Journal of Psycholinguistic Research (2023) 52:1–16
https://doi.org/10.1007/s10936-021-09791-2

Colour Categorization and its Effect on Perception: A Conceptual Replication

Lenka Štěpánková[1,2] · Tomáš Urbánek[3,4]

Accepted: 14 July 2021 / Published online: 24 July 2021

Abstract
The presented study examines the question of colour categorization in relation to the hypothesis of linguistic relativity. The study is based on research conducted by Gilbert et al. (2006) and their claim that linguistic colour categorization in a particular language helps colour recognition and speeds the process of colour discrimination for colours from different linguistic categories but only for the right visual field. Our study approached the research question differently. We used the same methodology as Gilbert's team et al. (2006), but we used different colour categories in the Czech language and significantly enlarged the number of participants to 106 undergraduate psychology students. Our results show that the fastest reaction times were in trials when the target was located in the left visual field, quite opposite from the Gilbert's et al. (2006) study. We argue that this finding is based on different processes than simple colour linguistic categorisation and attentional processes actually play an important role in the task.

Keywords Language · Colour categorization · The hypothesis of linguistic relativity

Introduction

The Hypothesis of Linguistic Relativity and Colour Categorization

The principles of colour naming and overall language development in different cultures is a scientific question that anthropology and other disciplines such as psychology have tried to solve for centuries. In nineteenth century anthropology, the prevailing opinion was

✉ Lenka Štěpánková
lenka.stepankova@mail.muni.cz

1 Department of Psychology, Faculty of Social Studies, Masaryk University, Joštova 218, 602 00 Brno, Czech Republic

2 The Institute for Research On Children, Youth and Family, Faculty of Social Studies, Masaryk University, Brno, Czech Republic

3 Department of Psychology, Faculty of Arts, Masaryk University, Brno, Czech Republic

4 Institute of Psychology, The Czech Academy of Sciences, Brno, Czech Republic

that unwritten non-European languages were not as rich as the known written European ones. Franz Boas and his student Edward Sapir had a different opinion (Kay & Kempton, 1984). They studied these non-European languages and tried to show how rich they could be. Sapir and his student Benjamin Whorf later formulated a hypothesis based on their research of various languages, written and non-written, that structural differences in languages are parallel to non-linguistic cognitive differences. This idea is a cornerstone of the hypothesis of linguistic relativity, also known as the Sapir-Whorf hypothesis.

The idea is that the structure of a language influences (weak view of the linguistic relativity hypothesis) or determines (strong view of the linguistic relativity hypothesis) the world view of a native speaker of that language (Kay & Kempton, 1984). This hypothesis suggests relativity in an observer's cognitive perception of the world, and this relativity is more or less dependent on language structure: the way people of a certain culture name the things around them influences their perception of their surroundings. One of the best ways to test this hypothesis is with colour categories. There are many colours visible to human eye, but not as many linguistic categories for naming them. Also, various languages differ in the number of words for colour categories. Children learn the names for colours relatively slowly, which is an interesting fact on its own and can serve as an example on the relation between language and thought (Roberson & Hanley, 2010). Colour is therefore a clear candidate for the research of linguistic relativity. But there are other domains that are obviously very differently named in various cultures, raising the question of whether the perception of these domains is different. Casasanto (2008) suggests studying different names for time periods in various languages and different language expressions about time. Gordon (2004) presented research from the Amazonian tribe of the Pirahã. Members of this tribe did not develop a counting system apart from words representing 'one' and 'two'. According to Gordon (2004), members of the Pirahã tribe are strongly affected by their lack of a numerical system in specific numerical tasks, leading Gordon to suggest the plausibility of the linguistic relativity hypothesis.

There is a strong body of opposition to the linguistic relativity hypothesis. Berlin and Kay (1991; originally published in 1969) studied 20 languages and determined that colour categorization is not random; some universal tendencies exist among the languages examined. Kay and McDaniel (1978) support the conclusion by Berlin and Kay (1991), suggesting that colours have universal categories based on the biology of the human vision. Lindsey and Brown (2002) examined this idea. They studied languages that do not linguistically distinguish between blue and green (known as 'grue' languages) and compared them with those that do. The authors claimed that people who live in parts of the world where they are exposed to high levels of ultraviolet-B light are prone to abnormally high optical density in the lens (yellow pigments accumulate in the lens as it ages, causing blue light to seem greenish). Their analysis of 203 languages showed that 'grue' language cultures live close to the equator (high UV-B exposure) and blue/green language cultures live in higher altitudes (relatively low UV-B values). There are of course some differences among individuals within a culture. The authors asserted that these results show that linguistic relativity is not needed to explain why certain cultures have different colour naming systems than others. Language categories seem to be formed based on perception specifications caused by the physiology of the eye (Lindsey & Brown, 2002). Kay and Kempton (1984) conducted an experiment with native speakers of English and native speakers of Tarahumara (a language spoken in northern Mexico) with only one category, *siyóname*, for green and blue. The study found that English speakers reported the difference between green and blue hues as larger than the Tarahumara speakers did, but this effect disappeared when the authors used a different methodology. The authors concluded that there could be a *naming* *strategy*, which suggests that the cognitively perceived distance between colours from different language categories is the same between languages that have and do not have these categories. The reason for their results could be an effect of *knowing the name of a colour category* in rating colours and subjectively perceived differences between them. The *naming strategy* is described also in the research conducted by Zhou et al. (2010), who let one group of their participants learn new made up colour names for two colour hues, that would both be considered "blue" and compared the results of the visual search task between the group that learned a new categories for two blue hues and the group that did not go through the learning phase and labeled both the hues as simply blue. They found out, that even with the made up colour categories, that were artificially created by the researchers and learned by participants, the results show that the discrimination was faster for the between-categories colour target-distractor pairs with target located in the right visual field. Winawer et al. (2007) report similar results that could be explained by the *naming strategy* for Russian and English speaking participants. Russian linguistically distinguishes between *dark blue* and *light blue* (*siniy* and *goluboy,* respectively), while English considers light and dark blue as a simple "blue". Results showed that for the Russian speaking participants, there was a difference in discriminating colour hues crossing the colour boundary (*siniy* and *goluboy*) and colours from the same colour category (either all of the colours were *siniy,* or all of them were *goluboy*). This difference was not found among English speakers. The colour *naming strategy* may affect not only the perceptual discrimination between presented colour hues, but also how participants remember the colours. Lowry and Bryant (2019) showed, that Russian participants tend to remember the colour of a presented eyes as more gray while English speakers tend to remember them as more blue. The naming strategy may actually trigger some kind of priming in participants, where knowing the name of the category and perceiving it as distinct from other categories may affect later perception of the representants of those categories. This is consistent with the conclusion by Wolff and Holmes (2011) that language might be a powerful tool that somehow affects thinking, but that does not automatically mean that linguistic determinism, in the form of either weak or strong linguistic relativity hypothesis, is a plausible explanation for the relation between language and thought.

Testing of Linguistic Relativity Hypothesis by Gilbert et al. (2006)

Gilbert et al. (2006) tested the linguistic relativity hypothesis using very interesting methodology (Gilbert et al., 2006; Regier & Kay, 2009). First, participants rated colour chips as either blue or green. The colour boundary between green and blue was thereby set for further experiments. Next, participants were presented with a circle made of 12 colour chips, eleven of which were of the same colour. A chip of a different colour represented the target. The participants' task was to determine as quickly and as accurately as possible whether the target was on the left or the right side of the circle.

All the pairs of four selected colour hues were used to test whether the participants' reaction time would be shorter for trials using colours that cross the colour boundary. According to the linguistic relativity hypothesis, subjects perceive differences between colours from different semantic categories easier and faster than among those from the same colour category (Gilbert et al., 2006). The results confirmed the authors' hypothesis, but only for the right visual field. They interpreted these findings as a result of the brain regions for language being located mainly in the left brain hemisphere. The left hemisphere has a direct connection to the right visual field, and that is why the results were only conclusive for the right visual field. In the second part of the study, the participants had to complete a verbal interference task while working on a focal colour task, and the effect disappeared. The authors interpreted this result as supportive of their original interpretation. If the brain regions for language are occupied by other task, they do not seem to affect the colour categorization and recognition task. Gilbert et al. (2006) present this study as evidence in favour of the linguistic relativity hypothesis, but only for the right visual field.

Drivonikou et al. (2007) replicated the effect found by Gilbert et al. (2006), in which the participants were able to detect a target faster in the right visual field if the target was from a different colour category than the distractors. On the other hand, Drivonikou et al. (2007) also found the same (but weaker) effect for the left visual field; Gilbert's et al. (2006) study showed no effect or very little effect on the left visual field. The Drivonikou's et al. (2007) finding of an effect for a left visual field may have suggested an interesting notion about the processes that play a role in the Gilbert's et al. (2006) discrimination task described above. Bartolomeo (2006) describes an attention process as the ability of *"…an individual to respond quickly and accurately to incoming information by selecting relevant and ignoring irrelevant stimuli*". The task may actually trigger an attention networks that have been located mainly in the right hemisphere (Shulman et al., 2010; Bartolomeo & Chokron, 2002; Bartolomeo, 2006; Chica et al., 2012), which could suggest, that the right hemispheric attention networks play a role in the colour discrimination task and that may be one of the reasons Drivonikou et al. (2007) found an effect in the left visual field. But Drivonikou and colleagues (2007) were not the only one, who explore the left visual field effect in colour discrimination task.

Franklin et al. (2008) used eye tracking to find similar results in children who had already learned colour categories; those children had faster differentiation of colours from different categories in the right visual field. But children only learning colour categorization at the time of the experiment had a reversed effect of faster differentiation between colours from different categories but for the left visual field. The authors (Franklin et al., 2008) concluded that their results suggest developmental changes in the localization of categorical perception of colours, and they consider their findings to be proof of language's influence on perception.

To back up their results about the effect of the right visual field, Gilbert et al. (2008) later tried the same method of left–right decisions about the target position but with different stimulus material. They used black silhouettes of cats and dogs on a white background; the results were the same. In the cross-category trials (dog among cats/cat among dogs), participants showed faster reaction times in the left–right decision about the target position in the right visual field; the effect disappeared during verbal interference trials.

The main reason we decided to conduct our study was the fact that most of the previous studies have tested very few participants and usually focused on using the blue/green colour categories. By conducting this replication we try to argue, that the linguistic categorizations processes are not the only ones playing an important role in the interpretation of Gilbert's et al. (2006) results. Also, the Central European languages (not spoken by many people around the world) are not often used as focal languages in the psycholinguistic research. We find it important to include Central European languages in the linguistic research and publish results that challenge the findings from research conducted on widely spoken languages (such as English, Russian etc.) or languages spoken by specific tribes in Amazonia.

Regarding the replication, we did not use the same participants and exactly the same conditions as Gilbert et al. (2006), it is not an exact replication. We also did not use the exact conditions with a different participant pool, so it is not an approximate replication.

We would consider our replication a conceptual one (more on replications in Appelbaum et al., 2018; Mackey, 2012). We tried to test the same concept and anticipated results that could be compared with the original study, but we used a different set of stimuli and different participants.

The Pilot Study

At the original pilot study conducted by Gilbert et al. (2006), participants were asked to place a colour category boundary in a particular place in a row of blue and green colour hues. Only those individuals, who placed the boundary in particular spot between two specific colour hues were then selected to take part in the main study. There were only 11 participants in the original Gilbert's et al. (2006) study. The aim in presented study was to create a single unified colour scheme that could be used for all the participants, therefore much more participants could take part in the main study. A pilot study was conducted to achieve this goal. The pilot study is described next.

Participants and Method

Thirty-four psychology students participated in the pilot study (10 men; 24 women). The study consisted of an introduction and instructions in Czech, one test trial, 44 experimental trials, and demographic questions. Each participant was given information about the study in written form and signed a consent form beforehand. The test took approximately 10 min. The participants were tested individually in the presence of a research assistant. The study situation was the same as during the main study, describe below in more detail, including the calibration of the device etc.

The task was to categorize colour hues one by one into six colour categories. There were 44 colour hues chosen for the pilot study. The distance between each pair of neighbouring colour hues in the RGB colour spectrum was standardized. Each trial started with a fixation cross on a computer screen, then a square of a specific colour hue appeared together with six colour categories written (in words) under it. The categories were: brown, yellow, red, green, blue, and purple (in Czech language). After the participant had chosen one of the categories and clicked on the word representing the category with the left mouse button, another trial started. There was no time limit for this task altogether, neither was there a time limit for each trial. After all 44 trials, we thanked the participant for helping and the participant then left.

Results and Discussion

The aim of the pilot study was to choose colours for which the participants agreed on their language category in at least 94% of the cases. It was necessary to choose colour hues that have at least approximately the same distance between them in the CIELAB colour space. First, we transferred the RGB code of each colour to the CIELAB colour space (based on the procedure described in Gilbert et al., 2006). We calculated the colour distance in

Table 1 Characteristics of colours chosen during the pilot study for the main study

r	g	b	Colour hue	Agree.%	L*	a*	b*	Distance
102	149	164		100 **blue**	58.974	− 11.914	− 12.921	11.8992
118	141	170		94.1 **blue**	57.85	− 1.189	− 17.951	7.9244
136	136	166		97.1 **purple**	57.651	6.447	− 15.842	8.1414
152	132	156		100 **purple**	57.635	12.107	− 9.99	

CIELAB colour space using the CIE 1976 formula and an online calculator.[1] The colours picked are listed in Table 1, with their RGB code, CIELAB coordinates, and their distance in the CIELAB system. The colour categories used in this study are blue and purple; Gilbert's et al. (2006) original categories were blue and green. As was said, the decision to choose blue and purple colour categories was based on the necessity to choose colour hues that are at least approximately the same distance from each other in the colour space and also have at least 94% agreement among participants about their colour category. These two criteria did not apply to any blue and green colour hues, therefore the next best solution was chosen. The decision to choose the blue and purple colour hues and its potential effect on the results of the study is thoroughly discussed in the limitation section of the main discussion.

To examine the potential difference between chosen colours and so-called focal colours of Czech linguistic colour categories, an analysis has been conducted based on the focal colour hues found for purple and blue in Czech language by Uusküla (2008). Focal colour is a typical colour hue for certain (linguistic) colour category, for example a typical colour hue for the colour category "blue" in certain language. The CIE coordinates for the focal colour hues found by Uusküla (2008) can be found in Davies and Corbett (1994). Then the CIE coordinates for blue and purple focal colour hues have been transferred to RGB codes and the distance between colour hues used in our study and the focal colour hues have been calculated (online Delta-E calculator).

The Delta-E number varies from 0 to 100, the closer the number is to 100, the more two colour hues are different, if the distance is 100, the compared hues are exact opposite. Number less than 50 means that two compared colour hues are more similar than opposite.

[1] Delta–E Calculator. (n.d.). http://colourmine.org/delta-e-calculator.

There are two focal colour hues for purple and one for blue (Uusküla, 2008) in Czech language. The distance for purple colour hues chosen for our study and both purple focal colour hues are satisfying, varying from 40.9266 to 43.3306. Our chosen purple hues and the purple focal colour hues are more similar than opposite. The blue colour hues chosen for our study are more complicated and the analysis revealed that their distance from the blue focal colour hue (Uusküla, 2008) is 63.8259 for the blue hue in the first row of Table 1 and 53.9162 for the blue hue in the second row of Table 1. Interestingly, the first blue hue had 100% agreement among participants about it belonging to the blue category, but at the same time it has the biggest distance from the blue focal colour hue.

To compare our results with Gilbert´s, we calculated the distance between colour hues used in Gilbert's et al. (2006) study and the blue and green focal colour hues of English language from Uusküla (2008). The distance from the focal colour hues for green hues were 22.6928 (colour A, Gilbert et al., 2006; p. 490) and 29.2296 (colour B, Gilbert et al., 2006; p. 490) and for blue hues were 61.4433 (colour C, Gilbert et al., 2006; p. 490) and 46.3316 (colour D, Gilbert et al., 2006; p. 490). Seeing that Gilbert and colleagues (2006) also chose blue hue that is not close to the blue focal colour hue suggests, that this is not the reason, why our results differ from the original Gilbert's et al. (2006) study.

Even though this analysis has revealed, that the distances between chosen colour hues and focal colour hues are not negligible, especially for the blue category, the hues have been chosen based on very specific criteria. First, we wanted the agreement of the colour category among participants to be at least 94%. Secondly, the colour hues had to be on the linguistic colour category boundary between blue and purple, which itself means, that they would not be close to colour hues in the middle of the colour category, where the typical example of certain category (focal colour hue) is located. Also, we had to choose colour hues, that have approximately similar distance between each other (as described above), therefore we could not choose colour hues closer to the focal colour hues. To conclude, we argue, that the most important criterion for choosing the colour hues for our study was the agreement among participants about the colour category for each hue. We suggest a potential further research based on these findings.

Main Study

Participants

The participants were recruited by word of mouth from the author's social circles or were recruited by other faculty members at the Masaryk University, Brno, Czech Republic. There were 113 participants: 72 females (63.71%) and 41 males (36.28%). The average age of the participants was 22.99 years ($Sd = 4.48$; max = 46; min = 19). Approximately one-third of the participants (39; 34.51%) had a sight impairment that was corrected with glasses or contact lenses; participants were explicitly asked to use these corrections throughout the whole testing session. There were no colour-blind participants (colour blindness was self-reported by participants).

Fig. 1 Example of a trial. The correct response is 'left' (pressing the 'Q' key on the keyboard), as the target is on the lower left-hand side (the approximate position of the numeral 8 on a clockface). The method was identical to Gilbert et al. (2006), the colours used were different (Table 1) (Color figure online)

The number of participants excluded from the data analysis is based on their having more than 10% of the trials answered incorrectly (due to pressing a wrong key on the keyboard). Seven participants were excluded from the analysis (5 male participants; 71.4%), average age 22.1 years ($Sd = 1.45$; max = 25; min = 20). No outliers were excluded from the analysis. The final number of participants was $N = 106$.

Procedure

The testing took place at the Masaryk University, Brno, Czech Republic. The computer screen was calibrated by a professional using the colour calibration device Spyder 3 (Elite version) by Datacolor. The participant sat in front of the computer screen at a standard distance from the screen, there was standard artificial lighting with no natural light variations during the day. Participants read the information about the study and signed the consent form. The testing took approximately 20 min; participants were informed about the test length beforehand.

The method was based on the methodology used by Gilbert et al. (2006).

The task was to spot a target between eleven distractors. The target was a square, which had a different colour than the rest of the squares (the distractors). The squares were arranged into a circle with a fixation cross in the middle (Fig. 1). The participant was instructed to fixate his/her gaze on a fixation cross. The task was to respond as quickly as possible (and with maximum accuracy), on which side of the circle of squares the target is located. It was a left–right judgment task, where participants responded pressing a corresponding key on a keyboard ("Q" for left and "P" for right).

The test consisted of 144 trials. Each trial contained a target and eleven distractors, which means that in each trial, one colour was the target and one colour was the distractor. We have chosen 4 colour hues during the pilot study, that made 12 colour pairs. Also, we wanted the target to randomly appear on all 12 positions around the circle (hence $12 \times 12 = 144$ trials). The sequence of colour pairs and the position of the target within the circle were randomized for each participant. There was no time limit to complete the task, but the participants were instructed to be as fast as possible with maximum possible accuracy. After responding by pressing Q or P for a trial, the next trial appeared and so on.

Accuracy and reaction time were measured.

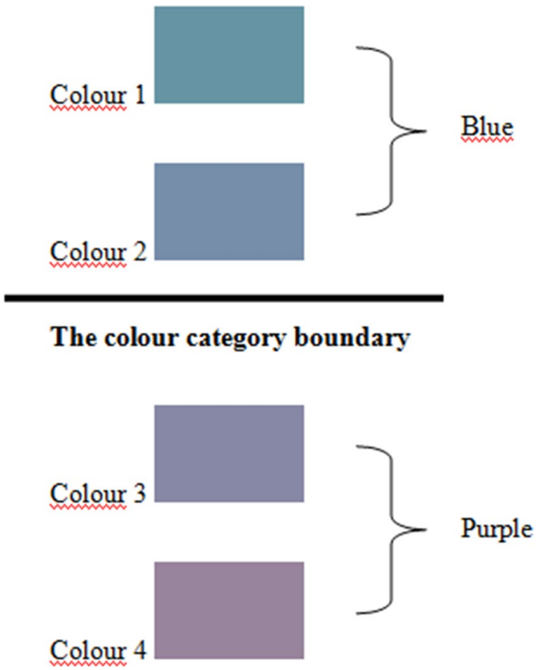


Fig. 2 Representation of colour categories (blue and purple) and the linguistic colour boundary (Color figure online)

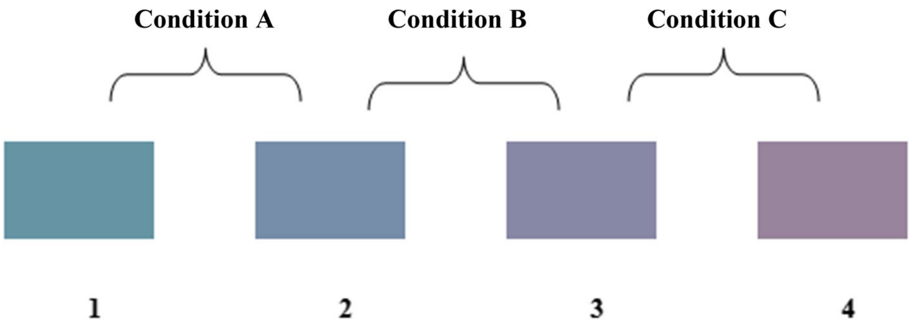


Fig. 3 The condition representing the compared colours. The linguistic boundary is between colours 2 and 3 (condition B) (Color figure online)

Research Question and Hypothesis

Our goal was to find out if the judgment would be faster for target–distractor pairs where the target colour and the distractor colour is from a different colour category (in case of Gilbert et al., 2006 green/blue boundary). Based on the findings from the study by Gilbert et al. (2006), we hypothesized that the reaction times in detecting the target's position in trials with a target and distractors from different colour categories would be shorter than the reaction times in trials with a target and distractors from the same colour category, but only in the right visual field.

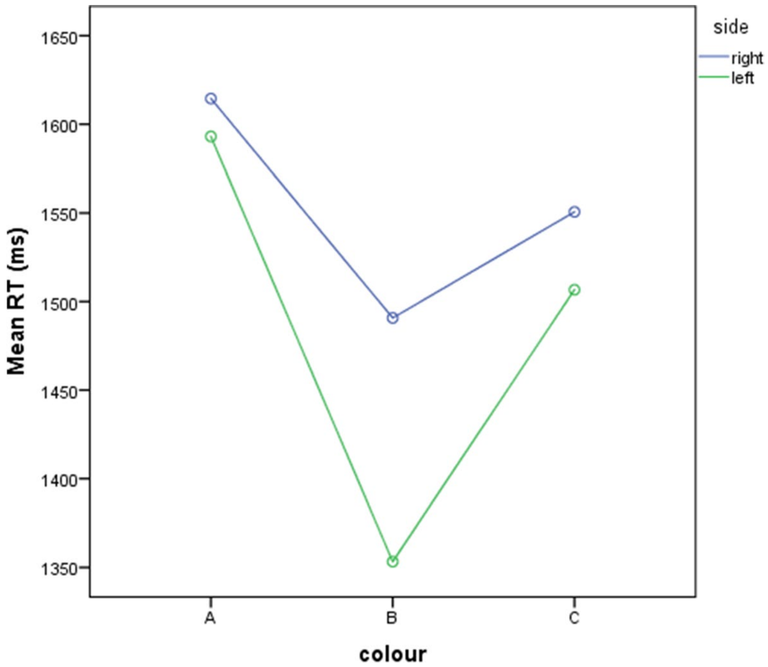
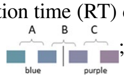


Fig. 4 The mean reaction time (RT) difference between three colour conditions (A, B, C) and positions of the target*. **Note*: colour: 

; side: position of the target (right/left visual field) (Color figure online)

Results

The main goal of the conceptual replication of Gilbert's et al. (2006) method was to determine whether the participants would find the target in trials containing colours from different colour categories faster than in trials containing colours from the same category. Figure 2 presents the colours used.

We also tested whether the hypothesis applies only to the right visual field. The conditions tested were as follows (Fig. 3):

First, we tried to compare conditions A, B, and C. The analysis performed was the factorial repeated-measures ANOVA (3 × 2), with factors of a colour condition (A, B, C) and of a target position (right/left).

Mauchly's sphericity test suggested a violation of the assumption of sphericity for colour, $\chi^2(2) = 42.108$, $p < 0.001$, and for the colour × side interaction $\chi^2(2) = 10.967$, $p < 0.05$. So, for reporting the main effect of colour and the interaction effect of colour × side, we used Greenhouse–Geisser correction.

The main effect of colour condition was significant at $p < 0.05$, $F(1.5, 157.5) = 5.239$. The main effect of side (left/right target position) was significant at $p < 0.05$, $F(1, 105) = 4.139$. This suggests that there is a significant difference in reaction times between the left and right positions of the target and that there is a difference between colour conditions. The contrasts for the colour conditions suggest that the mean reaction time in colour condition A is significantly higher than in colour condition B $F(1, 105) = 24.325$;

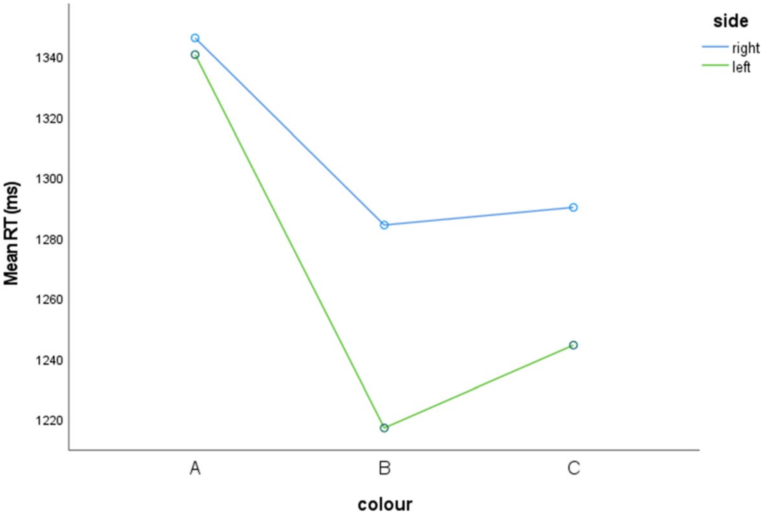
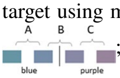


Fig. 5 The reaction time (RT) difference between three colour conditions (A, B, C) and positions of the target using median RTs for each participant in the factorial repeated-measures ANOVA *. **Note*: colour: 

; side: position of the target (right/left visual field) (Color figure online)

$p < 0.001$; the mean reaction time in conditions B and C do not differ significantly. Figure 4 represents the results for colour conditions and left–right target positions.

The contrasts of a significant main effect of position (left/right) $F(1, 105) = 4.139$, $p < 0.044$, suggest a difference between left and right positions of the target with significantly faster reactions if the target was on the left-hand side (in contrast to the hypothesis and to the results of Gilbert's et al., 2006 study).

The interaction effect between colour and side conditions was not significant, $F(1, 190.8) = 2.28$, $p > 0.05$.

The fact that in condition B the target was spotted faster if it was on the left side of the visual field (Wilcoxon signed-rank test, $Z = -4.094$, $p < 0.001$), which opposes the findings of Gilbert et al. (2006), is the most surprising finding; it raises many questions and calls for interpretation.

After following Drivonikou et al. (2007) and using median RTs for each participant instead of mean RTs and calculating factorial repeated-measures ANOVA (3×2), with factors of a colour condition (A, B, C) and of a target position (right/left), the results were very similar (Fig. 5), but only the main effect of side stayed significant $F(1,105) = 4.86$, $p < 0.05$ in favour of the trials with target located in the left side of the visual field. The contrasts revealed non-significant effect of colour and colour × side interaction ($p > 0.05$ for both). After using the Wilcoxon signed-rank test to analyse the difference between the trials with the left vs. right target location only for the condition B, the results showed significantly faster RTs for the trials with the target located in the left side of the visual field ($Z = -3.1$, $p = 0.002$). We would discuss these results further.

Discussion

The original study conducted by Gilbert et al. (2006) showed significantly faster reaction times in the task of spotting the target if the target and the distractor were from different colour categories (in their case, blue/green categories). This finding was significant only for the right visual field (the 'Whorf effect'). The authors explain this by noting that language brain regions in most individuals are located in the left hemisphere and claiming that language categorization affects visual perception.

This study is not the first to find such results (e.g. Gilbert et al., 2006, 2008; Regier & Kay, 2009; Xia et al., 2018).

Our findings did not support the results of the previous study. We did not find significantly faster reaction times in trials in which the target/distractor pair crossed the colour category boundary; we found that participants were significantly faster in the trials in which the target was located in the left visual field, not the right one. This applied for both kinds of trials—those in which the target/distractor crossed the colour category boundary and those in which the target/distractor pair was from the same colour category. The results we obtained are based on a much larger pool of participants than the results of studies with similar methodology, such as Gilbert et al. (2006) with 11 participants, Gilbert et al. (2008) with 12 participants, and Xia et al. (2018) with 30 participants. The advantage of a research study with fewer participants, in this case, lies mostly in the possibility to include only those participants, who placed the linguistic colour category boundary at the specifically defined place in a pilot study (Gilbert et al., 2006), which we did not have the chance to do. But this method of choosing participants for the main study can bring another question. The participants were exposed to the colour hues used in Gilbert's et al. (2006) main experiment beforehand, during the pilot study. Therefore, they could have been primed to the colours they then have seen during the main experiment. This priming could also explain why our mean reaction times were longer in general than those reported by Gilbert et al. (2006). Our way of conducting the pilot study did avoid the priming issue, because we did conduct a pilot study in which different participants than those included in the main experiment chose the colour categories. We then used the colour categories with the highest agreement among participants from the pilot study for all the participants in the main experiment. There is a chance that some individuals did not consider the colours we used as being from the same or different categories; this would affect their results. We argue that after choosing only the colours with more than 94% agreement about their category among participants in the pilot study, we did not anticipate that more than 6% of the participants in the main experiment would have different notions about the categories. The advantage of our process of choosing the colours is that we were able to have many more participants in the study, which increases the statistical strength of the results and gives us a better opportunity to generalize results compared to studies with fewer participants. The question is why we obtained results that are so different from the study we conceptually replicated. We would like to propose several possible explanations and interpretations of the differences in the results of the original study and our attempt.

Our results are more in line with Drivonikou et al. (2007) research, where the authors found an effect in the left visual field in trials with target from the same colour category as a distractor. This may suggest that the colour hues chosen for our experiment were in fact perceived by our participants as belonging to the same colour category (either all four of them as blue or as purple), even despite our effort to choose the colour from two different categories in our pilot study and the argumentation about the chosen colour hues presented above.

Also, Franklin et al. (2008) found an effect of the left visual field among children, who were only learning colour categories at the time of the experiment compared to those, who already were familiar with the colour categorization. Franklin et al. (2008) suggest, that there are developmental changes in the localization of categorical perception of colours, that with no linguistic categories present (or before the linguistic categories form), the right hemisphere play a prim in the colour discrimination and categorization. As was described, our participants taking part in the pilot study were not the same as those taking part in the main experiment. The participants taking part in a main experiment were not exposed to the colour categories during the pilot study. Therefore, their colour categorization processes were not primed by taking part in the pilot study, the linguistic colour categorization was not triggered. Maybe our main experiment´s participants can be (metaphorically) compared to the children in Franklin's et al. (2008) study, who only learn the linguistic colour categorization. They might have been figuring out the differences and categories of used colours during the experiment, therefore their linguistic colour categorization was not activated. But this may not be the only explanation of the different results that we obtained compared to Gilbert's et al. (2006) original study.

The explanation of the results suggesting an advantage of stimuli located in the left visual field may lie in the studies examining attention (e g. Chica et al., 2012). Attention is a complex cognitive function. The task that the participants were going through may trigger processes of a 'sustained attention', the ability to willingly control focused attention over a period of time (Chica et al., 2012; Sturm & Willmes, 2001), together with spatial attention (Shulman et al., 2010) focusing on the position of the stimuli. And as was said, many studies (Bartolomeo & Chokron, 2002; Bartolomeo, 2006; Chica et al., 2012) located centres of spatial selective attention in the right hemisphere, which would mean that when the attention networks are triggered, the target located in the left visual field would be recognized faster. Based on these findings and interpretations, the Whorf effect might not manifest when the participants are not primed with the linguistic categorization of colours before the experiment (during the pilot study) and the effect of right hemispheric attention networks location may interfere, which may cause a faster recognition of the target in the left visual field regardless of the colour categories of the target and the distractor.

Limitations and Further Research

The main concern that needs to be addressed is the difference between Gilbert's et al. (2006) methodology and the methodology used in presented study. Firstly, the pilot study methodology was different in a way that allowed us to include more participants in the main study. But this way, we cannot be sure, that all the participants would place the linguistic colour category boundary to the specific spot, that we anticipated it to be placed while Gilbert's et al. (2006) participants were only included in the main study, if they placed the boundary to the specific defined place between blue and green colour hue. Gilbert's et al. (2006) method of the pilot study could on the other hand prime the participants to think about and recognize the colours during the main experiment. We avoided this issue by including different participants in the pilot study and different ones in the main study. This may explain, why we report longer reaction times in the results than Gilbert et al. (2006), but also it may rise a question of the extent to what this process effects the results. It would be interesting to conduct a follow up study where the research question would ask whether the categorization of colours beforehand affects results of the colour discrimination in the main experiment.

Another important difference that could have affected the results is using different colours than the original Gilbert's et al. (2006) study. While Gilbert et al. (2006) used blue/green categories, we chose blue/purple categories based on the pilot study results. This may be the main reason that our results are different from Gilbert's et al. (2006). Among eleven universal categories that Berlin and Kay (1991) defined, blue and green are linguistically more represented in various languages, while purple does not occur as often. This reasoning should not play a significant part in our research, as the linguistic relativity hypothesis advocates the idea of non-universality of colour categories, suggesting that the effect should appear in case of any two distinctive colour categories. For example, Drivonikou et al. (2007) and Daoutis et al., (2006a, 2006b) used purple category in their experiments and found results comparable to Gilbert's et al. (2006), therefore there is no reason to anticipate, that using purple category would prevent finding results comparable to Gilbert's et al. (2006).

Another important fact that should be considered is the ecological validity of experiment of this kind. While in lab, participants usually see colour chips arranged on a computer screen, in the outside world we rarely see colour on their own, without any context. In everyday life we evaluate the colours we see based on the knowledge of the colours we are supposed to see (banana is typically yellow) and based on the colours surrounding the object we are evaluating (Hansen et al., 2006), so not only the perception processes, attention or linguistic categorization, but also top-down processes of past knowledge and memory play a role in colour evaluation and discrimination. The question is, to what extend are these processes involved in experiments in laboratory setting, that do not work with the contextual information and colour knowledge and memory.

Last but not least, the typicality of a colour hue (to what extent the blue colour hue is a typical representation of "blue" colour category?) may play a role in discriminating it among other colour hues. A follow up study may replicate the Uusküla (2008) effort to find focal colour hues and potentially confirm the focal colour hues for Czech language, or find new ones and use those in the Gilbert´s method.

Conclusion

Our conceptual replication of the study by Gilbert et al. (2006) did not provide the same results as the original study and therefore does not support the Whorf effect for the right visual field. It is difficult to say whether the difference in results was due to the differences in methodology or due to the differences in cultures and languages that were tested. We think that our altered methodology, which enabled a much larger participant pool, outweighed the limitations. This study was aimed to show how complicated the questions of language development in different cultures might be and mainly to indicate the complex attempts to test it.

Acknowledgements This research was supported by the research infrastructure HUME Lab Experimental Humanities Laboratory, Faculty of Arts, Masaryk University. We would like to thank our research assistants, the reviewers and the editor for their time.

Declarations

Conflict of interest The authors report no conflict of interest.

Ethical Approval This research was approved by the Ethics Committee for Research of Masaryk University.

References

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist, 73*(1), 3–25. https://doi.org/10.1037/amp0000191

Bartolomeo, P. (2006). A parietofrontal network for spatial awareness in the right hemisphere of the human brain. *Archives of Neurology, 63*(9), 1238–1241. https://doi.org/10.1001/archneur.63.9.1238

Bartolomeo, P., & Chokron, S. (2002). Orienting of attention in left unilateral neglect. *Neuroscience & Biobehavioral Reviews, 26*(2), 217–234. https://doi.org/10.1016/S0149-7634(01)00065-3

Berlin, B., & Kay, P. (1991). *Basic color terms: Their universality and evolution*. Univ. of California Press.

Casasanto, D. (2008). Who's afraid of the big bad Whorf? Crosslinguistic differences in temporal language and thought. *Language Learning, 58*, 63–79. https://doi.org/10.1111/j.1467-9922.2008.00462.x

Chica, A. B., de Schotten, M. T., Toba, M., Malhotra, P., Lupiáñez, J., & Bartolomeo, P. (2012). Attention networks and their interactions after right-hemisphere damage. *Cortex, 48*(6), 654–663. https://doi.org/10.1016/j.cortex.2011.01.009

Daoutis, C. A., Franklin, A., Riddett, A., Clifford, A., & Davies, I. R. (2006a). Categorical effects in children's colour search: A cross-linguistic comparison. *British Journal of Developmental Psychology, 24*(2), 373–400.

Daoutis, C. A., Pilling, M., & Davies, I. R. (2006b). Categorical effects in visual search for colour. *Visual Cognition, 14*(2), 217–240.

Davies, I. R., & Corbett, G. G. (1994). The basic colour terms of Russian. *Linguistics, 32*(1), 65–90.

Delta–E Calculator. (n.d.). ColorMine. http://colormine.org/delta-e-calculator.

Drivonikou, G. V., Kay, P., Regier, T., Ivry, R. B., Gilbert, A. L., Franklin, A., & Davies, I. R. L. (2007). Further evidence that Whorfian effects are stronger in the right visual field than the left. *PNAS, 104*(3), 1097–1102. https://doi.org/10.1073/pnas.0610132104

Franklin, A., Drivonikou, G. V., Clifford, A., Kay, P., Regier, T., & Davies, I. R. L. (2008). Lateralization of categorical perception of colour changes with colour term acquisition. *PNAS, 105*(47), 18221–18225. https://doi.org/10.1073/pnas.0809952105

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *PNAS, 103*(2), 489–494. https://doi.org/10.1073/pnas.0509868103

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2008). Support for lateralization of the Whorf effect beyond the realm of colour discrimination. *Brain and Language, 105*(2), 91–98. https://doi.org/10.1016/j.bandl.2007.06.001

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science, 306*(5695), 496–499. https://doi.org/10.1126/science.1094492

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience, 9*(11), 1367–1368.

Kay, P., & Kempton, W. (1984). What is the Sapir–Whorf Hypothesis? *American Anthropologist, 86*(1), 65–79. https://doi.org/10.1525/aa.1984.86.1.02a00050

Kay, P., & McDaniel, C. K. (1978). The linguistic significance of the meanings of basic colour terms. *Language, 54*(3), 610–646. https://doi.org/10.2307/412789

Lindsey, D. T., & Brown, A. M. (2002). Colour naming and the phototoxic effects of sunlight on the eye. *Psychological Science, 13*(6), 506–512. https://doi.org/10.1111/1467-9280.00489

Lowry, M., & Bryant, J. (2019). Blue is in the eye of the beholder: A cross-linguistic study on color perception and memory. *Journal of Psycholinguistic Research, 48*, 163–179. https://doi.org/10.1007/s10936-018-9597-0

Mackey, A. (2012). Why (or why not), when and how to replicate research. In G. K. Porte (Ed.). *Replication research in applied linguistics* (1st ed., pp. 34–69). Cambridge University Press.

Regier, T., & Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends in Cognitive Sciences, 30*(10), 1–8. https://doi.org/10.1016/j.tics.2009.07.001

Roberson, D. & Hanley, J. R. (2010). Relatively speaking: An account of the relationship between language and thought in the colour domain. In B. Malt & P. Wolff (Eds.). *Words and the Mind: How words capture human experience* (1st ed., pp. 183–198). Oxford University Press.
Shulman, G. L., Pope, D. L. W., Astafiev, S. V., McAvoy, M. P., Snyder, A. Z., & Corbetta, M. (2010). Right hemisphere dominance during spatial selective attention and target detection occurs outside the dorsal frontoparietal network. *Journal of Neuroscience, 30*(10), 3640–3651. https://doi.org/10.1523/JNEUROSCI.4085-09.2010
Sturm, W., & Willmes, K. (2001). On the functional neuroanatomy of intrinsic and phasic alertness. *NeuroImage, 14*, 76–84. https://doi.org/10.1006/nimg.2001.0839
Uusküla, M. (2008). *Basic colour terms in Finno-Ugric and Slavonic languages: Myths and facts*. Doctoral dissertation.
Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on colour discrimination. *PNAS, 104*(19), 7780–7785. https://doi.org/10.1073/pnas.0701644104
Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *Wires Cognitive Science, 2*(3), 253–265. https://doi.org/10.1002/wcs.104
Xia, T., Xu, G., & Mo, L. (2018). Bi-lateralized Whorfian effect in colour perception: Evidence from Chinese sign language. *Journal of Neurolinguistics, 49*, 189–201. https://doi.org/10.1016/j.jneuroling.2018.07.004
Zhou, K., Mo, L., Kay, P., Kwok, V. P., Ip, T. N., & Tan, L. H. (2010). Newly trained lexical categories produce lateralized categorical perception of color. *Proceedings of the National Academy of Sciences, 107*(22), 9974–9978.