A HUMAN MODEL OF COLOR TERM EVOLUTION

MIKE DOWMAN

General Systems Studies, University of Tokyo, 3-8-1 Komaba
Meguro-ku, Tokyo, 153-8902, Japan

JING XU, THOMAS L. GRIFFITHS

Department of Psychology, University of California, Berkeley,
3210 Tolman Hall, Berkeley, CA 94720, USA

When Berlin and Kay (1969) identified striking typological patterns in the denotations of basic color terms, they suggested that they arose through a process of cultural evolution. We explored the role of cultural evolution in the development of color term systems in a large scale study using 195 human participants. As in computer simulations of cultural evolution by "iterated learning" (Steels and Belpaeme, 2005; Dowman, 2007), color term systems were passed along a chain of people, who each tried to learn the color term system used by the previous person. We sought to investigate how the systems would be transformed by this process, and to what extent individual learners would shape the categories in accordance with their own prior expectations. Our results show clearly that, as color terms evolve, their denotations are transformed by the people who learn them, so that color term systems are products both of the psychological biases of the individual learners, and of the process by which language is transmitted from generation to generation.

For most languages we only have information about their current state, rather than a record of how they have changed over time. Proposals about how color term systems evolve are therefore based mainly on extrapolation from the range of color term systems observed in the world today. Recently this has been complemented by computer models that have investigated how color term systems evolve when passed along a chain of computational agents (Steels and Belpaeme, 2005; Dowman, 2007). Each agent in these models was capable both of learning color words and of using the words they had learned when speaking to another agent. These models allowed the consequences of the social transmission of color vocabulary via iterated learning to be studied, but it was necessary to make assumptions about how people learn and represent color words. By using human participants in place of computational agents, we removed the need to make such assumptions, as this change replaced artificial

learning and representation mechanisms with human ones. Using this same methodology, Kalish, Griffiths, and Lewandowsky (2007) revealed strong prior biases concerning function mappings, but the methodology has not previously been applied to color language.

In our experiments, we told the participants that they would be learning the color term system of a language unrelated to English. We then showed them made up words on a computer screen, together with a series of randomly selected examples of colors that could be named by each word. After training, we asked participants to name each of the 330 color chips in the standard World Color Survey Munsell array, using one of the words given in training. Examples were then randomly selected from these responses, and used as training data from which the next participant could try to reconstruct the color categories in the language. This process was repeated over 13 generations of learners.

We conducted experiments in which there were either 2, 3, 4, 5 or 6 basic color terms, and in which the color term system taught to the initial learner divided the color space up either on the basis of hue, or of lightness, or was simply completely random. The participants quickly imposed structure on the random color term systems by naming a relatively coherent range of colors with each term during the testing phase. Otherwise, the color term systems usually evolved gradually, but at some points participants would impose radically new categorizations. Systems with 2 or 3 color terms tended to alternate between dividing the color space primarily on the basis of hue or lightness. In systems with 4, 5 or 6 words, categories emerged that were based on both hue and lightness, as is the case with color terms in naturally occurring languages. Therefore, while the color term systems were based on the input received by language learners, unnatural systems were restructured to reflect participants' preferences for some kinds of category structure over others. As the experiment progressed, the color term systems increasingly came to reflect those seen in naturally occurring languages, suggesting that the structure of color term systems is largely the product of people's learning biases, brought to the surface through the process of cultural transmission.

References

Berlin, B., & Kay, P. (1969). *Basic color terms*. Berkeley: University of California Press.

Dowman, M. (2007). Explaining Color Term Typology with an Evolutionary Model. *Cognitive Science*, *31(1),* 99-132.

Kalish, M. L., Griffiths, T. L., & Lewandowsky, S. (2007). Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin and Review, 14,* 288-294.

Steels, L. and Belpaeme, T. (2005). Coordinating Perceptually Grounded Categories through Language. A Case Study for Color. *Behavioral and Brain Sciences*, *28(4),* 469-489.