Adaptive **Behavior**

Explaining Universal Color Categories Through a Constrained Acquisition Process

Tony Belpaeme[1], Joris Bleys[2]
[1]*University of Plymouth, UK*
[2]*Artificial Intelligence Laboratory, Vrije Universiteit Brussel*

Color categories enjoy a special status among human perceptual categories as they exhibit a remarkable cross-cultural similarity. Many scholars have explained this universal character as being the result of an innate representation or an innate developmental program which all humans share. We will critically assess the available evidence, which is at best controversial, and we will suggest an alternative account for the universality of color categories based on linguistic transmission constrained by universal biases. We introduce a computational model to test our hypothesis and present results. These show that indeed the cultural acquisition of color categories together with mild constraints on the perception and categorical representation result in categories that have a distribution similar to human color categories.

Keywords color · color categories · linguistic relativism · language game · universalism

1 Introduction

The world out there is what William James called one "blooming buzzing confusion." In order to handle this flood of impressions, humans rely on categories and concepts. These cut up our perception, making it possible to treat discriminable stimuli as equivalent. As we are a linguistic species, the words we utter all refer to categories and concepts and their relations. Of all categories that humans use, color categories have a special status for different reasons. Color categories are perceptual categories, they relate to the direct perception of chromatic stimuli. Color categories are also undoubtedly linked to color terms; lexicalized color categories are stronger than unlexicalized ones and are consequently more used, more rapidly recalled and better remembered. There is also a growing body of evidence suggesting that color categories are *universal*: Very different cultures seem to have surprisingly similar color categories.

In this paper we present the hypothesis that this universal character is the product of a number of constraints, most importantly the nature of human color perception and the influence of language on category acquisition, that act together during a cultural acquisition process. We present a computational model to illustrate and support our hypothesis.

1.1 Insights from Cross-cultural Studies

Some languages have 11 or more basic color terms[1], while other languages—typically spoken by non-industrialized societies—have fewer (the Dugum Dani of Papua New Guinea are reported to have only two

Correspondence to: Tony Belpaeme, University of Plymouth, School of Computing, Communication and Electronics, A318 Portland Square, Plymouth PL4 8AA, UK. *E-mail*: tony.belpaeme@plymouth.ac.uk
Tel.: +44-1752-232516, *Fax*: +44-1752-232540.

(2005), Vol 13(4): 293–310.
[1059–7123(200512) 13:4; 293–310; 059324]
Figure 3 appears in color online: http://adb.sagepub.com

color terms: one for dark/cool colors and one for light/warm colors; Rosch-Heider, 1972). Over the decades, research into the nature of color perception and color categorization has unveiled much about how humans perceive chromatic stimuli, how they categorize color and how language has an impact on color categorization. But the most striking aspect of color categories is that there is a remarkable cross-cultural similarity between categories. Papuans and Inuits for example, even though geographically and culturally isolated from each other, entertain roughly the same distinctions of the color spectrum (Kay, Berlin, Maffi, & Merrifield, 2003).

However, there is also experimental evidence showing that color categories are not determined at birth, but are to a certain extent influenced by our experiences and by the language we speak (see, for example, Roberson, Davies, & Davidoff, 2000). So on the one hand there are universal tendencies and on the other hand there is evidence of an effect of experience and language on color categorization. This article tries to reconcile both and shows that these positions need not be orthogonal.

The universal character of color categories was first reported by Berlin and Kay (1969). In their experiments Berlin and Kay elicited color terms from native speakers of 20 different languages and asked them to mark the color terms on a chart containing a spectrum of color chips. When comparing the results of the 20 languages, Berlin and Kay noted that large regions of the color chart remained unnamed, whereas a restricted area of the color chart contained the foci of the color terms of all 20 languages. Recently the World Color Survey, a large-scale replication of Berlin and Kay's experiments for which data has been collected from 110 pre-industrial societies, has reconfirmed the universal character of color categories (Kay & Regier, 2003; Kay et al., 2003; Regier, Kay, & Cook, 2005). In summary, the centres of color categories of most cultures tend to fall in approximately the same positions; these are the positions known in English by the basic color terms black, white, red, yellow, blue, green and so forth.

A straightforward explanation for this universal character is that color categories are innate. They are in a certain sense "hard wired" in the human brain. One view is that the categories are directly genetically encoded, and that every human being possesses a fully fledged repertoire of color categories, even though not all categories might be lexicalized (Rosch-Heider, 1972). More subtle innatist accounts argue that certain neurophysical structures might be responsible for universal color categories. Indeed, humans are a trichromatic species, meaning that anyone with normal color vision has the same three types of color-sensitive receptors in the retina. Or, humans invariably process color in an opponent manner, placing white against black, blue against yellow and green against red. These shared neurophysical properties of color perception could possibly explain the shared categorization of color.

Arguments for the nativist position abound. Shepard (1992) for example argues that color categories are internalized during human evolution because colors and their use in our survival have always been largely constant. Ecological constants—such as gravity, earth rotation, but also color—are likely to have been absorbed in our genome during millions of years of evolution. Durham (1991) considers color categories to be a function of neural constraints: "Regularities in the linguistic encoding of color result from regularities in the neural coding of color in the brain…" (p. 218). Kay and McDaniel (1978) make a case for six innate primary categories, resulting from opponent color processing. All other color categories are a fuzzy set recombination of these six primaries. However, this model has been disputed for making an unwarranted leap from neurophysiology to categorization (Dedrick, 1998). Also, the opponent colors on which the model is based do not seem to correspond to red/green and yellow/blue, but to cherry/teal and chartreuse/violet (Jameson & D'Andrade, 1997). Bornstein, Kessen, and Weiskopf (1976) and Davies and Franklin (2002) have shown that prelinguistic infants react to color categories much in the way adults do: This makes a case for color categories being innate or at least being available at an early age without language having an influence on the acquisition of color categories. Similar and other flavors of nativist positions can, among others, be found in Rosch (1975), Bornstein (1985), Hardin (1988), Kay, Berlin, and Merrifield (1991), Soja (1994), Boynton (1997), Valberg (2001), and Matsuno, Kawai, and Matsuzawa (2004).

Yet, the last decade has been witness to an increasing number of scholars that have been refining the nativist position or even blatantly rejecting it. Davidoff, Davies and Roberson (1999) and Roberson et al. (2000) report on results from experiments with natives from a Papua New Guinean tribe. The natives' performance during color memory tests does not support nativism; on the contrary it would seem that color categories are formed under influence of language. Kay and Kempton (1984) and Davies (1998) show how speakers of a language having only one color term covering both blue and green subjectively see the difference between green and blue stimuli as being smaller than speakers of a language having two color terms for blue and green. This demonstrates how language has an impact on color judgement. More arguments against an innatist account can be found in Saunders and van Brakel (1997), Jameson and D'Andrade (1997), Jameson (2005), Roberson, Davidoff, Davies, and Shapiro (2005), Steels and Belpaeme (2005), and Roberson (2005).

1.2 Insights from Infant Color Cognition

Next to the previously mentioned cross-cultural studies, there is also a body of studies of child color categorization which might help unveil the nature of color categories and of perceptual categories in general. Bornstein et al. (1976) show how 4-month-old (and thus prelinguistic) children respond categorically to monochromatic color stimuli: They habituated the 4-montholds to a color stimulus, and then showed either a stimulus from a different category or a stimulus from the same category, in both cases with the same difference in wavelength between the original and new stimulus. The children reacted more to a stimulus from a different category, lending support to the idea that color categories are innate. Recently Franklin and Davies (2004), using an improved version of this memory task using reflective color cards instead of monochromatic lights, reconfirmed the categorical color perception of prelinguistic children. Franklin, Pilling, and Davies (2005) show how adults and 20-week-old infants react similarly in a search task: A colored dot was shown on a green background and the eye gaze of subjects was recorded. The dot color was different from the background by a fixed color difference and was either of the same category (a different green) or from a different category (blue). Both adults and infants took longer to fixate on the dot when its color was in the same category, demonstrating how infants have adult-like color categorization. These findings are "problematic for the idea that color categorization is linguistically constructed" (Franklin et al., 2005, p. 244). However, when color categories are innate they can still be plastic and under the influence of language: Learning color terms could reorganize the already present color categories. There is abundant evidence for other domains that language indeed influences cognition (Bowerman & Levinson, 2001).

If color categories are indeed available at birth it is all the more puzzling why children have difficulties acquiring, comprehending and producing color terms. At an age when children actively use a few hundreds of words, they still struggle with the correct use of color terms. Charles Darwin reported on this and found it remarkable that his children knew the names of all common objects but could not correctly name colors. Darwin wrote: "I distinctly remember declaring that they were colour-blind" (Darwin & Seward, 1905, letter 416, p. 47). Recent generations of children have gained competence in color term usage, probably due to earlier and more frequent exposure to color and color terms (Shatz, Behrend, Gelman, & Ebeling, 1996), but still the relatively late development of linguistic color cognition is remarkable (Bornstein, 1985; Andrick, & Tager-Flusberg, 1986; Soja, 1994; Mervis, Bertrand, & Pani, 1995; Braisby & Dockrell, 1999; Pitchford & Mullen, 2001). Several suggestions have been made to explain this late development of color cognition. Andrick and Tager-Flusberg (1986) suggest that color categories are not sharply delineated, as opposed to naturalkind categories. Color categories tend to overlap, and this might hinder the correct usage of color terms. On a similar account, Braisby and Dockrell (1999) suggest that natural-kind lexicalization is facilitated by the presence of similar but non-member examples, while the color domain does not have similar but non-member examples: Color has vague semantics. For example, a cat is in many aspects similar to a dog, but is not a member of the concept DOG and cannot gradually change into DOG. This is not so for color: Green is not yellow, but can gradually change into yellow. Soja (1994) indicates that children might have the necessary concepts for colors, but lack a language-specific mechanism which hampers associating words with colors. Sandhofer and Smith (1999, 2001) suggest two steps in developing color cognition, where children first learn color words without making a proper mapping to the underlying color concepts and then learn to correctly use those color concepts non-linguistically. Pitchford and Mullen (2001) suggest that late color cognition might be due to a general delay in conceptualizing abstract object attributes. Children seem to have an equally hard time conceptualizing shape, speed and size (although Sandhofer and Smith (1999) did observe a different developmental pattern for color and size) but readily conceptualize and name objects with functional significance.

For the models presented in this work, we assume color categories to be learnt and plastic and not to be innate and fixed. Also, we assume that linguistic communication has an influence on color categories, as suggested in Sandhofer and Smith (2001).

2 Constraints on Color Categorization

If color categories are not genetically determined, but instead acquired by an ontogenetic process or by a learning process, one would naively expect color categories to be arbitrary between different cultures and different environments. As shown by the World Color Survey field data (Kay & Regier, 2003; Kay et al., 2003; Regier et al., 2005) this is not the case: The centers of color categories in different cultures seem to gravitate towards the centers of color categories named with the basic English color terms white, black, red, yellow, green, blue, brown, gray, pink, orange and purple.

The World Color Survey results are based on an analysis of anthropological data collected from 110 non-industrialized societies. Informants were asked to name 330 color chips. When the chips are ordered along their hue and value (the value corresponds to the intensity of the chip), the chips form a two-dimensional color chart, also known as a Munsell chart[2]. The chart goes from red (left side of the chart) over orange, yellow, green, blue, purple to pink (on the right side). Chips at the top of the chart are lightest, the lower the chip, the lower its intensity; chips at the bottom are darkest (note that the WCS data do not show achromatic—that is white, gray or black—chips). Figure 1 shows a histogram reproduced from Kay and Regier (2003)

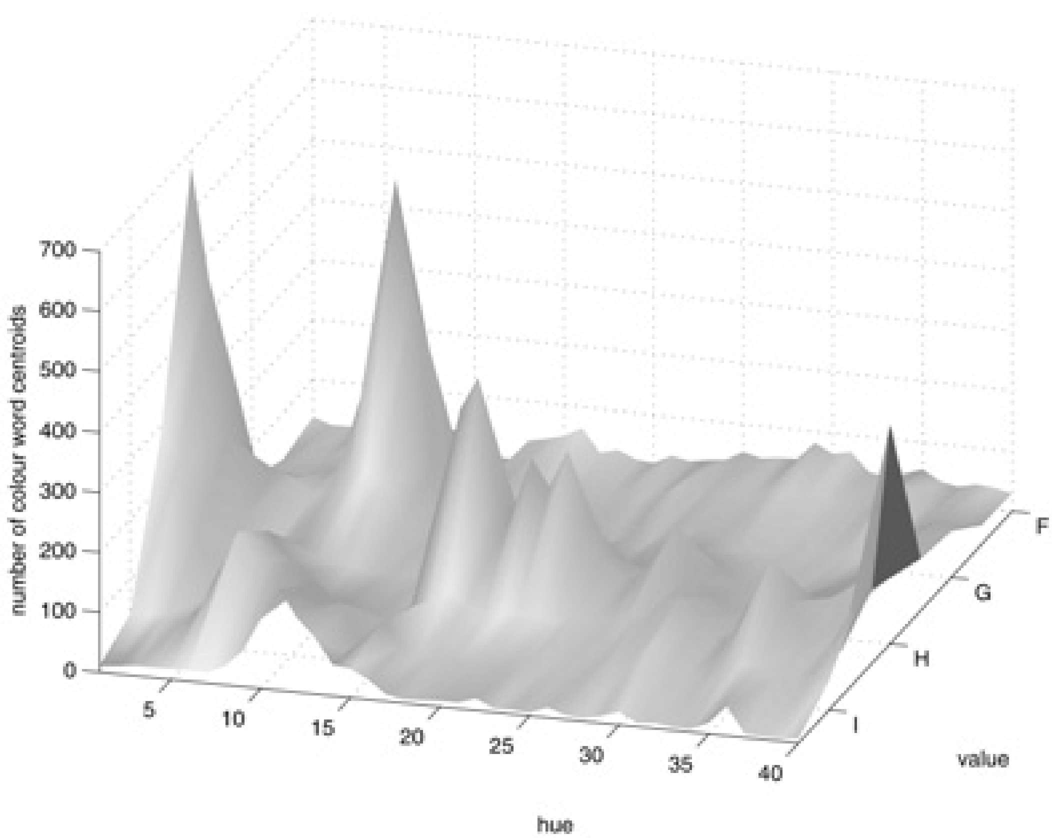

Figure 1 Histogram showing the linguistic color categories for 110 languages spoken in non-industrialised societies (data from Kay & Regier, 2003).

for the centroids[3] per color term of all languages. The floor plane of the histogram corresponds to the ordered Munsell chart (without the achromatic chips). If all cultures were to have arbitrary color categories the histogram would be expected to be flat. Instead, the histogram contains obvious peaks at regions close to the English color terms pink/red, brown, yellow, green, blue and purple. Kay and Regier (2003) showed that the probability of these peaks arising by chance is extremely low.

To explain the non-uniform distribution of color categories under ontogenetic development and learning, several "biases" have been suggested which drive color categories to an optimal configuration. A short overview is provided here.

The psychophysical and neurophysiological make-up of human color perception could form a bias on the acquisition of color categories. Jameson and D'Andrade (1997) have argued that the structure of the perceptual color space can explain the location of color categories. The perceptual color space, often presented as a three-dimensional space, is a representation of the psychological color experience of the average human. What is remarkable is that the space is neither spherical nor symmetrical, instead the color space contains a number of regions that are more pronounced. These are the regions in which color categories tend to end up during an optimization process. Yendrikhovskij (2001), using computational experiments, showed how the natural environments contained a certain distribution of chromatic stimuli that could explain the nature of human color categories. He used a *k*-means clustering algorithm to cluster the color information of pixels drawn from photographs of natural scenes. Yendrikhovskij reports how the extracted clusters corresponded to human categories. Griffin (2004) used a computational model in which he implemented an odd-one-out experiment: From three objects, one object is selected on the basis of color information as being least similar to the two others. Griffin showed that English color categories are better suited at performing this task than alternative color category configurations. Dowman (2003), also using a computational model, demonstrated how typological patterns in color terms could be the product of learning biases acting on a diachronic communication system.

Probably the most important bias on category acquisition, but also the most controversial, is language. Language is undeniably crucial in learning abstract concepts. But for many it has been hard to accept the influence of language on the acquisition of perceptual concepts. Nevertheless, linguistic relativity effects have been reported, among others, for numerical reasoning (Gumperz & Levinson, 1996), spatial relations (Gumperz & Levinson, 1996; Choi, McDonough, Bowerman, & Mandler, 1999), and time (Gentner & Boroditsky, 2001).

The effect of culture, or more specifically, language on color categories has been reported and defended by, among others, Kay and Kempton (1984), Davidoff (1999), and Roberson et al. (2000). However, if color categories are under direct influence of language with little or no other constraints, and culture is arbitrary, we would expect to see wildly different color categories in linguistically and culturally unrelated societies. Field data show this is not the case. So, if color category acquisition is influenced by language, what constraints are there to steer color categories towards the same configuration in every culture?

3 Computational Model

In this article we present a computational model of color category acquisition. Using this model we investigate whether the acquisition of color categories under certain circumstances exhibits a "universal character." In other words, can we explain the universality of color categories as resulting from an acquisition process? Specifically, we are interested in what general tendencies can be observed when the color categories are acquired using language.

3.1 Internal Structure of an Agent

The model is agent-based, meaning that individuals are implemented as software agents. Each software agent has the capacity to perceive color stimuli, to categorize percepts, to lexicalize categories and to utter color terms to other agents. Perception is implemented as a mapping from RGB to a perceptual color space. The categorization is implemented as a point representation together with a distance measure on that color space. And finally, the lexicalization is based on associating color terms with color categories using variable association strengths. We describe each of these abilities in detail in the following sections.

3.1.1 Perception of Color Stimuli Color stimuli are presented to an agent as *RGB* values. *RGB* is a technical color representation suited for reproducing color on display devices. It is quite remote from how humans would quantify colors and therefore an agent converts the *RGB* representation into three values in the CIE $L^*a^*b^*$ color space (see Appendix A). The L^* dimension corresponds to the lightness of the color, while the a^* and b^* dimensions respectively correspond to a red–green and a yellow–blue dimension. The CIE $L^*a^*b^*$ color space is a "perceptually equidistant" color space, meaning that distance in CIE $L^*a^*b^*$ corresponds to psychological dissimilarity (Fairchild, 1998). The similarity between two different colors can be quantified by taking the inverse of the Euclidean distance between their two CIE $L^*a^*b^*$ values. The CIE $L^*a^*b^*$ forms the color representation space for an agent and it is in this space that categories will be defined.

3.1.2 Categorization Color categories, as many perceptual categories, are prototypical in nature (Rosch & Lloyd, 1978). They have a maximal membership for one particular stimulus; for other stimuli the membership decays gradually. A color category is also only sensitive to only one region of the color space, and never to two or more unconnected regions. Furthermore, perceptual categories exhibit a magnet effect: Stimuli near a category are subjectively perceived to be closer to that category than they objectively are.

We have opted to represent a color category as a point in the CIE $L^*a^*b^*$ space; the membership function of a category is then the inverse of the Euclidean distance of a stimulus to the category. Previous implementations of color categorization (Belpaeme, 2001; Steels & Belpaeme, 2005) have used radial basis function networks (RBFN) to represent a category. A RBFN representation is more complex, and allows for non-symmetric and non-convex category memberships. However, a simple point representation is easier to implement and faster to compute without sacrificing the most essential properties of color categorization: Prototypicality and maximal sensitivity to only one location in the color space (Belpaeme & Bleys, 2005).

3.1.3 Lexicalization and Communication The agents lexicalize categories by associating a category with a color term: The strength between a category and a term is given by a value $s \in [0, 1]$. Categories can be associated with more than one term, allowing for synonymy, and a term can be associated with more than one category, allowing for homonymy. The lexicon of an agent can be seen as a matrix containing the strength of the association between categories and terms. An example of an association matrix is

$$\begin{array}{c c} & \begin{array}{cccc} t_1 & t_2 & \dots & t_m \end{array} \\ \begin{array}{c} c_1 \\ c_2 \\ \vdots \\ c_n \end{array} & \begin{pmatrix} 0.1 & 0.6 & \dots & 0.0 \\ 0.0 & 0.1 & \dots & 1.0 \\ \vdots & \vdots & s_{ij} & \vdots \\ 0.1 & 0.0 & \dots & 0.0 \end{pmatrix} \end{array} \tag{1}$$

When an agent needs to select a term to express a category, it picks the term with the highest strength. Vice versa, when a term needs to be interpreted, the agent uses the category having the highest strength. This implements a simple winner-takes-all strategy. Note that rows and columns in the association matrix do not have to add to one.

When agents communicate, it is a term t that is conveyed from one agent to the other.

3.2 Interactions Between the Agents

The dynamics of the simulation are implemented in two simple scenarios. The first scenario, dubbed the *discrimination game*, serves to let an agent build a repertoire of categories in order to discriminate color stimuli. In the second scenario, two agents play a *guessing game*; this serves to acquire a lexicon and adapt their categories. Both games are explained below.

3.2.1 The Discrimination Game During a discrimination game an agent A is shown a set of N color stimuli; this is called the *context* O. From the context one stimulus is selected; this is the *topic* o_t. If the topic is uniquely associated with a category—that is, if the category only matches the topic and no other stimulus in the context—then the discrimination game is a success. If not, this provides an opportunity to improve the agent's repertoire of categories (Steels, 1997a; Belpaeme, Steels, & van Looveren, 1998). The protocol for the discrimination game is summarized in algorithm 1.

1: Agent A chooses a topic o_t from the context $O = \{o_1, \ldots, o_N\}$
2: Agent A perceives each stimulus in the context by constructing an internal representation for it: $\{o_1, \ldots, o_N\} \to \{r_1, \ldots, r_N\}$
3: For each internal representation r_i, the best matching category is found. This is the category which has the highest output for r_i of all the categories available in the category repertoire of A and which we will denote by c_i: $\{r_1, \ldots, r_N\} \to \{c_1, \ldots, c_N\}$
4: If the best matching category for the topic is unique: $\text{count}(c_t, \{c_1, \ldots, c_N\}) = 1$, the game succeeded, otherwise it has failed.
5: If the game failed, the agent adds a new category or adapts the best matching category c_t (see text).

Algorithm 1 Discrimination Game (A, O)

When a game fails, the decision to add a category or adapt a category depends on a number of conditions. When the category repertoire is empty, a new category is created on the internal representation r_t of the topic. When no discriminating category can be found, there are two possible actions: (1) A new category is created on r_t or (2) the best matching category c_t is adapted to better represent the internal representation of the topic r_t ; this is done by shifting c_t towards r_t as in Equation 2; α is the learning rate, set by default to 0.7:

$$c_t \leftarrow c_t + \alpha(r_t - c_t) \qquad (2)$$

Option (1) is taken when the discriminative success (see below) of the agent is below a threshold $\theta_{\text{adapt}} = 0.90$, otherwise option (2) is used.

In a nutshell, during a discrimination game an agent adds and shifts categories to obtain a repertoire that can discriminate stimuli. If an agent plays a sufficient number of games, it will end up with categories that are sufficient for discriminating color stimuli with at least θ_{adapt} success. This learning rule is a variation on competitive learning (see for example Rumelhart & McClelland, 1986).

3.2.2 The Guessing Game A guessing game is played between two agents randomly selected from the population. One agent acts as *speaker*, the other as *hearer*. Both the speaker and hearer are offered the same set of color stimuli (the context O) of which only the speaker knows the topic o_t. The speaker now has to communicate to the hearer what the topic is. The guessing game allows agents to acquire a repertoire of lexical labels and serves to couple the categorical representations of the agents through their linguistic interactions (Steels, 1998; Steels & Kaplan, 1999; Belpaeme, 2001; Steels & Belpaeme, 2005).

During a guessing game, the speaker observes the topic o_t and, through playing a discrimination game (DG), finds a discriminating category c^S for the topic (with superscript S for the speaker and H for the hearer), it then communicates the term t^S that has the highest strength s^S. The hearer hears $t^S = t^H$ and finds the category c^H having the highest strength s^H. The hearer then points to the object o_h which is closest to the category c^H. If o_t and o_h are the same object, the game succeeds. The speaker and hearer now adapt their strengths s^S and s^H according to a learning rule (see below), and the hearer shifts category c^H towards the topic (as in Equation 2). The protocol for a successful guessing game is summarized in algorithm 2.

The guessing game can fail at many points; each failure is an opportunity to change the category repertoire and the lexicon of the agents:

(1) The speaker's discrimination game fails. In this case, a category is added or shifted, as described above. The guessing game stops here.
(2) The speaker does not have a term associated with c^S. In this case, it creates a term t^S consisting of characters drawn randomly from an alphabet. The term is added to the agent's lexicon and associated to the category c^S with a default strength s^S = 0.5. The game continues from here on.
(3) The hearer does not know the communicated term t^S. This is signaled to the speaker, whereupon the

	speaker A_S		hearer A_H
	chooses topic o_t		
	plays DG for o_t		
	DG succeeds and returns c^S		
	finds term t^S for c^S		
	utters t^S	$\rightarrow t^S \rightarrow$	hears $t^S = t^H$
			finds category c^H for t^H
			finds o_h closest to c^H
	sees o_h	$\leftarrow o_h \leftarrow$	points to o_h
	$o_t = o_h$		
	updates s^S using Equation 3		
	points to o_t	$\rightarrow o_t \rightarrow$	sees o_t
			updates s^H using Equation 3
			adapts category c^H to r_t using Equation 2

Algorithm 2 Guessing Game(A_S, A_H, O)

speaker points to the topic o_t. The hearer plays a discrimination game for o_t and creates an association between c^H and $t^H = t^S$ with default strength $s^H = 0.5$. The speaker updates s^S according to Equation 4.

(4) The hearer fails to point out the topic. Both speaker and hearer update respectively s^S and s^H as in Equation 4.

There are two possible updates for the association strength between a category and a term. In the case of a successful game, Equation 3 is used. The strength s_{ij} is increased by a value $\delta = 0.1$, the competing associations s_{kl} are decreased by δ. This mechanism, called *lateral inhibition*, serves to avoid synonymy and homonymy[4]:

$$\begin{cases} s_{ij} = \min(s_{ij} + \delta, 1) \\ s_{kl} = \max(s_{kl} - \delta, 0) \text{ in row i and column} \\ \qquad \text{j with } k \neq i, l \neq j \end{cases} \tag{3}$$

When the game fails, the strength of association s_{ij} is decreased:

$$s_{ij} = \max(s_{ij} - \delta, 0) \tag{4}$$

The guessing game gets its real worth from being played by a larger number of agents, organized in a population. The agents are then coupled to each other through the guessing game and through the shared environment. The number and nature of the categories and forms of the agents will depend on the environment: The number of color stimuli presented to the agents and the distance between the stimuli will influence the number of categories of the agents. When the task is made more difficult, by showing more stimuli or by showing stimuli that are less distinguishable, the agents will react by creating more categories.

The guessing game is one of many implementations of the *language game*. Language games are simple one-to-one communicative interactions between agents and have been used as computational models to study the evolution and dynamics of language (for example, Steels, 1997b; Zuidema & Westermann, 2003). Language games relying on cross-situational learning—a form of unsupervised learning, where no feedback is available on the outcome of the game—have been proposed as an alternative to guessing games (Vogt, 2000; Smith, 2001, and Smith, 2005). Language games have also been implemented on robotic agents to let the robots bootstrap a communication system (Vogt, 2000, 2003; Steels, 2001).

3.2.3 Performance To evaluate the performance of agents in successive language games, several measures exist. The discriminative success and communicative success report the performance of an agent's efficiency at discriminating and communicating respectively. The discriminative success gives the ratio of successful discrimination games over the total number of discrimination games for the last $n = 20$ games. The communicative success respectively reports the ratio of successful guessing games for the last *n* games.

Other measures exist which report on the quality of the lexicon in a population (Zuidema & Westermann, 2003). However, the goal of this work is not to showcase the dynamics and performance of language games, but to study the categorical repertoires that emerge from the agents' interactions. In previous work it has been demonstrated that the guessing game serves to let the agents acquire shared categories and shared lexicons, for more information the reader is referred to (Belpaeme, 2001; Steels & Belpaeme, 2005).

4 Results

We aim to study color categorization as the product of cultural/linguistic interactions in a group. If culture is largely arbitrary, then color categories are expected to be arbitrary as well (Roberson, 2005). In this view, constraints on color perception and categorization alone are not sufficient to steer categories of different cultures towards the same constellation: Two separate populations will end up with different color categories, even if the populations start out under the same conditions.

As the constraints formed by the embodiment and the environment are small, it might well be that their effect is only noticed on a larger scale. When comparing only two populations an observer might have the impression that their color categories are dissimilar and thus arbitrary. To avoid this pitfall we study a larger number of populations.

The yardstick for our experiments are the results from the World Color Survey (WCS) as reported in Kay and Regier (2003). In the WCS the linguistic color categories of 110 different languages were reported. A cross-cultural summary, where the color categories for each term of every language were combined into one histogram, showed "there are clear cross-linguistic statistical tendencies for named color categories to cluster at certain privileged points in perceptual color space." A contour plot[5] of the data presented in Figure 1 is shown in Figure 2. It shows how the 110 languages indeed have categories that cluster in specific regions of the color chart.

Two experiments are reported here. One centers on the acquisition of color categories *without* linguistic pressure, the other on acquisition of color categories *with* language.

4.1 Color Stimuli for the Agents

The color stimuli presented to the agents are pixels extracted from images. Two data series are used: One containing a *uniform* chromatic distribution and one a

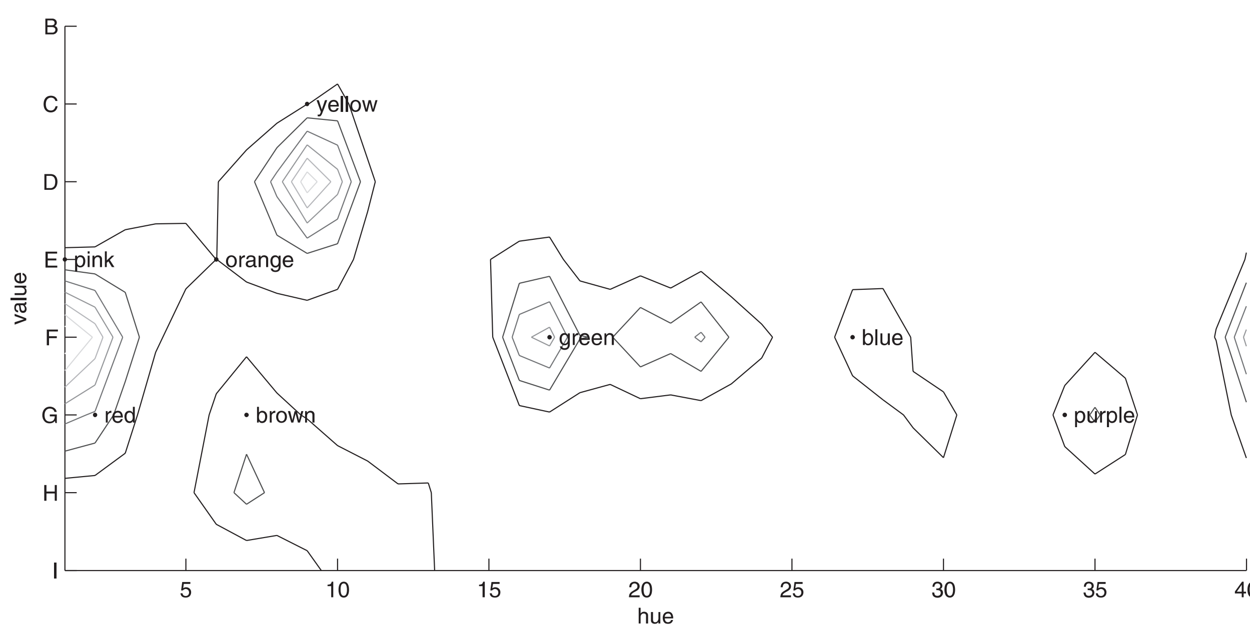

Figure 2 Contour plot of the WCS data.

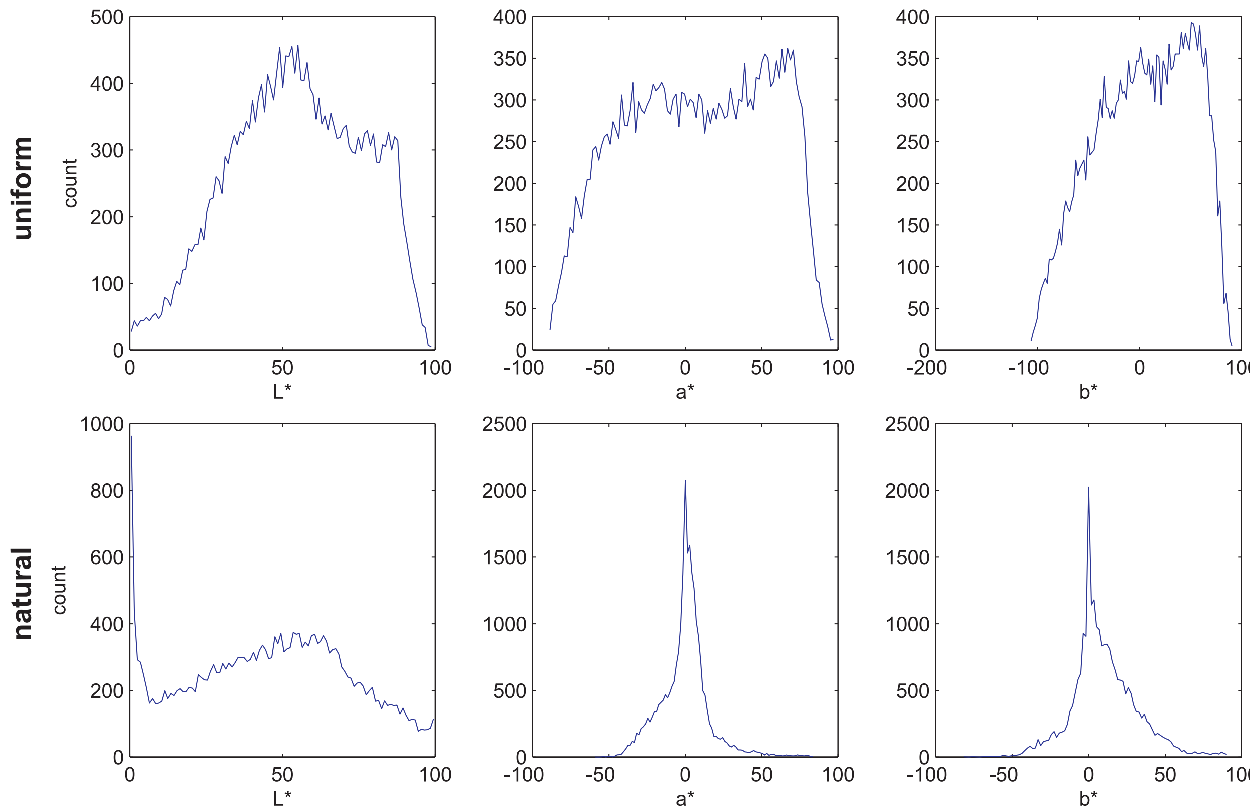

Figure 3 The 25,000 color stimuli fed to the agents plotted as histograms in the L^*, a^* and b^* dimensions. Top row of histograms shows the uniform chromatic distribution, bottom row shows the natural distribution. The natural distribution contains more unsaturated (i.e. brownish and grayish) colors and contains relatively more dark colors.

natural chromatic distribution. The uniform distribution consists of 25,000 pixels drawn with uniform probability from the *RGB* space[6].

For the natural distribution we start from 300 photographs (resolution 640 × 480 pixels) from on-line photo galleries, from which 25,000 pixels are randomly extracted[7]. Figure 3 shows histograms in the CIE $L^*a^*b^*$ space of the 25,000 color stimuli for both distributions. The natural distribution has an abundance in low saturated and achromatic stimuli.

4.2 Learning Color Categories Without Language

Here we study the color categories of agents as they learn only from a discrimination task. The only constraints present are (1) the neuropsychological constraints on chromatic perception as modeled by the CIE $L^*a^*b^*$ space, (2) the environmental constraints modeled by the number of stimuli presented simultaneously in a context to the agents and the minimal distance between these stimuli, and (3) constraints posed by the categorization process. The agents do not communicate, so there are no linguistic or communicative constraints.

Each population consists of 10 agents playing 15,000 discrimination games. We run 105 simulations, each with a different random seed. The context presented to the agents consists of three stimuli, the distance between the stimuli varies between 40 and 60 in the CIE $L^*a^*b^*$ space[8]. The different agent populations could be seen as 105 different societies where individuals acquire their color categories without language having an impact on the acquisition process. In each population all agents end up with a repertoire of categories which is sufficient to discriminate color stimuli with at least 90% success.

The results are presented as contour plots, the floor plane again being the Munsell color chart. For this, all categories of the agents are first mapped from the CIE

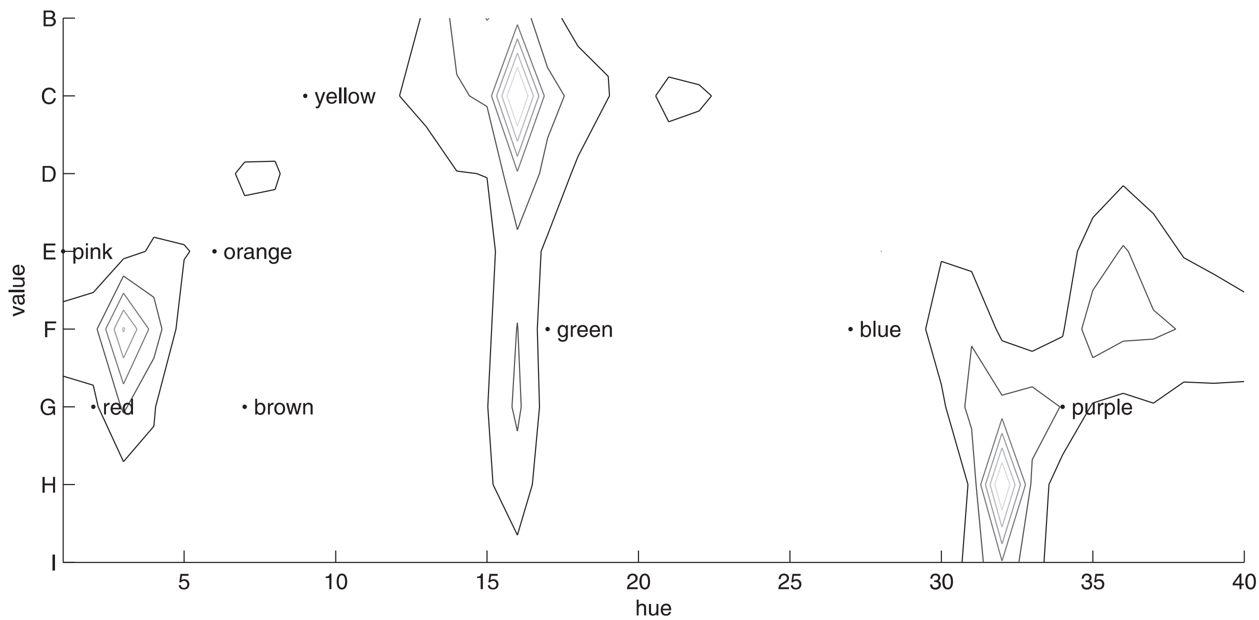

Figure 4 Contour plot of the categories acquired from a *uniform* chromatic distribution, *without* linguistic constraints.

$L^*a^*b^*$ space to their corresponding Munsell coordinates, consisting of a row r and column c (see Appendix B).

As mentioned in Section 3.1.1, the agents are exposed to two different chromatic distributions: A uniform distribution, containing no environmental biases, and a natural distribution, containing environmental biases as present in natural scenes. Figures 4 and 5 show contour plots of the histograms of the categories of all agents in 105 simulations. For reference, the contour plots are annotated with the location of eight English color terms. Figures 4 and 5 show the results for

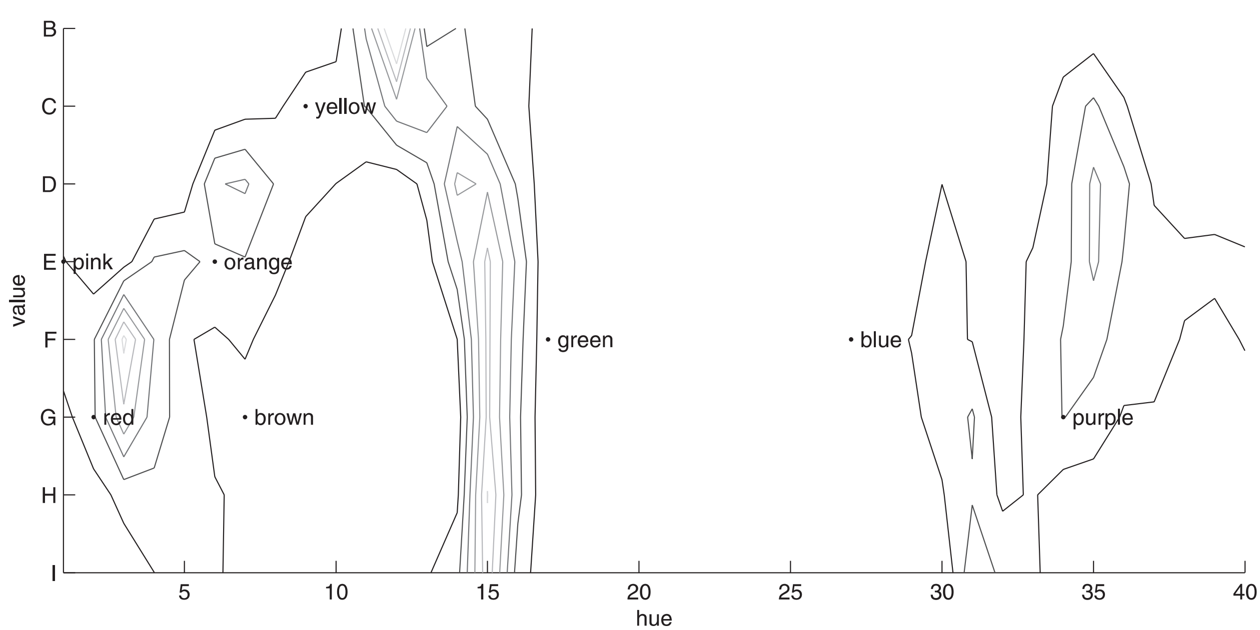

Figure 5 Contour plot of the categories acquired from a *natural* chromatic distribution, *without* linguistic constraints.

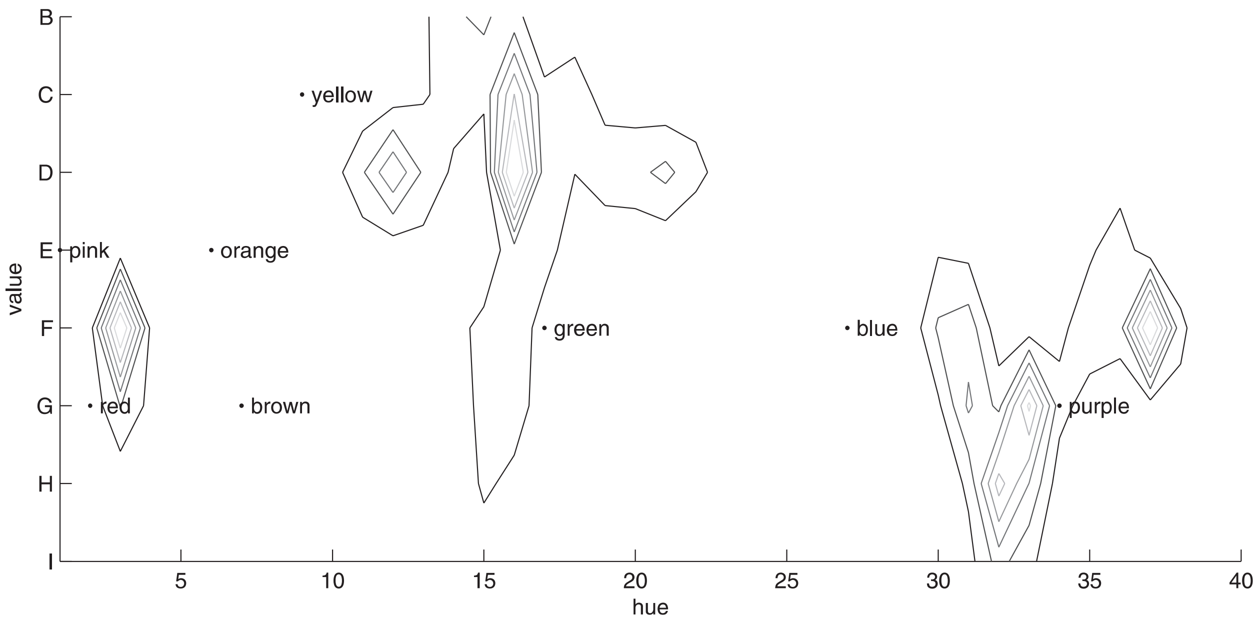

Figure 6 Contour plot of the categories acquired from a *uniform* chromatic distribution, *with* linguistic constraints.

stimuli drawn from respectively a uniform and a natural distribution.

4.3 Learning Color Categories with Language

In addition to the constraints posed by the agents' embodiment, environment and categorization, we now introduce communicative constraints. The agents now play guessing games instead of discrimination games. The parameter settings are identical to those in Section 4.2. The agents now end up with a repertoire of categories and associated color terms sufficient to communicate color stimuli with at least 83% success[9].

The results show contour plots of the categories of all agents of all 105 languages. Figure 6 shows the results when agents are offered stimuli drawn from the uniform chromatic distribution, and Figure 7 when drawn from the natural distribution.

4.4 Analysis

To compare the simulation results with the data from the World Color Survey, we first find the location of the highest peaks in all histograms. This is done by a search for local maxima in the histogram. Local maxima are connected components of histogram values with the same value v, whose external boundary values all have a value less than v. Next we compare the 10 highest peaks of the World Color Survey data with the 10 highest peaks of our simulation data. This is done by computing the undirected Hausdorff distance between the highest peaks. The undirected Hausdorff distance is used in computer vision to compute the distance between two images and is useful in this context as it is unsensitive to translations and scaling of data points (e.g., Rucklidge, 1997). The undirected Hausdorff distance $H(A, B)$ between two sets of coordinates A and B is computed as in Equation 5, with $d(s, t)$ being the Euclidean distance between coordinates s and t:

$$\begin{aligned} H(A, B) &= \max(h(A, B), h(B, A)) \\ h(S, T) &= \max_{s \in S}\left(\min_{t \in T}(d(s, t))\right) \end{aligned} \tag{5}$$

Table 1 shows the distances between the highest WCS peaks and the highest peaks of the simulation results. The best result is obtained when communicative constraints are present but environmental constraints are absent. This suggests that environmental constraints are more of a restriction than a blessing: If the agents are allowed to sample the whole color gamut, they form categories at locations that more closely resemble human color categories.

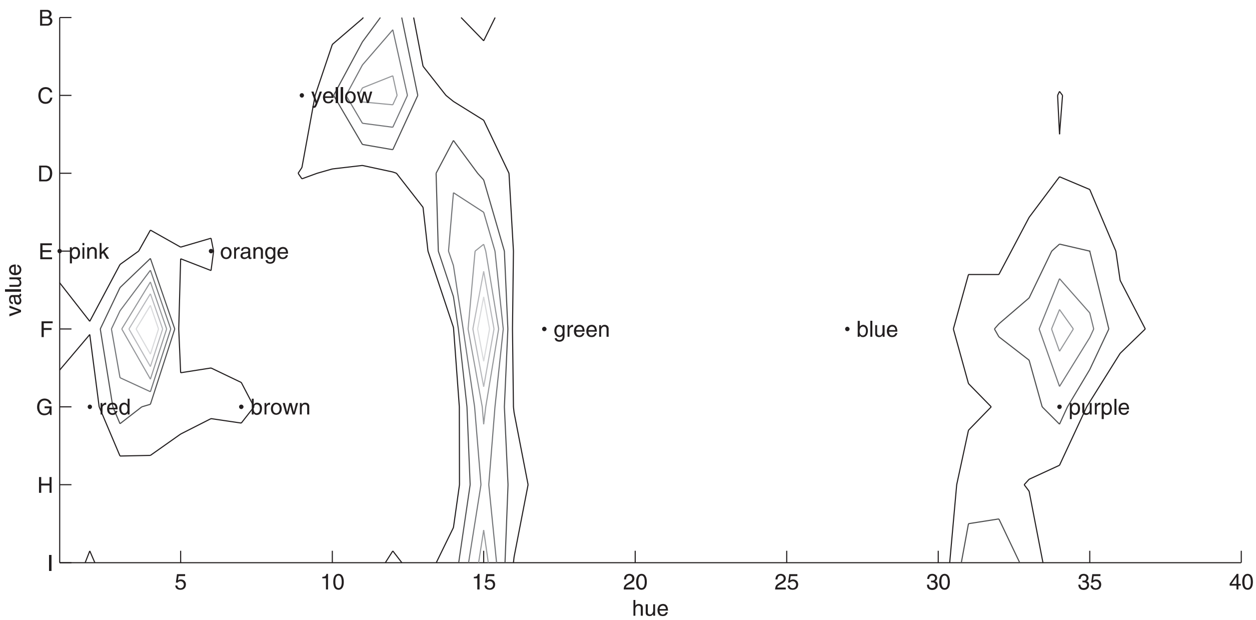

Figure 7 Contour plot of the categories acquired from a *natural* chromatic distribution, *with* linguistic constraints.

Table 1 Hausdorff distances between WCS data and simulation data.

x	$H(x, WCS)$
without language, uniform distribution	5.39
without language, natural distribution	7.00
with language, uniform distribution	5.10
with language, natural distribution	7.00

5 Discussion

The simulations serve to investigate whether it would be possible to learn a set of color categories (a) without resorting to innate, hard-wired color categories and (b) still reproducing the universal typology of color categorization. There are a number of constraints which could influence the nature of the color categories.

Constraints on embodiment. Biophysical constraints are implemented on the one hand by the CIE $L^*a^*b^*$ color model which restricts the positions of color categories and on the other hand by the categorization process itself. Categories serve to distinguish perceptual input and therefore, without any other constraints, will self-organize to be maximally distinct. Combined with the properties of the perceptual color space, categories end up at positions which are maximally distant from each other. This constraint is shared by all agents and already restricts the possible category configurations.

Ecological constraints. The ecological constraints are formed by two components. First, the task—i.e., discrimination of color categories—of the agents influences category acquisition. The size of the context or the distance between the stimuli in that context has a direct effect on the number of categories the agents acquire, which has an indirect influence on the configuration of the categories (Steels & Belpaeme, 2005). A second ecological constraint is formed by the environment, more specifically by the distribution of the color stimuli. In the experiments, two different distributions were offered to the agents: One uniform distribution (posing no environmental constraints) and one capturing a natural chromatic bias.

Communicative constraints. If agents are to communicate, their internal representations should be shared. If not, communicative interactions will fail as they are unable to convey the meaning of their perception to each other. This poses a third and important constraint on category formation. This constraint is implemented by the guessing game, which serves to couple the agents' categories through language.

Each one of these constraints is responsible for the structure observed in the histograms. Note that without any constraints the histograms would be flat (apart from some negligible structure due to noise). So, constraints on embodiment, ecology and communication can contribute to macroscopic effects. The question remains whether these constraints are enough to explain the universal character of human color categories.

Note that when comparing the results to the WCS data we observe a discrepancy: The peaks of the WCS data and those of the simulation results are not at the same positions (for example compare Figures 2 and 6). The result closest to the WCS data—with communicative constraints but without environmental constraints—has peaks at pink/red, green/yellow, blue and purple. Although this does not map one-to-one on the WCS data, it is nonetheless a structure reflecting an efficient configuration which cuts up the color continuum so as to maximize the distinctiveness of the categories (and consequently maximize the communicative accuracy).

Agents being fed natural chromatic stimuli end up with color categories that are less similar to human categories than agents being fed uniform chromatic stimuli. It seems that environmental constraints do not help in getting results near the WCS data. But at the same time this suggests that the distribution of colors presented to the agents can be optimized so that the agents end up with human-like color categories. Of course, this would be putting the cart before the horse.

The fact that agents when they are shown uniformly distributed colors end up with color categories that match better to the WCS data is actually not surprising. The WCS is a large collection of data geographically spanning the planet. As a whole, the subjects tested in the WCS will have been exposed to a broad spectrum of chromatic stimuli, and therefore one would expect that agents exposed to a uniform distribution will have a comparable color category typology.

To conclude, the results presented here do indeed suggest that an explanation for universal color categories does not need to resort to innate categories or innate mechanisms directly responsible for shaping categories. On the contrary, a cultural acquisition process on top of a combination of constraints or slight biases common to all humans might be enough to explain the universal nature of color categories and maybe even of other perceptual categories. Note that these common constraints alone are not enough to reach color categories that are similar between agents: The environmental and biological constraints are so weak that agents will end up with different color categories, so different that they would hamper communication (Steels & Belpaeme, 2005). A feedback mechanism is required to get the categories of different agents tuned to each other: We propose that category acquisition under linguistic influence fulfills this role.

Appendix A Conversion from *RGB* to CIE $L^*a^*b^*$

RGB values, ranging between [0, 1], are first converted to *XYZ* values using Equation 6; the conversion matrix is for PAL/SECAM viewing conditions with $\gamma = 2.5$. Next the *XYZ* values are converted to CIE $L^*a^*b^*$ values using Equation 7, with whitepoint $[X_n Y_n X_n]^T = [0.9504682\ 1.000\ 1.08883]^T$. For more information see Wyszecki & Stiles (1982) and Fairchild (1998):

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.430587 & 0.222021 & 0.0201837 \\ 0.341545 & 0.706645 & 0.129551 \\ 0.178336 & 0.0713342 & 0.939234 \end{pmatrix} \cdot \begin{pmatrix} R \\ G \\ B \end{pmatrix}^{\gamma} \tag{6}$$

$$\begin{aligned} L^* &= \begin{cases} 116\left(\frac{Y}{Y_n}\right) - 16 & \frac{Y}{Y_n} > \epsilon \\ 903.3\left(\frac{Y}{Y_n}\right) & \frac{Y}{Y_n} \le \epsilon \end{cases} \\ a^* &= 500\left(\mathrm{f}\left(\frac{X}{X_n}\right) - \mathrm{f}\left(\frac{Y}{Y_n}\right)\right) \\ b^* &= 200\left(\mathrm{f}\left(\frac{X}{X_n}\right) - \mathrm{f}\left(\frac{Z}{Z_n}\right)\right) \\ \mathrm{f}(x) &= \begin{cases} x^{\frac{1}{3}} & x > \epsilon \\ 7.787x + \frac{16}{116} & x \le \epsilon \end{cases} \\ \epsilon &= 0.008856 \end{aligned} \tag{7}$$

Appendix B Conversion from CIE $L^*a^*b^*$ to Munsell Chart Coordinates

We follow the procedure outlined in Kay and Regier (2003). Mapping a CIE $L^*a^*b^*$ value p is done by first finding the row r in the Munsell chart of which the L^* value is closest to the L^* value of p (L^* values are constant in each row of the Munsell chart). The chromatic chip having angle $C^* = \text{atan}\left(\frac{b^*}{a^*}\right)$ closest to p in row r is compared to the achromatic chip in row r. Eventually, the one having a radius $H^* = \sqrt{(a^{*2} + b^{*2})}$ closest to p is returned.

Acknowledgments

Tony Belpaeme has been in part sponsored as a postdoctoral fellow of the Fund for Scientific Research, Flanders (FWO Vlaanderen). The authors wish to thank Terry Regier, Paul Kay and Richard S. Cook for making the WCS data available. The authors would also like to thank the reviewers for their insightful comments.

Notes

1 Basic color terms are, among others, color words used solely for describing colors, are monolexemic and do not refer to anything else but color. For a full description see Berlin and Kay (1969).

2 A Munsell chart is a chart containing the 320 most saturated colors of the Munsell color space together with 10 achromatic chips (going from white over gray to black), making a total of 330 colors. The Munsell color space is a three-dimensional system to notate color, designed in the 1920s by artist Albert H. Munsell. Each color is represented by a hue, a value and a chroma: 5R 4/12 for example is a color with hue 5R, value 4 and chroma 12—a bright, saturated red. Anthropologists (among others Berlin & Kay, 1969) have used the most saturated colors of the Munsell space to make a convenient, ordered chart: The Munsell chart. The 330 colors on the Munsell chart are denoted with letters, from A to J, and numbers, from 0 to 40. For example, A0 is white, J0 is black, G4 is a bright red and F18 a grass green. A rendition of the Munsell chart can be found in Hardin and Maffi (1997) and Kay and Regier (2003).

3 The *centroid* of a color term t is found by averaging the colors denoted by t over all speakers of a language (see Kay and Regier, 2003).

4 Synonymy and homonymy—respectively associating one meaning with multiple words, and associating one word with multiple meanings—are very counterproductive when associating perceptual categories, such as color categories, with words. In human languages synonymy is rare, but homonymy occurs frequently, although homonyms almost always refer to meanings in unrelated semantic classes, e.g., *blue* can be a color or a mood. The authors have never encountered a report of color homonyms or even of color synonyms, and therefore the model avoids both. Some computational models on general vocabulary acquisition do not: For a study see Smith (2004).

5 Each contour plot shows seven contour lines at equally spaced intervals between 0 and the maximum value of the histogram.

6 For a uniform stimulus we randomly pick a *RGB* value and convert it to CIE $L^*a^*b^*$ space. However, this skews the distribution towards the greenish/bluish end of the spectrum (see top three histograms of Figure 3). A better mechanism would be to pick a L^*, a^* and b^* value, respectively in the range $L^* = [0, 100]$, $a^* = [-152, 152]$ and $b^* = [-127, 140]$; as these range values specify a cube, this would allow for some unrealistic CIE $L^*a^*b^*$ triplets. Therefore each stimulus should be checked to be within the Munsell color solid; if it is not it should be discarded. Even though the latter method produces a more accurate uniform distribution, we did not find any qualitative differences between the simulation results with one or the other uniform distribution.

7 Both datasets are available from http://www.tech.plym.ac.uk/SoCCE/staff/TonyBelpaeme

8 Five simulations are run with distance 40 between the stimuli, five with distance 41, and so on. The rationale behind this is that agent populations are exposed to different chromatic environments, some challenging, some less challenging. This is to discourage every population from being exposed to the same environmental "complexity." To give the reader an idea of the CIE $L^*a^*b^*$ distance between typical colors: Green–blue is 258, red–blue is 177, yellow–blue is 232, and yellow–green is 70.

9 Humans are also not perfect in communicating color. In a psychological experiment 55 Flemish (a Dutch dialect spoken in Northern Belgium) speakers were asked to communicate colors in a similar setting. On average, communicative success was 84% (Belpaeme, 2002).

References

Andrick, G., & Tager-Flusberg, H. (1986). The acquisition of colour terms. *Journal of Child Language, 13*(1), 119–134.

Belpaeme, T. (2001). Simulating the formation of color categories. In B. Nebel (Ed.). *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'01),* Seattle, WA, (pp. 393–398). San Francisco: Morgan Kaufmann.

Belpaeme, T. (2002). Communicating colour embedded in a colour context: Report on experiments on communicating colour meaning between Flemish informants. Technical Report 2002-10, Artificial Intelligence Lab, Vrije Universiteit Brussel.

Belpaeme, T., & Bleys, J. (2005). Colourful language and colour categories. In A. Cangelosi, and C. L. Nehaniv, (Eds.), *Proceedings of the Second International Symposium on the Emergence and Evolution of Linguistic Communication (EELC'05), Hertfordshire, UK*. The Society for the Study of Artificial Intelligence and the Simulation of Behaviour (pp. 1–7).

Belpaeme, T., Steels, L., & van Looveren, J. (1998). The construction and acquisition of visual categories. In A. Birk, & Y. Demiris (Eds.), *Proceedings of the Sixth European Workshop on Learning Robots*, Lecture Notes on Artificial Intelligence (Vol. 1545, pp. 1–12). Berlin: Springer.

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

Bornstein, M. H. (1985). On the development of color naming in young children. *Brain and Language, 26,* 72–93.

Bornstein, M. H., Kessen, W., & Weiskopf, S. (1976). Color vision and hue categorization in young human infants. *Journal of Experimental Psychology, 2,* 115–129.

Bowerman, M., & Levinson, S. C. (Eds.), (2001). *Language acquisition and conceptual development*. Cambridge, UK: Cambridge University Press.

Boynton, R. M. (1997). Insights gained from naming the OSA colors. In C. L. Hardin, & L. Maffi (Eds.), *Color categories in thought and language*. Cambridge, UK: Cambridge University Press.

Braisby, N., & Dockrell, J. (1999). Why is colour naming difficult? *Journal of Child Language, 26*(1), 23–48.

Choi, S., McDonough, L., Bowerman, M., & Mandler, J. (1999). Early sensitivity to language-specific spatial categories in English and Korean. *Cognitive Development, 14*(2), 241–268.

Darwin, F., & Seward, A. (Eds.), (1905). *More letters of Charles Darwin*. London: John Murray. 2 volumes.

Davidoff, J., Davies, I., & Roberson, D. (1999). Colour categories in a stone-age tribe. *Nature, 398,* 203–204.

Davies, I., & Franklin, A. (2002). Categorical perception may affect colour pop-out in infants after all. *British Journal of Developmental Psychology, 20,* 185–203.

Davies, I. R. (1998). A study of colour grouping in three languages: A test of the linguistic relativity hypothesis. *British Journal of Psychology, 98,* 433–452.

Dedrick, D. (1998). *Naming the rainbow: Colour language, colour science, and culture*, volume 274 of *Synthese Library*. Dordrecht: Kluwer Academic.

Dowman, M. (2003). Explaining color term typology as the product of cultural evolution using a multi-agent model. In R. Alterman, & D. Kirsh (Eds.), *Proceedings of the 25th Annual Meeting of the Cognitive Science Society, Boston, MA* (pp. 360–365). Cognitive Science Society.

Durham, W. H. (1991). *Coevolution: Genes, culture and human diversity*. Stanford, CA: Stanford University Press.

Fairchild, M. D. (1998). *Color appearance models*. Reading, MA: Addison-Wesley.

Franklin, A. & Davies, I. R. (2004). New evidence for infant color categories. *British Journal of Developmental Psychology, 22,* 349—378.

Franklin, A., Pilling, M., & Davies, I. (2005). The nature of infant color categorization: Evidence from eye movements on a target detection task. *Journal of Experimental Child Psychology, 91,* 227–248.

Gentner, D., & Boroditsky, L. (2001). Individuation, relational relativity and early word learning. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 215–216). Cambridge, UK: Cambridge University Press.

Griffin, L. D. (2004). Optimality of the basic colours categories. *Journal of Vision, 4*(8), 309a.

Gumperz, J. J., & Levinson, S. C. (1996). *Rethinking linguistic relativity*. Studies in the Social and Cultural Foundations of Language 17. Cambridge, UK: Cambridge University Press.

Hardin, C., & Maffi, L. (Eds.), (1997). *Color categories in thought and language*. Cambridge, UK: Cambridge University Press.

Hardin, C. L. (1988). *Color for philosophers: Unweaving the rainbow*. Indianapolis: Hacket.

Jameson, K. A. (2005). Why GRUE? An interpoint–distance model analysis of composite color categories. *Cross-Cultural Research: A Journal of Comparative Social Science, 39*(2) 159–204.

Jameson, K. A., & D'Andrade, R. (1997). It's not really red, green, yellow, blue: An inquiry into perceptual color space. In C. Hardin, & L. Maffi, (Eds.), *Color categories in thought and language* (pp. 295–319). Cambridge, UK: Cambridge University Press.

Kay, P., Berlin, B., Maffi, L., & Merrifield, W. R. (2003). *The world color survey*. Center for the Study of Language and Information, Stanford, CA.

Kay, P., Berlin, B., & Merrifield, W. (1991). Biocultural implications of systems in color naming. *Journal of Linguistic Anthropology, 1*(1), 12–25.

Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist, 86*(1), 65–79.

Kay, P., & McDaniel, C. (1978). The linguistic significance of the meanings of basic color terms. *Language, 54*(3), 610–646.

Kay, P., & Regier, T. (2003). Resolving the question of color naming universals. *Proceedings of the National Academy of Sciences USA, 100*(15), 9085–9089.

Matsuno, T., Kawai, N., & Matsuzawa, T. (2004). Color classification by chimpanzees (Pan troglodytes) in a matching-to-sample task. *Behavior and Brain Research, 148,* 157–165.

Mervis, C. B., Bertrand, J., & Pani, J. R. (1995). Transaction of cognitive-linguistic abilities and adult input: A case study of the acquisition of colour terms and colour-based subordinate object categories. *British Journal of Developmental Psychology, 13*(3), 285–302.

Pitchford, N. J., & Mullen, K. T. (2001). Conceptualization of perceptual attributes: A special case for color? *Journal of Experimental Child Psychology, 80,* 289–314.

Regier, T., Kay, P., & Cook, R. S. (2005). Focal colors are universal after all. *Proceedings of the National Academy of Sciences, USA, 102*(23), 8386–8391.

Roberson, D. (2005). Color categories are culturally diverse in cognition as well as in language. *Cross-Cultural Research, 39,* 56–71.

Roberson, D., Davidoff, J., Davies, I. R., & Shapiro, L. R. (2005). Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology, 50*(4), 378–411.

Roberson, D., Davies, I., & Davidoff, J. (2000). Color categories are not universal: replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General, 129*(3), 369–398.

Rosch, E. (1975). Universals and cultural specifics in human categorization. In R. Brislin, S. Bochner & W. Lonner (Eds.), *Cross-cultural perspectives on learning: The interface between culture and learning*. New York: Halstead Press.

Rosch, E., & Lloyd, B. (1978). *Principles of categorization, in cognition and categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rosch-Heider, E. (1972). Universals in color naming and memory. *Journal of Experimental Psychology, 93,* 10–20.

Rucklidge, W. J. (1997). Efficiently locating objects using the Hausdorff distance. *International Journal of Computer Vision, 24*(3), 251–270.

Rumelhart, D. E., & McClelland, J. L. (Eds.), (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*, Volume 1: Foundations. Cambridge, MA: The MIT Press.

Sandhofer, C. M., & Smith, L. B. (1999). Learning color words involves learning a system of mappings. *Developmental Psychology, 35*(3), 668–679.

Sandhofer, C. M., & Smith, L. B. (2001). Why children learn color and size words so differently: Evidence from adults' learning of artificial terms. *Journal of Experimental Psychology: General, 130*(4), 600–620.

Saunders, B. A., & van Brakel, J. (1997). Are there nontrivial constraints on colour categorization? *Behavioral and Brain Sciences, 20*(2), 167–228.

Shatz, M., Behrend, D., Gelman, S. A., & Ebeling, K. S. (1996). Colour term knowledge in two-year-olds: Evidence for early competence. *Journal of Child Language, 23*(1), 177–200.

Shepard, R. N. (1992). The perceptual organization of colors: An adaptation to regularities of the terrestrial world? In J. Barkow, L. Cosmides & J. Tooby (Eds.), *Adapted mind* (pp. 495–532). Oxford: Oxford University Press.

Smith, A. D. (2001). Establishing communication systems without explicit meaning transmission. In J. Kelemen & P. Sosik (Eds.), *Proceedings of the European Conference on Artificial Life (ECAL01), Prague* (pp. 381–390). Berlin: Springer.

Smith, A. D. (2005). The Inferential transmission of language. *Adaptive Behavior, 13*(4).

Smith, K. (2004). The evolution of vocabulary. *Journal of Theoretical Biology, 228*(1), 127–142.

Soja, N. N. (1994). Young children's concept of color and its relation to the acquisition of color words. *Child Development, 65*(3), 918–937.

Steels, L. (1997a). Construction and sharing perceptual distinctions. In M. van Someren, & G. Widmer (Eds.), *Proceedings of the European Conference on Machine Learning*, Berlin: Springer.

Steels, L. (1997b). The synthetic modeling of language origins. *Evolution of Communication, 1*(1), 1–34.

Steels, L. (1998). Synthesising the origins of language and meaning using co-evolution, self-organisation and level formation. In J. R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language: Social and cognitive bases*. Cambridge, UK: Cambridge University Press.

Steels, L. (2001). Language games for autonomous robots. *IEEE Intelligent Systems, Sept–Oct,* 17–22.

Steels, L., & Belpaeme, T. (2005). Coordinating perceptually grounded categories through language. A case study for colour. *Behavioral and Brain Sciences, 24*(8), 469–529.

Steels, L., & Kaplan, F. (1999). Situated grounded word semantics. In T. Dean (Ed.), *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI'99, San Francisco)*. San Francisco: Morgan Kauffman.

Valberg, A. (2001). Unique hues: An old problem for a new generation. *Vision Research, 41,* 1645–1657.

Vogt, P. (2000). Bootstrapping grounded symbols by minimal autonomous robot. *Evolution of Communication, 4*(1), 87–116.

Vogt, P. (2003). Anchoring of semiotic symbols. *Robotics and Autonomous Systems, 43*(2), 109–120.

Wyszecki, G., & Stiles, W. (1982). *Color science: concepts and methods, quantitative data and formulae* (2nd ed.). New York: Wiley.

Yendrikhovskij, S. N. (2001). Computing color categories from statistics of natural images. *Journal of Imaging Science and Technology, 45*(5), 409–417.

Zuidema, W., & Westermann, G. (2003). Evolution of an optimal lexicon under constraints from embodiment. *Artificial Life, 9,* 387–402.

About the Authors

Tony Belpaeme is a senior lecturer in intelligent systems at the University of Plymouth. He obtained a PhD in 2002 from the Vrije Universiteit Brussel (VUB), Belgium, where he then worked as a postdoctoral researcher at the Artificial Intelligence Laboratory headed by Luc Steels. He also held the position of guest professor at the same university. His current research interests include cognitive robotics, the evolution of language and its application to intelligent systems and concept formation. *Address*: University of Plymouth, School of Computing, Communication and Electronics, A318 Portland Square, Plymouth, PL4 8AA, United Kingdom. E-mail: tony.belpaeme@plymouth.ac.uk

Joris Bleys received a MSc in computer science in 2004 from the Vrije Universiteit Brussel (VUB), Belgium. He is a PhD student at the Artificial Intelligence Laboratory at the VUB. His research focuses on concept formation and emergent communicative systems in artificial systems. *Address*: Vrije Universiteit Brussel, Artificial Intelligence Laboratory, Pleinlaan 2, 1050 Brussels, Belgium. E-mail: jorisb@arti.vub.ac.be