 'in almost any instance of S among members of P'. If we want more precision, we can replace it by 'in a fraction of at least d_0 of all instances of S among members of p' with d_0 set slightly below one. Nor is there any harm in allowing some, even most, normal instances of S to contain a few abnormal agents who may [violate the conditions of S being a convention]" (see p. 77 of [18]).

The stationary equilibria achieved in the above simulations may be viewed as conventions containing "a few abnormal instances" — or put another way, of conventions that hold almost universally. However, having a convention to hold almost universally does not necessarily lead to that convention being an "almost equilibrium of a coordination game", which in our reading of Lewis is his intended interpretation of "convention". As shown in Sec. 4, there are discrimination–similarity games whose dynamics yield conventions with non-stationary equilibria. The existence of such conventions present a need for reformulation of Lewis' definition of "convention" and its use in providing an evolutionary account of meaning.

4. Discrimination–Similarity Game with Smoothing Updating

Komarova *et al.* [13] investigated various updating rules for discrimination–similarity games. One of these, called the *smoothing updating rule*, is like the 2-player teacher game in many respects, the principal difference being that it does not use reinforcement. Like previously, smoothing updating rule assumes that

- each player in the population starts the first round with a random naming strategy,
- and for presented chips i and j in the round under consideration, Player 1 names them respectively α and β, and Player 2 names them respectively μ and ν.

The updating part of the smoothing rule is divided into four cases. In each case, the players different from 1 and 2, perform a null updating, that is, they keep their naming strategies the same.

Case 1. *The round is a social* (*and therefore a personal*) *success.* Players 1 and 2 perform a null updating, that is, they keep their naming strategies the same.

Case 2. *The round is a social failure and both players have personal failures.* Player 1 changes her naming strategy by assigning each presented patch a randomly chosen name and leaves the rest of her naming strategy unchanged. Player 2 similarly changes his naming strategy.

Case 3. *The round is a social failure and one player has a personal success and the other has a personal failure.* This case is divided into two subcases:

- **Subcase A.** *Player 1 has a personal success and Player 2 has a personal failure.* In this subcase, Player 1 null updates her strategy, that is, leaves it unchanged. Because of the result of the round and Player 2's interaction with Player 1, Player 2 changes his strategy to follow Player 1's strategy for patches i and j, that is, Player 2 changes his name for i to α and j to β and leaves the rest of his naming strategy unchanged.
- **Subcase B.** *Player 1 has a personal failure and Player 2 has a personal success.* Same as Subcase A except that the roles of Players 1 and 2 in updating are reversed.

Case 4. *The round is a social failure and both players have personal successes.* One of the players of the round, called the *chosen player*, is chosen at random and performs a null updating of her naming strategy. The other player, called the *other player*, changes his strategy to follow the chosen player's strategy for patches i and j, that is, the other player changes his name for i to the name the chosen player used for i and makes a similar kind of change for j, and leaves the rest of his naming strategy unchanged.

Recall for a population playing the 2-player teacher game, the solution remained in the vicinity of a nearly optimal configuration for a very long time. However, for a population using the smooth updating algorithm, Komarova *et al.* [13] observed a non-stationary convergence to a near optimal categorization that slowly drifted outside the vicinity of that categorization and into the vicinity of another optimal categorization. They comment,

> "Non-stationary conventions like those observed for the populations of smoothing and reinforcement learners behave *locally* like conventions based on almost equilibria — that is, behave like conventions based on almost equilibria for appropriate intervals about times t, where t is some time after the convention has been established — but *globally* behave differently from almost equilibria in that the conventional meaning of signals changes with time". (See p. 380 of [13].)

They observed (but did not report in their article) that for each name, its meaning — i.e. the color denoted by it — slowly drifted around the entire color circle *with the order of the names' meanings remaining intact.*

Not all meanings in a non-stationary signaling system need to drift with time. Consider the case of a population discrimination–similarity game with the color names *yellow*, *orange*, *red*, *purple*, *blue*, and *green*. Suppose, as in the above simulated game with a smooth learning algorithm, a non-stationary convergent solution is reached at time t, where the meanings for the color names are also organized in the same circular pattern as the colors they name, say counterclockwisely as: *yellow*, *orange*, *red*, *purple*, *blue*, and *green*. Then the proposition, "Θ: *orange* is immediately between *yellow* and *red*", is true for a very long time for times beyond time t, even though the conventional meanings of the individual color names are slowly changing with time. We interpret the proposition Θ to be an example of what Lewis [18] called "a consequence of the convention", while noting that Lewis formulated his concept for conventions that did not drift.

A deeper consequence of the above convention is that hue color names have circular structure that is invariant for a very long time after time t. (This follows from consequences of conventions involving the color names and the between relation.) This circular structure of names, which agrees with English color semantics, is not derived from cognitive representations of the players, because individual players in the simulations are not capable of anything remotely corresponding to such representations. Instead, it is a result of the players' attempts, and their ultimate achievement, of producing *pragmatically* meaningful communications in an evolutionary scenario that selects among *coordinated* naming strategies.[b]

[b]It is interesting that many in the social sciences assume that the organization of category words, derived by methods such as multidimensional scaling, reflect how the denotations of the words — or more sophisticatedly, the individual's perceptions of the denotations of the words — are organized in the world. In our smoothing algorithm simulation, the organization of the words result from their pragmatic use in communication and not through use of internal psychological structure. A similar result is implicit in a famous human experiment conducted by Shepard and Cooper [27]. In that experiment, normal and impaired color vision participants were given nine English color names and nine color chips corresponding to names dictated by English semantics. Using standard

Komarova *et al.* (see [13], p. 380) comment the following about drifting conventions:

> "Conventions are at the heart of concepts like social contracts, norms, and conformative behavior. To our knowledge formulations of conventions that allow for drifting meaning have not appeared in the literature, although such "drifting conventions" are important in the modeling of some forms of social and institutional change".

It should be noted that in the case discussed just above concerning hue naming, the drifting convention result depends on the circular nature of the hue circle.

5. Simulations Involving von Neumann Neighborhoods on a Torus

In the 2-player teacher game, it is very likely that each player in the population will play with each other player. At the other extreme are interactions among geometrically distributed players, where each player plays only with its geometric neighbors. This section investigates the analog of the 2-player teacher game restricted to such geometrically determined interactions.

A torus is divided into a grid of n^2 isomorphic rectangular-like regions, each region being identified with an unique player from a population of n^2 players. A *von Neumann 4-neighborhood*, abbreviated as *VN4*, of a player α on the torus consists of the four players who are adjacent to α according to the grid — that is the players that share a "side" with α.

The *torus 2-player teacher game* is the same as the 2-player teacher game described previously except for the selection of players in a round: The first of the two players is selected at random and the second is selected from the first's VN4 neighborhood.

The torus simulations presented in this section use the following parameters: names $= 6$, color patches $= 50$, $k_{\text{sim}} = 6$, stack size $M = 20$, and population size 121.

methods, the sets of color names and color chips were separately multidimensionally scaled for each participant. The chips were chosen so that it was expected that the normals (normal color vision participants) would multidimensionally scale the names and the chips in hue circles, and this turned out to be the case. The color deficient participants (dichromats and monochromats) scaled the color names in the same manner as the normals, but scaled the color chips in a different manner — a manner that reflected the nature of their color perception deficiency. Thus for the deficient participants, their cognitive organizations of the color names did not agree with the organization of their perceptions of the denotations for those names. In other words, for the abnormal participants, linguistic meaning based on similarity (as measured through multidimensional scaling) was inconsistent with perceptual meaning based on similarity (as measured through multidimensional scaling). In general, most psycholinguistic approaches to meaning do not attempt to account for the obvious fact that language is used to communicate among members of a population and its semantics is to a large extent based on the communications being pragmatically effective, *and such pragmatic effectiveness does not require homogeneity in cognitive or perceptual meaning or in linguistic shared denotative meaning for the population.*

We ran 50 simulations of the 2-player teacher game on a VN4 torus with 121 players (grid 11×11). In all these simulations, the population of agents converged to a single solution prior to 2.5×10^6 rounds and remained stable until 10^7 rounds when the simulation was stopped. Of the 50 simulations, 33 solutions consisted of 4 continuous categories, 9 solutions had 5 continuous categories, 1 solution had 6 continuous categories, and 7 solutions were idiosyncratic in that there were somewhat continuous categories but with some small regions of discontinuity that was present in all agents.

We note that earlier results by Komarova and Jameson [14] involving the 2-player teacher game with comparable numbers of agents, used 100 agents instead of our 121, employed a larger number of rounds (10^8) instead of our 10^7, had global population random interactions instead of local VN4 neighborhood random interactions, and produced no idiosyncratic solutions. To address these differences and to give the torus simulations additional opportunities to reach a stationary equilibrium, we ran an additional 10 simulations with 10^8 rounds for the 2-player teacher game on the torus with 121 players and VN4 neighborhoods and, for comparison purposes, 10 simulations with 10^8 rounds for the 2-player teacher game with 121 players using random interactions.

For games with the torus VN4 neighborhoods, we found in all of the simulations that multiple spatial regions developed in which the agents within a region converged to a single optimal solution (5 continuous categories) and the solutions across regions were noticeably different from one another. The size, boundaries and solutions of each region continued to change until either the entire population converged to a single optimal solution or until the simulation was stopped at 10^8 rounds with multiple regions still discernible. Of the 10 simulations, half converged to a single optimal solution prior to 10^8 rounds and half had multiple regions at 10^8 rounds, when the simulations were stopped. For the random interaction scenario we found convergence, but to a sub-optimal population solution, usually consisting of 4 categories. Conclusions involving the random interaction and torus versions of the 2-person teacher game are presented in Sec. 7.

6. Extensions Involving Similarity and Memory

It 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. For better and more realistic modeling, the learner needs to be able to generalize her experience so that she can extend her learned semantic categories to colors never previously encountered.

The need for such generalization is common in categorization (e.g. see Sokolov [31]), and is relevant to evolutionary situations depending on categorization involving a large number of objects. However, generalization processes are rare or non-existent in evolutionary game theory. One type of generalization we are investigating is a streamline version of "generalization" that has been much studied in psychology, particularly by the psychologist Roger Shepard. He writes in [26] about the importance of similarity and generalization as cognitive processes: "Recognition that similarity is fundamental to mental processes can be traced back over 2000 years to Aristotle's principle of association by resemblance. Yet, the experimental investigation of generalization did not get under way until the turn of this century, when Pavlov [19] found that dogs would salivate not only at the sound of a bell or whistle that had preceded feeding but also at other sounds — and more so as they were chosen to be more similar to the original sound, for example, in pitch. Since then, numerous experimenters have obtained empirical 'gradients of stimulus generalization', relating the strength, probability, or speed of a learned response to some measure of difference between each test stimulus and the original training stimulus". (See p. 1317 of [26].) Various methods for describing similarity, e.g. multidimensional scaling, have been used by psychologists to describe mathematically the "difference between each test stimulus and the original training stimulus". Shepard in [26] considered the psychological basis for similarity to be a product of evolution that had universal characteristics across people and animals. He provided some empirical support for this idea.

Cognitive psychology assumes that a name gets attached to an item through perception and memory: The percept of an item is matched to a perceptual memory that is associated with a name. The matching and association takes place in long-term memory. More precisely, consider an item c presented to the participant. This invokes, with probability $p(c)$, the perceptual memory $m(c)$ in long-term memory. Attached in long-term memory, with probability $q(c)$, is the name $n[m(c)]$ that the participant gives to c. In place of the perceptual memory $m(c)$, we use an item from the set of items called an *icon item* that we denote by $I(c)$. $I(c)$ need not be an item that was or will be presented to the participant. Thus the name $n[I(c)]$ has probability $q(c)p(c)$ of being used by the participant to name c when it is presented. Our previous modeling of color naming degenerates to the special case where $I(c) = c$ and $p(c) = 1$. In icon item modeling, only icon items $I(c)$ and the probabilities $q(c)$ and $p(c)$ associated with them get updated. Additionally and importantly, icon items can move: That is, updating can change the icon item associated with a given item. This captures an important property about memories: They can drift. Additional properties of memories can be incorporated if needed. For example, psychological laws describing the probability distribution of items having a given name can be incorporated into evolutionary algorithms. Much is empirically known about such laws for the categorization of physical stimuli, and in particular for color. Theoretically, there are reasons to believe that these laws extend more generally (e.g. see [26]). Thus they can, in principle, be used outside of perceptual categorizations to apply to various social concepts that play central roles in evolutionary game theory, e.g. reputation. Such uses will provide a more realistic psychological basis in the evolutionary modeling while simultaneously providing richer modeling possibilities.

One approach to extending our color chip naming algorithms to cover these cases is to evolve for each name used by player α an *icon chip*. Intuitively an icon chip approximates a feature of human long-term memory in the naming of a newly or previously presented color chip (this is similar to Hering's view of color memory as described in [1]). This feature is based on the concept of $k_{\rm sim}$: Two color chips are said to be *k-similar* if and only if they are within $k_{\rm sim}$ of each other. Recall that $S^{\alpha}_{c,n}$ stands for player α giving color chip c the name n. Formally, any color chip c occurring in a game is named by player α as follows:

(i) If there is no icon chip to which c is k-similar, play the game, give c the name dictated by the result of the game, compute $S^{\alpha}_{c,n}$ for each name accordingly, and make c an icon chip.

(ii) If i is the only icon chip to which c is k-similar, play the game for c with c having i's name, recompute $S^{\alpha}_{i,n}$ for each name n to reflect the result of the success or failure of the game (even though the game was played with c), remove i as an icon chip and replace it with a new icon chip i' that is a chip that is nearby i in the direction of c, and set $S^{\alpha}_{i',n} =$ the recomputation of $S^{\alpha}_{i,n}$ for each name n.

(iii) If there are more than one icon chip to which c is k-similar, randomly select one of those icon chips i, play the game for c with c having i's name, recompute $S^{\alpha}_{i,n}$ for each name n to reflect the success or failure of the game (even though the game was played with c), and retain i as an icon chip.

In this approach, chips that are judged are only given names of icon chips, and only icon chips are updated. The approach may be viewed as a means of incorporating a primitive form of perceptual memory into the evolutionary process, with the presented chips being analogous to "perceptions" and the icon chips to "memories".

There are many approaches to the use of icon chips. The first to be carried out by our UC Irvine color research group is described in an article by Steingrimsson [33]. In that article, simulations of the 2-player teacher game, modified to include icon chips and their updatings, yield the same sort of near optimal equilibria as the 2-player teacher game. Interestingly, these simulations show that structural differences can occur in players' organization of icon chips, producing situations where players' homogeneous color perceptual experience produces heterogeneous cognitive color organization while producing near optimal population naming behavior.

7. Discussion and Summary

Before the advent of evolutionary game theory, Lewis in 1969 [18] advocated an evolutionary game-theoretic approach to meaning for simple language systems. He argued that language was a self-organized convention that arose from evolutionary processes that were not dependent on previous language based agreements. His views on language and convention had — and continue to have — much impact in philosophy. As evolutionary game theory developed, the evolution of meaning in simple communication systems was studied by biologists, mathematicians, computational scientists, and others. Most of these researchers were apparently unaware of Lewis' contributions. An exception was the philosopher Skyrms [28] who reformulated and extended Lewis' ideas about language and convention in terms of the evolutionary game theory of the time.

The early game-theoretic applications of the evolution of shared meaning involved simple, language-like communication systems where names were assigned to individual states of the world. In each round of such a game, some players, called *senders*, see the state of the world, and they communicate it to certain other players, called *receivers*, by assigning it a name from a list of names. At the beginning of the game, these names had no meanings. Each receiver acts on the senders' information, that is, acts on the names they receive. If a receiver's action is effective for the state of the world seen by the sender, then both the sender and receiver are rewarded; otherwise they are punished. In such games, depending on various conditions and the evolutionary dynamics employed, shared meaning can evolve in the sense that the actions taken by the receivers are always effective. The earliest applications that the authors are aware of involving game-theoretic evolution of the shared meaning of categories (as opposed to the game-theoretic evolution of shared meanings of individual objects) are Belpaeme's 2002 Ph.D dissertation, [5], Dowman's 2003 article [8], and the well-known 2005 article of Steels and Belpaeme [32]. These applications involve the categorization of color. Subsequent to the Steels and Belpaeme article, several other evolutionary game-theoretic approaches to color categorization have appeared in the literature (e.g. [3, 11–14, 21]).

In our game-theoretic, evolutionary approach, we choose the reduced color space of the hue circle because (i) it simplifies calculations and simulations, and (ii) from various theoretical and mathematical modeling perspectives, it is the most natural one-dimensional reduction of the color space. [Footnote *a* discusses (ii).] We chose to include as little psychology as we could into our simulations in order to emphasize that pragmatic and simple communication constraints were enough to drive the evolutionary dynamic to produce stationary equilibria that are similar to those observed in similar situations in cross-cultural empirical data. We do not interpret this as saying that the kinds of results observed in empirical situations did not involve more complicated psychological or social processes in their evolutions, but only that the use of such richer psychological or social processes for explaining data require, at a minimum, richer data than those of the kind that can be explained by simple evolutionary processes like those we use.

Our group's naming studies extend the evolution of meaning to the categorization of color. With the few exceptions described above and our work, the evolution of meaning was generally restricted to the naming of individual states and actions in simple signaling systems. Results of our group, starting with [13] showed that the conventions that were the end results of their simulated social reinforcement dynamics were near optimal in terms of individual and group success. [13] also showed that varying the pragmatic importance of portions of the hue circle produced optimal categories having certain specific sizes. When such pragmatic importance was included as a part of their simulations, near approximations to the optimal categorizations were achieved. These results are in contrast to the Lewis–Skyrms theory and most other evolutionary naming and signaling studies, which are based on homogeneous items to be named or signaled. Also other studies by our group [11, 12, 14, 33] considered non-homogeneous players in terms of perceptual or cognitive abilities and again achieved near optimal results — again, in contrast to most of literature that use homogeneous players.

Previous research of ours showed the 2-player teacher game yielding near optimal solutions under a number of sampling of parameters.

For the intended applications of the interpretations of the discrimination–similarity game, it would be unrealistic and unfortunate if the results of the simulations critically depend on a player interacting with another player randomly chosen from the population. The simulations of the 2-player teacher game on a torus, where the players only interacted with other players in von Neumann 4-neighborhoods, alleviated this concern by showing that in some situations playing the 2-player teacher game on a VN4 torus appears to be more likely to yield an optimal solution than playing the 2-person teacher game on the same size population but with random interactions.

Another modification of the 2-player teacher game was the use of icon chips to incorporate a simple model of players' memory into modeling. Simulations by Steingrimsson [33] demonstrated using icon chip modeling that players with homogeneous perceptual systems and evolved heterogeneous cognitive systems yielded near optimal solutions similar to those in the 2-player teacher game. (Note that the situation in [33] is in contrast to those of [11, 12] involving the 2-player teacher game in which players had heterogeneous perceptual systems and evolved homogeneous cognitive systems.)

In other research, we have used discrimination–similarity game simulations to provide counterexamples to substantive theories in the literature and to demonstrate how variables of interest in color research trade-off or interact. For such purposes, the number of rounds to achieve a near optimal categorization is not of interest, and thus for such applications we are not bothered that some of our simulations involve a typical player having several hundred of thousands interactions with other players.

The dynamics of our algorithms are based on a form of reinforcement learning. In this article, it is applied to two extreme conditions: random interaction in a population, and random interaction in a von Neumann 4-neighborhood. Neither condition begins to approximate interactions involved in human or other forms of naturally occurring communication, and thus they cannot take advantage of evolved communication structures or social evolutionary updating rules. The discrimination–similarity game's reinforcement updating procedure is very slow in changing a player's strategy, especially if compared to social evolutionary updating rules based on imitation. In future work, we plan to investigate the discrimination–similarity game using social evolutionary modeling on networks with populations of tens of thousands of players. We expect to find that for the kind of issues considered in this article, successful population categorization can be achieved with each player interacting with another player a realistic number of times.

In his early definitions of "convention" in his 1969 book, Lewis had a convention being a perfect, coordinated equilibrium. He had in mind what is today called a "Nash equilibrium". In such a perfect equilibrium, no player can do better by choosing another strategy, and a player deviating from her strategy will make her and the other players worse off. In his final definition of "equilibrium", Lewis strives to make matters better conform to actual conventions, by allowing minor deviations that produce a non-perfect equilibrium. An optimal solution to a discriminabilty–similarity game that penalizes for personal failures will always produce a non-perfect equilibrium, because there will always be color patches within k_{sim} that straddle a category boundary that will produce a personal failure. Our discriminabilty–similarity games based on reinforcement algorithms achieve optimal, non-perfect equilibria that remain near one another for a very long time, i.e. they yield results that are, for a long time, optimal stationary equilibria. However, there is a type of non-perfect, coordinated equilibrium that does not fall under Lewis' definition: a "drifting equilibrium". We showed that the discriminability–similarity game with a smoothing algorithm produced a solution that drifted in the following sense: The meanings assigned by individual players to names rotated in a coordinated manner over time while maintaining near optimal categorization for a very long time after a specified time t. The meanings of these color names drifted to a degree that for each name n and each color patch c, there would be a not-too-distant time after t when c will be named with n. Of particular note is that even though the meanings of the color names drifted over time, the semantic structure among the names did not drift, e.g. if English names were used and *a solution* had English naming semantics, in particular, the "orange" patches were between the "red" and "yellow" patches on the hue circle, then in *all subsequent solutions* (in the not-too-distant future after t), the patches designated as "orange" will remain between those on the hue circle designated as "red" and "yellow", even though in some cases the meanings, that is the colors on the hue circle assign to "red", "orange", and "yellow" change dramatically.

Acknowledgments

This research was funded by a National Science Foundation award (NSF#07724228) from the Methodology, Measurement, and Statistics (MMS) Program of the Division of Social and Economic Sciences (SES), and by the Air Force Office of Scientific Research (AFOSR) (#FA9550-08-1-0389). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the Air Force Office of Scientific Research.

References

[1] Adams, G. K., An experimental study of memory color and related phenomena, *Am. J. Psychol.* **34** (1923) 359–407.

[2] Alexander, J. M., *The Structural Evolution of Morality* (Cambridge University Press, Cambridge, 2007).

[3] Baronchelli, A., Gong, T., Puglisi, A. and Loreto, V., Modeling the emergence of universality in color naming patterns, *Proc. Natl. Acad. Sci. USA* **107** (2010) 2403–2407.

[4] Barrett, J., Dynamic partitioning and the conventionality of kinds, *Philos. Sci.* **74** (2007) 527–546.

[5] Belpaeme, T., Factors influencing the origins of colour categories, Ph.D. thesis, Vrije Universiteit Brussel, Artificial Intelligence Lab (2002).

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

[7] Cook, R. S., Kay, P. and Regier, T., The world color survey database: History and use, in *Handbook of Categorisation in Cognitive Science*, Cohen, H. and Lefebvre, C. (eds.) (Elsevier, 2005), pp. 223–242.

[8] Dowman, M., Explaining color term typology with an evolutionary model, *Cognitive Sci.* **31** (2007) 99–132.

[9] Huttegger, S., Evolution and the explanation of meaning, *Philos. Sci.* **74** (2007) 1–27.

[10] Huttegger, S., Evolutionary explanations of indicatives and imperatives, *Erkenntnis* **66** (2007) 409–436.

[11] Jameson, K. A. and Komarova, N. L., Evolutionary models of categorization. I. Population categorization systems based on normal and dichromat observers, *J. Opt. Soc. Am. A* **26** (2009) 1414–1423.

[12] Jameson, K. A. and Komarova, N. L., Evolutionary models of categorization. II. Investigations based on realistic observer models and population heterogeneity, *J. Opt. Soc. Am. A* **26** (2009) 1424–1436.

[13] Komarova, N. L., Jameson, K. A. and Narens, L., Evolutionary models of color categorization based on discrimination, *J. Math. Psychol.* **51** (2007) 359–382.

[14] Komarova, N. L. and Jameson, K. A., Population heterogeneity and color stimulus heterogeneity in agent-based color categorization, *J. Theor. Biol.* **253** (2008) 680–700.

[15] Kuehni, R. G., Color spaces, Conference presentation in the *Workshop on Colour Ontology and Colour Science*, University of British Columbia (2003).

[16] Kuehni, R. G., *Color Space and its Divisions: Color Order From Antiquity to the Present* (Wiley-Interscience, 2003).

[17] Kuehni, R. G. and Schwarz, A., *Color Ordered: A Survey of Color Systems From Antiquity to the Present* (Oxford University Press, New York, 2008).

[18] Lewis, D., *Convention* (Harvard University Press, Cambridge, MA, 1969).

[19] Pavlov, I. P., *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex*, Translated and edited by Anrep, G. V. (Oxford University Press, London, 1927).

[20] Pointer, M. R. and Attridge, G. G., The number of discernible colors, *Color Res. Appl.* **23** (1998) 52–54.

[21] Puglisi, A., Baronchelli, A. and Loreto, V., Cultural route to the emergence of linguistic categories, *Proc. Natl. Acad. Sci. USA* **105** (2008) 7936–7940.

[22] Quine, W. V., Truth by convention, in *Philosophical Essays for A. N. Whitehead*, Lee, O. H. (ed.) (Longmans, New York, 1936).

[23] Quine, W. V. (ed.), Two dogmas of empiricism, *From a Logical Point of View*, 2nd edn. (Harvard University Press, Cambridge, MA, 1961), pp. 20–46.

[24] Quine, W. V. Carnap and logical truth, in *The Philosophy of Rudolf Carnap*, Schilipp, P. A. (ed.) (Open Court, LaSalle, Ill., 1963), pp. 385–406.

[25] Quine, W. V., *Word and Object* (MIT Press, Cambridge, MA, 1960).

[26] Shepard, R. N., Toward a universal law of generalization for psychological science, *Science* **237** (1987) 1317–1323.

[27] Shepard, R. N. and Cooper, L. A., Representation of colors in the blind, color blind, and normally sighted, *Psychol. Sci.* **3** (1992) 97–104.

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

[29] Skyrms, B., *The Stag Hunt and the Evolution of Social Structure* (Cambridge University Press, Cambridge, 2004).

[30] Skyrms, B., Signals, Presidential Address, Meeting of the Philosophy of Science Association (2006).

[31] Sokolov, E. N., *Perception and the Conditioned Reflex* (Macmillian, New York, 1963).

[32] Steels, L. and Belpaeme, T., Coordinating perceptually grounded categories: A case study for colour, *Behav. Brain Sci.* **28** (2005) 469–529.

[33] Steingrimsson, R., Evolutionary game theoretical model of the concept of hue, a hue structure, and color categorization in novice and stable learners, *Adv. Compl. Syst.* **15** (2012) 1150022.

[34] White, M., The analytic and the synthetic: An untenable dualism, in *John Dewey: Philosopher of Science and Freedom*, Hook, S. (ed.) (Dial Press, New York, 1950), pp. 316–330.

[35] White, M., *Toward Reunion and Philosophy* (Harvard University Press, Cambridge, MA, 1956).

[36] Wilson, M. H. and Brocklebank, R. W., Complementary hues of after-images, *J. Opt. Soc. Am.* **45** (1955) 293–299.