COMPARATIVE COGNITION & BEHAVIOR REVIEWS

2010 Volume 5, pp 100 - 116

Cross-species Assessment of the Linguistic Origins of Color Categories

Jules Davidoff
University of London

Joël Fagot
CNRS-Université de Provence

This article considers the relation between language and categorical perception (CP) of color. Two opposite theories are reviewed, the universalist position arguing that categories are universal with an essentially biological origin, and the relativist position that holds that color categories are essentially arbitrary and derive from color terms of the speaker's language. A review of the human literature presents developmental, neuropsychological, cross-cultural, neuro-imaging and computer simulation evidence that CP of colors has at least partly linguistic origins. As animal studies also contribute to this debate, we then review evidence of CP in the visual and auditory domains, and pinpoint the inconsistencies of the literature. To make a direct comparison between humans and monkeys, experimental studies compared humans and baboons for their color thresholds and in a recognition memory task designed to assess CP of colors. Only humans showed better between-category than within-category discrimination performance, suggesting species differences in the processing of a color continuum. That study along with some of our previous research supports the theory of a linguistic origin for color categories in humans.

Keywords: color, perception, categorization, language

The role of language in promoting symbolic reasoning is often given as critical in considerations of the intellectual difference between humans and other primates (e.g., Premack, 1983; Thompson & Oden, 2000). However, if language is the critical factor, there is a non-obvious consequence. One ought to find cross-species differences in the apparently non-verbal task of color categorization because recent cross-lingual and neuroimaging investigations have shown a linguistic component to its performance (Davidoff, 2001; Davidoff, Davies & Roberson, 1999; Roberson & Davidoff, 2000; Roberson, Davidoff, Davies & Shapiro, 2004, 2005; Roberson, Davies & Davidoff, 2000; Siok, Kay, Wang, Chan, Chen, Luke & Tan, 2009; Tan, Chan, Kay, Khong, Yip & Luke, 2008; Thierry, Athanasopoulos, Wiggett, Dering & Kuipers, 2009; Winawer, Witthoft, Frank, Wu & Boroditsky, 2007; see also Kay & Kempton, 1984). In one of the neuroimaging studies, Tan et al. (2008) even showed activation of language areas of the brain for the simple, apparently purely visual, task of judging whether two colors were the same or different. We could therefore argue that one of the functions of language or more generally of symbols must be to allow within- and between-subject agreement on the range of colors to be defined by a color category. Computer simulations of color category learning come to a similar conclusion because robots with human color vision do not arrive at the same color categories as each other unless they communicate (Steels & Belpaeme, 2005). This language-based or "relativist" view about the formation of color categories is contrasted to an alternative view that color categories are universal (Berlin & Kay, 1969). In a large scale-review of many languages, Berlin and Kay (1969) showed that there were systematic regularities in

Jules Davidoff, Department of Psychology, Goldsmiths, University of London, Lewisham Way, London SE14 6NW, UK. Joël Fagot, Laboratory of Cognitive Psychology, CNRS-Université de Provence, Marseille France.

Correspondence concerning this article should be addressed to Joël Fagot at joel.fagot@univ-provence.fr.

ISSN: 1911-4745 doi: 10.3819/ccbr.2010.50005

the way colors were categorized. It is these two views that are contrasted as a distinction between nature and nurture.

For the relativist position, language is important for color categorization because that task is different to most other categorizations as has been argued by philosophers for centuries in their discussion of the Sorites paradox (Dummett, 1975; Wright, 1975; see also Roberson, Davidoff & Braisby, 1999). The essential difference is that color is categorized within a continuous dimension. The philosophical contention is that dividing a continuous variable into categories is not possible by observation alone; rather it requires some non-perceptual justification. A classic example concerns the color spectrum. Take two colors that are perceptually indistinguishable (say the two colors 630nm and 629.9nm that we would call red). Suppose the 629.9 nm color has its wavelength altered to 629.8 nm. As this color is perceptually indistinguishable from the previous, it would also have to be called red. There is soon a paradox because, if we continue with the procedure, eventually we will have reached the opposite end of the spectrum and have to agree that these (blue) colors should also be called red. Thus, in order to resolve the paradox, we need some additional and non-perceptual mechanism such as color labels to mark a distinction between two colors.

Philosophical issues aside, the most critical aspect of color categories is that our perception of stimuli from continuously varying dimensions changes when it is categorical. Categorical Perception (CP; Harnad, 1987) entails that stimuli from the same category become to look more similar to each other and more different to stimuli from other categories. The centers of categories are classified faster than those at the edges and consequently discrimination of stimuli is better across than within categories. It is such behavioral consequences that we look for when deciding whether an individual or species possesses a color category.

Two lines of evidence from humans have converged to the conclusion that categorical color perception is, at least, modulated by language. The first and most direct line comes from studies that have related the categorical division of stimuli directly to experience of language (e.g., Davidoff et al., 1999; Repp, 1984; Roberson et al., 2000; Roberson et al., 2005; Strange & Jenkins, 1978). Both the Berinmo language (Davidoff et al., 1999) in Papua New Guinea, and the Himba language (Roberson et al., 2005) which is spoken in Northern Namibia, contain only five basic color terms compared to the eleven in English. When recognition memory for color was examined in both these remote cultures, the Himba and Berinmo showed no memory advantage for English best examples (Roberson et al., 2005). Instead, speakers of each language recognized good examples of their own linguistic color categories better than poor examples, regardless of the status of these items in English color categories (Roberson et al., 2005; see also Jameson & Alvarado, 2003 for data on Vietnamese speakers).

Roberson et al. (2000, 2005) also investigated CP for color in speakers of English, Berinmo and Himba. For English speakers, pairs of colors that cross the boundary (e.g., between blue and green) were discriminated faster and more accurately than pairs of colors with equal physical separation that are either both green or both blue . These authors also investigated whether speakers of Berinmo and Himba would show CP at the boundaries of the English categories green and blue, or whether the Berinmo and Himba show CP at category boundaries within their own language that do not exist for English speakers. Participants were shown a colored target and had to decide which of two stimuli, presented five seconds later, was identical to the target. For each language tested, performance was facilitated when the target and distractor stimuli had different color names (e.g., in English, a blue target with a green distractor) relative to the same name (e.g., in English, two different shades of blue). The results indicated that all three groups of participants showed CP, but only at color boundaries that were explicitly marked in their own language. Crucially, there was no effect of the proposed universal boundary (Berlin & Kay, 1969) between green and blue for speakers of Himba and Berinmo whose languages do not make this distinction.

A criticism of this cross-lingual research is that it used memory tasks to investigate the relationship between language and color categorization. A strong link between naming and recognition might have emerged because individuals chose to rely on verbal coding to retain information about color during the retention interval (Munnich & Landau, 2003). Subsequent research has more commonly used a visual search task (Daoutis, Pilling & Davies, 2006; Gilbert, Regier, Kay & Ivry, 2006) that makes little or no demands on memory. Participants were told to fixate on a cross in the center of the computer screen. They were then asked to report the location of an oddball colored target appearing among an array of identically colored distractors. Participants showed clear evidence of CP on this task. They were faster to detect a difference between the target and distractors when they came from different categories (e.g., blue target, green distracters). Again, the critical question is whether CP on the visual search task occurs only at boundaries between colors in the putative universal set or whether it also occurs at boundaries that are not marked in English. The question has been answered in recent studies with speakers of Russian (Winawer et al., 2007), and with speakers of Korean (Roberson, Pak and Hanley, 2008). Russian participants showed CP at the boundary between *sinii* (dark blue) and

goloboy (light blue), which are basic color terms for speakers of Russian. Korean participants showed CP at the boundary between *yeondu* (yellow-green) and *chorok* (green), which are basic color terms for Korean, but not for speakers of English. In consequence, these two studies provide evidence for superior discrimination of stimuli that cross a category boundary (such as that found for English speakers at the boundary between blue and green) and does not support the view that there is a set of universal color categories that are hard-wired in the human visual system.

Thus, in summary, the cross-cultural results have indicated that linguistic training affected low-level perception (see also Goldstone, 1994). Moreover, the results uphold the view that the structure of linguistic categories distorts perception by stretching perceptual distances at category boundaries (Goldstone, 1994; 1998). It would appear that the internal color space (Davidoff, 1991; 2001) is not static; some distances within it are 'stretched' or 'distorted' by the influence of color labels.

The second line of evidence for the involvement of language concerns experimental studies with human adults showing that categorically controlled performance became non-categorical by the introduction of verbal interference tasks (Gilbert et al., 2006; Hermer-Vazquez, Spelke & Katsnelson, 1999; Roberson & Davidoff, 2000; Winawer et al., 2007). For example, Roberson and Davidoff (2000) contrasted verbal interference with visual interference in a color recognition task. Both interference tasks decreased recognition performance but, surprisingly, the categorical difference was retained with visual interference but completely disappeared under verbal interference.

Both lines of evidence implicating language in color categories return us to the widely held view of fifty years ago, when color categories were thought derived from color terms. The belief was, in accord with the linguistic relativity hypothesis of Whorf (1956) that "We dissect nature along lines laid down by our native language".

Contrary to the Whorfian view, Berlin and Kay (1969) followed by Rosch Heider and Olivier (1972) argued that the perceptual/cognitive division of color space was universal and therefore independent of language. The universal view has, at least until recently, become the established position in the field with the implication that there are innate physiological mechanisms underlying color categories. In support of the universalist view, Kay and McDaniel (1978) suggested that primary categories (red, green, yellow and blue) derive from the output of opponent-process cells. These cells selectively respond to different ranges of wavelength, and produce the well-known phenomenon of colored after-images. For example, after habituating to a long ("red") wavelength a neutral surface appears to be colored green. Although these cells might appear to be categorical, this is not the case (de Valois & de Valois, 1990). In fact, opponent-process cells do not output categorical information per se, but only respond to specific ranges of wavelength. Opponent-process cells do show particular wavelengths at which discrimination is clearly more sensitive, but these discontinuities are not at the required places to produce the boundaries of primary red, yellow, green and blue color categories. Nor do colors that we see as uniquely red, yellow, green or blue have any clear relation to these discontinuities (Kuehni, 2004; Saunders & van Brakel, 1997). In the same way, the remarkable wavelength specificity found for some neurons located in the early visual cortex of the macaque (Yoshioko, Dow & Vautin, 1996) cannot account for color categories. To comply with our definition of CP, what would be needed is to have neurons that were more sensitive to cross- than within-category distinctions. To our knowledge, no such neurons have been described in the literature.

No better success is achieved by trying to relate color category boundaries to spectral sensitivities, When spectral sensitivity thresholds have been measured along the color continuum, the minima or maxima for color sensitivity show no clear relationship with color category boundaries. For example, in Heywood, Cowey and Newcombe (1991), the minimum threshold in the green range was at 525 nm (inferred from their graphs) and does not occur either at the location of the actual green to blue boundary (495 mm) or at the location of the green to yellow boundary (550 mn). A similar conclusion comes from a cross-cultural study by Roberson et al. (2009) where thresholds and boundaries were comparatively assessed in English and Korean speakers. From the universalist position, we are only left with the possibility that primary color prototypes (foci) have a physiological basis (Kay & Regier, 2006; Philipona & O'Regan, 2006), but it would still be necessary to show that there is a behavioral consequence restricted only to those focal colors, there was no such restriction in Roberson et al. (2004).

Recent theories of color categories formation have rejected the physiological basis as the unique explanation, but suggest a compromise instead between the universality and linguistic hypotheses (e.g., Kay, Regier & Cook, 2005). On the one hand, it is acknowledged that language processing affects colors categories. On the other hand, it is acknowledged that the physiological system must put some constraints on color category formation. For example, Kay & Regier (2006) point out that languages like Berinmo (and also Himba) divide up color space in very similar ways, as illustrated in Figure 1. One may note that all the languages listed in Figure 1 have a term for dark colors, one for light colors, one for the "red"

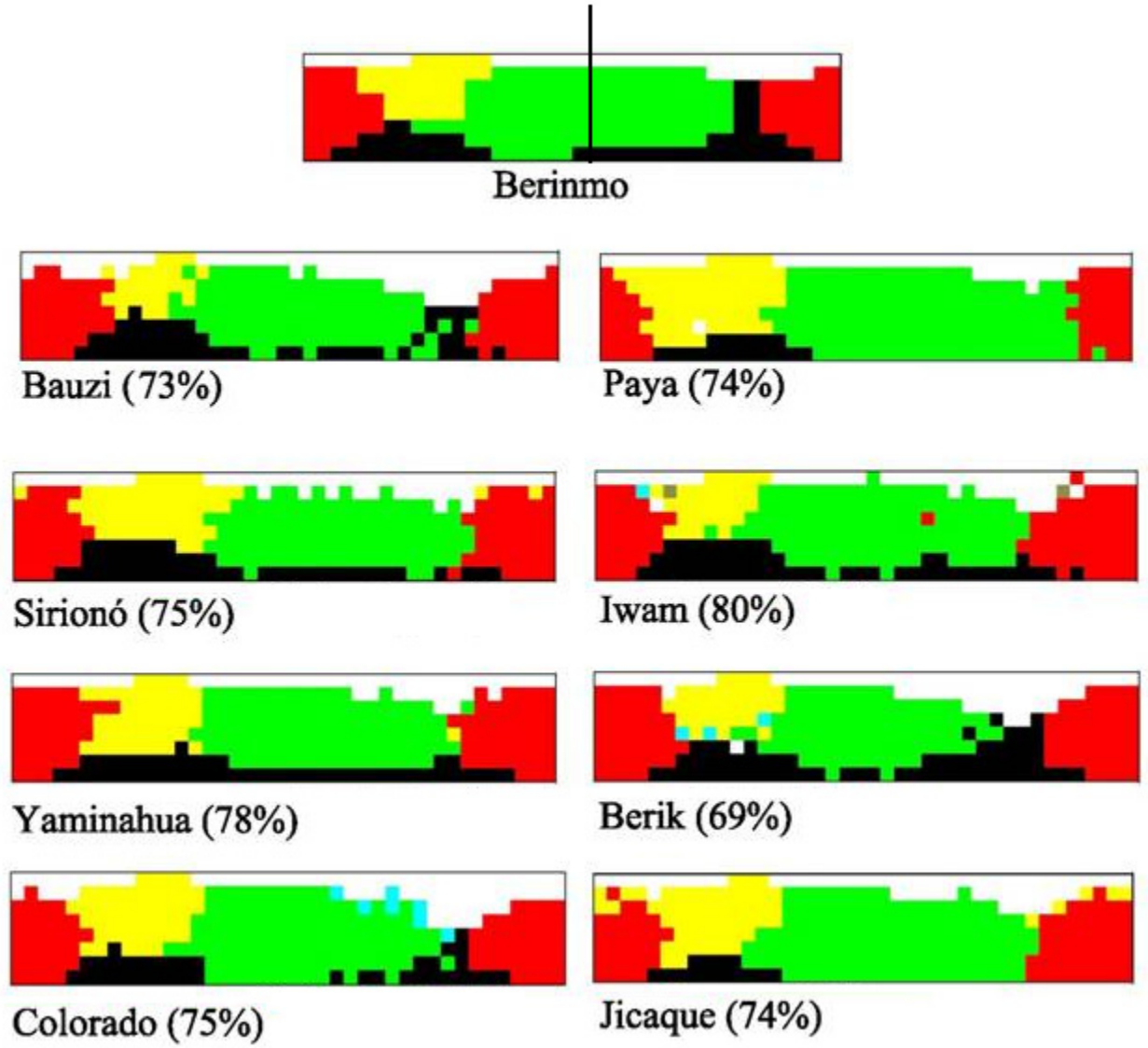

Figure 1. *Modal naming of responses of Berinmo speakers and those of eight languages from the World Color Survey, each with five basic color terms. The percentage following each language name is the boundary match with Berinmo. Adapted from Kay and Regier, 2007. The vertical line on the Berinmo color space indicates the green-blue boundary for English speakers.*

area of color space, one for the "yellow" area, and one that encompasses both "blue" and "green".

Regier, Kay and Khetarpal (2007) have elaborated on these regularities and have given a biological explanations for their origin. They propose that these optimal partitions could be based on some universal discrimination differences useful for the efficiency of verbal communication (see also Komorova, Jameson & Narens, 2007). However, there remain variations across cultures in color categories even when they have the same number of color terms (Roberson et al., 2005), suggesting an additional contribution of language on color categories (see also Jameson & Alvarado, 2003). From the "relativist" perspective, it should be noted that the optimal solution proposed by Regier et al. (2007) does not provide any boundary between the supposed primary colors of green and blue.

Before considering how the argument might be examined in the non-human primate, we first examine a couple of unresolved issues from the human evidence concerning the role of language in the origins of color categories.

Contrasted findings on hemispheric asymmetry

Hemispheric asymmetry with respect to color categories could be expected from the involvement of language. This hypothesis is supported by neuropsychological research

showing that language impairments produced by brain damage may produce failures in color categorization (Goldstein, 1948; Roberson et al., 1999). Despite preserved color vision, in addition to a retained ability to categorize other stimuli, patients in these studies failed to understand the meaning of color terms, and failed to sort colors, suggesting that colors do not automatically form categories once color terms have been lost. Further, neuroimaging studies showed left hemisphere activations in known language areas during color comparison tasks (e.g., Siok et al , 2009; Tan et al., 2008; Thierry et al., 2009). For example, in Siok et al (2009) it was only categorical color decisions in a visual search task that were accompanied by activation in language areas.

Again supporting ideas of hemispheric asymmetry, adult studies using a visual search paradigm found that it was only (or mainly) right visual field presentations that produced a cross-category advantage (Drivonikou, Kay, Regier, Ivry, Gilbert, Franklin & Davies, 2007; Gilbert et al., 2006; Roberson et al., 2008). When the left-hemisphere system was suppressed by a concurrent task that prevented access to verbal codes in normal adults, or when it cannot be reached in split-brain patients, no trace of categorical organization remains in the left hemisphere (Gilbert et al., 2006). These studies have been used to distinguish between different hemispheric codes for the representation of color differences (see Roberson et al., 2008).

In contradiction to the above findings, a notable aspect of recent ERP studies is that they find no evidence for hemispheric asymmetry for color categories (Fonteneau & Davidoff, 2007; Holmes, Franklin, Clifford & Davies, 2009) . The same conclusion comes from a behavioral study (Danilova & Mollon, 2009). It is clear that these hemispheric asymmetries require further research not least because Franklin, Drivonikou, Bevis, Davies, Kay and Regier (2008) found that left visual field presentations produced cross-category advantages in human infants for many color categories. They therefore argued that only the right hemisphere of the brain –presumably the right visual cortex- is imbued with innate color categorical properties. The results of Franklin et al. (2008) certainly pose problems in understanding the development of color categories. While there is now general agreement that language (left hemisphere) networks are involved in the production of color categories, it is difficult to understand how those would co-exist with a right hemisphere with possibly conflicting innate color categories.

Contrasted findings on infant color categories

As well as the converging evidence for the importance of color terms in establishing CP in adult human populations, there is also evidence that naming draws the attention of infants as young as 9-13 month of age to commonalities among objects and facilitates the formation of categories (Plunkett, Hu & Cohen, 2008; Waxman & Markow, 1995; Younger, 1985). Categorization is also possible in even younger infants in the absence of language (Quinn, Eimas & Rosenkrantz, 1993) but this is restricted to the special case of the categorization of discontinuous stimuli where say, dogs vs. cats, can be distinguished simply by features (French, Mareschal, Mermillod & Quinn, 2004). However, there are several reports showing color categorization to be present as early as 4 months (Bornstein, Kessen & Weiskopf, 1976; Franklin & Davies, 2004; Franklin, Pilling & Davies, 2005; Franklin et al., 2008). These studies give a powerful argument in favor of innate mechanisms, and hence in favor of the universal color categories proposed by Berlin and Kay (1969). In their classic study, Bornstein et al. (1976) showed novelty preference for a stimulus from a different color category but not from the same category. The study used a habituation-dishabituation paradigm testing dishabituation for either between-category or within-category spectral sources. The result showed greater dishabituation for cross-category stimuli.

Doubts were raised at the time about the methodology in Bornstein's studies (Banks & Salapatek, 1983; Werner & Wooten, 1985). Cross-and within-category distances were measured simply by wavelength differences. From this procedure, one cannot claim for example that a 450nm (blue) and 510nm (green) stimuli are equally different from a 480nm training stimulus. The more recent studies have used more appropriately spaced Munsell stimuli but obtained essentially the same results as Bornstein et al. (1976). Indeed, in Franklin and Davies (2004), CP was found in 4-month-olds between secondary as well as between primary colors, in places where even universalist theories (e.g., Kay & Regier, 2003) would not predict them. There is a contrary finding of Gerhardstein, Renner and Rovee-Collier (1999), but Davies and Franklin (2002) contend that the paper's conclusions are unsafe due to unintended category boundaries.

Franklin et al. (2005, 2008) used visual search to study color categorization but this methodology has its own problems. In testing infants, Franklin et al. (2008) used only 3 widely-separated colors (one green and one blue target and a background color that was just on the green side of the boundary between green and blue), because those used for adults were too difficult for infants to discriminate. Thus, no direct comparison with adults can be made because the adults in the study saw a different set of colors. Furthermore, given the difficulties of carrying out eye-tracking studies with infants, the data came from only 13 infants. Nor have any infants to-date been tested for the presence of other category boundaries (such as that between the Russian categories sinii

and goloboy) that are not proposed to be part of any universal innate set. It is therefore premature to conclude that all pre-linguistic infants show the same pattern of categorization as (English-speaking) adults across the full range of visible colors. Their conclusion also stands at odds with the findings of Xu and Carey (2000) where 6-months-of age children were found to be insensitive to much larger changes in the color of stimuli than used by Franklin et al. (2008).

For a strong relativist position on the origin of color categories (Davidoff, 2001) these infant data are hard to accommodate. If color category divisions are established through language, one would not expect any evidence for CP before the acquisition of color terms. However, Bornstein (1985) recognized that some difficulties arise from the proposal of innate color categories. The first is explaining the well-known difficulty children have in learning color names. If the physiological apparatus is already in place at 4 months, it seems odd that it takes another 18 months to learn the first color word during the time the child shows a spurt of word learning. Bornstein (1985) suggested that there must be some maturational delay perhaps of callosal fibers; a proposal that does not explain the dramatic increase in the acquisition of color terms of present day 4-year-olds compared to those of 100 years ago. Bornstein (1985) was also aware that a color vision system in line with Western color terms would, if innate, pose problems for the many languages in the world that did not possess these terms. Children in those languages ought to find color naming even more difficult. Thus, the infant data present a challenge to researchers who maintain a relativist position for the issue of color categorization.

Experiment 1: Comparative assessment of color thresholds in humans and baboons

A consequence of language involvement in learning color categories could be that they are restricted to humans. However, a different prediction is offered by Notman, Sowden and Özgen (2005) from the CP in their study on learned orientation categories. They argued for a change to neural sensitivity at V1. Notman et al. (2005) investigated discrimination thresholds for orientation category boundaries. They taught human participants to distinguish between two categories of 45° Gabor patterns each composed of gratings that differed by spatial phase. Their learned category showed an enhanced sensitivity at the category boundary in a same/different discrimination task. The enhanced sensitivity did not generalize to a 90° change of orientation, only applied to a narrow orientation bandwidth estimated to be 6.5°, and was restricted to the retinal locus of stimulation. From these data, Notman et al. (2005) reasonably argue that category learning has changed perceptual sensitivity probably at V1. But, as Fahle (2004) notes, enhanced sensitivity for such a narrow bandwidth is an improbable basis for perceptual categorization as we would need to learn to categorize the same stimuli separately at many different orientations. The same would apply to the specificity of retinal location. For other reasons, Mollon and Danilova (1996) have also warned against interpreting data similar to Notman et al. (2005) as necessarily implying the lack of cognitive intervention.

In a recent paper, Roberson et al. (2009) explicitly tested whether there was an increase in sensitivity at color category boundaries as would be predicted by the approach taken by Notman et al. (2005). They examined a blue-green range that contained category boundaries for both English and Korean speakers. The boundaries are at different points for speakers of the two languages as the color terms are different in English and Korean. However, neither language group showed enhanced sensitivity at either of the two boundaries. Comparative data could speak to this issue because if color categories were innate and produced from neurons at early visual cortex, one ought to see threshold differences at color boundaries for all species with our trichromatic vision. We examine this issue in Experiment 1 and subsequently consider the issue of CP in non-human primates in Experiment 2.

Experiment 1 assessed threshold values for the range of colors used in Roberson et al. (2009), but with a non-human primate species. If category boundaries are associated with enhanced sensitivity one ought to find lowered threshold around 495 nm for spectral stimuli and close to 7.5BG when using the Munsell system (Roberson et al., 1999).

Method

Participants

They were two Guinea baboons (Papio papio) who lived at the CNRS (Marseille) and two young human female adults. The baboons had already taken part in many different experiments using the matching-to-sample procedure and joystick computerized systems (e.g., Fagot, Goldstein, Davidoff & Pickering, 2006).

Apparatus

The experiment took place in a darkened room. The baboons were tested in an experimental enclosure (60 x 50 x 72 cm) facing a joystick, a metal touch pad, and a 14-inch color monitor. On the front of the enclosure were a view-port, a hand-port, and a food dispenser delivering banana-flavored food pellets into the enclosure. Manipulation of the joystick induced isomorphic displacements of a cursor on the monitor. The eye/screen distance was approximately 49 cm.

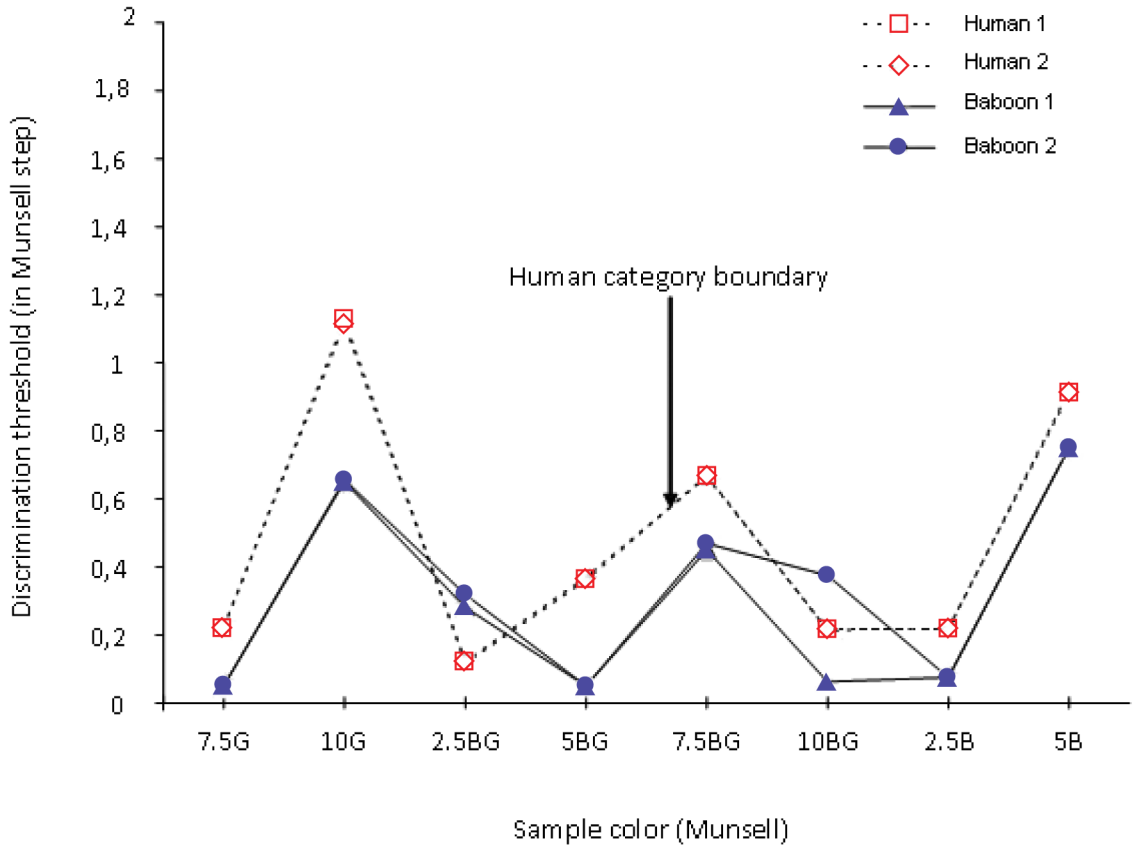

Figure 2. *Discrimination thresholds in humans and baboons, as a function of the sample stimulus color.*

For humans, the monitor and joystick were placed on a table so that viewing distance remained equal to 49 cm. Control and randomization of conditions were achieved through purpose-made programs written in E-Prime V 1.0 (Psychology Software Tools, Inc). Color calibration was achieved as in Roberson and Davidoff (2000). Accuracies and latencies were recorded.

Procedure

Discrimination thresholds were measured with an adaptive staircase procedure (Levitt, 1971) leading to a rapid convergence near the asymptotic threshold level. Thresholds were obtained in separate blocks for the following reference (sample) stimuli: 7.5G-10G-2.5BG-5BG-7.5BG-10BG-2.5B-5B, avoiding the training stimuli used in Fagot et al. (2006). The same sample stimulus was systematically employed in all trials of a single session. All stimuli were of constant brightness (Value 5), saturation (Chroma 6) and of 6.4° by 6.4° visual angle.

Each trial started with the display of a .5° circular green cursor along with a white .5°x.5° square-shaped stimulus, located 1.5° above or below the cursor. In response to this display, the participant had to place the green dot on the white square so as to initiate the trial. Once done, the sample stimulus was displayed at random on the left or right of the screen for 500 msec and was immediately followed by a pair of stimuli arranged vertically on the mid-line of the screen. One of the stimuli was identical to the sample. The task was to point with the cursor to the stimulus matching the sample. Baboons received a short inter-trial interval (ITI) of 1 second and a food pellet inside the enclosure in case of correct response. When their response choices were erroneous, responses were followed by a three second timeout during which the screen remained black. For humans, the feedback consisted of a short ITI for a correct response, and a three second timeout for an incorrect choice.

The session started with trials having a high probability of positive response. In these early trials, the foil was two Munsell steps different from the sample. When a run of three positive responses was obtained, the next trial reduced the match/foil color difference by 1/20th of a Munsell step. When a single negative response was obtained, the color difference was increased by 1/20th of a Munsell step. The procedure continued until 35 reversals (defined as either a color difference increment or color difference decrement) were obtained; this compares favorably to most other usage of the staircase procedure (Garcia-Pérez, 1998).

Humans received three series of test sessions with each of the eight possible sample colors. Baboons received 6 series with each sample color. The order of color testing within a series was randomized. For baboons, the test sessions lasted approximately from 200 to 350 trials. Humans often reached the asymptotic discrimination level more quickly than baboons, and needed from 150 to 200 trials to perform a session. The total testing time was thus of approximately 12,000 trials for each baboon and 4,800 for humans.

Results

The increment thresholds were defined, for each sample color, as the minimal average color difference obtained in any run of 8 consecutive reversals in any session (Garcia-Pérez, 1998). In the context of very low thresholds for both species, there was generally superior performance from the baboons (see Figure 2). Thus, on average for the 8 sample colors, discrimination thresholds for the two baboons were of .29 (baboon 2) and .34 (baboon 1) Munsell steps. It was of .48 Munsell steps for each of the two human participants. Correlations between the thresholds were extremely high and significant within species (humans: r = .99; baboons: r = .93) and between individuals from the two species (range of rs : .81-.89). In addition, each baboon indicated a lower discrimination threshold than the humans (two-tailed pair t-tests, all ps < .02). Regardless of these similarities and differences, it is clear from Figure 2 that there is no evidence for superior sensitivity at the category boundary in the region of 7.5BG.

Discussion

Our study allowed two conclusions. First, in spite of lower discriminations thresholds in baboons than in humans, baboons show the same pattern of results as humans with similar threshold variations along the color continuum. Second, in line with Roberson et al. (2009), there is no evidence from either human or baboon for an enhanced sensitivity at the known human green/blue color (i.e., 7.5BG) boundary. Therefore it seems safe to conclude that categorical similarity in color appearance can not derive from a differential threshold sensitivity at the boundary. Nevertheless, humans have sharp color category boundaries (e.g., Fagot et al., 2006) and one may ask how that comes about if it is not a result of threshold differences.

An important feature of categorical color judgments is that they are invoked automatically (Fonteneau & Davidoff, 2007; Holmes et al., 2009; Tan et al., 2008; Thierry et al., 2009) and even in a patient who has lost color terms (Roberson et al., 1999). For example, Fonteneau and Davidoff (2007) found that ERP activity to a color presented in a series of colors reflected its category relationship to the other colors. The ERP was categorical despite no response being made to the color and observers not even realizing that any aspect of color was under investigation. The categorical response was early in the ERP trace but even earlier (i.e., P1) components have been recently claimed for color categories (Holmes et al., 2009; Thierry et al., 2009). Given the activation of language areas in color judgments, these recent ERP studies suggesting categorical activity in (extra)striate cortical areas would appear to require a rapid activation of feedback networks during color processing (Siok et al., 2009). Presumably, these feedback loops are disrupted by concurrent verbal activity (Gilbert et al., 2006; Roberson & Davidoff, 2000) and hence color perception is no longer categorical

Experiment 2: Comparative assessment of color CP: A human vs. baboon experiment

Given such clear implication for language in human color CP, we examined the cross-species research that might lead one to believe that non-humans also demonstrate CP for colors. Herrnstein and Loveland (1964) were the first to demonstrate that an animal species (i.e., the pigeon) could learn to sort visual stimuli on a categorical basis. In their study, pigeons learned to provide behavioral responses consistent with the categorical structure of the stimuli and, even more impressively, generalized to novel, previously unseen exemplars of the categories. Since Herrnstein and Loveland (1964), categorization has been demonstrated in a variety of species and with a variety of experimental procedures (D'Amato & van Sant, 1988; Dépy, Fagot & Vauclair, 1997; Fagot, Wasserman & Young, 2001; Martin-Malivel & Fagot, 2001). While these studies show that some forms of categorization are possible in animals, and consequently that language might not be necessary condition for categorization, the cognitive procedures permitting those categorical responses still remain largely unclear. It is particularly unclear in which context the animals solve the task by way of procedures, such as feature analysis or prototype learning, implying an analysis of the physical dimensions of the stimuli, and when they rely on more abstract thought.

Among the possible reasons limiting the heuristic values of categorical studies in animals is the general use of non-continuous variables to define differences between categories. For example, studies have tested the capability of animals to discriminate between categories such as tree, monkey or people (e.g., Herrnstein & Loveland, 1964). Use of these types of categorical structures makes it very hard to identify the cognitive procedure employed by the animals to solve the task, partly because the stimuli are multidimensional, and therefore to make conclusions about

the level of abstractness achieved by the subjects.

Notable exceptions for the use of non continuous variables with animal species are the studies on CP of auditory stimuli in chinchillas and monkeys (e.g., May, Moody & Stebbins, 1989; Ohlemiller, Jones, Hedibreder, Clarck & Miller, 1999). Most of these authors consider the Voice Onset Time (VOT) as the critical variable in their studies. VOT corresponds to the interval between consonant onset and the start of rhythmic vocal-cord vibrations. Perception of consonants such as /d/ and /t/ appear indeed categorical in humans, with a boundary lying at a VOT of 20-40 ms. Note however that boundaries in perceptual categorization studies do not always converge between humans and nonhuman primates, suggesting that this effect is not purely perceptual (e.g., /ra-la/ contrast: Sinnott & Brown, 1997).

Application of the VOT paradigm has suggested that the discrimination of such auditory stimuli is also categorical in animals (e.g., May et al., 1989; Ohlemiller et al., 1999; Steinchneider, Fishman & Arezzo, 2003), and therefore that there is no need of language for CP. This conclusion might however be premature for the following three reasons. First, it should be noted that the stimuli employed in VOT paradigms are not unidimensional, as they vary in a number of dimensions, such as spectral, temporal, or overall energy. It is therefore difficult to ascertain with this procedure the actual mechanism at the origin of CP effects. Listeners in VOT studies were noticeably found to attend to different phonemic cues (Ohlemiller et al., 1999), therefore opening the possibility that different cognitive procedures might be applied by participants with VOT procedures. Second, not all experiments could demonstrate CP with the VOT procedure (e.g., Hopp, Sinnot, Owren & Petersen, 1992). Third, it remains possible that the stimuli used by these researchers are of special ecological value for the animals, and therefore that CP would not necessarily generalize to other continua of stimuli.

Other evidence for CP in animals has resulted from categorization studies using unidimensional visual stimuli. Thus, Wilson and Debauche (1981) tested CP of length, orientation and texture in macaques. After training with the extremes of a set of stimuli, category boundaries were identified for each dimension as the stimulus eliciting 50 choices over all comparisons. After training, normal macaques (in comparison with animals with inferotemporal lesions) had improved discrimination performance for between-category compared to within-category discrimination. However, inspection of individual monkey data does not show clear categorization responses.

Very few studies (Fagot et al., 2006; Matsuno, Kawai & Matsuzawa, 2004; Matsuzawa, 1985; Sandell, Gross & Bornstein, 1979) have assessed color categorization in primates such as the chimpanzee or the baboon, despite that they are known to have the same wavelength discrimination function as humans (Bowmaker, Astell, Hunt & Mollon,

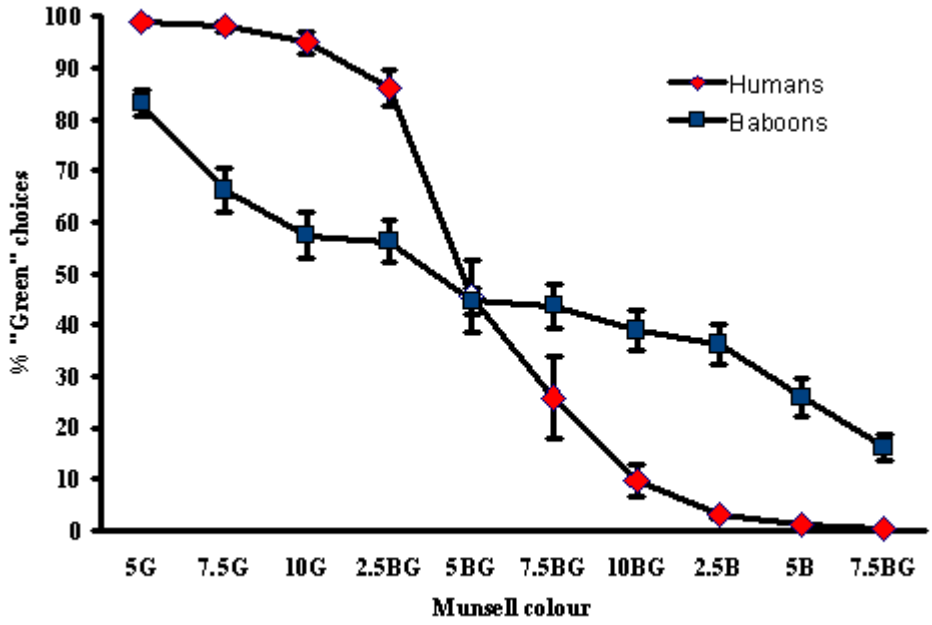

Figure 3. *Comparison between baboons and humans on a matching to sample color task. The y-axis gives the number of "green" responses for targets between the two training samples 2.5G (a good green) and 10BG (a good blue). The data are from Fagot et al. (2006).*

1991; Bowmaker, Mollon & Jacobs, 1983). The lack of research might be due to the widespread belief that color categories are universal in the human (Berlin & Kay, 1969) and therefore should be present in all primates. A further disinclination to investigate color categories could be that it has already been thought shown in the chimpanzee Ai (Matsuzawa, 1985). Ai, after being trained to use symbols to name the so-called 11 basic colors, generalized these names to other colors much in the way one would expect of a Western or Japanese speaker. However, there was no evidence that such generalization would be possible in the absence of specific training. Indeed, these categories do not even seem to be well-formed in a chimpanzee with two years of experience with color symbols (Matsuno et al., 2004). The most recent investigation of color categories in the monkey (Fagot et al., 2006) could also be judged inconclusive.

Fagot et al. (2006) trained baboons to match color patches from the green-blue continuum. In the training phase, the baboons were required to only match the extremes of the continuum, namely the 2.5G and 10B stimuli. In the transfer test, they were requested to match the intermediate stimuli with one of the two extremes of the continuum, for instance match the 10G stimulus with the 2.5G stimulus. Figure 3 illustrates the findings from 8 baboons, and compares the results obtained in baboons with those of a group of 8 humans tested in the same conditions. A visual inspection of Figure 3 promptly reveals that humans expressed a clear-cut boundary in that continuum. The boundary corresponded to the one obtained when asked to name the stimuli as being green or blue. There was by contrast no such boundary in baboons, in other words, they process the continuum as a continuum, contrary to humans who process the stimuli as belonging to two clearly defined green and blue categories. However, as pointed out in several studies with human observers, the matching to sample (MTS) procedure can produce artificial categories not related to a perceptual decision (Angeli, Davidoff & Valentine, 2008; Massaro, 1987; McKone, Martini & Nakayama, 2001; Viviani, Binda & Borsato, 2007). So, while the absence of clearly delimited color categories in the baboon shows a difference to humans, one cannot be sure that the difference is essentially one related to color categorization.

The one published attempt to show CP for colors in nonhuman primates using a method that could give a definitive answer was by Sandell et al. (1979). However, there are concerns regarding the stimuli used in their study. The stimuli were derived from those used in Bornstein et al. (1976) where separation was based on wavelength differences rather than psychologically equal interval steps between colors. When Sandell et al. (1979) turned to monkeys, they correctly changed from spectral to Munsell stimuli but, still having in mind wavelength differences, got the separations wrong. They made the cross-category comparisons substantially easier than the within-category comparisons. In fact, their within-category stimuli were only 5.23 Munsell steps apart on average whereas the between category-stimuli were considerably greater (18.7 on average). Thus, Sandell et al.'s (1979) conclusion of categorical color perception in monkeys warrants serious reconsideration.

We conducted an experiment to further investigate CP of colors in humans and monkeys, using a procedure that directly compared cross- and within-category pairs. CP was therefore studied in a group of 6 Guinea baboons (Papio papio) housed in the animal facility of the INCM institute in Marseille, France.

Methods

Participants

Four male and two female baboons of approximately 16 years, which were already familiar with the MTS procedure (e.g., Fagot & Deruelle, 1997, 2002), were employed in our research on CP. Eight English-speaking humans (3 male, 5 female) aged between 19-27 years (mean age 21.6 years) were paid for their participation in the study at Goldsmiths, University of London. All were found to have normal color vision on the Ishihara plates (Ishihara, 1998).

Stimuli

As these animals had taken part in the MTS study of Fagot et al. (2006) and are known to have a prodigious visual memory (Cook & Fagot, 2009; Fagot & Cook, 2006), we used a different color range to the blue/green used in Fagot et al. (2006). Four pairs of stimuli were generated comprising two cross-boundary pairs and two within-color pairs, equally spaced in Munsell steps (Munsell, 1905) around the blue/purple boundary at 10PB (Franklin & Davies, 2004). The cross-boundary pairs were (8PB & 1P) and (9PB & 2P). The within-color pairs were (5P & 8P) and (2PB & 5PB). All stimuli were of constant brightness (Value 2) and saturation (Chroma 8) and subtended 6.4º x 6.4º visual angle, and were generated and calibrated on screen using a Minolta CS100 color gun.

Procedure

Each trial started with the display of a .5° circular green cursor along with a white .5°x.5° square-shaped stimulus on the screen, located 1.5° above or below the cursor. In response to this display, the animal participant had to manipulate the joystick in order to place the green dot on the white square so as to initiate the trial; then a square-shaped

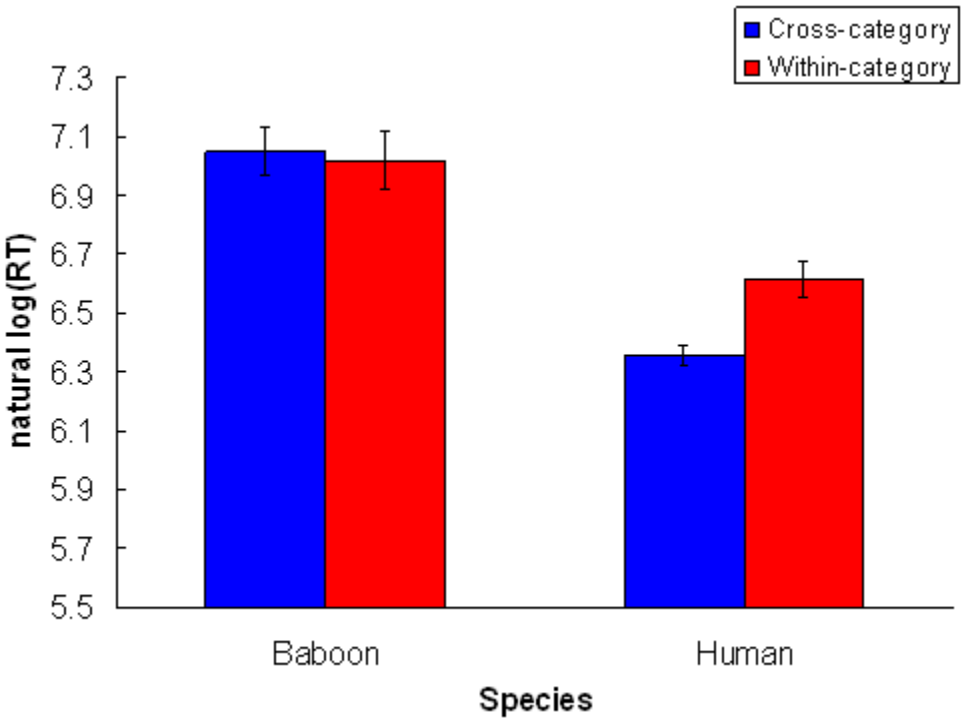

Figure 4. *Natural log (RT) for baboons and humans for cross- and within-category pairs using the blue/purple boundary.*

sample stimulus appeared with 4.5° of lateral eccentricity on the right or the left of the screen. The sample stimulus was displayed for 500 ms and immediately replaced by a display comprising a cursor and two patches of color one of which was identical to the preceding sample. In balanced order, one color square appeared on the top and the other on the bottom half of the screen. During the response period, participants had to point with the cursor to the comparison stimulus matching the sample. There was no time limit for responding. Response choices and response times were recorded. Baboons received a short ITI of 1 second and a food pellet inside the enclosure in case of correct response. When their response choices were erroneous, responses were followed by a three second timeout during which the screen remained black. For humans, the feedback consisted of a short ITI for a correct response, and a three second timeout for an incorrect choice.

The testing consisted of three sessions of 64 randomly ordered trials for each participant, using the stimulus pairs 5P – 8P and 2PB - 5PB for within-category comparisons, and the stimulus pairs 8PB - 1P and 9PB - 2P for cross-category comparisons. All stimuli were used equally often as sample and comparison, and their location on the screen was completely balanced. Prior to testing, the participants from the two species received 16 randomly ordered and differentially reinforced training trials using two colors (green and yellow) different from the test colors. These trials were repeated until subjects reached a criterion level of 80% correct.

The basic procedure was identical for the human participants except that the joystick was replaced by a response box with two buttons arranged vertically to direct the cursor either ‘up’ or ‘down’.

Results

Trials associated with latencies less than 150 ms or greater than 4 sec (less than 1% of the total number of trials) were omitted from statistical analyses because they likely reflect anticipation or inattention. Only correct responses were used in the latency analysis for both species. The average baboon performance collapsed over category was 58.56%, and the average human performance was higher at 80.07%. Baboons responded in 1156 ms on average, humans were faster with an average response time of 666 ms. Response times were log transformed prior to statistical analyses (see Figure 4). Application of a Species (baboons, humans) by Category (within, between) ANOVA on transformed response time revealed a significant effect of Species, $F(1,12) = 35.83$, $p < .001$, a significant effect of Category, $F(1,12) = 13.90$, $p < .005$, and a significant Species by Category interaction, $F(1,12) = 23.45$, $p < .001$. Analysis of this interaction with a Tukey HSD test indicated that mean response times were reliably shorter for humans in cross- compared to within-category trials (log RT for cross = 6.36, for within = 6.61, $p < .05$). However, there was no significant difference between cross- or within-category response times for the

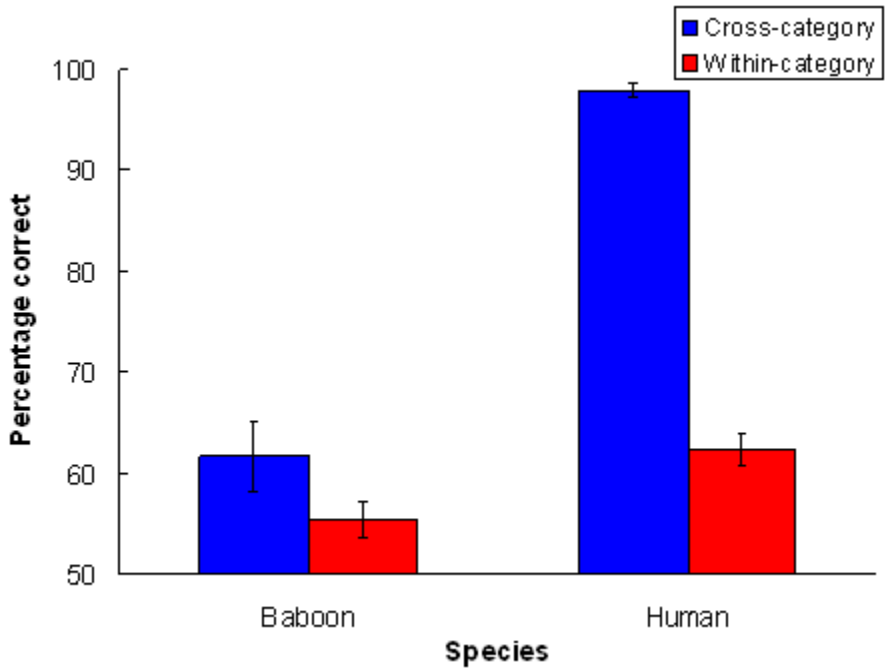

Figure 5. *Percentage of correct responses for baboons and humans, and for cross- and within-category pairs using the blue/purple boundary.*

baboons (log RT for cross = 7.05, for within = 7.01).

Accuracy scores were also submitted to an ANOVA again using the Species (baboons, humans) as between-subject factor and the Category (within, cross) as within-subject factor. Again, (see Figure 5) there was a significant effect of Species, $F(1,12) = 115.86$, $p < .001$, a significant effect of Category, $F(1,12) = 115.05$, $p < .001$, and a significant interaction between Species and Category, $F(1,12) = 56.89$, $p < .001$. Analysis of this interaction with a Tukey HSD test indicated that accuracies were significantly higher for humans in cross- compared to within-category trials (mean cross accuracy = 97.88%, mean within accuracy = 62.25%, $p < .001$). However, there was no significant difference between categories for the baboons (mean cross accuracy = 61.66%, mean within accuracy = 55.47, $p > .15$). Further, the baboon performance in either category did not differ from the human within-category performance ($p = .997$ compared to baboon cross-category and $p = .098$ compared to baboon within-category). Paired t-tests showed that baboons performed significantly above chance in both cross- and within-categories, $t(5) = 3.33$, $p < .03$ and $t(5) = 3.02$, $p < .03$ respectively.

Discussion

The experiment produced a clear cross-species difference in CP. It was only the humans that showed the cross-category advantage. We note that the baboons did not perform well at the recognition memory task though they were above chance. However, that poor performance needs to be seen in comparison to the equally poor performance of the humans for the within-category pairs. Indeed, poor human performance in the within-category condition is the norm in our studies (Roberson & Davidoff, 2000). Clearly it is some aspect of the cross-category comparison that produces the enhanced performance for humans. The advantage could be simply attributed to labelling (Munnich & Landauer, 2003) but this is unlikely because visual search paradigms (Daoutis et al. 2006; Gilbert et al., 2006) that have no memory component produce equivalent cross-category advantages to the 2AFC paradigm. All in all, our findings suggest strong human-monkey differences in the processing of color categories. In line with the argument presented in introduction, that difference might be accounted by a linguistic origin of color categories in human.

General Discussion

Could monkeys be trained to learn color categories?

We do not know whether baboons might acquire color categories with training. Humans can acquire color categories even without overt naming (Özgen & Davies, 2002), but there has been no attempt to so train monkeys. Although the findings in Matsuzawa (1985) suggest similar human/chimpanzee color classifications, they must be contrasted with Matsuno et al. (2004) that revealed a much less stable color classification in a different chimpanzee with reduced abilities for symbol use. The fact that the monkeys in Fagot

et al. (2006) were only prepared to match colors that were very similar to the training stimuli must make the human type of broad color categories difficult for them to achieve.

One further potential limitation to training color categories in the monkey might be their supposed inability to perform analogical reasoning. It might be thought a likely limitation as analogical reasoning has been argued to be promoted by linguistic competencies (Gentner & Christie, 2008), and that color categorization also requires linguistic ability. However, the human ability to categorize colors does not rely on the ability to reason analogically (Davidoff & Roberson, 2004). Furthermore, while it has been argued that only language-trained (Premack, 1983) or token-trained (Thompson, Oden & Boysen, 1997) apes can perform analogical reasoning tasks, recent work by Fagot and Parron (2010) suggests that, even in the absence of language- or token-training, monkeys seem to be able to judge the relation between relations.

An inconsistency: Avian color categories

Wright and Cumming (1971) examined MTS for color in pigeons. Their pigeons produced cross-over points (i.e., boundaries) between the training stimuli in the MTS though not at the same points as for human primary colors. They suggested that these boundaries correspond to divisions of the spectrum analogous to human color naming (Beare, 1963). Furthermore, Wright and Cumming (1971) also carried out MTS experiments where the range of colors was altered but the cross-over points remained stubbornly at the same wavelengths. Hence, they titled their paper "color naming in the pigeon". They suggested that humans, like their pigeons, would divide the continuum at the same boundary point irrespective of the shift of the training stimuli. We have recently conducted a similar experiment with humans and baboons (Davidoff, Goldstein, Tharp, Wakui & Fagot, submitted) but with a quite different outcome for the two species. The baboons again showed no evidence of a color boundary. It is somewhat surprising that pigeons may express CP of colors, but not the baboons. One answer could come from Jones, Osorio and Baddeley (2001) who found that chicks would interpolate between two training stimuli but not extrapolate outside the range. The outcome of the chick's behavior could be to provide boundaries that apparently divide their color space into categories.

Conclusions

Berlin and Kay (1969) classically proposed that color categories are innate; a hypothesis clearly consonant with studies that have found color CP in pigeons (Wright & Cumming, 1971), monkeys (Sandell et al., 1979) and in human infants (Bornstein et al., 1976; Franklin et al., 2008). Thus, the previous nonhuman and infant data pose a considerable question to those who propose that color categories derive from the color terms in the speaker's language. Despite these reports, the arguments for a linguistic, and thus learned, origin of most color categories are now at least incorporated into the universal theories of color category formation (Kay et al., 2005) widely accepted from the human data. Here, we enhance the claim for a linguistic contribution to color CP by demonstrating that only humans, and not baboons, show a cross-category advantage in a color recognition CP task. This finding converges with our previous demonstration of human-baboon differences in color matching (Fagot et al., 2006). Categorization of continuous types of stimuli, such as colors, seems both facilitated and constrained by the acquisition and use of linguistic terms. We might even presume that the need to solve these unidimensional categorization problems was a driving force for the evolution of language (Davidoff, 2001).

References

Angeli, A., Davidoff, J., & Valentine, T. (2008). Face familiarity, distinctiveness, and categorical perception. *The Quarterly Journal of Experimental Psychology, 61*, 690-707. doi:10.1080/17470210701399305

Banks M. S., & Salapatek P. (1983). Infant visual perception. In M. M. Haith & J. J. Campos (Eds.), *Handbook of child Psychology* (pp. 435-571).

Beare, A.C. (1963). Color-Name as a Function of Wave-Length. *The American Journal of Psychology, 76*, 248-256 doi:10.2307/1419161

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

Bornstein, M. (1985). On the development of naming in young children: data and theory. *Brain and Language, 26*, 72-93. doi:10.1016/0093-934X(85)90029-X

Bornstein, M.H., Kessen, W., & Weiskopf, S. (1976). Color vision and hue categorization in young human infants. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 115-129. doi:10.1037/0096-1523.2.1.115

Bowmaker, J. K., Astell, S., Hunt, D. M., & Mollon, J. D. (1991). Photosensitive and photostable pigments in the retinae of Old World monkeys. *Journal of experimental Biology, 156*, 1-19.

Bowmaker, J.K., Mollon, J.D., & Jacobs, G.H. (1983). Microspectrophotometric measurements of Old and New world species of monkeys. In J.D. Mollon & L.T. Sharpe (Eds.), *Color Vision-Physiology and Psychophysics* (pp. 56-68). London: Academic Press.

Cook, R. & Fagot, J. (2009). First trial rewards promote 1-trial learning and prolonged memory in pigeon and baboon. *Proceedings of the National Academy of Sciences*

of the USA, published online before print May 26, 2009, doi:10.1073/pnas.0903378106

D'Amato, M.R., & van Sant, P. (1988). The person concept in monkey (Cebus apella). *Journal of Experimental Psychology: Animal Behavior Processes, 14*, 43-55. doi: 10.1037/0097-7403.14.1.43

Danilova, M. V., & Mollon, J. D. (2009). The symmetry of visual fields in chromatic discrimination. *Brain and Cognition, 69*, 39-46. doi:10.1016/j.bandc.2008.05.003

Daoutis, C.A., Pilling, M., & Davies, I.R.L. (2006). Categorical effects in visual search for color. *Visual Cognition, 14*, 217-240. doi:10.1080/13506280500158670

Davidoff, J. (1991). *Cognition through color*. Cambridge, M.A.: MIT Press.

Davidoff, J. (2001). Language and perceptual categorization. *Trends in Cognitive Science, 9*, 382-387. doi:10.1016/S1364-6613(00)01726-5

Davidoff, J., Davies, I., & Roberson, D. (1999). Color categories of a stone-age tribe. *Nature, 398*, 203-204 & 402, 604. doi: 10.1016/j.cogpsych.2004.10.001

Davidoff, J., Goldstein, J., Tharp I. Wakui, E., & Fagot, J.(submitted). Categorical divisions within perceptual (color) continua: A human tendency but with considerable variation.

Davidoff, J., & Roberson, D. (2004) Preserved thematic and impaired taxonomic categorisation: a case study. *Language and Cognitive Processes, 19*, 137-174. doi:10.1080/01690960344000125

Davies, I.R.L., & Franklin, A. (2002). Categorical similarity may affect color pop-out in infants after all. *British Journal of Developmental Psychology, 20*, 185-203. doi:10.1348/026151002166406

Dépy, D., Fagot, J., & Vauclair, J. (1997). Categorization of three-dimensional stimuli by humans and baboons (Papio papio): Search for prototype effects. *Behavioral Processes, 37*, 299-306. doi:10.1016/S0376-6357(96)00757-7

de Valois, R. L., & de Valois, K.K. (1990). *Spatial vision*. Oxford: Oxford University Press.

Drivonikou, V., Kay, P., Regier, T., Ivry, R., Gilbert, A., Franklin, A., & Davies, I.R.L. (2007). Further evidence of Whorfian effects to the right visual field. *Proceedings of the National Academy of Sciences, 104*, 1097-1102 doi:10.1073/pnas.0610132104

Dummett, M. (1975). Wang's Paradox. *Synthese, 30*, 301-324. doi: 10.1007/BF00485048

Fagot, J. & Cook, R. (2006). Evidence for large long-term memory capacities in baboons and pigeons and its implications for learning and the evolution of cognition. *Proceedings of the National Academy of Sciences of the USA, 103*, 17564-17567. doi:10.1073/pnas.0605184103

Fagot, J., & Deruelle, C. (1997). Processing of global information and local visual information and hemispheric specialization in humans (Homo sapiens) and baboons (Papio papio). *Journal of Experimental Psychology: Human Perception and Performance, 23*, 429-442. doi:10.1037/0096-1523.23.2.429

Fagot, J., & Deruelle, C. (2002). Perception of pictorial eye-gaze by baboons (Papio papio). *Journal of Experimental Psychology: Animal Behavior Processes, 28*, 298-309. doi:10.1037/0097-7403.28.3.298

Fagot, J. Goldstein, J., Davidoff, J., & Pickering, A. (2006). Cross-species differences in color categorization. *Psychonomic Bulletin & Review, 13*, 275-280.

Fagot, J., & Parron, C. (2010). Relational matching in baboons (Papio papio) with reduced grouping requirements. *Journal of Experimental Psychology: Animal Learning and Behavior*, *36*, 184-193. doi:10.1037/a0017169

Fagot, J., Wasserman, E. A., & Young, M. E. (2001). Discriminating the relation between relations: The Role of entropy in abstract conceptualization by baboons (Papio papio) and Humans (Homo sapiens). *Journal of Experimental Psychology, 27*, 316-328. doi: 10.1037/0097-7403.27.4.316

Fahle, M. (2004). Perceptual learning: a case of early selection. *Journal of Vision, 4*, 879-890 doi:10.1167/4.10.4

Fonteneau, J., & Davidoff, J. (2007). Neural correlates of color categories. *Neuroreport, 18*, 1323-1328. doi:10.1097/WNR.0b013e3282c48c33

Franklin, A., & Davies, I.R.L. (2004). New evidence for infant color categories. *British Journal of Developmental Psychology, 22*, 349-377. doi: 10.1348/0261510041552738

Franklin, A., Drivonikou, G.V., Bevis, L., Davies, I.R.L., Kay, P., & Regier, T. (2008). Categorical Perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults. *Proceedings of the National Academy of Sciences, 105*, 3221-3225. doi:10.1073/pnas.0712286105

Franklin, A. Pilling, M., & Davies, I.R.L. (2005). The nature of infant color categorization: Evidence from eye movements on a target detection task. *Journal of Experimental Child Psychology, 91*, 227-248. doi:10.1016/j.jecp.2005.03.003

French, R. M., Mareschal, D., Mermillod, M., & Quinn, P.C. (2004). The role of bottom-up processing in perceptual categorization by 3- to 4-month-old infants: Simulations and data. *Journal of Experimental Psychology: General, 133*, 382-397. doi:10.1037/0096-3445.133.3.382

Garcia-Pérez, M. A. (1998). Forced-choice staircases with fixed step sizes: Asymptotic and small-sample properties. *Vision Research, 38*, 1861-1881. doi:10.1016/S0042-6989(97)00340-4

Gentner, D. & Christie, S. (2008). Relational language supports relational cognition in humans and apes. *Behavioral and Brain Sciences, 31*. 137-183. doi:10.1017/S0140525X08003622

Gerhardstein, P., Renner, P., & Rovee-Collier, C. (1999). The roles of perceptual and categorical similarity in color

pop-out in infants. *British Journal of Developmental Psychology, 17*, 403-420. doi:10.1348/026151099165366

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences, 103*, 489-494. doi:10.1073/pnas.0509868103

Goldstein, K. (1948). *Language and language disturbances*. New York: Grune & Stratton.

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General, 123*, 178-200. doi:10.1037/0096-3445.123.2.178

Goldstone, R.L. (1998). Perceptual learning. *Annual Review of Psychology, 49*, 585-612. doi:10.1146/annurev.psych.49.1.585

Harnad, S. (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. In: S. Harnad (ed.) *Categorical perception: The groundwork of cognition*. Cambridge: Cambridge University Press.

Hermer-Vazquez, L., Spelke, E.S., & Katsnelson, A.S. (1999). Sources of flexibility in human cognition: dual task studies of space and language. *Cognitive Psychology, 39*, 3-36. doi:10.1006/cogp.1998.0713

Herrnstein, R. J., & Loveland, D. H. (1964). Complex visual concepts in the pigeon. *Science, 146*, 149-151. doi:10.1126/science.146.3643.549

Heywood, CA, Cowey, A & Newcombe, F. (1991). Chromatic discrimination in a cortically color blind observer. *European Journal of Neuroscience, 3*, 802-812. doi:10.1111/j.1460-9568.1991.tb01676.x

Holmes, A., Franklin, A., Clifford, A., & Davies, I. (2009). Neuropsychological evidence for categorical perception of color. *Brain and Cognition, 69*, 426-434. doi:10.1016/j.bandc.2008.09.003

Hopp, S.L., Sinnott, J.M., Owren, M.J., & Petersen, M.R. (1992). Differential sensitivity of Japanese macaques (Macaca fuscata) and Humans (Homo sapiens) to peak position along a synthetic coo call continuum. *Journal of Comparative Psychology, 106*, 128-136. doi:10.1037/0735-7036.106.2.128

Ishihara, S. (1998). *Ishihara's Tests for Color Blindness*. London : Hodder & Stoughton Educational.

Jameson, K.A. & Alvarado, N. (2003). Differences in Color Naming and Color Salience in Vietnamese and English. *COLOR Research & Application, 29*, 128-134. doi: 10.1002/col.10131

Jones, C.D., Osorio, D., & Baddeley, R. J. (2001). Color categorization by domestic chicks. *Proceedings of the Royal Society B: Biological Sciences, 268*, 2077-2084. doi:10.1098/rspb.2001.1734

Kay, P., & McDaniel, C.K. (1978). The linguistic significance of the meanings of basic color terms. *Language, 54*, 610-646 doi:10.2307/412789

Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist, 86*, 65-78. doi:10.1525/aa.1984.86.1.02a00050

Kay, P., & Regier, T. (2003). Resolving the question of color naming universals. *Proceedings of the National Academy of Science, 100*, 9085-9089. doi:10.1073/pnas.1532837100

Kay. P., & Regier, T. (2006). Language, thought and color: recent developments. *Trends in Cognitive Sciences, 10*, 51-54 doi:10.1016/j.tics.2005.12.007

Kay, P., Regier, T. & Cook, R. S. (2005) Focal colors are universal after all. *Proceedings of the National Academy of Science, 102*, 8386-8391. doi:10.1073/pnas.0503281102

Komorova, N.L., Jameson, K.A. & Narens, L. (2007) Evolutionary models of color categorization based on discrimination. *Journal of Mathematical Psychology, 51*, 359-382. doi:10.1016/j.jmp.2007.06.001

Kuehni, R.G. (2004). Variability in unique hue selection: A surprising phenomenon. *Color Research and Application, 29*, 158-162. doi:10.1002/col.10237

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49*, 467-477. doi:10.1121/1.1912375

Martin-Malivel, J., & Fagot, J. (2001). Perception of pictorial faces by baboons (Papio papio): Effects of stimulus orientation on discrimination performance. *Animal Learning and Behaviour, 29*, 10-20.

Massaro, D. W. (1987). Categorical partition: a fuzzy-logic model of categorization behavior. In S.Harnad (Ed.) *Categorical perception: The groundwork of Cognition* (pp. 254-283). New York: Cambridge University Press.

Matsuno, T., Kawai, N., & Matsuzawa, T. (2004). Color classification by chimpanzees (Pan troglodytes) in a matching-to-sample task. *Behavioral Brain research, 148*, 157-165. doi:10.1016/S0166-4328(03)00185-2

Matsuzawa, T. (1985). Color naming and classification in a chimpanzee. *Journal of Human Evolution, 14*, 283-291. doi:10.1016/S0047-2484(85)80069-5

May, B., Moody, D. B., & Stebbins, W. C. (1989). Categorical perception of conspecific communication sounds by Japanese macaques, Macaca fuscata. *Journal of Acoustic Society of America, 85*, 837-847. doi:10.1121/1.397555

McKone, E., Martini, P., Nakayama, K. (2001). Categorical perception of face identity in noise isolates configural processing. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 573-599. doi:10.1037/0096-1523.27.3.573

Mollon, J. D., & Danilova, M. V. (1996). Three remarks on perceptual learning. *Spatial Vision, 10*, 51-58. doi:10.1163/156856896X00051

Munnich, E., & Landau, B. (2003). The effects of spatial language on spatial representation: Setting some boundaries. In D. Gentner & S. Goldin-Meadow, *Language*

in mind: Advances in the study of language and thought. Cambridge, Mass: MIT Press.

Munsell, A. H. (1905). *A Color Notation*. Boston: G. H. Ellis Co.

Notman, L., Sowden, P. T., & Özgen, E. (2005). The nature of learned categorical perception effects: a psychophysical approach. *Cognition, 95*, B1-B14 doi:10.1016/j.cognition.2004.07.002

Ohlemiller, K. K., Jones, L. B., Heidbreder, A. F., Clark, W. W., & Miller, J. D. (1999). Voicing judgements by chinchillas trained with a reward paradigm. *Behavioural Brain Research, 100*, 185-195. doi:10.1016/S0166-4328(98)00130-2

Özgen, E., & Davies, I. R. L. (2002). Acquisition of categorical color perception: A perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology: General, 131*, 477-493 doi:10.1037/0096-3445.131.4.477

Philipona, D. L., & O'Regan, D. K. (2006). Color naming, unique hues, and hue cancellation predicted from singularities in reflection properties. *Visual Neuroscience, 23*, 331-339. doi: 10.1017/S0952523806233182

Plunkett, K., Hu, J., & Cohen, L. (2008). Labels can override perceptual categories in early infancy. *Cognition, 106*, 665-681. doi:10.1016/j.cognition.2007.04.003

Premack D. (1983). The codes of man and beast. *Behavioral and Brain Sciences, 6*, 125-137. doi:10.1017/S0140525X00015077

Quinn, P.C., Eimas, P.D., & Rosenkrantz, S.L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception, 22*, 463-475. doi:10.1068/p220463

Regier, T., Kay, P. & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences, 104*, 1436-1441. doi:10.1073/pnas.0610341104

Repp, B.H. (1984). Categorical perception: Issues, methods, findings. *Speech and Language: Advances in basic research and practice, 10*, 243-335.

Roberson, D., & Davidoff, J. (2000). The 'Categorical Perception' of colors and facial expressions: the effect of verbal interference. *Memory and Cognition, 28*, 977-986.

Roberson, D., Davidoff, J., & Braisby, N. (1999). Similarity and categorisation: Neuropsychological evidence for a dissociation in explicit categorisation tasks. *Cognition, 71*, 1-42. doi:10.1016/S0010-0277(99)00013-X

Roberson, D., Davidoff, J., Davies, I., & Shapiro, L. (2004). The development of color categories in two languages: a longitudinal study. *Journal of Experimental Psychology: General, 133*, 554-571. doi:10.1037/0096-3445.133.4.554

Roberson, D. Davidoff, J., Davies, I., & Shapiro, L. (2005). Color categories: evidence for the relativity hypothesis. *Cognitive Psychology, 50*, 378-411. doi:10.1016/j.cogpsych.2004.10.001

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*, 369-398. doi:10.1037/0096-3445.129.3.369

Roberson, D., Hanley, J.R. & Pak, H. (2009). Thresholds for color discrimination in English and Korean speakers. *Cognition, 112*, 482-487. doi:10.1016/j.cognition.2009.06.008

Roberson, D., Pak, H & Hanley, J.R. (2008).Categorical perception of color in the left and right hemisphere is verbally mediated: Evidence from Korean. *Cognition, 107*, 752-762. doi:10.1016/j.cognition.2007.09.001

Rosch Heider. E., & Olivier, D.C. (1972). The structure of the color space in naming and memory for two languages. *Cognitive Psychology, 3*, 337-354. doi:10.1016/0010-0285(72)90011-4

Sandell, J.H., Gross, C.G., & Bornstein, M.H. (1979). Color categories in macaques. *Journal of Comparative and Physiological Psychology, 93*, 626-635. doi: 10.1037/h0077594

Saunders, B.A.C., & van Brakel, J. (1997) Are there non-trivial constraints on color categorization. *Behavioral & Brain Sciences, 20*, 167-232.

Sinnott, J.M., & Brown, C.H. (1997). Perception of the American English liquid /ra-la/ contrast by humans and monkeys. *The Journal of the Acoustical Society of America, 102*, 588-602. doi: 10.1121/1.419732

Siok, W. T., Kay, P., Wang, W. S. Y., Chan, A. H. D., Chen, L., Luke, K.-K., & Tan, L. H. (2009). Language regions of the brain are operative in color perception. *Proceedings of the National Academy of Sciences, 106*, 8140-8145. doi:10.1073/pnas.0903627106

Steinschneider, M., Fishman, YI., & Arezzo, JC. (2003). Representation of the voice onset time (VOT) speech parameter in population responses within primary auditory cortex of the awake monkey. *Journal of Acoustica Society of America, 114*, 307-321. doi:10.1121/1.1582449

Steels, L., & Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: A case study for color. *Behavioral and Brain Sciences, 28*, 469-489. doi:10.1017/S0140525X05000087

Strange, W., & Jenkins, J.J. (1978). Role of linguistic experience in the perception of speech. In R.D. Walk & H.L. Pick (eds.), *Perception and Experience*. New York: Plenum Press.

Tan, L.H., Chan, A.H.D., Kay, P., Khong, P-L, Yip, L.K.C., & Luke, K-K (2008). Language affects patterns of brain activation associated with perceptual decision. *Proceedings of the National Academy of Science, 105*, 4004-4009. doi:10.1073/pnas.0800055105

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., &

Kuipers, J. R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Science U S A, 106*, 4567-4570 doi:10.1073/pnas.0811155106

Thompson, R.K.R., & Oden, D.L. (2000). Categorical perception and conceptual judgments by nonhuman primates: The paleological monkey and the analogical ape. *Cognitive Science, 24*, 363-396. doi:10.1207/s15516709cog2403_2

Thompson, R.K.R. & Oden, D.L., & Boysen, S. T. (1997). Language-naive chimpanzees (Pan troglodytes) judge relations between relations in a conceptual matching-to-sample task. *Journal of Experimental Psychology: Animal Behavior Processes, 23*, 31-43. doi:10.1037/0097-7403.23.1.31

Viviani, P., Binda, P., & Borsato, T. (2007). Categorical perception of newly learned faces. *Visual Cognition, 15*, 420-467. doi: 10.1080/13506280600761134

Waxman, S.R., & Markow, D.B. (1995). Words as invitations to form categories: evidence from 12-month-old infants. *Cognitive Psychology, 29*, 257-302. doi:10.1006/cogp.1995.1016

Werner, J.S., & Wooten, B. R. (1985). Unsettled issues in infant color vision. *Infant Behavior & Development, 8*, 99-107 doi:10.1016/S0163-6383(85)80020-5

Whorf, B.L. (1956). The relation of habitual thought and behavior to language. In J.B. Carroll (ed.), *Language, Thought and Reality: Essays by B. L. Whorf*. Cambridge: M.I.T. Press.

Wilson, M., & Debauche, B.A. (1981). Inferotemporal cortex and categorical perception of visual stimuli by monkeys. *Neuropsychologia, 19*, 29-41. doi: 10.1016/0028-3932(81)90041-5

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Science, 104*, 7780-7785. doi:10.1073/pnas.0701644104

Wright, A. A., & Cumming, W. W. (1971). Color-naming functions for the pigeon. *Journal of the Experimental Analysis of Behavior, 15*, 7-17. doi: 10.1901/jeab.1971.15-7

Wright, C. (1975). On the coherence of vague predicates. *Synthese, 30*, 325-365. doi: 10.1007/BF00485049

Xu, F., & Carey, S. (2000). The emergence of kind concepts: a rejoinder to Needham and Ballargeon (2000). *Cognition, 74*, 285-301. doi:10.1016/S0010-0277(99)00076-1

Younger, B. (1985). The segregation of items into categories in ten-month-old infants. *Child Development, 56*, 1574-1583. doi:10.2307/1130476

Yoshioko. T., Dow, B.M., & Vautin, R.G. (1996). Neural mechanisms of color categorization areas V1, V2 and V4 of macaque monkey cortex. *Behavioral Brain Research, 76*, 51-70. doi:10.1016/0166-4328(95)00183-2