Training New Colour Boundaries through Linguistic and Perceptual Methods

D. E. Jensen and Anja Tørnqvist

Cognition and Communication, Cognitive science, Aarhus University 2016.

1.0 Abstract (A.T)

Many scientists have discussed mental representation, where representations are often divided into two extreme poles. Supporters of the Sapir-Whorf hypothesis would say that linguistics form one's perception, while multi-modalists argue that cognition has no need for linguistics, and thoughts are a product of perceptual representation converted from one's sensorimotor experience (J. R. Anderson, 2015). Former studies have shown how linguistics can form colour categories depending on one's native language (Winawer et al., 2007). Is it possible to acquire the same results as a non-native merely by exercise, or is it a simple question of biology? This experiment attempted to find an answer by training non-native Russians the two Russian colour terms for blue, siniy and goluboy, in hopes of teaching participants to discriminate between new category boundaries. Results showed that part of the original hypothesis was directly contradicted. It was proven that training of both linguistic and perceptual colour discrimination equally influenced participants' post-training trials. Further, it was discovered that the training sessions had a dramatic influence on within-category discrimination.

2.0 Introduction (A.T.)

Cognitive processes are often perceived as strongly connected to spoken language. In fact, early behaviorists argued that thinking was merely sub-vocal speech (Harley, 2010). It can be challenging to imagine only thinking in perceptual terms; yet a significant amount of thoughts operate in exclusively pictures and emotions. Abstract thinking plays a major role in human cognition. One can, for instance, taste the sweet strawberries when longing for summer in the darkness of winter, or even see images of one's beloved when being far away from home; but a person can also converse with themselves and hear their own voice in their minds when thinking. This is the coexisting relationship between perceptual and linguistic cognition. However, it is still to be determined how they individually function and affect cognition, and if it is possible to see their effects.

The meaning of thought is often vague and ungrounded. There are many definitions to find and one could easily get lost in the many classifications. For instance, on the website *Dictionary.com*, there are 12 different definitions of the word "thought". A few of them are as follows: "*The product of mental activity; that which one thinks"*, *"A consideration or reflection"* and *"The intellectual activity or the ideas, opinions, etc., characteristics of a particular place, class, time"* (Retrieved December 13, 2016, from http://www.dictionary.com/browse/thought). As for Harley (2010), his definition of thought is: "*the manipulation of ideas with an outcome that can enter consciousness"*. Even this definition of thought does not give a hint on how cognition is related to language and perception. Early behaviorists would argue that language and thought are the same thing, but were eventually faced with a problem. They could not deny their own thought processes. Jacobsen (1932) (as cited by Harley, 2010) provided evidence for the common behaviorist argument. He detected electrical activity in participants' throats when they were asked to think, which led him to conclude that thinking must be speech "stopped in time". Smith et al (1947), (as cited by Harley, 2010) proved Jacobsen wrong. They used tubocurarine to paralyze the muscles in the body, logically thinking if the throat muscles were paralyzed then cognition could not administer if the two were dependent. The experiment showed that the subject could in fact think in words and even solve mathematical problems while paralyzed (Smith (1947) as cited by Harley, 2010). Although Smith's experiment showed that the body does not necessarily need to be activated for the cognitive processes, there are still scientists who argue that there is no definitive separation of the mind and body (Johnson, M., & Lakoff, G., 2002).

Cognition also involves categorization. Humans categorize the world in objects that are separate from other categories (Harley, 2010). Words are symbols for the concept they represent, therefore some argue language facilitates cognitive processes and its development (Harley, 2010). This is the ongoing discussion of the nature of cognitive representations.

Some would even argue that the form of our language has an influence on the way one individually perceives the world. This is known as the Sapir-Whorf hypothesis. Whorf studied Native Americans and pondered on how their language could be so different from his own, English. Their language was not only different in syntax, but also in how the perceived world was structured (Whorf (1956) as cited in Harley, 2010). *"How utterly unlike our way of thinking!"* - was one of his phrases, which has become famous (Whorf, 1956). Whorf claimed that thoughts are constrained by the form of the individual's language, and the differences between languages will therefore lead to differences in ways of thinking. However, the Whorfian hypothesis has been heavily criticized. The primary criticism reads that although subjects speak different languages, it does not necessarily mean they think differently (Harley, 2010). To confirm such an extreme hypothesis, one must have a way to measure thought. Whorf claimed that the contrasting languages lead to contrasting categorizations because languages structure the world differently. Once again, this was criticized since it is hard to tell how subjects perceive the world differently (Harley, 2010). It may be due to different cultures being more observant to specific details.

Gordon (2004) studied the Piraha tribe in Brazil. The tribe would struggle to remember specific numbers, since their language consisted of only a word for one, two and many. The tribe simply did not need words for larger numbers - and since they had no words, they would not practice counting higher than two. This shows that culture has a big influence on mathematical ability, presumably representations too, and that language for exact numbers is most likely a cultural invention (Gordon, 2004; Frank et al., 2008).

Although there is a lot of criticism to find on the Whorfian hypothesis, a paper by Regier and Kay (2009) reviews the recent findings on the naming and perception of colours and concludes that Whorf was “half right”. The two ways were that language influences perception but primarily in the right half of the visual field, and that colour naming across languages is shaped by universal and language-specific impact (Regier & Kay, 2009).

However, it is self-explanatory that cognition does not exclusively involve speech and language. Cognition also uses visual imagery, involved in multimodal processes. Multimodal representation describes images without translation into a syntactical representation. People in favour of amodal representation argue that cognition is propositional; that we think in linguistically structured information. Different theories such as *the language of thought, the theory of perceptual symbol systems* and *the neo linguistic relativism* have risen throughout the years, yet still there is no absolute answer to the question of how information is structured and stored in the mind (Barsalou et al., 2003; J. R. Anderson, 2015).

A professor of psychology and a specialist in cognitive psychology, Eleanor Rosch, has done research on categories using colours as a method to study cognitive processes. She believed categories were not all bound entities whose relationship was defined by an item’s features (Rosch, 1973). She has expressed that the domain of colours works as an adequate tool since colours are visual properties and because some colours are perceived as better examples of a specific category than others. In her research on natural categories, she hypothesized that the domain of colour is structured in nonarbitrary semantic categories, and that these categories would develop around perceptually salient *natural prototypes (*Rosch, 1973). She taught 162 subjects from a Stone Age culture, which had no colour concepts other than (roughly) bright and dark the names of the different colour hues chosen. She found that sets in which the natural prototype was central were learned with significantly fewer errors, alongside their names (Rosch, 1973).

Although Rosch's research speaks a clear message, there are still skeptics that say language is essential for categories. However, research by Ozturk et al. (2012) included studies of pre-linguistic children's response to categorical effects of colour perception. They found through their research that it could very well be that the categories in the study were results of perceptual learning.

Ozturk et al. (2012) might have had a point, but the research done by Thierry et al. (2009) has striking results. Studying the brain potentials in Greek and English native speakers when presented with blue and green colours, they tested the effects of colour terminology in the two languages on early stages of visual perception. They tested the degree to which pre-attentive and unconscious aspects of perception was affected by an individual's native tongue, by recording electrical activity in the brain using EEG. Since Greek distinguishes blue by light blue, *ghalazio*, and a dark blue, *ble*, the researchers hypothesized that the existence of two basic colours in Greek would lead Greek native subjects to perceive luminance deviation as more diverse in the blue than in the green blocks. Significance was found in the interactions between participant, group, colour and deviancy. They confirmed that the interactions in fact were generated by different vMMN response patterns in Greek and English native speakers. Although not significant, they also found numerically greater deviance for green than blue for English participants but significantly greater deviance for blue than green in Greek participants. Therefore, the results showed a greater distinction between different shades of blue than green for

Greek participants. No such findings were obtained with English participants. Thierry et al. (2009) therefore concluded that language specific distinctions does indeed affect early visual processing. They were proud to demonstrate the conclusion above for the first time going beyond the observation of high-level categorization, by using a form of measurement.

Winawer et al. (2007) also set out to investigate how categorization could lead to differences in colour discrimination. They tested Russian and English natives and measured colour discrimination performance in the respective language groups. Like Greek, Russian also discriminates blue with two words, *siniy* for dark blues and *goluboy* for light blues. Winawer et al. (2007) criticized some of the former studies for their way of conducting the experiment. It was argued that testing cross-linguistic differences in individual colour similarity judgements could be confused with colour memory. If so, there would be a high risk of subjects making perceptual discriminations where linguistics had no influence at all. Therefore, the question of whether a participant's normal ability to distinguish colours in a subjective procedure that is reshaped by language, would be left unanswered (Winawer et al., 2007). Since the stimuli in their own experiment was constantly present until subjects answered, it was argued that there was a minimal effect on memory. Their results revealed that Russian natives had a category advantage. English natives showed no such advantage. Furthermore, the effects of language were most obvious when difficult within-category discriminations were made. Across-category distinctions were easily discriminated by both language groups, but Russians had a clear advantage when it came to difficult within-category discriminations. There was no significance found for reaction time, nor accuracy. Significance was found in a partial correlation between language groups and a fused measure of the reaction time effect (Winawer et al., 2007). The study implies that linguistic categories have an effect on performance of simple perceptual colour judgements. Further, it hints that colour discriminations vary across language groups as a consequence of what distinctions are customarily made in a particular language. However, this study does not demonstrate that English natives cannot distinguish between the light and dark blue colour, rather it suggests that Russians cannot avoid distinguishing them. With this study Winawer et al. (2007) proposed that language specific categorical representations in a specific language are habitually activated in perceptual decisions.

There are many studies on how language affects perception with colours, but can it be studied in a different way? Casasanto et al. (2013) studied how mental representations differed in speakers of different languages. Musical pitch is described spatially diverse in different languages. The researchers asked if people who use different metaphors for pitch also mentally represent pitch differently. If so, then does language shape the non-linguistic representations used when perceiving or producing pitches even when not using language (Casasanto et al., 2013)? The researchers argued that if mental representations of pitch were similar and irrespective of the spoken language of participants, then the performance should not differ between Dutch and Farsi participants (Casasanto et al., 2013). Results for cross-domain effects showed for Dutch speakers that spatial height of the stimuli did affect the pitch estimates. Further, it showed the thickness of stimuli had no significant effect. The exact opposite results occurred for Farsi speakers, indicating that the two language groups in fact do have diverse mental representation of pitch. The results for within-domain effects indicated that the observed cross-dimensional interference effects could not be explained by unexpected variances in within-domain performance. Although the results prove the hypothesis, it does not determine whether language plays a role in the diversity in the results. Further, Casasanto et al. (2013) conducted a two-part training study with Dutch participants. They argued that if the training was successful, it would result in the same findings as Farsi speakers the previous experiment. The results suggested that the effect of thickness on pitch reproduction in thickness-training participants was statistically indistinguishable from native

Farsi's results. The results from the training experiment indicated that from experience using linguistics, one *can* change non-linguistic mental representations of musical pitch. This demonstrates that using the habitual space-pitch metaphors may shape one's mental representation of pitch via learning mechanisms similar to those that changed participant's representation in their experiment. Casasanto et al. (2013) suggested with their research that differences in space-pitch mappings are encoded in the language people speak. These expressions are so highly conventional that one no longer notices that one is in fact using spatial metaphors.

This study aims to answer some of the same questions as the research above. Does language shape the way people perceive the world? Alternatively, does the world we perceive shape our language? Looking at experiments like Winawer et al. (2007), one cannot help but wonder if Russian natives are more attuned to making distinctions for the colour blue, and that the results of their study is a mere question of expertise rather than diverse categorical perceptions. Is it possible to teach a person a new colour category? If so, will one be more accurate to distinguish a colour, if one learns the respective colour names, hence develop a new perceptual category? The theories and concepts listed above eventually led this study to the hypothesis: *Linguistics have an influence upon categorical perception*. Similar to Casasanto et al. (2013), participants will go through a training session, either a linguistic-training condition or a perceptual-training condition. Furthermore, it is expected that; *a) linguistic-training participants will have a shorter reaction time for both within and across-category distinctions, and not accuracy, and will improve from the first trial,* and *b) perceptual-training participants will have a higher accuracy of distinguishing within and across-category distinctions, and not reaction time, and will improve from the first trial.* A shorter reaction time, but not necessarily a higher accuracy is expected from the linguistic-training, since it is anticipated when participants have learned a new category, they will be faster at distinguishing the colour categories apart. A higher accuracy is expected from the perceptual-training, since participants will have more experience distinguishing the colours, but not necessarily faster since they do not distinguish the colours by category. Like former studies, this experiment will use the Russian colour categories for blue, siniy and goluboy.

3.0 Materials and Methods (A.T)

3.1 Participants

Most participants in this study were students at Aarhus University, although a few were exceptions. Originally, 27 participants contributed to the study, but 7 of the 27 participants' data had to be removed from the final analysis, since it was discovered that the coding was not optimal. The gender distribution of the final analysis had 11 females, 8 males, and 1 person who identified with "other" as their gender. The overall mean age was 22.95, ranging from 19 to 46 years. Participants had different native languages, although none were Russian natives, nor had any connection to the Russian language. None of the participant's native languages distinguish blue in two colour terms. All participants read and agreed to a consent form.

3.2 Materials and Stimuli

The experiment was conducted on a computer. It was designed and coded in the computer language Python by the researchers. Winawer et al. (2007) inspired the setup. Similarly, the experiment had a triad of coloured squares that were used, where the bottom square had to be matched with one of the above squares (see appendix). This was a trial in the main experiment. The colours consisted of 10 different hues from the blue colour spectrum. The blue colours were divided in two categories according to the Russian division between siniy and goluboy, which was later in the analysis used as category partition. The standard experiment looped randomly through the colour spectrum twice. There were two conditions, a linguistic training (condition 1) and perceptual training (condition 2). Both training trials lasted 5 minutes. The experimental trials ran twice, once before and once after the training session. In the linguistic training, participants had 10 coloured squares in the top of the screen and 1 square at the bottom. The 10 squares were split in the two categories, with the Russian names beneath them (siniy for the light colours and goluboy for the dark) (see appendix). The bottom square then had to be matched to one of the two categories. The top colour squares and names were visible throughout the whole training session. On the other hand, the perceptual training consisted only of one big square in the center of the screen. Pressing either the left or right arrow key to identify if the colour had changed to a lighter or darker colour than the one before (see appendix). For every 10th trial in both standard and training trials, participants were given the opportunity to rest their eyes as long as necessary. To resume the experiment, one had to press space. The variables accuracy and reaction time were measured only in the main experiments (pre-training and post training). The training trials were not fundamental to the hypothesis. The variables were analyzed within the R environment.

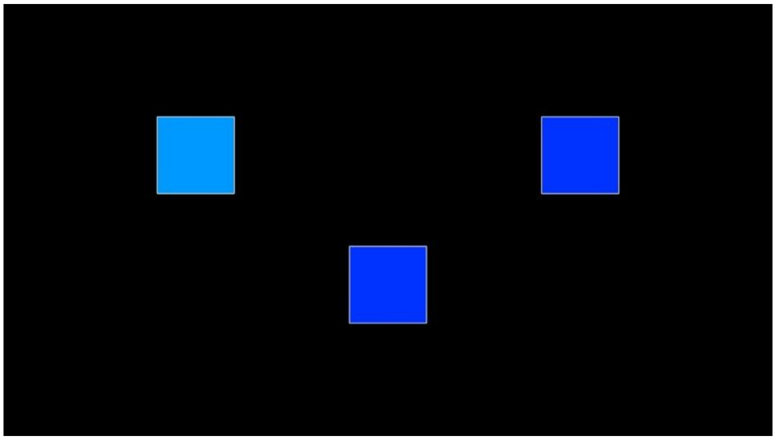

Figure 1: A standard session the main experiment

3.3 Procedure

After reading and agreeing to the consent form, participants were placed in front of a computer that ran the experiment. Participants were given a random ID and one of the conditions by the researchers. Instructions were shown on the screen before the experiment started. When the first trial began, the triad of blue squares appeared, with the bottom square matching one of the two squares above. Participants were instructed to press either the left or right arrow key according to which of the two squares matched the bottom square. After every 10th trial, participants were given a break. If participants were in condition 1, the linguistic training would begin. Participants then had to press either left or right arrow keys to match the bottom square to one of the categories above the square. After pressing, the bottom square would change colour, and participants then had to rematch the new colour. After every 10th trial participants were given a break. This went on for 5 minutes. If participants were given condition 2, the perceptual training would begin. A big coloured square would appear in the center of the screen. After the first click, the square would change colour. If the present colour was lighter than the previous, participants had to press left. If the present colour was darker than the previous, participants had to press right. Once participants had clicked, the square changed colour, and participants again had to judge again. This went on for 5 minutes. After every 10th trial, participants were given a break. When the training session was over, they had to do the standard experiment again. The first and last main experiments were identical to each other. When participants were finished, one of the researchers were called to close the test. The experiment ran on the program Spyder or Psychopy, according to which computer was used. When the experiment was over, the data was imported into R for further analysis.

4.0 Analysis (D.E.J.)

The data collected from the experiment was the participants' information (i.e. age, gender, etc.), along with their reaction time to respond to the stimuli in the pre-training phase and post training phase (recorded in milliseconds). Their score of accuracy in both phases (number of trials they identified colours correctly) and whether the stimulus presented was a within or an across category colour for both phases was also recorded. This was recorded for both conditions, although the participants only ran either one condition or the other (a between-subject experiment). The performance of the participant during training was not recorded, as it was not relevant to contribute to the hypothesis.

Special care was taken to avoid a type I error in the repeated measures experiment. Because participants went through multiple trials, the assumption of independence was violated. To fix this, a linear mixed model fit by maximum likelihood t-tests was used along with a Laplace approximation in a glmer (general linear mixed effects regression inverse gamma) model. An ANOVA (Analysis of Variance test) was also done to compare the model to a null hypothesis to produce the associated Chi Squared statistic. This model accounted for the violation of independence and the discovered violation of normality, and thus lowered the probability of a type I error. The data was analysed within the R environment.

The first few participants who ran the experiment had a slightly different experience than the rest of the subjects. Shortly after the first couple of participants began, it was discovered that the code was not optimal for what was needed to achieve accurate results. The time limit for the first condition (linguistic training) and the second condition (perception training) were not equal. The training would last fifteen minutes for condition one, whereas condition two would only last thirty seconds. After this was discovered, the experiment was improved so both conditions ran for five minutes. Because of this, the data for the first participants with the faulty experiment was removed so the rest of the reliable data could be analysed separately. This was done to ensure that the significance of an effect would not be missed or misinterpreted. This will be examined further in the discussion section.

In the study, the focal point of interest was the comparison of the subjects' performance pretraining and post training, and how the reaction time and accuracy varied from condition one to condition two. The orientation of the colour (within or across-category) was also considered. Therefore, the analysis was carried out to compare a) reaction time pre-training and reaction time post training for conditions one and two, b) compare pre-training accuracy to post-training accuracy for conditions one and two, and c) the relationship colour category plays according to the two training conditions and the two experiment phases.

5.0 Results (D.E.J.)

The results indicated that there was a significant systematic affect pertaining to reaction time, ($X^2(2) = 4894.7$, $p(>X^2 = <.0001)$. Reaction time was longer on average in condition one (M=1116.6ms, STD=955.33ms) than condition two (M=900.85ms, STD=581.15ms). This contradicted the first part of the hypothesis: $\beta = 0.35$ (SE= 0.28), $t(21)=1.25$, $p=.21$, although the fixed effect of conditional state was not found to be a significant factor of variance. Additionally, the contrast between the pre-training experiment vs. the post-training experiment was significant. The pre-training reaction time was longer (M=1119.56s, STD=963.6ms) than the post-training reaction time (M=929.12ms, STD=639.35ms). This coincides with the hypothesis: $\beta = 0.43$ (SE=0.04), $t(21)=12.07$, $p=<.0001$). Also, reaction time was highly influenced by whether the colour was across categories or within the same category. Reaction time was longer when the stimulus was a within-category colour (M=1235.61ms, STD=1001.31ms) verses an across-colour category (M=854.96ms, STD=594.43ms).

This was an interesting new observation that was not predicted in the hypothesis: β =-0.88 (SE=0.04), *t*(21)=-24.04, *p*=<.0001. Moreover, a significant systematic effect also influenced accuracy, ($X^2(2)=331.74$, p=<.0001). Accuracy decreased from condition one (M=92%, STD=28%) to condition two (M=87%, STD=33%) (linguistic training vs. perceptual training), although conditional state was not shown to be a significant factor of variance. This was in contradiction to the hypothesis: β =-0.50 (SE=0.26), *z*(21)= -1.94, *p*=.05. The result from the pre-training experiment (M=89%, STD=31%) to the post-training experiment (M=90%, STD=29%) showed that accuracy increased in the latter, however only in the slightest. This aligned with the hypothesis: β =0.17 (SE=0.11), *z*(21)=1.50, *p*=<.14, but was found to be insignificant. Finally, accuracy depended highly on whether the colour was across categories or within the same category. Accuracy decreased when the colour was within the same category (M=80%, STD=40%) compared to across categories (M=97%, STD=16%). Although not a part of the original hypothesis, it was an interesting discovery: β =-2.28 (SE=0.15), *z*(21)=-15.11, *p*=<.0001.These findings and their implications will be explored in the following section.

Table 1 Average Reaction Time

Reaction Time (milliseconds)	Condition One	Condition Two	Pre-training	Post-training
Pre-training	1020.50	943.10	n.a	n.a
Post-training	888.81	837.32	n.a	n.a
Across-category	790.59	749.54	821.67	716.30
Within-category	1159.72	1066.04	1177.38	1043.46

Table 2 Average Accuracy

Accuracy (%)	Condition One	Condition Two	Pre-training	Post-training
Pre-training	89	87	n.a	n.a
Post-training	92	88	n.a	n.a
Across-category	98	96	97	97
Within-category	81	76	77	81

6.1 Interpretation

In relation to the analysis, the results were shocking, as they clearly contradicted the original hypothesis. One of the surprising findings was the ultimate insignificant role that conditional state played on reaction time and accuracy. Although there was a difference in the reaction time and accuracy for the two conditions, the effect was minimal. There was a shortening on average of reaction time for the perceptual condition compared to the linguistic condition, and a lower average accuracy for the perceptual training compared to the linguistic training. Although it is humorous enough that these outputs conflict with the predicted hypothesis, it should be considered that the conditions were not found to be a significant agent of variance (p=0.21 for reaction time and p=0.14 for accuracy). The original thought process that drove the hypothesis was that the perceptual training would have a more finely tuned discrimination between two colours, and therefore participants would learn to discriminate between colours more easily. This would lead to a higher rate of correct answers, not necessarily having an influence on reaction time. However, the results would suggest that the finely tuned perceptual training provided actually taught the participant to discriminate between colours *faster*, without necessarily acquiring a new perceptual expertise. Moreover, the insignificance of the conditions was an outcome of either two things: first, it could be that linguistics plainly did not share a role in perception, or secondly, that the conditions were not idiosyncratic enough to show distinguishable differences in effects. Although the probability value for the conditions suggested that the alternative hypothesis must be rejected, an improved experiment could give deeper insight into whether there really was no effect, or if the insignificance was due to an ineffective training method. Because of previous research on this topic (overviewed in the introduction), it is believed that there are reasonable grounds to suspect that it was an ineffective training method that was the root of the insignificance. Therefore, it is suggested that improvements to the linguistic training, particularly to make the training exclusively linguistic without any perceptual training, would result in a different effect on reaction time and accuracy than the perceptual training. This idea will be explored later in the 6.2 Improvement section.

Another interesting implication the data brought to light was that the pre-training and post-training phase was only significant for reaction time, and not for accuracy (p=>.0001 for reaction time, and p=.14 for accuracy). Accuracy for the linguistic condition only saw a 3% increase, while the perceptual condition stayed level at 87% accuracy from the pre-training experiment to the post-training experiment. Expressed differently, reaction time lowered in the post-training phase compared to the pre-training experiment, whereas accuracy improved only minimally for the post-training phase; even less impact than the conditions had on accuracy. This would imply that the extra training constitutes for an improvement for reaction time but not for accuracy. The assumption from the results would be that participants did not necessarily improve at visually depicting colours from the pre-training experiment to the post training experiment, only that they became faster at doing so. This however could be due to noise in the experiment; for instance if the participant grew tired and inattentive later in the experiment and therefore responded quicker without acutely being attentive to differentiating colours. Or, as the probability value suggests, it may be due solely to the universal force called chance. To solve this, an improved experiment that would give participants more time to rest between trials and phases is suggested. This would then eliminate any noise that may have been present during the experiment before.

Finally, for the most interesting finding of all, reaction time and accuracy depended greatly on whether the stimulus was within the same category or across colour categories. There was a significant increase in reaction time when the colour stimulus was within-category. This left the implication that it was much harder to perceptually discriminate between two colours in the same category than two colours in different categories. There was a large difference between the average reaction time for both across and within-categories before the training session; where there was approximately a 380 millisecond increase for within-category colours compared to across category colours. Then, there was approximately a 75 millisecond decrease in reaction time for both categories from the pre-training to post training, which was not so substantial. However, in regards to accuracy, there was no difference from the pre-training to post-training for across-category colours. Pre-training accuracy and post-training accuracy were both 97% for across category colours. Likewise, for within-category colours the improvement was minute, only 3% from pre-training to post-training (79% to 82%). The conclusion to draw from this would be that the training on average affected reaction time more than it affected accuracy. This however may be due to a ceiling effect. It is not

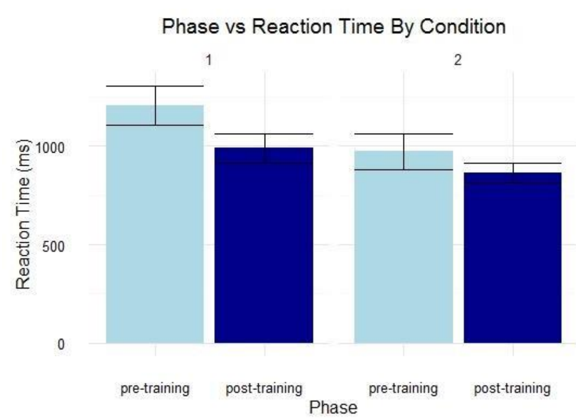

Figure 2: A bar graph showing phase vs. average reaction time split by condition

possible for a participant to achieve greater than 100% accuracy. This may be the reason the pre-training phase and post-training phase both saw an accuracy of 97% for across category colours. An answer may be that participants were already attuned to make discriminations across categories, whereas it is less demanded that inter-category discriminations be made. Because of this, it is suggested for further studies to take interest in the relationship between perception and reaction time pertaining to colour category. For more graphical representations, see appendix.

6.2 Improvement

Before the results are too heavily considered, some healthy criticism should be taken into account. There were multiple aspects of the experiment which may have led to non-accurate results. An example is the demographics of the subjects who participated in the experiment. The age ranged from 19 to 46, with a mean age of 22.95, most of which were attending the University of Aarhus in Denmark. This limits the results to this sample of people in a geographical region and age group, and possibly other unaccounted for characteristics. To generalize the results to the population, a more diverse and balanced sample should be taken.

As discussed before, there were inconstancies in the experiment itself. The people who performed the experiment first experienced a longer or shorter training length than the participants to follow. The problem was eventually addressed and fixed. For this reason, the data of the first few subjects were removed, and the reliable data was analysed without the tainted data.

There were also certain aspects pertaining to the method of the experiment which should be improved in order to secure more accurate results. Although it may seem like a long training session to the participants, a training session of five minutes might not be sufficient to show the different effects linguistic and perceptual training had on the neural plasticity of category boundaries. For this reason, a longer training session is suggested, and/or variance in the amount of time a subject is trained (e.g. five minutes, ten minutes, fifteen minutes). This would directly show the extent to which timing of a training session has on the participants' accuracy and reaction time to stimuli.

Another manner of improvement should be considered for condition one, the linguistic training. Although the categories siniy and goluboy were both displayed as labels in the training session, participants may have tended to overlook the colour category names and only compare the target colour to the colour spectrum which the linguistic label represents. This complication can be overcome by having the participant type the name of the category when the target colour is displayed; or half way through the training session the colour spectrum can be taken off the display so participants would only see the names of the two colour categories to judge which category the target colour belongs to. Alternatively, the entire training sessions could be varying in training methods, which are all linguistically or perceptually linked, to ensure the participant does not grow bored and inattentive during the tedious training period. This would ensure that the participants are using a linguistic representation of the categories to determine their answer and therefore be more supportive of the hypothesis.

As was mentioned before, the assumption of normality was violated in the distribution. To overcome this, a general linear mixed effects model was used using the inverse gamma family. This however, may not have been necessary. Contrary to popular literature, some researchers even go as far as to suggest that the assumption of normality is unimportant, and should be ignored (Gellman and Hill 2007). Although this opinion is exciting, the violation of normality in the data was taken into consideration and actions were carried out to remedy the violation.

A last final note in accordance of improvement of the experiment: the well-known confirmation bias may have taken place when conducting the experiment. To ensure that it was indeed a linguistic aspect of colour representation that was in play, an alternative study with different circumstances to disprove the theory should be conducted for absolute security and integrity.

6.3 The Beauty of Science

As a final note, the topic of science as a whole will be touched upon. As it has been well known in the scientific community for centuries, it should not exclusively be the significant results and findings in support of theories that is strived for when conducting research. In fact, it can be said that there will be more insignificant findings than significant ones in the complexity of the world (there are more things that do not have an influence on an outcome than do). The purpose of science always is to find the truth. Therefore, it is the duty and obligation of a scientist to report findings of all kinds; whether in support of a hypothesis or a theory, or (like in this case) not. Likewise, unfashionable findings, despite what affect they may have on a scientist's ego, should be weighed equally to findings that are more trendy to accept. Scientific approach is a neutral standpoint, where scientists are merely observing and recording the natural phenomena of the universe, not a beauty pageant to dress up hot air.

7.0 Conclusion (D.E.J.)

As was mentioned before, the nature of the findings sparked some interesting implications. It is first interesting that the results contradicted the original hypothesis. Because of this, it is encouraged for researchers to look more into the role that linguistic and perceptual training has on the reaction time and accuracy for colour categories. However, based on this study, it cannot be conclusively stated that linguistic training plays a different role in colour discrimination than perceptual training. The effects simply did not meet the criteria for being another force other than chance. However, it has been proven that both types of training did have an influence on participant performance during the post-training experiment, suggesting that it is possible to train a person's perceptual representation in a brief period of time. Additionally, it has been discovered that the training sessions did not have a sizable effect on across-category discrimination, but did have a dramatic influence on within-category discrimination. In short, both types of training sessions played a role in improving participant colour discrimination, but the individual idiosyncrasies of the two training sessions were not exploited. An improved experiment would be needed to fully prove the original hypothesis.

8.0 Acknowledgments (D.J.)

At this point, we would like to acknowledge Galileo Galilei, for his contributions to the many fields of science and mathematics, and most importantly to observational astronomy. He embodied the phrase: *Just because everyone else believes it, does not make it right*.

9.0 Bibliography (A.T.)

Anderson, J. R. (2015). *Cognitive Psychology and It's Implications* (8th ed.). New York, NY: Worth Publishers.

Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences,7*(2), 84-91. doi:10.1016/s1364-6613(02)00029-3

Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The Thickness of Musical Pitch: Psychophysical Evidence for Linguistic Relativity. *Psychological Science,24*(5), 613-621. doi:10.1177/0956797612457374

Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition,108*(3), 819-824. doi:10.1016/j.cognition.2008.04.007

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models* (Analytical Methods for Social Research). Cambridge: Cambridge University Press.

Gordon, P. (2004). Numerical Cognition Without Words: Evidence from Amazonia. *Science,306*(5695), 496-499. doi:10.1126/science.1094492

Harley, T. A. (2010). *Talking The Talk*. New York, NY: Psychology Press, 90 - 115.

Johnson, M., & Lakoff, G. (2002). Why cognitive linguistics requires embodied realism. *Cognitive Linguistics*,*13*(3). doi:10.1515/cogl.2002.016

Ozturk, O., Shayan, S., Liszkowski, U., & Majid, A. (2012). Language is not necessary for color categories. *Developmental Science,16*(1), 111-115. doi:10.1111/desc.12008

Regier, T., & Kay, P. (2009). Language, thought and color: Whorf was half right. *Cel Press*. Retrieved December 20, 2016.

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*,*4*(3), 328-350. doi:10.1016/00100285(73)90017-0

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences,106*(11), 4567-4570. doi:10.1073/pnas.0811155106

Whorf, B. L., & Carroll, J. B. (1956). *Language, thought, and reality: selected writings*. Cambridge, MA: Technology Press of Massachusetts Institute of Technology (p241).

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of langage on color discrimination. *Stanford University, Stanford, CA.*

(1995). Retrieved December 13, 2016, from http://www.dictionary.com/browse/thought

9.0 Appendix (D.J.)

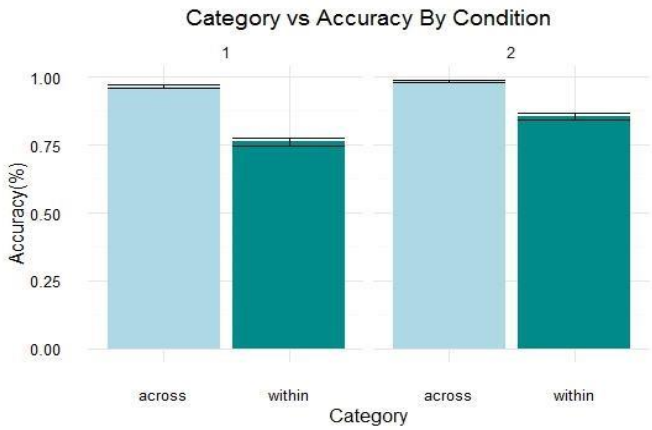

Figure 2

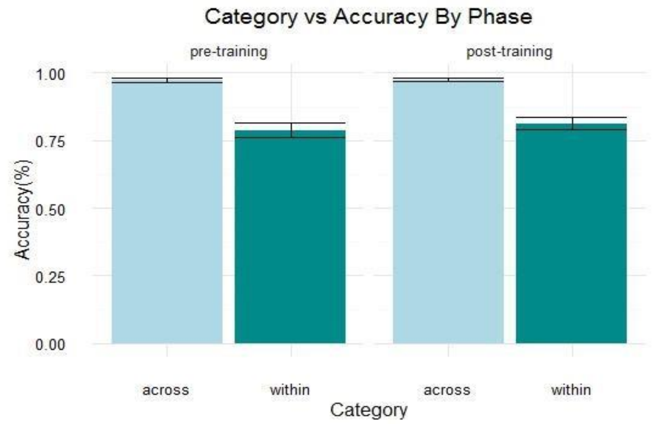

Figure 3

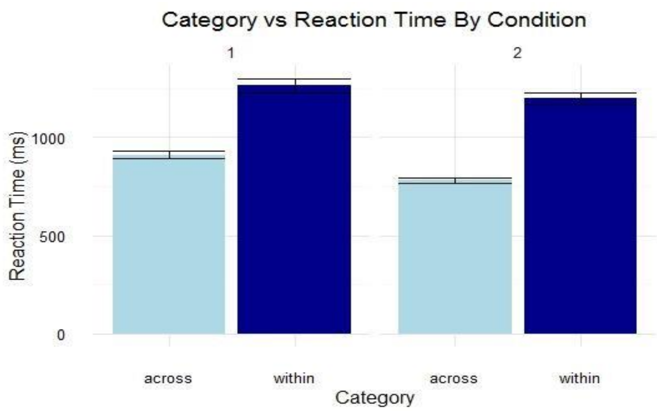

Figure 4

Figure 5

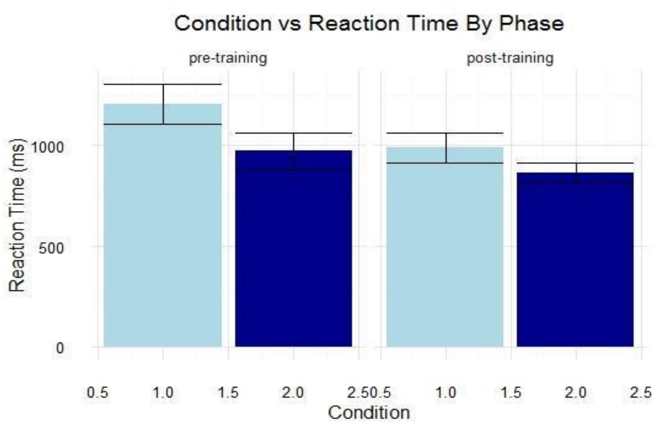

Figure 6

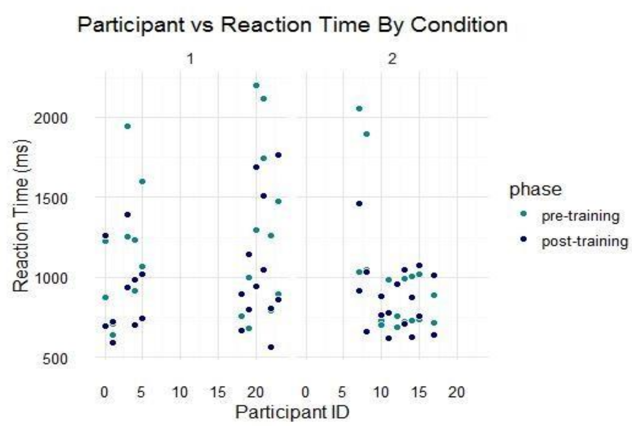

Figure 7

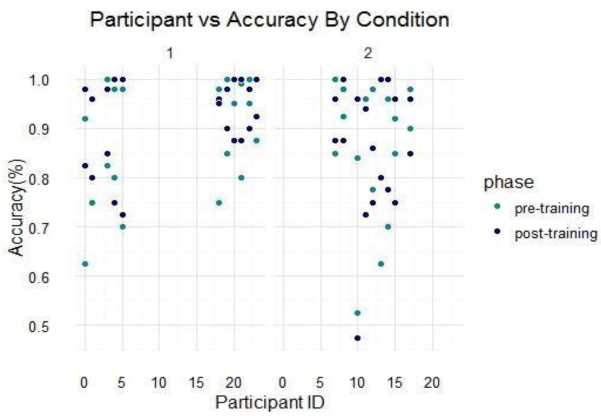

Figure 8

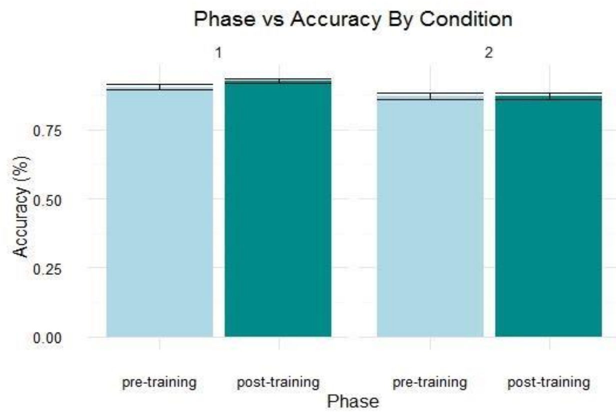

Figure 9

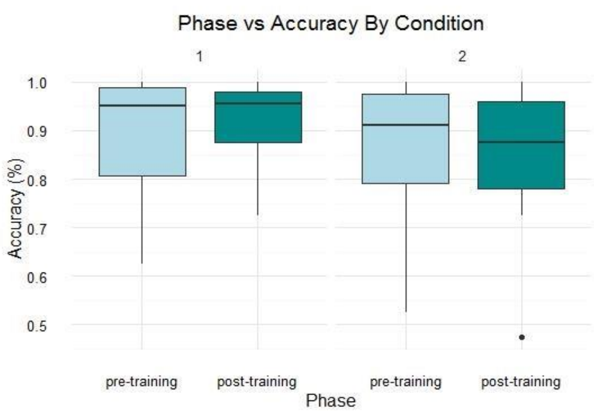

Figure 10

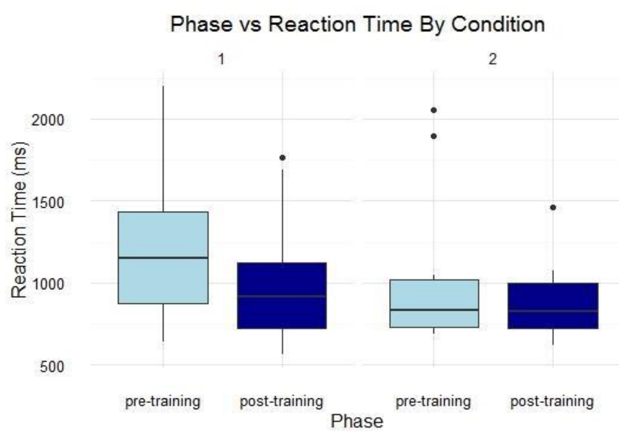

Figure 11

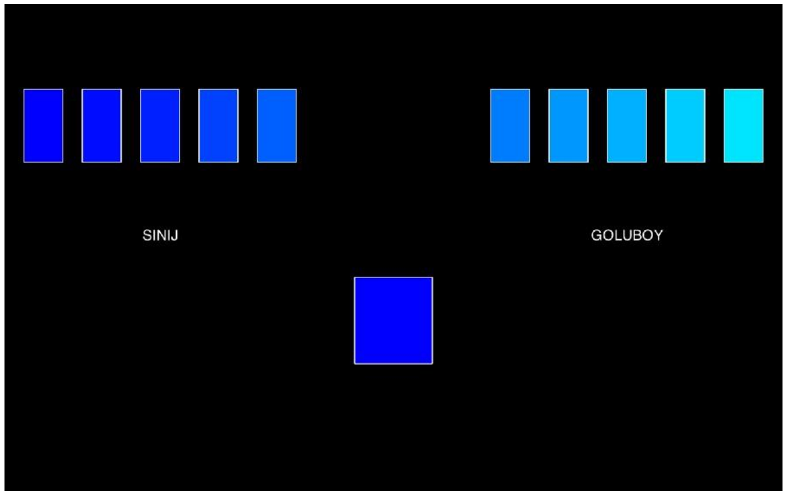

Figure 12: Condition 1 linguistic training

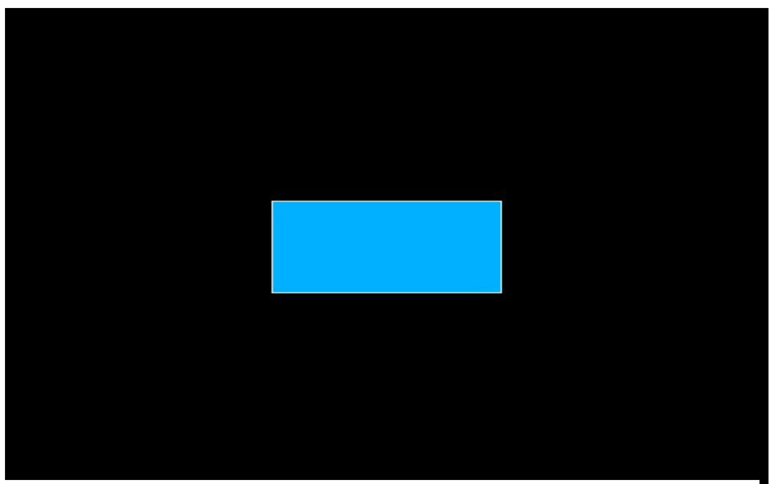

Figure 13: Condition 2 perceptual training

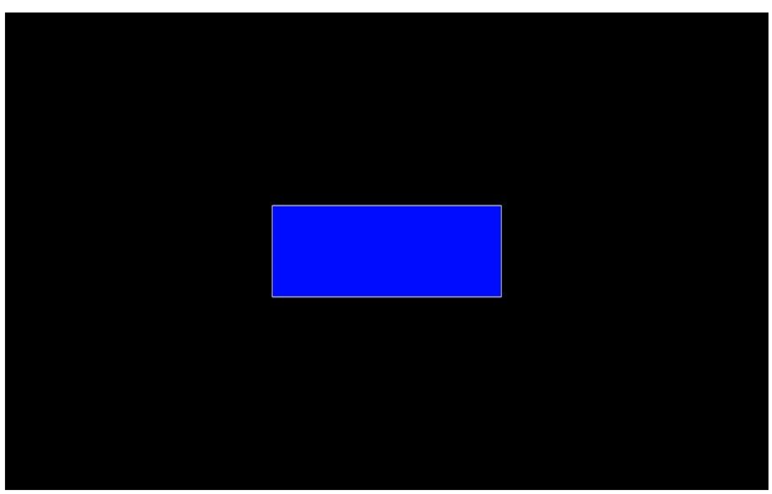

Figure 14: Condition 2 perceptual training. After pressing an arrow key, the square changes colour for the next trial.