

## Categorization in artificial agents: Guidance on empirical research?

William S.-Y. Wang and Tao Gong

Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong, China. [wswywang@ee.cuhk.edu.hk](mailto:wswywang@ee.cuhk.edu.hk)  
<http://www.ee.cuhk.edu.hk/~wswywang/> [tgong@ee.cuhk.edu.hk](mailto:tgong@ee.cuhk.edu.hk)

**Abstract:** By comparing mechanisms in nativism, empiricism, and culturalism, the target article by Steels & Belpaeme (S&B) emphasizes the influence of communicational constraint on sharing color categories. Our commentary suggests deeper considerations of some of their claims, and discusses some modifications that may help in the study of communicational constraints in both humans and robots.

The article by Steels & Belpaeme (S&B) presents a multiagent model that adopts many prevalent mechanisms used in other similar cognitive or self-organizing models, such as neural networks (e.g., Munroe & Cangelosi 2002), associative networks (e.g., Smith et al. 2003), and strength-based competition (e.g., Steels et al. 2002). The authors refrain from any judgment on mechanisms that might be more realistic. However, further discussions are required to assess some of their conclusions.

Their article summarizes the categorical repertoire-sharing process by using four types of simulations: (a) acquisition of repertoires with the same learning mechanism (individual learning), (b) individual learning and adjustment of acquired repertoires during language communication (cultural transmission), (c) genetic transmission of repertoires with occasional mutation (genetic evolution), and (d) genetic evolution and cultural transmission. The comparisons of Category Variance (CV) of (a) and (b), as well as (a) and (c) lead to the compelling conclusion that “both a cultural learning hypothesis . . . and a genetic evolution hypothesis . . . could explain how agents in a population can reach a shared repertoire of categories. . . . The difference between the two models appears to be in terms of the time needed to adapt to the environment or reach coherence” (sect. 5). Then, the authors suggest “the collective choice of a shared repertoire must integrate multiple constraints, including constraints coming from communication” (Abstract). However, deeper discussions of these claims are necessary.

First, in their model, the rate of genetic evolution is controlled by adjusting the parameters in the neural network. The rate of cultural transmission is determined by a different set of parameters that associate categories and their symbols. Although there is a general consensus that cultural transmission operates at a much higher rate, it is not clear how the two sets of parameters can be made commensurate with each other and meaningfully compared.

Second, to support the authors’ suggestion, is it necessary to show why we must integrate cultural transmission, because genetic evolution alone can already achieve category sharing? Can cultural transmission influence genetic evolution, and if so, what is the influence? The answers to these questions lie in the comparison of the CV difference between (c) and (d), or between (b) and (d). In fact, this topic is touched on by Munroe and Cangelosi (2002) in their mushroom-foraging model (M & C model). Based on the *Semiotic Square* (Steels 2002, see Fig. 1), in the M & C model, genetic evolution adjusts the sensorimotor tools (neural network’s connection weights, Sensation aspect, aspect A), and cultural evolution introduces changes to the outputs of neural networks in the previous generation, the combination of input times being the connection weights (Representation aspect, aspect B). The M & C model shows that cultural transmission can assist genetic evolution; the learning time under cultural transmission and genetic evolution is much shorter than that under only one of these mechanisms. However, in the M & C model, both mechanisms work on the internal aspects (aspects A and B), and it neglects the Symbol aspect (aspect D). The framework of the target article covers all four aspects of the semiotic square. The neural network handles the color representation, and genetic evolution

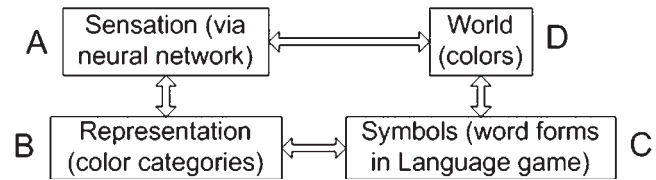


Figure 1 (Wang & Gong). Semiotic Square modified from Steels (2002).

adjusts these representations; the associative network handles the mappings between semantics and symbols; and cultural transmission adjusts these mappings. Therefore, besides demonstrating that both cultural transmission and genetic evolution can achieve the sharing task, this model can also explore whether these two mechanisms, separately working on different aspects, can affect each other by comparing CV or the number of games necessary to acquire certain CV in different simulations.

Finally, S&B should state clearly what “sufficiently shared” means in their claim that “a perceptually grounded categorical repertoire can become sufficiently shared among the members of a population to allow successful communication” (sect. 6). In their case study, because of identical learning mechanisms and limited features considered in creating categories, the categories created by different agents for the same color may be very similar, and successful communication may require the sharing of identical categories. It implies that only by sharing identical categories can successful communication be possible.

However, in general, this is not the case. Considering heterogeneous sensorimotor systems or learning mechanisms adopted by agents and the multiple features contained in world items, it is possible for agents, through different learning mechanisms, to create different categories for the same world item based on its different features. Besides, if both the categories partitioned in semantics inside one agent and the word forms partitioned in symbols can distinguish world items, it is possible that each agent will develop its own associative network between word forms and its categories, and successful communication is still possible even though there are no shared categories. This is more obvious when humans perceive abstract concepts like “friend,” “loyalty,” “game,” and so forth. Different criteria are developed to represent these concepts, and communication is still available on a certain level.

In addition, some of their methods need modification if the authors want their results to be “relevant to a much broader audience of cognitive scientists” (sect. 6). First, the language game (Steels 2001a) in this model adjusts the association between words and categories, and the association between symbols and world items. No matter whether or not it can represent the speaker’s word form, the hearer always gets the association between the symbol and the world item that this symbol represents in the speaker’s mind. However, this method, similar to mind-reading, is too strong to be realistic, because in actual conversations there are communications in which the hearer gets no hints or even gets wrong ones. This indicates that language, or other communicational constraints, is not always reliable. Besides, as Quine’s (1960) question about *gavagai* shows, nonlinguistic feedback only provides limited confirmation. Therefore, even without noise, misunderstanding is inevitable, and mind-reading does not simulate the actual influence of communicational constraints. Whether or not communicational constraints still have similar effects on sharing categories when occasional misunderstanding is allowed is worth studying, and this is already discussed in some models (e.g., Gong et al. 2004).

Also, this model adopts a Genetic Algorithm (GA) (Holland 1995) without crossover, in which, mutations, “happen with a probability inversely proportional to discriminatory success” (sect. 3.4). This method will undoubtedly accelerate the acquisition of

common categories because categories that are not successfully used will undergo more mutations. Therefore, this GA introduces a selective force though the mutation itself has no intelligence about what is good change. Genetic operations, like mutation, should be independent of certain factors outside the genome. Besides, the main driving force for evolution is the reorganization of the available materials (crossover), instead of the occasional mutation (Holland 2005). However, in this model, asexual reproduction does not incorporate crossover, and the low mutation rate may not explicitly represent the speed of the genetic evolution.

ACKNOWLEDGMENTS

The authors of this commentary would like to thank James W. Minett and Wong Chun-Kit for their useful discussions and resourceful suggestions. Our work is supported in part by grants from the Research Grant Council of the Hong Kong SAR: CUHK-1224/02H and CUHK-1127/04H.

Variations in color naming within and across populations

Michael A. Webster<sup>a</sup> and Paul Kay<sup>b</sup>

<sup>a</sup>Department of Psychology, University of Nevada – Reno, Reno, NV 89557;

<sup>b</sup>International Computer Science Institute and University of California – Berkeley, Berkeley, CA 94704. [mwebster@unr.nevada.edu](mailto:mwebster@unr.nevada.edu)

[kay@icsi.berkeley.edu](mailto:kay@icsi.berkeley.edu) <http://www.icsi.berkeley.edu/~kay/>

**Abstract:** The simulations of Steels & Belpaeme (S&B) suggest that communication could lead to color categories that are closely shared within a language and potentially diverge across languages. We argue that this is opposite of the patterns that are actually observed in empirical studies of color naming. Focal color choices more often exhibit strong concordance across languages while also showing pronounced variability within any language.

Steels & Belpaeme (S&B) use theoretical simulations to explore the potential role of physiological, environmental, and cultural (linguistic) constraints on the acquisition of shared color categories. Although their stated aim is to identify principles that could guide the design of communication among artificial intelligence systems, they emphasize that the results are also relevant for understanding color categorization in human observers. Our commentary focuses on the extent to which the trends they observe are evident in actual studies of color naming.

In S&B’s simulations, whether or not a factor provides a loose or tight constraint is evaluated by measuring the variance in color categories across observers. In all cases, they find the variance to be greater for agents drawn from separate populations than for those drawn from the same population, yet this difference becomes dramatic when the categories are learned through language, in which case, the within-group variance approaches zero.

This, in theory, points to a strong potential for cultural relativity in color naming.

What are the patterns of variance in empirical measures of color naming? There are two striking patterns. First, there are strong universal tendencies across languages. These tendencies were originally suggested by Berlin and Kay (1969) and have been confirmed by Kay and Regier (2003) in a recent analysis of the World Color Survey (WCS), which provides color-naming responses from an average of 24 primarily monolingual speakers from each of 110 unwritten languages. Specifically, they showed that the centroids of color-naming responses for different languages exhibit much stronger clustering than would be predicted by chance. This is qualitatively consistent with S&B’s analyses, showing that physiological and/or environmental constraints can support some degree of consistency among speakers. Whether it is quantitatively consistent could potentially be evaluated by applying the authors’ variance metric to the WCS data (which is available on-line at <http://www.icsi.berkeley.edu/wcs/data.html>). This might allow one to assess whether different languages show more concordance in color categories than would be expected from their models of physiological and environmental factors. Without such comparisons, it is difficult to interpret the relevance for human behavior of the values they derive from simulations.

The second prominent property of actual color-naming data is the pronounced variation among speakers of the same language. Individual differences in unique hue and focal color choices have been widely documented, though their causes remain poorly understood (Webster et al. 2000). For example, the wavelengths that individuals select for unique green within a linguistically homogeneous group span a range of more than 80 nm; these variations are in fact so large that the same wavelength might be chosen as unique green by one observer and unique yellow or blue by another (Kuehni 2004). Individual differences in focal color choices remain large for more naturalistic spectra like the Munsell chips and represent another obvious feature of the WCS data (as well as for most other data sets on color naming). Moreover, comparable differences persist even when the samples are restricted to individuals who select colors with the highest reliability (Webster et al. 2000). In sum, in actual measures of color naming, as contrasted to simulations, within-group variance is very large.

This fact appears difficult to reconcile with the minimal variance predicted by S&B to arise from adding communication to the simulated agents. Actual agents do not show the close agreement that language could potentially support. As an illustration of this, Table 1 compares the average within-language variance to the variance in mean foci across languages for “red,” “green,” “blue,” or “yellow” terms for the WCS respondents, based on an analysis by Webster and Kay (in press). (These are calculated from the raw distances in the Munsell palette for the Hue and Value dimensions separately.) For each language, terms corresponding to the English terms were determined by finding the focal choices for con-

Table 1 (Webster & Kay). Average variance in individual foci within a WCS language compared to the variance of mean foci between languages, computed for the hue or lightness of “red,” “green,” “blue,” or “yellow.” F-tests compare the between-language variance to the variance predicted by randomly sampling speakers of different languages. The hue scale runs from 1 = Munsell 2.5R, in 40 steps, to 40 = Munsell 10RP. The lightness scale is Munsell Value

Term	#	Focal Hue					Focal Lightness				
		Mean	Variance	Predicted variance	F	p	Mean	Predicted Variance	variance	F	p
red	103	1.77	.46	.25	1.81	<.002	4.25	.095	.040	2.41	< e-5
green	73	18.9	3.01	.96	3.16	< e-8	4.74	.41	.099	4.12	< e-10
blue	50	27.7	2.45	.93	2.56	< e-5	4.30	.46	.093	4.84	< e-10
yellow	86	9.46	.65	.31	2.13	<.0002	7.79	.13	.038	3.38	< e-8