C

Color Category Learning in Naming-Game Simulations

Tony Belpaeme
School of Computing and Mathematics,
University of Plymouth, Plymouth, UK

Synonyms

Computer modelling; Linguistic relativism; Simulation

Definition

Color categories are inextricably linked to language as color categories are typically, though not necessarily, associated with color terms. It is believed that the acquisition of categories, including color categories, is influenced by language [1]. Prelinguistic infants do seem to have a set of color categories, which are then either consolidated or modified through observing or engaging in linguistic interactions about color.

Language has been shown to have an influence on a range of modalities, such as time, space, and color. This phenomenon, known as linguistic relativism, shows how the language one uses, and by extension the culture one lives in, has an impact on perception and cognition. It has also been shown that language has an influence on color perception: having a particular color word speeds up spotting a chip of that color among distracting color chips [2, 3]. What is not entirely clear is how language influences the acquisition of color categories. As data on color category acquisition in infants is hard to come by, we can resort to computer simulations to learn more about how language impacts the acquisition of color categories.

A language is a communication system that is shared within a group of language users. As such, a language can be seen as an agreement between all language users on the words and rules of a language, and their meaning. Color words are also subject to this agreement: in the English language speakers agree to use "red" for, among others, the chromatic perception of a ripe tomato, blood, and a light with dominant spectral wavelength of 780 nm. There is no central authority insisting on this: language users themselves agree on this convention. When a new language user, such as a newborn child, enters a linguistic community, it will to varying degrees adopt this convention.

If language is a convention that is agreed upon by a linguistic community, and if language impacts category acquisition, then it follows that categories are to a certain extent also agreed upon by the community. Computer simulations can help us understand how a community can arrive at an agreement on linguistic conventions and how language shapes concepts and categories.

There are a number of simulation models that can elucidate the process of language acquisition. The Language Game model [4, 5] has proven to be both effective and popular and can be used to study how a group of individuals reach a consensus on linguistic forms and associated categories. If, for example, a new word is introduced in a language, the Language Game model can simulate how that word spreads through the population. The model as such serves to study the dynamics of language change, and by setting parameters of the model, one can study what conditions make a language change. Iterated Learning Models, an alternative model to Language Games, study the sequential transfer of language [6]. Individuals are placed in a chain, and each individual's output is used as input for the next individual. Language Games study horizontal transmission, while Iterated Learning Models study vertical transmission of language. Both models demonstrate how small biases and communication bottlenecks can have a large effect on the language and conceptual structures that arise. A third class of simulated Language Game models are based on Evolutionary Game Theory (e.g., [7, 8]) or Statistical Mechanics (e.g., [9]). These are in essence mathematical models which start from a minimal set of parameters and study the influence of different settings of these parameters.

Language Game Models

In Language Game models, a community of language users is modeled as a population of N software agents. Each agent can store and recall words (or other linguistic information, such as rules) and categories. In the domain of color, the agents store color terms and color categories. In addition, each agent stores associations between color terms and color categories. An association typically is a value showing how strong the association between a term and category is. Agents start with empty inventories and gradually fill these with words and categories.

Various Language Game models represent color in different ways. Color can be modeled

as a point on a single circular dimension [7, 8, 10, 11]; a color stimulus is then a real number in the interval $[0,1]$. For a more realistic model of color, one can endow the artificial agents with a color appearance model, such as the CIE $L^* a^* b^*$ color space [12]. In this space, each color is represented by three real numbers L^*, a^*, and b^*, with L^* representing lightness, a^* the amount of green or magenta, and b^* the amount of yellow and blue. The CIE $L^*a^*b^*$ color appearance model aims to provide an accurate representation of color perception differences and allows for a similarity measure to be calculated between two colors, which is done by taking the Euclidean distance between two color values, permitting a good first approximation to categorical color perception (see [13] for an experimental appraisal and extension of CIE $L^*a^*b^*$).

In addition to the color categories and color terms used by an agent, simulations also need to prescribe what agents do when interacting with each other. In one form of a Language Game model interaction two agents are selected at random from the population; one acts as a *speaker*, the other as a *hearer*. Both agents are presented with a *context*; this is a set of M random color stimuli, each at a distance d from each other. The distance d guarantees that colors are not too similar or identical. From the context one color stimulus is selected, this will be the topic, and the speaker will attempt to communicate what the topic is to the hearer.

The speaker first finds a category that best matches the topic (often, but necessarily, this category is a unique match, meaning that it matches no other stimuli in the context). If no category can be found, the speaker will adapt its category set by adding a new category. Next it finds a word associated with the category and communicates it to the hearer. The hearer will attempt to guess the topic by looking up the word and the associated category in its inventory. It will check which stimulus matches the category best and will "point out" the stimulus. The hearer will then signal if this stimulus is indeed the intended topic. If it is, the game is successful. If the hearer points out any other stimulus, the game fails. When successful, categories and word-

500 games

15000 games

a*

b*
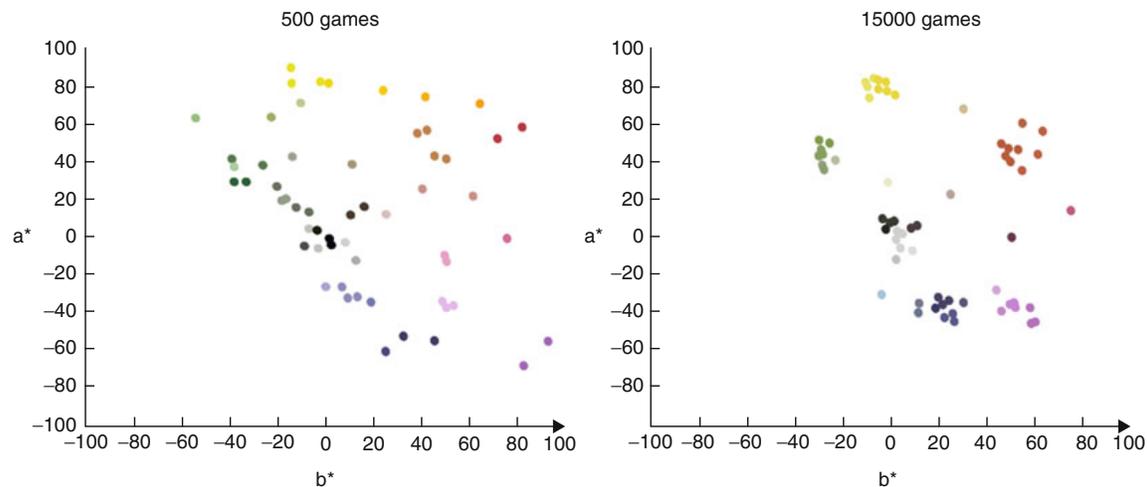
Color Category Learning in Naming-Game Simulations, Fig. 1 Both plots show all color categories of ten agents plotted in the $a*b*$ plane of the CIE $L*a*b*$ color space (the $L*$ lightness dimension is projected onto the $a*b*$ plane; as such *white* and *black* categories are plotted on *top* of each other). The *left* figure shows the state after 500 Naming Games; the categories are still scattered in the color space. The *right* figure shows the state after 15,000 games when the communicate success of the agents is over 85 %; the categories have gravitated to a limited number of clusters

category associations in both agents are reinforced, with the categories used in the interaction adapting such that they match the topic closer. When the game fails, the associations are weakened [4, 5].

During the iterative playing of Language Games [6] the agents create novel categories and words and change existing categories and category-word associations to optimize the communication. Only when communication is good enough (as determined by a preset threshold, e.g., $\tau = 90$ % of all games end in success) will the agents' internal representations stop changing. It is important to note that the internal representation of all agents at this moment will not be identical and that the agents do not necessarily have the same number of words and categories. Their internal representations are merely sufficiently coordinated for communication to succeed with a success rate of τ. A population typically has a size N between two and several thousand agents and will play tens of thousands or perhaps millions of Language Games before stabilizing. An interesting observation is that the lexical and category systems of the agents adapt until they are "good enough" to successfully communicate; the agents do not need identical words and identical categories; they only need to sufficiently overlap to allow successful communication in most interactions. As such the semantics of words differ between agents: indeed your red is different from my red. As such, word and category are not true descriptions of the world; they are merely useful [14].

The dynamics of Language Game models have been extensively studied, as they inform research into the diachronic evolution of language. For example, the conditions under which a new word or a new linguistic construction is taken up by a language community can be modeled using Language Games [15, 16], and model predictions have been confirmed in studies with human participants [17]. Language Games have been used to clarify the minimum constraints needed to evolve a shared color category system by populations of agents [7], how varying agent perceptual properties impact color category system evolution [18–20], and how varying color salience affects color category evolution [20]. In the case of color, however, the Language Game model serves a different purpose. It helps us to understand how relatively small biases present in color communications can have large-scale effects on the evolution of color category

systems. Small biases are amplified through repeated interactions between language users. Specifically, simulated Language Game models help us formally investigate factors likely to influence color communications and help us understand why color categories appear to be universal and the degree to which pragmatics of communication or culture may contribute to color category evolution.

Explaining the Universal Character of Color Categories

It has been suggested that human color categories exhibit a universal pattern: many cultures have color categories that are seemingly similar. This was first suggested based on tenuous evidence in 1969 by Berlin and Kay [21] and later refined in the World Color Survey [22–24]. As such, color categories are not arbitrary, and this infused a principled search for the basis of their universal character. There are cultures which deviate sufficiently from the universal pattern, virtually ruling out the possibility that color categories are genetically determined. Other processes must be at work, and computer simulations can help us elucidate these.

When agents in a Language Game are communicating about color stimuli, their categories soon converge on a limited set of categories (Fig. 1). The locations in which the categories converge are not predetermined; instead they result from the slight biases introduced by the perceptual system of the agents [12, 18–20, 25]. Two different runs of a simulation will result in two different end results. However, repeated simulations do exhibit a pattern in which some color categories consistently emerge from the interaction between the agents. Categories such as red, green, yellow, blue, white, black, and so forth tend to almost always emerge. This matches the observations in the World Color Survey: these basic color terms and categories are found in the majority of the world's languages.

As the repeated playing of Language Games forces agents to coordinate their color categories and color terms, small biases in the agents' color perception will have a large influence. Belpaeme, Bleys and Steels [12, 23] showed how the bias of the CIE $L*a*b*$ color appearance model together with a repeated negotiation of linguistic conventions results in the emergence of universal patterns of color categories. Baronchelli et al. [11] refined this; again using a Language Game model they showed how the human Just Noticeable Difference (JND) function, a function which shows the wavelength differences that are just about distinguishable to the human eye for each wavelength in the visual spectrum, also provides a small but important bias that can explain the universal character of color categories.

Language Games show how a variety of factors may contribute to the universal character of color categories without the need for color categories to be explicitly genetically determined. They permit the evaluation of, for example, neurophysical properties of human color perception as well as other small biases which, through repeated linguistic negotiations, amplify and can contribute to the similarities seen across groups of languages that have roughly similar color categories.

References

1. Bowerman, M., Levinson, S. (eds.): Language Acquisition and Conceptual Development. Cambridge University Press, Cambridge, UK (2001)
2. Gilbert, A., Regier, T., Kay, P., Ivry, R.: Whorf hypothesis is supported in the right visual field but not the left. Proc. Natl. Acad. Sci. **103**(2), 489–494 (2006)
3. Roberson, D., Pak, H., Hanley, J.R.: Categorical perception of colour in the left and right visual field is verbally mediated: Evidence from Korean. Cognition **107**(2), 752–762 (2008)
4. Steels, L.: Evolving grounded communication for robots. Trends Cogn. Sci. **7**(7), 308–312 (2003)
5. Steels, L.: Modeling the cultural evolution of language. Phys. Life Rev. **8**((4), 339–356 (2011)
6. Smith, K., Kirby, S., Brighton, H.: Iterated learning: a framework for the emergence of language. Artif. Life **9**(4), 371–386 (2003)
7. Komarova, N.L., Jameson, K.A., Narens, L.: Evolutionary models of color categorization based on discrimination. J. Math. Psychol. **51**, 359–382 (2007)
8. Steingrimsson, R.: Evolutionary game theoretical model of the evolution of the concept of hue, a hue

structure, and color categorization in novice and stable learners. Adv. Complex Syst. **15**(03n04), 1150018 (2012)

9. Castellano, C., Fortunato, S., Loreto, V.: Statistical physics of social dynamics. Rev. Mod. Phys. **81**(2), 591 (2009)

10. Puglisi, A., Baronchelli, A., Loreto, V.: Cultural route to the emergence of linguistic categories. Proc. Natl. Acad. Sci. **105**(23), 7936–7940 (2008)

11. Baronchelli, A., Gong, T., Puglisi, A., Loreto, V.: Modeling the emergence of universality in color naming patterns. Proc. Natl. Acad. Sci. **107**, 2403 (2010)

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

13. Komarova, N.L., Jameson, K.A.: A quantitative theory of human color choices. PLoS One **8**, e55986 (2013)

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

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

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

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

18. Jameson, K.A., Komarova, N.L.: Evolutionary models of categorization. I. Population categorization systems based on normal and dichromat observers. J. Opt. Soc. Am. A **26**(6), 1414–1423 (2009)

19. Jameson, K.A., Komarova, N.L.: Evolutionary models of categorization. II. Investigations based on realistic observer models and population heterogeneity. J. Opt. Soc. Am. A **26**(6), 1424–1436 (2009)

20. Komarova, N.L., Jameson, K.A.: Population heterogeneity and color stimulus heterogeneity in agent-based color categorization. J. Theor. Biol. **253**(4), 680–700 (2008)

21. Berlin, B., Kay, P.: Basic Color Terms: Their Universality and Evolution. University of California Press, Berkeley (1969)

22. Kay, P., Berlin, B., Maffi, L., Merrifield, W.: The World Color Survey. Center for the Study of Language and Information, Stanford (2003)

23. Kay, P., Regier, T.: Resolving the question of color naming universals. Proc. Natl. Acad. Sci. **100**(15), 9085–9089 (2003)

24. Lindsey, D.T., Brown, A.M.: World color survey color naming reveals universal motifs and their within – language diversity. Proc. Natl. Acad. Sci. **106**(47), 19785–19790 (2009)

25. Steels, L., Belpaeme, T.: Coordinating perceptually grounded categories through language. A case study for colour. Behav. Brain Sci. **24**(8), 469–529 (2005)