

A Colour Sorting Task Reveals the Limits of the Universalist/Relativist Dichotomy: Colour Categories Can Be Both Language Specific and Perceptual

Nicolas Claidière*
Yasmina Jraissati
Coralie Chevallier

Institut Jean Nicod, Pavillon Jardin – Ecole Normale Supérieure,
29, rue d'Ulm, 75005 Paris, France
Corresponding author, e-mail: nicolas.claidiere@normalesup.org

Abstract

We designed a new protocol requiring French adult participants to group a large number of Munsell colour chips into three or four groups. On one, relativist, view, participants would be expected to rely on their colour lexicon in such a task. In this framework, the resulting groups should be more similar to French colour categories than to other languages categories. On another, universalist, view, participants would be expected to rely on universal features of perception. In this second framework, the resulting groups should match colour categories of three and four basic terms languages. In this work, we first collected data to build an accurate map of French colour terms categories (Experiment 1). We went on testing how native French speakers spontaneously sorted a set of randomly presented coloured chips and, in line with the relativist prediction, we found that the resulting colour groups were more similar to French colour categories than to three and four basic terms languages (Experiment 2). However, the same results were obtained in a verbal interference condition (Experiment 3), suggesting that participants rely on language specific and nevertheless perceptual, colour categories. Collectively, these results suggest that the universalist/relativist dichotomy is a too narrow one.

Keywords

Colour categories, sorting task, Whorf, dual task, verbal shadowing

English speakers use the terms 'green' and 'blue' to refer to colours in the blue-green region of the colour space. In contrast, several languages around the world do not mark this distinction. For instance, Berinmo speakers only have one word, 'nol', to refer to the same blue-green area of the colour space (Roberson *et al.*, 2000). Such discrepancies in the colour lexicon have nurtured a century-long debate regarding colour cognition. This debate has traditionally

opposed the proponents of linguistic relativism (culminating in the Sapir-Whorf hypothesis: Ray, 1952; Whorf, 1956) according to whom language determines categorisation, and the proponents of a perceptual universalism (Berlin and Kay, 1969; Heider, 1972) according to whom colour categories are innate. These two positions, however, have been moderated over time (Kay and Regier, 2006). Proponents of linguistic relativism acknowledge the minimal thesis according to which colour categories are not completely arbitrary divisions of the colour space, while proponents of the universalist view acknowledge the fact that language does affect colour cognition in non trivial ways.

For instance, the seemingly trivial difference in the colour lexicon of Berinmo compared to English, produces important differences in colour cognition (Roberson *et al.*, 2000, 2005a). Indeed, the fact that English speakers have a word for 'blue' and another one for 'green' when Berinmo speakers refer to these two adjacent colours using one single term 'nol' has far reaching consequences. For instance, it has been found that English speakers will tend to judge them as more dissimilar than Berinmo speakers and they will also be less likely to confuse them in memory. This phenomenon of "categorical perception" (henceforth CP, Harnad, 1987), which causes the colour continuum to be perceived as discontinuous, has been reported in numerous studies involving colour similarity judgements or colour memory. Yet, the mechanisms underlying this discontinuity are still open to debate. Indeed, on the one hand, relativists have argued that these results, obtained in a vast number of experiments using different settings and languages (Kay and Kempton, 1984; Roberson *et al.*, 1999, 2000, 2005a; Roberson and Davidoff, 2000; Ozgen and Davies, 2002; Franklin and Davies, 2004; Pilling and Davies, 2004; Franklin *et al.*, 2005; Kay and Regier, 2006; Winawer *et al.*, 2007), fit their framework extremely well and provide strong evidence in favour of the influence of language on cognition. On the other hand, proponents of universalism have argued that this effect of language was compatible with their thesis. Indeed, although they admit that language has some effect on colour cognition, they also contend that colour categories are nascent, and consequently, that some CP is innate (Kay and Kempton, 1984; Franklin and Davies, 2004; Franklin *et al.*, 2005; Gilbert *et al.*, 2006; Drivonikou *et al.*, 2007).

In recent years, the study of the CP effect has thus become central in the debate regarding the effect of language on colour perception. Thus, the general question of how language influences colour perception is often tackled in the light of a more specific issue, namely the extent to which language influences CP. In what follows, we present the two alternative forced choice task (2AFC task), which is a standard paradigm testing CP effects in colour cognition, and we summarise some of the most striking findings that it uncovered. In the

2AFC task, participants are shown a colour probe and asked to remember it. They are subsequently presented with two test colours, one identical to the probe and one closely related to it, and asked which of the two is similar to the probe. If the test and the probe colours belong to the same lexical category they are more easily confused than if they belong to two different lexical categories. There are two possible explanations for this result. One is that participants *perceive* less difference between colours belonging to the same lexical category than between colours belonging to two different lexical categories. The other is that naming has a direct effect on *memory* and not on the perception of colours. Participants, instead of relying on the precise colour of the probe may use the name of that colour as a cue to recover it in the test phase. In cases where the two test colours belong to the same lexical category, relying on the name of the probe may induce confusion whereas in cases where the two test colours belong to two different lexical categories confusion may be limited. Thus, CP could be the result of the direct influence of the colour lexicon on memory, not genuinely affecting perceptual processes. This question is in fact quite central to the debate. Indeed, if CP is merely due to the influence of language on memory, a relatively modest version of linguistic relativism would be put forward. If, however, CP results from the influence of language on the *perception* of colours, a much stronger version of relativism would be supported. For this reason, a number of recent experiments have addressed this question by limiting the use of memory (e.g., in simultaneous judgment tasks). In line with strong relativism, results indicate that CP persists even when the use of memory is very limited (Winawer *et al.*, 2007), suggesting that CP is not the effect of colour naming on memory.

The effect of language on colour cognition was further demonstrated in the context of verbal interference tasks which manipulate the participant's access to the colour lexicon (Roberson and Davidoff, 2000; Pilling *et al.*, 2003; Winawer *et al.*, 2007). Typically, such paradigms involve dual tasks where the participant engages in one task related to colour cognition and, simultaneously, in another task requiring her to constantly repeat numbers for instance, thereby preventing her to use her colour lexicon. In such paradigms, participants are led to mostly rely on their *perception* of colours and not on colour *terms*. Results indicate that CP disappears when verbal interference is used (Winawer *et al.*, 2007), suggesting that CP crucially depends upon colour terms use. In line with linguistic relativism, these findings support the idea that CP results from a direct effect of colour naming on perception (Davidoff, 2001). In other words, it appears that, if the use of colour terms is prevented, individuals do not perceive the colour continuum categorically at all.

This set of results strongly favours a relativist perspective. However, CP is just one aspect of colour categorisation. Indeed, CP studies do not address the categorisation phenomenon as a whole, but only one of its possible factors. For instance, CP is usually observed in studies involving judgment tasks of a very specific area of the colour continuum; at the linguistic border between two colour categories. Colour categorisation, however, may not be exclusively understood by studies assessing the perceived similarity between adjacent colours. More generally, what needs to be understood, is how a physically continuous stimulus is partitioned into several discrete categories (Davidoff, 2001). In this respect, CP is only one aspect of the question. The more general issue of the categorisation behaviour goes beyond CP and may be addressed through a range of other tasks.

In particular, naming tasks and colour sorting tasks have also been used and, unlike CP studies, have provided strong support for the universalist position. Studies of categorisation using naming tasks have usually involved a large set of colours ranging over the whole colour continuum (Berlin and Kay, 1969; Kay and Regier, 2003, 2007; Regier *et al.*, 2005, 2007). For instance, in the World Color Survey (WCS) participants from 110 non-industrialised societies were asked to name 330 colours (the 320 most saturated colours of the Munsell atlas, and 10 achromatic colours) which are used to build the so-called 'Munsell map' (see Figure 2A). The aim of the WCS was to determine the extension of colour categories in various languages, based on the name given to each colour on the map. Although this method does not address the question of the relationship between categorisation and language, the data from the WCS have nevertheless been put forward in the debate and presented as strong support for the universalist view. Indeed, the large language database provided with by the WCS shows a surprising regularity across languages with the same number of basic colour terms. For instance, most three terms languages have a "red" category, a "light" and a "dark" one; most four terms languages have a "red", a "yellow", a "white" and a "dark" one and so on. These data support the idea that basic colour categories emerge, in each language, according to a specific and universal evolutionary sequence. For instance, the evolution from a three-term language to a four-term language is likely to differentiate the category "light", which includes white and yellow colours, into two distinct categories, "white" and "yellow". This evolutionary sequence is thus meant to reflect the most common path along which languages differentiate (see Figure 1).

Although the WCS data are of great support for the universalist theory, they cannot be used to assess the extent to which language influence colour categorization. Indeed, the participant is explicitly asked to make use of the colour

Stage	Basic Colour Categories					
1	White	Red	Yellow	Black	Blue	Green
2	White	Red	Yellow	Black	Blue	Green
3	White	Red	Yellow	Black	Blue	Green
4	White	Red	Yellow	Black	Blue	Green
5	White	Red	Yellow	Black	Blue	Green

Figure 1. The first five stages of the evolutionary sequence as modelled in by Kay and McDaniel (1978). In stage 1, the 6 universal colour categories are referred to by two terms (one referring to White, Red and Yellow, the other to Black, Blue and Green). In stages 2, 3 and 4, they are respectively referred to by 3, 4 and 5 terms. Thus, more basic colour terms appear and the universal foci which were formerly grouped (e.g., Red+Yellow in stage 2) are differentiated (e.g., Red and Yellow in stage 3). In stage 5, each universal colour category is referred to by a distinct colour term. It should be noted that although this evolutionary sequence is most frequently observed, it is not the only possible one (Kay and McDaniel, 1978; Kay *et al.*, 1991).

lexicon, which therefore makes it impossible to manipulate her access to the colour lexicon (e.g., by resorting to verbal interference). Sorting tasks, where no explicit mention of the colour lexicon is made, address this issue and provide crucial data for the question which is at stake here (Boster, 1986; Davies and Corbett, 1997; Davies, 1998; Pilling and Davies, 2004; Roberson *et al.*, 2005b). In a sorting task, participants are presented with colours covering the whole continuum and are asked to group them according to a specific criterion (e.g., the colour which could be said to belong to the same family). Using such a paradigm, Boster (1986) asked English speaking adults to group 8 focal colours (white, black, red, yellow, green, blue, orange, purple) into 2 groups, then to divide one of these two groups to form 3 groups, then again to divide one group to form 4 groups and so on until all the colours appeared in eight separate groups. Boster's aim was to test whether or not individuals would reproduce the evolutionary sequence postulated by the BCTT (see Figure 1). Quite strikingly, Boster found that the evolutionary sequence was indeed followed by the participants. From there, he concluded that the sorting behaviour of individual participants could recapitulate the global evolution of colour

terms in language, thereby providing an eloquent argument in favour of the universalistic view.

In 1998, however, Davies and Corbett called Boster's interpretation into question (Davies and Corbett, 1998). Their study, unlike Boster's, involved a large number of colours (67 colours), and three different languages with a variable number of basic colour terms (from 5 to 12). Similarly, participants in this study were asked to sort colours into groups (from 2 to 12). Davies and Corbett predicted that the colour groups would correspond to the categories found at each evolutionary stage of the sequence. They reported that the categories formed are often consistent with the theory, but, opposite to Boster, that the order of their appearance is not.

In sum, Boster's and Davies and Corbett's results both support the idea that adult participants, when engaged in a sorting task, group colours in a specific, universal, way. This claim leads to another prediction which will be the focus of this paper: participants speaking different languages and having a variable number of basic colour terms, should group colours in the same way, following a pattern predicted by the evolutionary sequence. More precisely, if French participants are asked to group a large number of colours into three or four groups, they should produce colour groups corresponding to languages with three and four basic terms respectively (i.e., stages 2 and 3 of the evolutionary sequence shown in Figure 1). Furthermore, if we perform a sorting task with a verbal interference condition, preventing them to access their colour lexicon, we should observe the same universal categories as in the non interference condition.

In our experiments, French participants were asked to perform a constrained sorting task under two different experimental conditions: with or without verbal interference. In this experimental context, universalists and relativists make radically different predictions. On the one hand, universalists predict that the colour groups formed by the participants will match the second and third stages of the evolutionary sequence; on the other hand, relativists predict that the groups will be mostly similar to French colour categories. This paradigm thus enables one to discriminate between the two theories.

General Outline of the Experiments

The goal of the present experiments is to address the question of the influence of language on adults' colour sorting behaviour and to discriminate between the universalist and the relativist position. We first built an accurate description of French colour terms categories and collected data to constitute a map representing the extension of each French colour term (Experiment 1). Using

a constrained sorting task, we asked French participants presented with 87 randomly distributed colour chips of the Munsell array to sort them into three or four groups (for previous experiments using free or constrained sorting tasks see Boster, 1986; Davies and Corbett, 1997; Davies, 1998; Davies and Corbett, 1998; Pilling and Davies, 2004; Roberson *et al.*, 2005b).

According to the relativist position, participants should rely on their colour lexicon during the sorting task and the categories they will form should closely match French colour categories. In contrast with this view, the universalist position predicts that participants, being required to constitute 3 or 4 groups of colours, should form the categories matching those found in languages with three or four basic terms respectively. Using the WCS online database, we thus identified languages with three and four colour terms with the goal of comparing them to the spontaneous sorting of French participants asked to divide a set of colour in 3 or 4 groups.

To sum up, we compared our participants' sorting behaviour to two different sets of categories: French colour categories and those of 3 and 4 colour term languages. Furthermore, two experimental settings were used to manipulate the access of participants to their lexicon: without and with verbal interference (Experiment 2 and 3 respectively).

Experiment 1: Assessing the French Basic Colour Categories

French colour categories have been described by MacLaury on the basis of a single person report only (McLaury, 1992). For the purpose of this study, we needed a reliable survey of French colour categories which could be subsequently compared to studies of three and four term languages reported in the WCS. To do so, we gathered data from French speaking participants according to the standard procedure used in the World Colour Survey (which, so far, has only surveyed languages from unindustrialised societies).

Method

Participants. 20 (11 men, 9 women) participants voluntarily took part in the study. The participants were students in cognitive sciences at the Institut des Sciences Cognitives in Lyon (France), they were aged between 22 to 33 years old (mean = 27) and were all native monolingual speakers of French with no specific knowledge of colours (painters, photographers, infographists were excluded). All participants had normal or corrected to normal vision (based on self-reports) and none were colour blind (Ishihara test, 1998).

Stimuli. In this experiment, we used the Munsell Stimulus Array which depicts 330 samples from the Munsell Book of Colour (Munsell Colour Company, 1966). In the Munsell Stimulus Array, the colours are organised on a grid following a lightness continuum (Munsell value) along the vertical axis and a colour (or “chromatic”) continuum along the horizontal axis. We used all 330 samples from the Munsell Stimulus Array. During the naming phase of the experiment, the stimuli we used were 1×2 cm glossy chips individually inserted in a square transparent glass slide case. During the elicitation of the best example of each colour category, the colour chips were mounted on a light grey piece of cardboard in their Munsell order (hues horizontally and lightness vertically).

Procedure. The World Colour Survey is a project aiming to compare colour categories across many languages and a fixed procedure is used in order to ensure the comparability of the data (see <http://www.icsi.berkeley.edu/wcs/>). We thus followed the guidelines available on the project’s website, and prepared the stimuli material in accordance to a personal communication by Terry Regier and Paul Kay, so that our data could be best compared with the other languages. The procedure includes two tasks: a naming task and a mapping task. In the naming task, the participants were shown each of the 330 colour chips (see Figure 2A) one at a time, in a fixed random order (corresponding to the one used in the usual procedure of the WCS). The participants were instructed to name the chips using the simplest term that first came to their mind. The following day, each speaker came back and indicated which chip in the array represented the best example, or focus, of each colour term in French (“mapping task”).

Both phases of the experiment took place during daylight, in a naturally lit room of the Institut des Sciences Cognitives.

Analysis. From the naming task we obtained a name for each of the 330 colour chip for each participant. As in other studies, we determined the modal colour term for each chip, i.e. the most frequent term associated to that particular chip (Davidoff *et al.*, 1999; Kay and Regier, 2007; Regier *et al.*, 2007). Next, we constructed a map which consisted in a visual representation of the most frequent term used for each colour chip (see Figure 3A).

Results

47 terms were spontaneously produced by our 20 participants. 11 of them are basic colour terms according to (Kay and McDaniel, 1978). Those terms are ‘noir’, ‘blanc’, ‘rouge’, ‘jaune’, ‘vert’, ‘bleu’, ‘marron’, ‘gris’, ‘rose’, ‘violet’ and

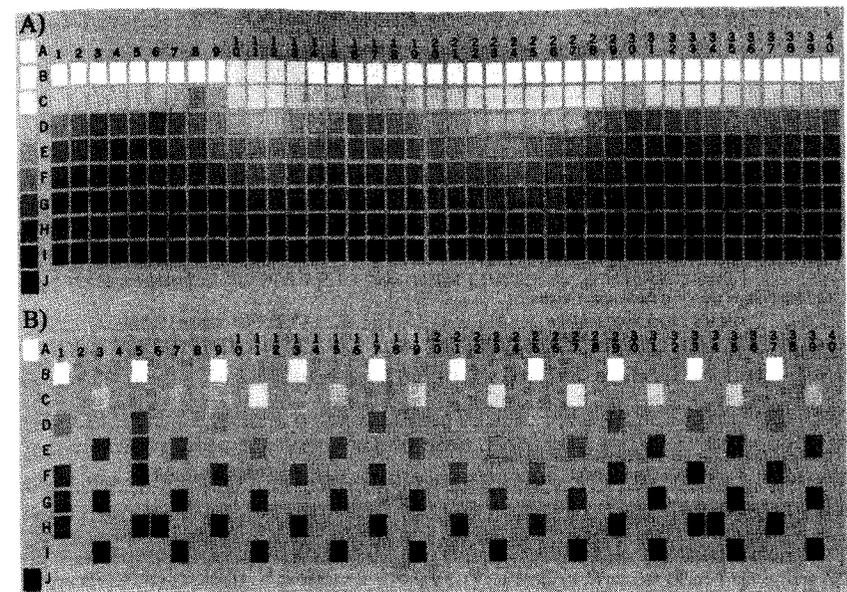


Figure 2. The Munsell colour chips used in the experiments. (A) The complete set of the Munsell Stimulus Array comprising 330 colour chips at maximum saturation. (B) The subset used during the sorting experiment. We eliminated every other column of the grid and every other chip on each line of the grid and added the missing focal colours identified by Regier *et al.* (2005)

‘orange’ – which correspond, respectively, to the English ‘black’, ‘white’, ‘red’, ‘yellow’, ‘green’, ‘blue’, ‘brown’, ‘grey’, ‘pink’, ‘purple’ and ‘orange’.

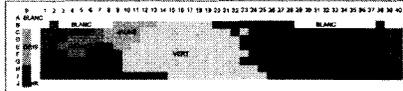
Discussion

We found that French had 11 basic terms by WCS standards (see Figure 3A), an unsurprising finding, which corresponds to previous results (McLaury, 1992) and in particular which is comparable to the English colour terms map (Heider, 1972; Roberson and Davidoff, 2000). This modal map is the one on which we ground the interpretation of our results in the rest of our study.

Experiment 2: The Constrained Sorting Task without Verbal Shadowing

In this experiment, we aimed to characterise how participants spontaneously sort a large number of colours and to assess the respective influence of

A: Modal map of French with 11 basic colour terms obtained from 20 native French speaking participants.

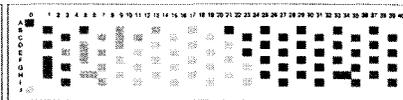


B: Modal map in the non verbal shadowing condition

B1: In the three group condition

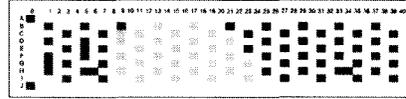


B2: In the four group condition



C: Modal map in the verbal shadowing condition

C1: In the three group condition

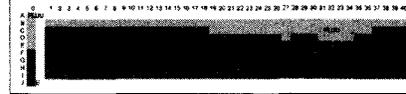


C2: In the four group condition



D: Modal map of non industrialised societies' languages

D1: Wobé, Ivory Coast



D2: Culina, Peru/Brazil

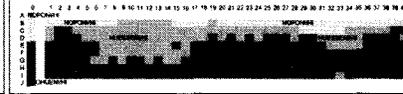


Figure 3. This figure represents the results we obtained in the three experiments we conducted along with representative maps of languages with 3 and 4 basic terms (Regier *et al.*, 2007). (A and D) Modal maps of the WCS' studies representing the most frequent name given to each of the 330 colour chips of the Munsell Array. (B and C) Modal maps representing the most frequent groups made by participants in the sorting task. (A) Result of the study of French colour categories. (B1) Result of the sorting task in the 3-group condition without verbal shadowing. (B2) Result of the sorting task in the 4-group condition without verbal shadowing. (C1) Result of the sorting task in the 3-group condition with verbal shadowing. (C2) Result of the sorting task in the 4-group condition with verbal shadowing. (D1) Modal map of the Wobé, a language with three basic terms. (D2) Modal map of the Culina, a language with four basic terms. All colours were arbitrary selected.

perceptual and linguistic factors on their sorting behaviour. Following the hypothesis presented in the introductory part of this paper, the aim of this experiment was to determine how much the grouping behaviour of participants could be predicted by the structure of the French lexicon on the one hand and by the evolutionary sequence on the other. With this goal in mind, we were interested in two things. First, we compared our participants' performances to those of speakers of languages with fewer colour terms (3 or

4 colour terms) and second, we compared the performances to the modal map of French colour terms created in Experiment 1.

Method

Participants. 36 (16 men, 20 women) participants voluntarily took part in the study. The participants were students in biology at Pierre and Marie Curie University, in Paris. They were aged between 18 to 38 years old (mean = 23) and were recruited according to the same criteria as in Experiment 1.

Stimuli. We selected our stimuli from the Munsell Stimulus Array which contains 330 colour chips. We selected 87 out of the 330 available Munsell colour chips to limit the duration of the task. The selection process involved the systematic elimination of every other column of the grid and every other chip on each line of the grid. This led to the selection of 80 chips with all lightness levels and half the chromatic dimensions represented. This first selection uniformly sample colour chips in the Munsell array but removes some of the universal focal colour. Because in the theory of Kay *et al.* universal focal colours are highly salient, we chose to include in the set those focal colour that were removed in the selection. Seven chips were then added: black (J0), white (A0), orange (E5), red (G1), brown (H6), violet (H34) and yellow (C9). The selected chips thus varied in hue and lightness and included all the universal focal points identified by Regier *et al.* (Regier *et al.*, 2005, Figure 2B).

Procedure. Participants were first asked a few questions about their age, knowledge of colour and were tested for colour blindness. The 87 colour chips were randomly spread out in front of the participant who heard the following set of instructions:

"On this table, there are 87 coloured chips. Your job is to form N ($N= 3$ or 4 depending on the experimental condition) sets of colour using all the chips. You need to constitute the colour sets as fast as possible because your performance will be timed".

The experiment took place during daylight, in a naturally lit room of Pierre and Marie Curie University, in Paris.

Analysis. We performed a hierarchical clustering analysis to see whether colour groups made by the participants were similar to each other. The raw data we obtained from our experiment were N ($N= 3$ or 4 depending on the condition)

groups of colour for each participant. For each group of colour we coded the presence or absence (1 or 0 respectively) of each of the 330 colours of the Munsell array. We then calculated the distance between colour groups using the 'jaccard distance', i.e., the percentage of non zero coordinates that differ. For instance if the 'Jaccard distance' is 0.6 between group A and B, it means that 60% of all the colours present in group A and B are not present in both A and B. Last we used the UPGMA (Unweighted Pair Group Method with Arithmetic mean) technique, which use the unweighted average distance between groups, to build a hierarchical cluster tree. We used the cophenetic correlation coefficient to check that the obtained dendrogram faithfully represent the data. In this way we obtained a hierarchical dendrogram representing the similarity between the colour groups made by each participant.

Once the dendrogram was obtained, we used the fact that the colour groups were not independent of each other, one participant making 3 or 4 colour groups, to find a distance below which colour groups were considered as similar (we used a distance of 0.6 in all experimental conditions).

Finally, we constructed a modal map which represents for each colour chip the most frequent colour group in which it used to be put. So, for instance, suppose among 10 participants, 6 make a blue group and a green group and 4 a blue-green group. Further suppose that the colour say number 12 is in the blue group for 4 participants, in the green group for 2 participants, in the blue-green group for 3 of them and in another group for one of them. Colour number 12 is most frequently associated with the blue group, hence, in the modal map it appears as belonging to the blue group.

Results

3-group condition. The first thing to note is that participants behave very consistently. Four types of sorting behaviours may be identified, one of them being adopted by up to 7 participants (out of 15). Thus, the majority grouped together [rouge, rose, orange, violet, jaune, blanc, marron], [vert, noir] and [bleu] – which correspond, in English, to [red, pink, orange, purple, yellow, white, brown], [green, black] and [blue]. Another 4 participants grouped together [violet, rose, blanc], [rouge, marron, orange, jaune] and [vert, bleu, noir] – [purple, pink, white], [red, brown, orange, yellow] and [green, blue, black]. These two alternative behaviours were the most frequent but 2 participants grouped colours on the basis of 3 lightness levels and the remaining 2 participants grouped the colours in a variable way.

Accordingly, the modal map we obtain from the data reflects the most frequent type of behaviour. The clustering analysis of the data yielded a dendrogram whose cophenetic correlation coefficient was 0.86 and we used a jaccard

distance of 0.6 to associate colour groups together and to construct the modal map (see Figure 3B1).

When we compare the obtained map with the map of a language with only three colour terms (compare 3B1 with 3D1), we see that the divisions between colour groups are along the hue dimension for the sorting task and mostly along the lightness dimension for the three term language. If the same criterion was guiding the behaviour in the two different experiments, we would at least expect divisions along the same dimension.

Moreover, when we compare the map obtained in the sorting task with the one obtained in the first experiment (compare 3B1 with 3A), we find a good fit between the two maps. More precisely we tested whether modal colour groups were more similar to a group of French categories than to those of three- or four-term languages by calculating a superposition coefficient. The superposition coefficient maximises the percentage of colours which share the same name in a particular language and are grouped together in our task. We tested whether the superposition index in French was significantly different from the superposition index of all three- or four-term languages surveyed by the WCS. In the three group condition we found that the superposition index was 81.6% for French and 67.5% (SD = 0.065) on average for the 10 three terms languages reported in the WCS (this difference approaches significance, *t*-test, *df* = 9, *P* < 0.068).

French colour categories divide the colour space mostly on the hue dimension and three terms languages mostly on the lightness dimension. Our participants grouped colours along the hue dimension and not the lightness dimension. Furthermore, the colour groups we obtain are more similar to groups of French categories than to the categories of languages having three basic terms.

4-group condition. Again we find that individuals behave very consistently, even more than in the three group condition. 8 individuals out of 20 grouped together: [rose, violet, blanc], [rouge, orange, marron, jaune], [vert, noir] and [bleu] – [pink, purple, white], [red, orange, brown, yellow], [green, black] and [blue]. Another 8 individuals made three of these four groups and one individual divided the colours along the lightness dimension.

Accordingly, the modal map we obtain from the data reflects the most frequent type of behaviour. The clustering analysis of the data yielded a dendrogram whose cophenetic correlation coefficient was 0.92 and we used a Jaccard distance of 0.6 to associate colour groups together and to construct the modal map (see Figure 3B2). In this condition we also observe that participants group colours along the hue dimension and not the lightness dimension, as in the French modal map and opposite to four terms languages (compare 3B2, 3A and 3D2).

We also tested whether the sorting map is more similar to the French map than to four terms languages map of the WCS studies. We found that in the four group condition the superposition index with French was 93.1% and with four terms languages 61.2% (SD = 0.072) on average for the nine four term languages reported in the WCS (this difference being significant, *t*-test, *df* = 8, *P* < 0.01).

Thus, the comparison between the map in the four group condition, the map of four terms languages and the modal map of French colour terms leads to the same result than previously; namely that participants grouped the colour in a way which is more similar to French than to other language's colour lexicon.

Discussion

Based on the arguments of the proponents of the universalistic view, individuals would have been expected to reproduce the stages 2 and 3 of the evolutionary sequence. This result would have been a strong argument in favour of the universalistic position, showing that perception crucially intervenes in the grouping of the colour chips. However, a comparison of the modal maps of the 3- and 4-group conditions with the maps of languages having 3 and 4 basic colour categories clearly do not support this hypothesis (see Figure 3). We therefore cannot conclude that individuals reproduce these stages of the evolutionary sequence.

Based on relativist expectations, participants would have been expected to use language specific colour names to group colours. Hence, participants, when asked to group colours may (un)consciously reason as follows: 'this chip is pink, it goes in the group with other pink chips; this one is yellow, it goes in the group with other yellow chips; so on and so forth'. Using a colour naming strategy would explain why the colour groups match the colour categories.

We have seen in the introduction that previous studies on CP clearly indicate that language specific colour names affect colour cognition. Some of these studies also show that CP may disappear under verbal interference, suggesting that our participants may have some difficulties to perform the task under verbal interference. If the stages of the evolutionary sequence are not recovered because participants rely on colour terms, one could expect their behaviour to be influenced by universal categories when their naming strategy is disrupted. We test these two alternative hypotheses in the next experiment.

Experiment 3: The Constrained Sorting Task with Verbal Shadowing

In this experiment, we performed a verbal shadowing task designed to interfere with a potential strategy based on language. Verbal shadowing is a form

of strong verbal interference where participants are asked to repeat a continuous stream of speech without pausing more than one second, while performing the sorting task (for a similar task in other domains see Hermer-Vazquez *et al.*, 1999; Newton and de Villiers, 2007). This obviously requires extreme concentration and arguably affects memory, attention and lexical access.¹ In such a situation, participants are thus expected to rely mostly on their perception of colours and not on their colour lexicon. We predicted that if linguistic cues were so pregnant in the non shadowing task, verbal shadowing would alter the participants' behaviour. Indeed, similar tasks to the one we used here have been resorted to in spatial recognition tasks and have forcefully shown that verbal shadowing masks the participants' access to their verbally represented spatial orientation (Hermer-Vazquez *et al.*, 1999). In Experiment 3, participants were thus asked to perform the same task as in Experiment 2 while repeating a continuously recorded prose passage that was played throughout the session. The aim of the experiment was to assess whether simultaneous verbal shadowing had an impact on the sorting strategy used by the participants.

Method

Participants. 33 participants (18 men, 15 women) voluntarily took part in the study. The participants were all undergraduate students in biology at the University of Jussieu in Paris, they were aged between 18 and 28 years old (mean = 21.5) and were recruited according to the same criteria as in Experiment 2.

Stimuli. The colour stimuli were identical to Experiment 2. For the verbal shadowing task, a prose text (giving facts about some vegetables) was read by a native French female speaker and recorded in a quiet room (Sennheiser microphone) using the software WaveStudio. The speech signal was digitised at a 22 kHz/16 bits sampling rate.

Procedure. The procedure was identical to Experiment 2 but the sorting task was preceded by a training phase during which participants were taught how to perform verbal shadowing. Participants were asked to repeat the verbal material played through their earplugs as it was spoken, syllable by syllable or word by word, rather than waiting for larger chunks.

"You will be asked to repeat what you hear through the headphones as you hear it. You should try not to wait for the end of the sentences but you should

¹ Note, however, that because all stimuli were laid in front of participants the sorting task in itself required only a limited amount of memory.

rather focus on the words, or even the syllables that are pronounced without aiming to understand the meaning of the text. What truly counts is that you manage to keep on talking, even if what you say makes no sense to you”.

The training was terminated when the participant was able to shadow for about one continuous minute without pausing more than a second at any time. If a participant failed to train for verbal shadowing, she did not take part in the study (this happened for 20 participants). Once the participant reached this training criterion, the experimental phase started. The participants were told that they would have to perform continuous shadowing whilst sorting 87 colour chips in N ($N = 3$ or 4 depending on the condition) groups of colours. They then started shadowing again, using a different prose text, as the experimenter lifted the sheet initially covering the colour chips so that the participant could start the sorting task. The experiment took place during daylight, in a naturally lit room of Pierre et Marie Curie University, in Paris.

Analysis. We performed exactly the same analysis as in Experiment 2.

Results

3-group condition. The participants' sorting behaviour is still highly consistent. 9 participants, out of 16, grouped together [rouge, rose, orange, violet, jaune, blanc, marron], [vert, noir] and [bleu] – [red, pink, orange, purple, yellow, white, brown], [green, black] and [blue] and 4 participants grouped together [violet, rose, blanc], [rouge, marron, orange, jaune] and [vert, bleu, noir] – [purpe, pink, white], [red, brown, orange, yellow] and [green, blue, black]. The remaining 3 participants grouped the colours in a variable way.

Accordingly the modal map reflects the most frequent behaviour. The clustering analysis of the data yielded a dendrogram whose cophenetic correlation coefficient was 0.86 and we used a jaccard distance of 0.6 to associate colour groups together and to construct the modal map (see Figure 3C1). We found that the superposition index was 87.4% for French and 65.9% on average (0.073 SD) for the 10 three-term languages reported in the WCS (this difference being significant, t -test, $df = 9$, $P < 0.05$).

This results are similar to the non shadowing condition, the colour groups made during the sorting task are more similar to French colour categories than to three-term language's colour categories.

4-group condition. Again, the participants sorting behaviour is very consistent. Out of the 13 participants, 9 grouped together: [rose, violet, blanc], [rouge, orange, marron, jaune], [vert, noir] and [bleu] – [pink, purple, white], [red,

orange, brown, yellow], [green, black] and [blue]. Among the 4 remaining participants, all but one made at least one of these four groups.

The modal map is again very clear. The clustering analysis of the data yielded a dendrogram whose cophenetic correlation coefficient was 0.92. We used a distance of 0.6 to associate colour groups together and to construct the modal map (see Figure 3C2). We found that in the 4-group condition the superposition index with French was 90.8% and with four-term languages 60.4% on average (0.076 SD) for the nine 4-term languages reported in the WCS (this difference being significant, t -test, $df = 8$, $P < 0.01$). Again the behaviour of participants is closer to the French colour categories than to four term languages.

Discussion

Given the regularity of our first experiment, and the fact that the colour groups made by the participants were more similar to French categories than to other languages categories, our second experiment aimed to test whether or not a naming strategy was used by the participants during the sorting task. We therefore performed a strong verbal shadowing task. We predicted that if linguistic cues were so pregnant in the non shadowing task, verbal shadowing would alter the participants' behaviour. For instance, Winawer *et al.* showed that verbal interference affect the categorical perception of the distinction between 'goluboy' (light blue) and 'siniy' (dark blue) in Russian (Winawer *et al.*, 2007). It therefore seems reasonable to assert that participants were not accessing their colour lexicon when performing the sorting task under verbal shadowing.

However, opposite to our expectations, we found that verbal shadowing had little influence on people's behaviour. In this context, a reasonable conclusion would be that neither the universalistic nor the relativistic theoretical frameworks allow a satisfying explanation of these results. We will develop this point further in the general discussion.

General Discussion

This paper reports data from three experiments aiming to gain insights into how French speakers spontaneously categorise colours. This work brings further evidence in the debate opposing relativists and universalists.

Taking a universalistic stance, we predicted that, when asked to sort 87 colour chips into three (or four) groups, participants' spontaneous groupings would correspond to the second (or third) stage of the evolutionary sequence

and reproduce the colour categories most widely found in languages with three (or four) colour terms. The results of the second experiment indicate that participants sorted the colours in groups that did not correspond to the predicted colour categories. For instance, when asked to form 4 groups, participants produced what corresponded to a green group, a blue group, a red group and a pink group,² which clearly differs from what is predicted by the evolutionary sequence (see Figure 3). One possible explanation for these results is that participants were actually using a verbal strategy. In Experiment 3, we thus used the same procedure as in Experiment 2, together with a verbal shadowing task to disrupt a potential verbal strategy. The results we obtained in Experiment 3 were highly consistent with those of Experiment 2. Thus, in the verbal shadowing condition, the universalistic prediction was not met.

So far, existing colour sorting studies have yielded contradictory results (Boster, 1986; Davies and Corbett, 1998). For instance, Boster took his results as showing that the evolutionary sequence is recapitulated by individuals (Boster, 1986) but his conclusion was called into question by Davies and Corbett (1998). Davies and Corbett took their results to be compatible with the universalist view but once associated to a weak relativist position, according to which the sorting behaviour reproduces universal categories, but not the stages of the evolutionary sequence (Davies and Corbett, 1997, 1998; Davies, 1998; Davies *et al.*, 1998). In comparison, the results of our study provide strong evidence against the universalist thesis. One of the main issues with Boster and Davies and Corbett's results is that the participants' behaviour was compared to the way the 6 universal colour foci were grouped in the 1978 model of the basic colour term theory. In other words, the interpretation of the results presented in these experiments is heavily dependent upon the *a priori* hypothesis the theory makes regarding the nature, number, position and role of the focal points. Unlike Boster and Davies and Corbett, we used actual three and four term languages as a test of the universalist prediction. This is indeed a more appropriate comparison because it is entirely faithful to the way each individual colour was sorted, therefore not depending on a commitment to the notion of colour foci. In contrast, it rests on a basic universalist postulate, common to all the variations of the universalist framework. This postulate states that all individuals, irrespective of their language, share the same cognitive mechanisms guiding colour categorisation. In our experiment, this leads to the clear prediction that the colour groups formed by the participants should be more similar to categories of three- (or four-) term languages than

² Here we use colour names to designate those groups for convenience; please refer to the difference in as shown in Figure 2.

to French colour categories. We showed that, contrary to the universalist minimal hypothesis, colour groups were more similar to French colour categories than to the categories found in three- (or four-) term languages.

From a relativist point of view, the fact that participants are able to consistently group colours according to the categories available in their language is rather unsurprising (Experiment 2). However, the fact that participants still formed language specific colour groups in the verbal shadowing task is quite unexpected (Experiment 3). Indeed, CP studies show that categorical perception disappears under verbal interference (Winawer *et al.*, 2007). Note that studies of the influence of language on CP claim to address the more general question of the influence of language on colour categorisation and cognition. The experimental study of categorical perception thereby presupposes that the same cognitive capacities intervene both in colour discrimination (tested in CP studies) and in colour categorisation (tested in sorting tasks). To this extent, it makes sense to suppose that colour discrimination tasks and colour sorting tasks behaviours will be equally disrupted by verbal interference. Our results, however, indicate that our verbal shadowing task did not change the participants' sorting behaviour. From there, we can conclude that our French speakers categorised the colours according to the colour categories available in their language in the absence of a direct access to colour terms.

One possible explanation of the discrepancies between CP studies and our sorting task is that the verbal shadowing task was not efficient enough to fully hinder the participants from using their lexicon. In favour of this argument is the fact that verbal shadowing has been used to limit the formation of propositions (such as 'at the left of the blue wall', Hermer-Vazquez *et al.*, 1999), not the simple access to words which, arguably, is easier. However, previous studies in colour categorisation have used very simple verbal interferences tasks, such as remembering a digit string and repeating it mentally (Winawer *et al.*, 2007), or reading colour words aloud before the test phase (Roberson and Davidoff, 2000) and they nevertheless found an effect of verbal interference. In this context, it seems unlikely that such a stringent condition as verbal shadowing had no effect on the cognitive mechanism which is impaired by much easier tasks.

A more plausible explanation is that colour discrimination in the context of CP studies and in that of sorting tasks, involve different cognitive capacities. For instance, CP could disappear under verbal interference, while the participants' sorting behaviour would remain unaffected. This would suggest that results based on CP studies and on sorting tasks cannot be generalised to 'categorisation' as a whole but must be confined to one aspect of 'categorisation'. Following this view, our results show that for some aspects of categorisation,

namely CP, colour terms are crucially involved whereas for other, such as colour grouping, colour terms are not indispensable. At present this possibility cannot be ruled out.

However, if one assumes that colour sorting tasks and CP studies rely on the same cognitive mechanisms, as seems to be generally the case, the discrepancy between CP studies and our sorting tasks requires some additional explanation. One possibly crucial difference between the two paradigms is that they investigate different colour categories. In the sorting task, the precise colour categories used by participants to group the colours are partially indeterminate. For instance, the 'red' group in the three-group condition (see Figure 3) includes both the red and orange French categories. Therefore we cannot determine whether participants considered two independent categories and grouped them together or used only one category made of red and orange. Our conclusion is that the language specific colour categories influencing the sorting behaviours of our participants are deeply rooted in perceptual mechanisms, but we cannot precisely determine which categories are deeply perceptual and which are not. One possibility is that some categories such as (red+orange) are perceptually salient whereas orange alone is not. One could therefore argue that the CP effect between 'goluboy' and 'siniy', two Russian terms for light and dark blue, could disappear under verbal interference because those two colour categories are more linguistically determined than perceptually rooted (Winawer *et al.*, 2007). In that case, our experiment would show that the role of language and perception vary according to the category, ranging from more perceptually to more linguistically determined categories – where 'perceptually determined categories' also include linguistic categories that are perceptually rooted.

Conclusion

Experiments addressing the universalist/relativist debate in the field of colour naming and colour cognition have been of three kinds mainly. First, the effect of language on colour perception has been studied by categorical perception studies; usually involving a judgment task and a restricted range of colours. Second, the universal properties of colour cognition have been addressed by naming tasks involving a large number of colours (typically 330). Finally, a third, less popular, type of study has focused on sorting tasks and given rise to ambiguous results in the pasts that, we believe, is mostly due to the theoretical framework in which the experiments have been elaborated.

We designed a new experimental paradigm by using a constrained sorting task associated with the data available in the WCS. This allowed us to compare

our results with the structure of languages studied according to the WCS protocol. The results of our first sorting experiment were clear-cut: French adults do not sort colours according to the universal perceptual features determining the categorisation behaviour of speakers of 3- and 4-term languages. They sort colours according to the colour categories of their own language. This result could be read as supporting the relativistic view, however, the fact that French participants' sorting behaviour remained consistent in a verbal shadowing condition, suggests that their sorting behaviour did not solely rely on a direct access to their lexicon.

Taken together, our results thus seem to indicate that colour categories can be both language specific and perceptual, a view challenging previous results of colour grouping tasks and the generality of the results obtained in CP studies. This further suggests that the dichotomy between the universalist and the relativist view is too narrow, echoing recent theoretical development (Kay and Regier, 2006; Regier *et al.*, 2007).

Acknowledgement

We would like to thank Roberto Casati, Paul Kay and Dan Sperber for their comments on earlier versions of the manuscript.

References

- Berlin, B. and Kay, P. (1969). *Basic color terms: their universality and evolution*. University of California Press, Berkeley, CA.
- Boster, J. (1986). Can individuals recapitulate the evolutionary development of color lexicons? *Ethnology* 25, 61-74.
- Davidoff, J. (2001). Language and perceptual categorisation. *Trends in Cognitive Sciences* 5, 382-387.
- Davidoff, J., Davies, I. and Roberson, D. (1999). Colour categories in a stone-age tribe. *Nature* 398, 203-204.
- Davies, I. (1998). A study of colour grouping in three languages: a test of the linguistic relativity hypothesis. *British Journal of Psychology* 89, 433-452.
- Davies, I. and Corbett, G. (1997). A cross-cultural study of colour grouping: evidence for weak linguistic relativity. *British Journal of Psychology* 88, 493-517.
- (1998). A Cross-Cultural Study of Color-Grouping: Tests of the Perceptual-Physiology Account of Color Universals. *Ethos* 26, 338-360.
- Davies, I., Swoden, P. T., Jerrett, D. T., Jerrett, T. and Corbett, G. (1998). A cross cultural study of English and Setswana speakers on a colour triads task: a test of the Sapir-Whorf hypothesis. *British Journal of Psychology* 89, 1-15.
- Drivonikou, G. V., Kay, P., Regier, T., Ivry, R. B., Gilbert, A. L., Franklin, A. and Davies, I. R. L. (2007). Further evidence that Whorfian effects are stronger in the right visual field than the left. *Proceedings of the National Academy of Sciences of the United States of America* 104, 1097-1102.

- Franklin, A., Clifford, A., Williamson, E. and Davies, I. (2005). Color term knowledge does not affect categorical perception of color in toddlers. *Journal of Experimental Child Psychology* 90, 114-141.
- Franklin, A. and Davies, I. (2004). New evidence for infant color categories. *British Journal of Developmental Psychology*, 22, 349-377.
- Gilbert, A. L., Regier, T., Kay, P. and Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences of the United States of America* 103, 489-494.
- Harnad, S. R. (1987). *Categorical perception: the groundwork of cognition*. Cambridge University Press, Cambridge.
- Heider, E. R. (1972). Universals in color naming and memory. *Journal of Experimental Psychology* 93, 10-20.
- Hermer-Vazquez, L., Spelke, E. S. and Katsnelson, A. S. (1999). Sources of flexibility in human cognition: dual-task studies of space and language. *Cognitive Psychology* 39, 3-36.
- Kay, P., Berlin, B. and Merrifield, W. (1991). Biocultural Implications of Systems of Color Naming. *Journal of Linguistic Anthropology* 1, 12-25.
- Kay, P. and Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist* 86, 65-78.
- Kay, P. and McDaniel, C. K. (1978). Linguistic Significance of Meanings of Basic Color Terms. *Language* 54, 610-646.
- Kay, P. and Regier, T. (2003). Resolving the question of color naming universals. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 9085-9089.
- (2006). Language, thought and color: recent developments. *Trends in Cognitive Sciences* 10, 51-54.
- (2007). Color naming universals: The case of Berinmo. *Cognition* 102, 289-298.
- McLaury, R. E. (1992). From brightness to hue: an explanatory model of color-category evolution. *Current Anthropology* 33, 137-186.
- Newton, A. M. and de Villiers, J. G. (2007). Thinking while talking: adults fail nonverbal false-belief reasoning. *Psychological Science* 18, 574-579.
- Ozgen, E. and Davies, I. (2002). Acquisition of categorical color perception: a perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology: General* 131, 477-493.
- Pilling, M. and Davies, I. (2004). Linguistic relativism and colour cognition. *British Journal of Psychology* 95, 429-455.
- Pilling, M., Wiggert, A., Ozgen, E. and Davies, I. R. L. (2003). Is color "categorical perception" really perceptual? *Memory and Cognition* 31, 538-551.
- Ray, V. F. (1952). Techniques and Problems in the Study of Human Color Perception. *Southwestern Journal of Anthropology* 8, 251-259.
- Regier, T., Kay, P. and Cook, R. S. (2005). Focal colors are universal after all. *Proceedings of the National Academy of Sciences of the United States of America* 102, 8386-8391.
- Regier, T., Kay, P. and Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences of the United States of America* 104, 1436-1441.
- Roberson, D. and 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. and Braisby, N. (1999). Similarity and categorisation: neuropsychological evidence for a dissociation in explicit categorisation tasks. *Cognition* 71, 1-42.
- Roberson, D., Davies, I. and 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.
- Roberson, D., Davidoff, J., Davies, I. and Shapiro, L. R. (2005a). Color categories: evidence for the cultural relativity hypothesis. *Cognitive Psychology* 50, 378-411.
- Roberson, D., Davies, I., Corbett, G. and Vandervyver, M. (2005b). Freesorting of colors across cultures: Are there universal grounds for grouping? *Journal of Cognition and Culture* 5, 349-386.
- Whorf, B. L. (1956). *Language, thought, and reality; selected writings*. MIT Press, Cambridge, MA.
- Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R. and Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences of the United States of America* 104, 7780-7785.