

1-1-2014

Blue is in the Eye of the Beholder: a Cross Cultural Study on Color Perception and Memory

Mark Douglas Lowry

University of South Florida, mlowry@mail.usf.edu

Follow this and additional works at: <http://scholarcommons.usf.edu/etd>

 Part of the [Psychology Commons](#)

Scholar Commons Citation

Lowry, Mark Douglas, "Blue is in the Eye of the Beholder: a Cross Cultural Study on Color Perception and Memory" (2014). *Graduate Theses and Dissertations*.

<http://scholarcommons.usf.edu/etd/5360>

This Thesis is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.

Blue is in the Eye of the Beholder: a Cross Cultural Study on Color Perception and Memory

by

Mark Lowry

A thesis submitted in partial fulfillment
of the requirements for the degree of
Masters of Arts
Department of Psychology
College of Arts and Sciences
University of South Florida

Co-Major Professor: Judith B. Bryant, Ph.D.
Co-Major Professor: Nathan Maxfield, Ph.D.
Thomas Sanocki, Ph.D.

Date of Approval:
October 31, 2014

Keywords: Linguistic, Relativity, Sapir-Whorf, Cognition

Copyright ©2014, Mark Lowry

Dedication

I dedicate this thesis to my wife, Olya. She is the one who told me that Russian-speakers are more likely to describe eyes as being grey, which gave me the idea for this study. She also helped me find Russian and English participants and has supported me and my research every step of the way.

Acknowledgments

I would like to express my appreciation and gratitude to my advisor, Dr. Judith B. Bryant for guiding me through my thesis. Dr. Bryant was instrumental in helping me set up the study's design. Without her support, guidance and supervision, this thesis would not have been possible. I would also like to thank my committee members, Dr. Thomas Sanocki and Dr. Nathan Maxfield. Their contributions and input were invaluable.

In addition, I would like to thank all my good friends and family in both Florida and St. Petersburg, Russia who took part in the study. They unselfishly volunteered their time, and for that I am truly grateful. Especially important were my friends at Eclectic Translations, who helped me find many Russian participants.

Table of Contents

List of Tables	ii
List of Figures	iii
Abstract	iv
Chapter One: Introduction	1
The Linguistic Relativity Hypothesis	1
Early Cross Cultural Color Studies Favoring Universalism	3
The Neurophysiological Basis of Color Perception.....	6
Problems with the Universalist Approach and Evidence in Favor of the Linguistic Relativity Hypothesis Regarding Color	11
Russian and English Descriptions of Eyes.....	15
Pilot Study.....	18
Objectives and Hypotheses of this Thesis	21
Chapter Two: Method.....	23
Participants.....	23
Materials	23
Stimuli.....	23
Response Scale.....	24
Apparatus	25
Procedure	26
Chapter Three: Results.....	28
Descriptive Analyses	28
Demographic Analyses	29
Tests of Predictions.....	29
Exploratory Analyses.....	32
Chapter Four: Discussion.....	39
Conclusion	44
References.....	46
Appendix: IRB Approval Letter	49

List of Tables

Table 1	Eye color options on Russian dating sites.....	17
Table 2	Eye color options on English dating sites.....	17
Table 3	Means and other descriptive statistics of pilot study	20
Table 4	Means and other descriptive statistics for eyes and tiles	30
Table 5	Model testing the relationship between the covariate (mean tile score) and dependent variable (mean blue eye score) at each level of the independent variable	36

List of Figures

Figure 1: A representation of the color component channels.....	8
Figure 2: Color illusion where red squares look like different colors.....	10
Figure 3: Interaction graph of pilot study.....	20
Figure 4: An example of an eye used in the study on the left, and its corresponding tile used as a control on the right	24
Figure 5: An example of the blue color scale (R) and brown color scale (L).	26
Figure 6: Means for blue tile ratings across conditions.....	35
Figure 7: Regression slopes of the dependent variable and covariate at all levels.....	36
Figure 8: Adjusted mean estimates from the ANCOVA model.....	37
Figure 9: Estimates of deflections at each level of the independent variable.	37
Figure 10: Mean difference of blue eye ratings between Russian and English speakers.....	38
Figure 11: Mean differences of blue eye ratings between conditions	38
Figure 12: Mean differences of blue eye ratings for all levels of language and condition	38

Abstract

According to the linguistic relativity hypothesis, the language one speaks affects how one thinks. Because languages differ in how they categorize color, linguistic relativity has often been tested by conducting experiments on color perception and memory. This study examines the linguistic relativity hypothesis using ecologically valid stimuli: pictures of eyes. Because Russian-speakers are more likely to describe blue/grey eyes as grey, whereas English speakers are more likely to describe them as blue, English and Russian participants were asked to match the overall color of blue eyes to a color scale. There were three conditions. In the first condition (perception), participants saw the color scale and an eye picture simultaneously and then chose the color that best matched the picture. In the second condition (memory), participants matched the color of an eye to the color scale from memory. The third condition (label) was similar to the second, except participants labeled the eye orally before matching the color from memory. A 3 (condition) x 2 (language) ANCOVA and Bayesian analysis were used to analyze the data. Overall, the ANCOVA and Bayesian analysis indicated that there was a main effect of language. Russian-speaking participants were more likely to rate the eyes as greyer than English-speaking participants. The Bayesian analysis also suggested that there may also have been an interaction, with Russian and English-speaking participants rating the eyes similarly in the perception condition, but not the memory or labeling conditions. Overall, the findings provide novel evidence for the linguistic relativity hypothesis.

Chapter One

Introduction

The Linguistic Relativity Hypothesis

I have tested the linguistic relativity hypothesis by examining how native Russian and English speakers perceive and remember eye color. The linguistic relativity hypothesis is an idea that evolved over time based largely on the work of Benjamin Whorf in the first half of the twentieth century. Whorf was never formally trained in linguistics. He studied chemical engineering at the Massachusetts Institute of Technology and worked at the Harford Fire Insurance Company after his graduation until his death. Despite not having an education in linguistics, he became interested in how language may influence thought in his late twenties. As a result, he developed ties with several prominent linguists of his time including Edward Sapir. These ties helped him study Native American languages (Whorf & Carroll, 1998).

Whorf's ideas have been condensed into the linguistic relativity hypothesis. The hypothesis posits that the language one speaks influences the way one thinks. According to Whorf, grammar, word frequency and linguistic categories all shape thought. He said, "We dissect nature along lines laid down by our native languages.... The world is presented in a kaleidoscopic flux of impressions which has to be organized by our minds—and this means largely by the linguistic systems in our minds" (Whorf & Carroll, 1998, p. 213). For example, Whorf argued that in the Hopi language, verbs treat time differently than they do in English. According to Whorf's research, brief events that are described in English as nouns (e.g., lightning, wave, flame, meteor) are always verbs in Hopi. This is due to the fact that Hopi

grammar is more likely to classify nouns and verbs in terms of duration than is English (Whorf & Carroll, 1998, p.215). Whorf argued that variations in categorization such as the one just described train speakers to attend to different phenomena and thus think about such events differently.

Another famous example that is commonly used to explain the linguistic relativity hypothesis is the fact that certain Eskimo languages have several words for snow. The phenomenon was first described by Franz Boaz, a German-American anthropologist (Boaz, 1911, pp. 145-146). Whorf claimed that if Eskimos have more than one word for snow, then they think about snow differently. There has been a lot of debate about whether differences between languages really do shape thought and perception.

Among those most critical of the linguistic relativity hypothesis are cognitive psychologists. In my opinion, there are two main reasons why this is the case. First, linguistic relativity gained in popularity during a time when there were few if any empirical studies that examined it. When the cognitive revolution took place, many cognitive psychologists naturally questioned claims about thought that lacked data to back them up. Secondly, cognitive science deals with finding universal patterns in memory, perception and thought in general. Linguistic relativity states that cognition may vary based on the language one speaks. Researchers who are trying to find universal rules may feel challenged by such an idea.

Those who favor a universalist explanation of how thought and language are linked argue that thinking is more or less independent of language. For example, Noam Chomsky distinguished between the deep structure and surface structure of grammar. Deep structure refers to the components of grammar that give an utterance its underlying meaning. Surface structure refers to the outward manifestation (which is similar to the phonetic expression) of the utterance

(Chomsky, 1965, p. 17). Chomsky argued that although surface structure may vary across languages, deep structure does not (Chomsky, 1965, p. 118). In other words, languages are very similar in terms of the types of rules that govern them, but they are different in how they are conveyed verbally. To Chomsky, any differences observed between two languages are just different ways of saying the same thing. If this is true, there would be minimal (if any) influence of one's language on cognition.

I aimed to examine how language influences perception and memory in my thesis by examining the differences in how English and Russian-speakers describe eye color. Russian-speakers are more likely to describe an eye as grey whereas English-speakers are more likely to describe it as blue. If the linguistic relativity hypothesis is valid, then one would expect the way the two cultures describe eyes will influence the way they perceive and remember them. It should be noted that, although this study deals with color perception, Whorf never actually talked much about color. He was more interested in grammar, how people make categorizations based on their native language and how those categorizations influence thought.

Nevertheless, there has been a substantial amount of research done on color in an attempt either to support or refute Whorf's views. Both opponents and advocates of linguistic relativity point to different studies that examine color in order to provide evidence for or argue against the hypothesis. The experiments that have been done are useful in understanding color perception. However, most were done by using simple Munsell chips or stimuli that are uniform in color rather than real world objects. My study was an attempt to use stimuli one would find in the real world to test Whorf's ideas.

Early Cross Cultural Color Studies Favoring Universalism

In order to study how language influences thought, many researchers have studied color. Color is a common way to study the linguistic relativity hypothesis because the stimuli can be controlled. In other words, a color can be described objectively in terms of wavelength, saturation, brightness, etc. Different languages divide color into different categories. At first glance these categories seem to be culturally specific and non-universal. English has the word *blue*, whereas Russian divides blue up into two terms: *goluboy* (*light blue*) and *siniy* (*dark blue*). The fact that color categories are partitioned differently across languages supported the idea of the linguistic relativity hypothesis in the early twentieth century.

Berlin and Kay (1969) were among the first to question whether languages/cultures divide color into arbitrary categories. They systematically studied color terms across 20 different languages. Participants were asked to indicate the total number of 'basic color terms' their native language had and where the boundaries were for each basic color term. Basic color terms had to meet certain criteria: they had to be mono-lexemic (e.g., *lemon-colored* would not be acceptable), they could not be represented by any other color term (e.g. *crimson* is a type of *red*, and therefore not a basic color term), they could not be restricted to only a narrow class of objects (e.g., *blond* is unacceptable) and they had to be psychologically salient, meaning native speakers had to know what the color terms were and agree that they were indeed major color categories.

Once participants named their language's basic color terms, they were given a board with 329 standardized color chips that were arranged by hue and brightness. Participants were asked to indicate every chip that represented each color term as well as the best example of each. The data indicated that languages categorize colors systematically.

Despite the fact that one language may have a different number of basic color terms, a distinct pattern emerged. Each language had at least two color terms. If a language only had two, then the terms represented *black* and *white*. All languages with only three basic color terms had color categories for *black*, *white*, and *red*. If a language had four color terms, then they divided the color chips into *black*, *white*, *red* and either *yellow* or *green*. Languages with five color terms categorized colors into *black*, *white*, *red*, *yellow* and *green*. If a language had six, then they had the same categories as languages with five categories but added a *blue* term. If a language had seven or more color terms, it had some combination of purple, pink, orange and grey. Berlin and Kay's work was influential and helped turn the tide against linguistic relativism to universalism.

A couple of years later, Heider (1972) found that Berlin and Kay's "basic colors" are more easily encoded and better remembered across different languages than are "inter-nominal" or "boundary" colors. Inter-nominal colors are colors that do not fall into any of the basic color ranges. However, inter-nominal colors were similar in hue, value or saturation to basic color terms. Boundary colors were those that were adjacent to basic colors. In a series of experiments, Heider tested several dependent variables including the number of letters colors have, reaction time for naming colors, and memory for colors. In each experiment a general trend emerged: regardless of language, basic colors had fewer letters, were named more quickly and were remembered better than inter-nominal and boundary colors.

Other factors also contributed to universalism gaining popularity. Heider (1971) found that children chose basic colors more frequently when allowed to choose any color, matched basic colors better than non-basic ones, and chose basic colors to represent color terms. Bornstein, Kessen, and Weiskopf (1976) discovered that infants who are habituated to a certain hue will look at a different hue longer if it comes from a different adult color category than if it

comes from the same category even if the new colors are equally distant in terms of wavelength from the original hue. That is, they categorized colors as adults do. For example, the researchers found that infants who are habituated to a blue hue of 480 nanometers look longer at a green hue of 510 nm than a darker blue hue of 450 nm. On the basis of such research, linguistic relativity looked like a flawed hypothesis, at least in terms of color perception. It was replaced with a universal, neurophysiological explanation that was popular in cognitive science. A quote from Pinker (1995, p. 136) sums up the “consensus” that was formed:

[The linguistic relativity hypothesis] asserts that the categories and relations that we use to understand the world come from our particular language, so that speakers of different languages quite literally conceptualize the world in different ways....This is an intriguing hypothesis, but virtually all modern cognitive scientists believe it is false.

However, understanding the neurophysiological basis of color perception not only helps one understand how universal color terms develop, but may also help support the linguistic relativity hypothesis.

The Neurophysiological Basis of Color Perception

There are three points to be made in this section. The first is to explain the biological argument behind the universalist explanation of color perception. The second, which may seem contradictory to the first, is to demonstrate that neurophysiology does not preclude Whorfian effects on color perception and memory. The third is to demonstrate that even if the differences in perceiving color are small across cultures, those differences would be significant given the strong, universal, biological component of color perception.

Biologically speaking, it is difficult to explain how people experience color. Light enters the eye, is absorbed by light-sensitive cells, gets converted into electrochemical signals and is

sent to the brain where it is interpreted. However, there is still a ‘black box in our brain’ into which the signals enter and out of which vision is produced (Kuehni, 2012, p. 23). Despite this, much is known about the cells and parts of the brain that are involved.

Humans can see wavelengths of light that range roughly from 400 nm to 700 nm. Once light hits the retina, it passes through several layers of cells, including ganglion cells, amacrine cells, bipolar cells, rods, and cones (Kuehni, 2012). I only discuss cones in detail here because it is most responsible for color vision. There are three types of cones, and each type has a different absorption spectrum (Cornsweet, 2012). In other words, each type has different wavelengths of light to which they are most sensitive. There is debate as to what to call the types, but for this thesis they are named *L*, *M*, or *S* in reference to their sensitivities. If one were to label colors according to wavelength, red would correspond to long wavelengths, green would correspond to medium wavelengths and blue would correspond to short wavelengths. However, different types of cones can be activated at the same time. For example, light that creates a response in an *L* cone may also create a response in a nearby *M* cone. (Stone, 2012).

It should be noted at this point that this explanation of color perception is an oversimplification of what actually happens. Information from all three cones is used in order to process color. However, it is more complicated than just mixing input from the *L*, *M* and *S* cones. There appears to be an ‘opponent system.’ Color opponency refers to the fact that output data from different types of cones are subtracted from each other. There are two ways this happens: *L-M* and *L+M-S* (Kuehni, 2012). These form two color dimensions or channels. The *L-M* subtraction creates a color channel of red and green whereas the *L+M-S* subtraction creates a color channel of yellow and blue (see Figure 1). In other words, output from specific types of cones is compared to the output from others. In the *L+M-S* dimension, the *L* and *M* cones are

compared to S cones. If light hitting the retina is above 500 nm, the yellow side of the dimension dominates and the blue side is inhibited. In the L+M dimension, green dominates from about 475 nm to 575 nm, while red dominates above and below those values (Kuehni, 2012). The brain then takes the information from the two channels and perceives other colors. See Figure 1 for a representation of the two color channels.

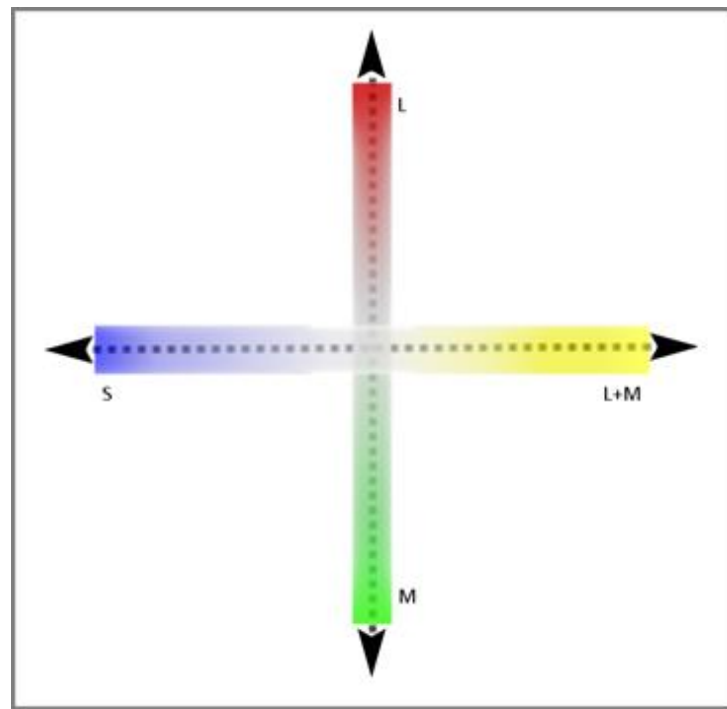


Figure 1. *A representation of the color component channels.*

A couple of points should be made. First, because of the opponent system, some pairs of colors can never be perceived as being mixed. When the L and M cones activate yellow, blue is inhibited. Thus, there is no “yellowish-blue” color term or vice versa. The same is true of red and green. Second, there are four unique hues that correspond to the excitation of only one half of one color channel. For example, if there is only L cone stimulation, then a unique hue of red is perceived that one would say is not a mix of any other colors (i.e., it is not perceived as being reddish-orange, purplish-red, etc.). There are also unique hues for yellow, green and blue where

unique yellow is produced by L and M cones being activated with no S cone input, unique green is created by having only M cone stimulation and unique blue corresponds to only S cone stimulation. These hues correspond to the colors that make up the two channels (Kuehni, 2012). As one may have noticed, these are the main four colors (other than black and white) that Berlin and Kay (1969) found in most languages. Biology certainly plays a role in how colors are categorized across peoples.

Taken together, it is clear that there is a strong universal neuropsychological component to color perception. The first colors (after black and white) that are common to most cultures are red, yellow, green and blue, which match the colors in the two color channels. Additionally, there are no words for the physiologically impossible hues of yellow-blue or red-green. If languages arbitrarily create color categories, one would expect to find at least some languages with such color terms. The fact that they do not exist is evidence that color perception is limited by biology. However, there is another piece of the puzzle that has not yet been discussed in this section of the proposal: the brain itself.

The previous description of cone type and color perception seems straightforward. In reality, there are multiple steps in processing the information in the brain that lead to color opponency. There is a lot of uncertainty as to the specifics, but in general ganglion cells in the retina play an important role in creating initial color opponency. Information is then sent to the lateral geniculate nucleus, the primary visual cortex (V1), the prestriate cortex (V2) and on to higher visual areas (e.g., extrastriate visual cortex, Visual Area V4), which are responsible for perceiving color, orientation and movement (Spillmann, 1990, pp. 179-185). It should be noted that by the time color information reaches the primary visual cortex, initial color opponent information has been modified, and by the time it reaches V4, color information is averaged

across cells and combined with brightness information that has been processed separately (Kuehni, 2012). Complicating matters further is the fact that the brain combines visual information with previous knowledge about the world. Previous knowledge that is used includes information from memory and expectations about what one expects to see.

A logical conclusion is that eyes do not see. The brain does. Color is not perceived based on wavelength alone. For example, the same color may look different depending on the color of its background. For an example of this, see Figure 2 where the red boxes are the same wavelength of light, but the ones on the right look lighter than the ones on the left. This happens because the brain uses the surrounding context in its interpretation of color.

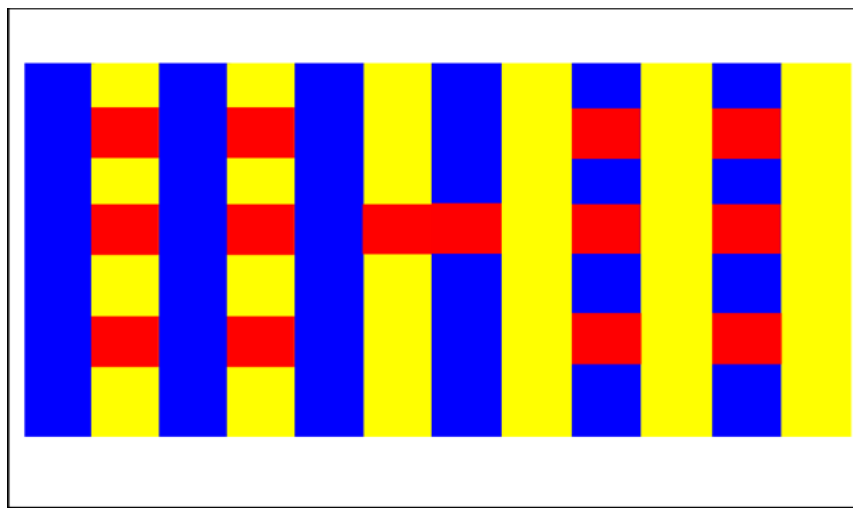


Figure 2. *Color illusion where red squares look like different colors.*

Color constancy is another phenomenon that demonstrates that the brain does not rely only on the wavelength of light to perceive color. Color constancy refers to the fact that the color of an object is perceived the same even if the illuminant changes (Ebner, 2007). For example, an apple looks red in bright sunlight and under a fluorescent light. Stone (2012, p. 156) pointed out that what we see follows a simple formula: a perceived image = image data + prior expectations. If this is true, then one could speculate as to how the language one speaks affects and/or reflects

one's prior expectations. If one's culture describes the color of an object differently than does another culture, would that affect the perception of that object? What about how one remembers it? These are questions that are directly related to the linguistic relativity hypothesis and this thesis.

Problems with the Universalist Approach and Evidence in Favor of the Linguistic Relativity Hypothesis Regarding Color

It is clear that the brain interprets input from the retina and combines it with prior expectations. It is unclear whether linguistic information is part of those prior expectations. However, an important point must be made first: the way Whorf's hypotheses were framed by cognitive scientists in the mid and late twentieth century was different than what Whorf actually proposed. Before one can examine evidence about the linguistic relativity hypothesis, one has to frame it correctly.

When researchers started criticizing what they called "linguistic relativity" or the "Sapir-Whorf hypothesis" they made three arguments: linguistic relativity implies that language is deterministic, it is hermetic, and there can be infinite variability between languages (Leavitt, 2010, pp. 167-168). Determinism refers to the idea that language (and only language) determines how one thinks and perceives. Hermeticism states that it is impossible to translate from one language to another because each language has its own unique way of conveying ideas. Infinite variability is the idea that languages are different from each other in every aspect imaginable (e.g., how they categorize color, phonemes, grammar). In fact, Whorf never made any of these arguments. They were extreme positions used in order to strengthen universalism's claims. Proverbial straw men were set up that could easily be knocked down by any one of three possible findings: language is not a prerequisite of thought, one can translate a language into another

language, or language universals exist. For example, Michaels (1977) oversimplifies Whorf's ideas by stating the following:

Whorf's claims that individual languages structure reality differently seem to fail on his own evidence. The fact that languages are inter-translatable... suggests that linguistic relativity reduces simply to the claim that to speak grammatically in any given language one must use the grammar of that language (p. 333).

The fact of the matter is that such criticism is an oversimplification and misrepresentation of Whorf's position. Whorf believed in the universality of cognitive abilities and that one can say anything one wants to in any language. If people can say whatever they want to in their native language, then translations are no longer impossible. In reality, the linguistic relativity hypothesis can be summed up by what Whorf really thought. He believed that different languages predispose speakers to attend to certain domains over others based on how they normally convey information, and that is the basis for how language affects thought (Leavitt, 2010). For example, in English there are different ways to describe water. One can talk about water as rain, dew, snow, ice, mist, steam, hail, frost, and so forth. Each term takes on a specific meaning, and the way each is used contextually trains native English speakers to notice specific properties of water that they otherwise may not have detected.

It should also be briefly noted that many researchers did not sufficiently consider evidence that supported the linguistic relativity hypothesis, and their methods may have been culturally biased. In Berlin and Kay's study (1969) on basic color terms, there was substantial variation in the number of color terms across cultures, and that variation was not studied in depth. When Heider (1972) published a series of experiments that found basic colors were more easily encoded and remembered across cultures, she found that Dani speakers were significantly

worse at remembering colors than English speakers. The memory task consisted of seeing a uniform colored chip, waiting five seconds and then picking out the color just seen in a 160-chip array. The fact that they did significantly worse than English speakers is worth mentioning because the Dani language only has two color terms. In fact, the discrepancy between the two groups had the largest effect size in the study. One interpretation of these results could be that having words for the so-called basic color terms aids in perceiving and remembering those colors. This explanation was not taken seriously by many researchers at the time because Dani speakers still remembered basic colors better than other hues (such as inter-nominal colors).

The studies also relied on characterizing color by hue and brightness. This may have biased the results. For example, Conklin (1955) found that Hanunoo, a language spoken in the Philippines, describes color based on hue, brightness as well as “wetness.” Based on Berlin and Kay’s criteria, Hanunoo has four basic color terms: black, white, red and light green. However, since colors also seem to be paired with information about their “wetness,” researchers may be prone to overlook how those categories are applied in practice. For example, in Hanunoo a shiny, wet brown colored section of bamboo is labeled as green because it is wet. Theoretically, it should be described as red because brown is closer in hue to red than green. The important point here is that if one wants to examine the linguistic relativity hypothesis and color perception critically, one must look at how language variation affects cognition, and not just how physiology (e.g., cones in the retina) affects language.

As mentioned previously, the main argument behind linguistic relativity is that one’s language helps determine what people attend to. If different languages do in fact make speakers pay attention to certain domains more than others, then there are a number of ways to test this proposition. Many researchers have turned to how language affects color coding and reaction

time. Kay and Kempton (1984) were among the first to examine coding, and they were able to find so-called “Whorfian effects.” They conducted a cross-cultural study between English and Tarahumara speakers. Unlike English, Tarahumara only has one word for what English speakers would consider green and blue. Participants were given three colors and asked to pick the one that was most different. They found that English speakers were more likely to exaggerate the subjective distances of colors close to the green-blue boundary. In other words, they tended to say that the overall difference between green and blue was greater than did Tarahumara speakers.

In the past 20 years, there has been a resurgence of researchers examining Whorfian effects on color. A few are mentioned briefly here. Goldstein, Davidoff and Roberson (2009) found that English-speaking toddlers did better at recognizing and remembering colors that were between categories of blue and green than did Himba-speaking children, but only if the toddlers had blue and green in their vocabulary. It should be noted that Himba does not have green and blue color terms. This suggests that having specific color terms may aid in memorizing colors of objects.

Winawer et al. (2007) discovered that Russian speakers were faster than English speakers at discriminating between light blue and blue (*goluboy* and *siniy* in Russian, respectively) in a reaction time task, but their advantage disappeared when they had to simultaneously do a verbal task. It should be noted that the Russian color terms for light blue and blue could both be considered basic color terms based on Berlin and Kay’s criteria, whereas light blue in English would not be considered a basic color term. Winawer et al. also discovered that Russian speakers’ advantage returned when doing a non-verbal spatial task. This indicates that language may aid in perceiving colors because when language resources in the brain were being used (i.e., when participants had verbal interference), Russian speakers were slower at discriminating

between light blue and blue hues. The effect reversed when language resources in the brain were not used.

Kwok et al. (2011) recorded changes to the visual cortex as Mandarin-speaking participants were trained to distinguish between four made-up categories over three days. They observed that there was an increase of grey matter in the V2 and V3 areas of the left visual cortex, providing a possible structural basis in the brain for Whorfian effects.

Russian and English Descriptions of Eyes

Researchers studying cross cultural color perception have primarily used Munsell color chips as stimuli for participants. Munsell color chips represent a color space that defines color by three dimensions: hue, lightness and chroma. Because each color chip is a uniform color, they are useful in studying categories of color in general and have been used by most of the researchers mentioned in this introduction. Most researchers gave participants an array of color chips and had them organize them by color for example. Other researchers had participants indicate the best example of a color term (e.g., blue). However, I have not found any studies that looked at how people of different cultures perceive colors of real world objects. In contrast, I aimed to take the linguistic relativity hypothesis one step further. If it is valid, I would expect different cultures to think about the color of the same object differently if their languages use different color terms to describe that object. One just has to find an object that would meet such criteria.

Fortunately, Russian and English describe eye color differently. Based on my own experience while living in Russia, I observed that what I describe as a blue eye, many Russians describe as a grey eye. This led me to believe that the two cultures label the same colored eye differently in at least some cases. There appears to be some evidence to back up my conclusion.

Examining how academics study eye color classification is one method of determining how different cultures describe eye color. In the West, in the early 20th century, Davenport and Davenport (1907) published an article in *Science*. In it, they stated that eye color falls into two categories: blue and brown. They also proposed that blue eyes were recessive to brown. Even today, many educators in English-speaking schools teach that people with blue eyes have recessive genes, even though such an explanation is overly simplistic and not entirely correct (Sturm & Frudakis, 2004). Many anthropological studies of eye color in the U.S. and Europe use a three-point scale consisting of combinations of blue, green-hazel and brown in order to classify eye-color (Sturm & Larsson, 2009).

There are major differences in categorizing eye color in Russia. It should be noted that for much of the 20th century, Russia was under Soviet control and cut off from Europe and the West. Communication between the two blocs was limited. Soviet researchers often developed their own methods of doing research, and eye-color classification was no exception. V. V. Banak was a Russian anthropologist who studied, among other things, eye color. Banak developed a way to categorize eyes. In his system, there are three main types of eye-color: dark, mixed and light (Banak, 1965). Each main type is subdivided into categories. For brevity, I will only consider light colored eyes. Light colored eyes are divided into colors that would be translated into English as grey, light-blue, greyish light-blue and blue. Banak's method is still popular in Russia today. For example, on one of Russia's main social networking sites VK, there is a page dedicated to figuring out one's eye color based on Banak's scale (WotanJugend - INFO).

Another way to look at differences between native Russian and English speakers in how they describe eye color is to look at dating sites aimed at Russian-speaking audiences and compare them to dating sites for English-speaking audiences. In my analysis of over 15 Russian

dating sites, seven had an option to specify one's eye-color. See Table 1 for the eye-color options found on these Russian dating sites. There were five color options given on those sites: *grey*, *light-blue*, *green*, *brown/hazel* and *black*, which were translated from Russian color terms *seriy*, *goluboy*, *zelyoniy*, *kariy* and *chyorniy*, respectively (in Cyrillic: *серый*, *голубой*, *зелёный*, *карий* and *чёрный*). Of the seven sites, six contained the term grey. Of those six sites, four had grey as the first or second choice. More than ten English sites were also analyzed. Of those, seven gave users options to choose eye-color. See Table 2 for the eye color options on dating sites aimed at English speaking audiences. Of the seven sites, all of them had an option for blue, and five of them had an option for grey. Of those, blue was always in the first or second position, while grey never appeared higher than the third position.

Table 1.
Eye color options on Russian Dating Sites

Russian Dating Website	Eye Color					
	1st Option	2nd Option	3rd Option	4th Option	5th Option	6th Option
mylove.ru	Grey	Light-Blue	Green	Brown/Hazel	Black	Multicolored
ru.fdating.com	Black	Grey	Light-Blue	Green	Brown	Brown/Hazel
badoo.com	Brown/Hazel	Green	Light-Blue	Grey	Hazel	Black
loveplanet.ru	Brown/Hazel	Grey	Light-Blue	Green	N/A	N/A
rusdate.net	Brown/Hazel	Light-Blue	Green	Black	N/A	N/A
znakomstva.ru	Light-Blue	Green	Grey	Brown/Hazel	Dark	N/A
komuza40.ru	Brown/Hazel	Grey	Light-Blue	Green	N/A	N/A

Table 2
Eye color options on English Dating Sites

English Dating Websites	Eye Color					
	1st Option	2nd Option	3rd Option	4th Option	5th Option	6th Option
datehookup.com	Blue	Green	Brown	Hazel	N/A	N/A
pof.com	Blue	Hazel	Grey	Green	Brown	N/A
match.com	Black	Blue	Brown	Grey	Green	Hazel
chemistry.com	Black	Blue	Brown	Green	Grey	Hazel
spark.com*	Blue	Blue-Grey	Blue Green	Grey	Green Grey	Turquoise
seniorsmeet.com	Blue	Green	Hazel	Brown	Grey	N/A
christianmingle.com	Black	Blue	Brown	Green	Hazel	Red

Examining word frequencies can also shed light on how speakers of English and Russian describe eye color. I used the American Corpus (Davies, 2014) to compare two relevant phrases in English: *Blue eyes* and *grey eyes* (the term *gray eyes* was also included in the search). I searched the Russian National Corpus (2014) for three terms in Russian: *grey eyes*, *light blue eyes*, and *blue eyes*. Overall, in U.S. English, the term *blue eyes* occurred 3577 times and *grey eyes* occurred 666 times. In Russian, there were 1569 results for the term *grey eyes*, 1937 results for *light blue eyes* and 697 results for *blue eyes*. It should be noted that singular and plural forms of each term for each language were part of each search. Also, Russian is morphologically complex. The endings change by case (genitive, prepositional, etc.) depending on the noun or adjective's role in the sentence. Word order in Russian is flexible as well. To compensate, all combinations of case and word order are included in the aforementioned results.

There are two possible explanations for the above findings. One is that Russian and English speakers classify eye-color differently and may perceive and/or remember eye color differently. The other is that there are genetic differences between speakers of the two languages and ethnic Russians' eyes are indeed greyer than native English speakers'. The truth may lie somewhere in between, and the present study helps answer such questions.

Pilot Study

Although the difference in describing eye color between the two cultures may have a genetic cause (e.g., ethnic Russians may have eyes that look greyer in general), it is unlikely that genetics is the only factor in determining how Russians describe eyes. In fact, a quasi-experiment would help determine whether language influences perception and memory. By providing native Russian and English speakers a picture of a blue or grey eye along with a color scale, any difference between the languages would indicate that language may play a role in perception.

One could also study whether language affects memory of eye color by showing participants the color scale separately after presenting a picture of an eye, or by having the participants label the color of an eye and match the color from memory. Having participants label the eye's color would theoretically prime them to remember the eye as they described it, which would give strong evidence of language affecting memory. In a pilot study, I tested these possibilities.

Ninety-seven participants (52 English speakers and 45 Russian speakers) rated the color of two eyes in one of three conditions: *perception*, *memory*, and *label*. In the perception condition, participants saw the color scale and eye simultaneously. Participants in the memory condition had to rate the color of the eye from memory, and the label condition was similar to the memory condition except those participants were asked to “describe the color of the eye” before rating the eye's color.

The results were mixed. Overall there was a main effect of language: Russian speakers rated the eyes as more grey (see Table 3). However, there was an unexpected interaction: both language groups were similar in their color rating for the label condition (see Figure 3). There are a couple of possible explanations for why this happened. The first is that the scale was flawed. The center of the scale was an estimated average of the eyes' color. The luminance was increased by equal intervals on one side while the other side increased in blueness in the same manner. This may have confounded the results, exaggerating the effects between language groups. The scale was meant to start from grey and go to blue, but two variables (luminance and color) were manipulated instead of one. It is therefore hard to know what was determining a participant's choice in any one condition, whether it be how light a color looks, or how blue it looks. A scale that systematically changes both dimensions at the same time to give shades of blue and grey may be a better choice.

Table 3

Marginal Means and Other Descriptive Statistics of Pilot Study. Higher numbers indicate more blue, while lower numbers indicate more grey.

Condition	Language	<i>n</i>	<i>M</i>	<i>SD</i>	Median	Min	Max	Skew	Kurtosis
Label	English	17	4.59	1.88	4.5	1.5	8.0	0.25	-0.70
Memory	English	18	6.06	2.13	6.0	2.0	9.0	-0.22	-1.36
Perception	English	17	5.53	2.55	5.5	2.0	9.0	0.03	-1.68
Label	Russian	15	5.20	2.19	5.5	1.5	9.0	-0.16	-0.96
Memory	Russian	16	3.94	1.81	3.5	2.0	8.5	1.03	0.16
Perception	Russian	14	4.25	2.15	4.0	1.0	8.0	0.36	-1.05

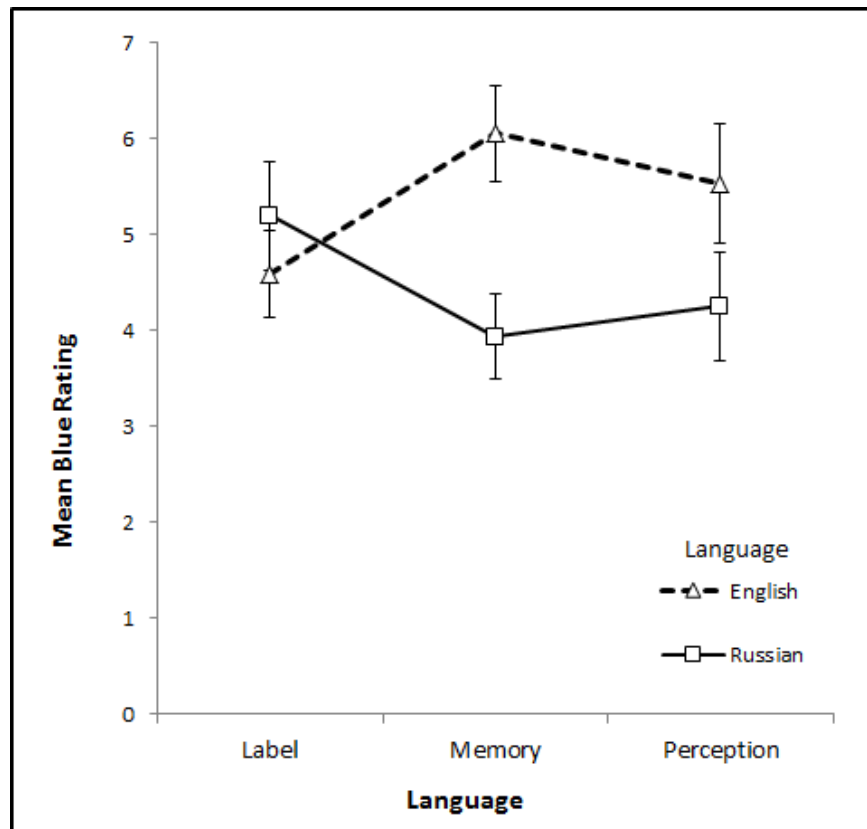


Figure 3. *Interaction graph of pilot study. Error bars represent standard errors.*

A second problem with the pilot study applies only to the label condition. Participants were instructed to “describe the color of the eye.” Such instructions prompted both language groups to be as accurate as possible. In fact, both language groups described the eyes similarly. Terms like *bluish-grey* and *light-bluish-grey* were given. I searched the Russian National Corpus and American Corpus for such terms, and there were very few results (less than 30 for each language). It would seem that asking participants in the label condition what category the eye belonged to rather than having them describe the color may have yielded more language-specific results.

The pilot study was informative. It provided novel evidence of the linguistic relativity hypothesis: there was an overall main effect of language. However, the color scale and instructions for the label group were flawed. These were corrected in the present study.

Objectives and Hypotheses of This Thesis

This study followed the same format as my pilot study, but it corrected its problems and included more stimuli to increase reliability and control for variables. I hypothesized that there would be an overall main effect of language on eye color ratings relating to how each language categorized eyes, which would provide evidence for the linguistic relativity hypothesis. Russian speakers were expected to rate the blue/grey eyes as being greyer on the color scale than would English speakers. I also expected Russian-speakers to be most different from English-speakers in the label and memory conditions and least different in the perception condition. I predicted this because color perception has a strong universal biological component, whereas color memory may rely more on how one categorizes objects. I did not believe there would be differences in the way Russian and English speakers rated colors of non-eye objects (e.g., tiles). The results of this ecologically valid study should help researchers better understand whether languages influence

how people perceive and remember the world. If the hypotheses are supported by the data, then it is plausible that language does influence memory and thought.

Chapter Two

Methods

Participants

Fifty native Russian-speaking participants from St. Petersburg, Russia, and 50 native English-speaking participants living in central Florida participated in this experiment. All participants were screened by self-report prior to inclusion in the study for having normal color vision.

Russian-speaking participants were recruited through snowball sampling of friends and family during May 2014. The mean age of the Russian speaking participants was 29.80 years ($SD=10.35$). 32 of these Russian speaking participants were women and 18 were men. Ages ranged from 18 to 60 years old.

English-speaking participants were recruited either through snowball sampling or through the University of South Florida Student Psychology Participant Pool from May 2014–October 2014. The mean age of the English speaking participants was 25.58 ($SD=10.83$). Ages ranged from 18 to 60 years old. 29 of the English speaking participants were women and 21 were men. Participants recruited through the USF participant pool were given one point of extra credit that they could apply to one of their psychology courses.

Materials

Stimuli. 32 pictures were presented to participants. 16 were pictures of various eyes. The eyes in each of the pictures were centered and zoomed in on. Of the 16 eye pictures, 10 were

blue (or grey) eyes, and six were brown eyes. The six brown eyes were included as foils to mask the intent of the study. See Figure 4 for an example of an eye picture used in the experiment.

The remaining 16 pictures were of tiles. Each picture contained nine tiles, eight of which were white that surrounded a colored tile in the center. The color of each center tile was set to match the color of one of the eye images as follows: The color of the center tile was calculated by sampling 10 pixels from its corresponding eye and averaging them together. I chose each pixel at random using PicPick software. The 10 tiles associated with blue eyes were used as a control for how participants perceive and remember color. The six tiles associated with brown eyes served to mask the intent of the study. See Figure 4 for an example of a tile used in the experiment.

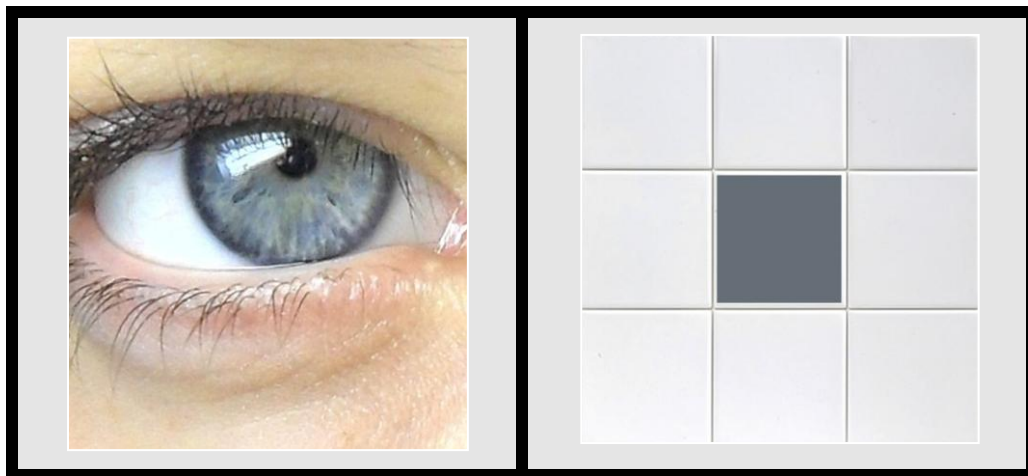


Figure 4. *An example of an eye used in the study on the left, and its corresponding tile used as a control on the right*

Response Scale. Participants were required to respond using a color scale comprised of a grid of color tiles. Two different response scale grids were presented to determine how participants perceived and remembered the color of the stimuli based on whether the eyes (and corresponding tiles) were blue/grey or brown. Each response scale consisted of 121 colors arranged in an 11x11 grid. 10 pixels from each picture were sampled from each eye and their

CIELAB values were determined. CIELAB is a color space that defines color based on three values: L, A and B. L refers to luminance, A corresponds with red and green (negative values indicate green, while positive values indicate red) and B corresponds to blue and yellow (negative values indicate blue while positive values indicate yellow) (Kuehni, 2003).

Once each pixel's color was defined in CIELAB color space for each picture, averages were computed. The average CIELAB color space values of the blue eyed pictures were then used as the center of the 11x11 color scale for the blue/grey eyes (and their corresponding tiles). The same procedure was done for the brown eyes, except samples of pixels from brown eyes were taken.

CIELAB is an approximately uniform color space, meaning that uniform increases in the L, A or B components correspond to uniform changes in perceived color ("CIE 1976 L*a*b* colour space"). Thus, in order to make intervals on the blue scale relatively equal, the B or L components were increased or decreased from the center. Colors that were below the center of the blue eye color scale increased in the B component by a fixed amount, increasing blueness. Colors above the center decreased in the B component by fixed amount, making shades of grey. Colors to the right of the center increased in the L component, and colors to the left of the center decreased in the L component. The brown scale was created in a similar fashion, except both the A and B components changed along the vertical dimension to create shades of brown. See Figure 5 for an example of both scales.

Apparatus. A 15.6-inch Dell 1564 laptop with an LED screen was used to present stimuli and response scales. Participants recorded their responses using a mouse and the laptop's keyboard. Stimuli and response scales were presented on OpenSesame software (Mathôt &

Theeuwes, 2012). The laptop was always plugged in to an outlet, and the same display settings were used for each participant.

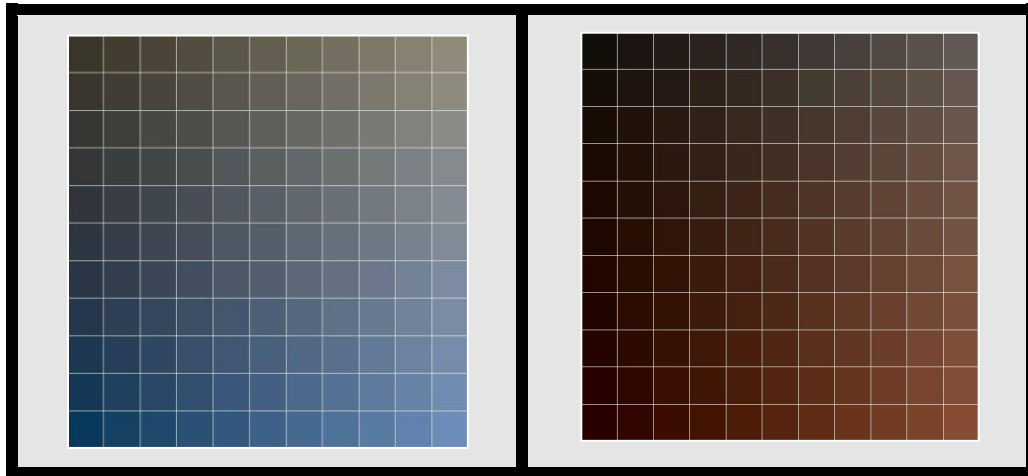


Figure 5. An example of the blue color scale (R) and brown color scale (L).

Procedure

Participants were randomly assigned to one of three conditions: a perception condition (n=32), memory condition (n=34) and label condition (n=34). I tested each Russian participant individually in St. Petersburg, Russia in various locations. I tested each Russian speaker in Russian. English speakers were tested in a lab setting at the University of South Florida or in his or her home, and each was tested in English. Participants first read instructions on the computer in either Russian or English, depending on one's native language. The stimuli and response scales were presented in a dark room after I was sure they understood the procedure. Stimuli were randomized and presented to participants in the same order each time. Participants were asked to match the color of part of a stimulus (e.g., the iris of the eye for the eye pictures or middle color of the tile pictures) to the picture's corresponding color scale. The color scale was presented either with the stimulus or after it depending on condition.

In the perception condition, participants saw each picture and color scale simultaneously and were asked to indicate which color on the scale best represented the color of the eye or tile. Pictures of the eyes and tiles were presented on the left half of the screen and color scales were presented on the right. In order to pick a color, participants clicked on the color scale with the mouse. They had as much time as they wanted to respond.

Participants in the memory condition looked first at the stimuli. Stimuli were presented in the middle of the screen. They had as much time as they needed to memorize the color of the stimuli. Participants then pressed the space bar with their left hand and a blank screen was presented for 2500 milliseconds. They then saw the response scale and were asked to indicate what color on it best represented the previous stimulus. In order to pick a color, participants clicked on the color on the color scale with the mouse. The response scale was also centered on the screen. They had as much time as they needed to respond.

The label group followed the same procedure as the memory condition, except the blank screen that was presented between the stimulus and color scale was replaced with a screen with the phrase “Color Category?” centered in the middle that lasted 2500ms. Participants viewed the eye or tile and pressed the space bar. After the eye/tile picture disappeared, participants orally responded by saying what color category the eye or color in the middle tile belonged to. Participants were told to say whatever first came to mind. Oral responses were recorded, and participants then matched the color of the eye/tile to the scale from memory.

Chapter Three

Results

Descriptive Analyses

Data were calculated by averaging each participant's responses on the blue color scale for the 10 blue/grey eyes and their corresponding tiles. Because I was interested in the blue–grey dimension, only the vertical dimension of the blue color scale (see Figure 5) was used. Colors on the top row were coded 1, colors on the row immediately below it were coded 2, and so forth through 11. Therefore, if a participant matched an eye or tile to a color in the top row, the score for that item would be one. All in all, there were 20 data points for each participant (10 data points for the eyes and 10 for their corresponding tiles). 18 pairs of data points out of the 2000 total data points were removed due to participant error (e.g., clicking or pressing the space bar too quickly resulting in not seeing a stimulus). Overall, internal reliability of the 10 individual responses for the blue eyes was good, $\alpha = 0.84$. Having good internal reliability indicates that the mean of the ten items is a good measure of how participants were rating the eyes in general. Means, standard deviations and other descriptive statistics are given in Table 4 on page 30.

Responses for how participants in the label condition categorized the eyes were recorded. Three responses from each language group were discarded due to participant error (e.g., clicking or pressing the space bar too quickly resulting in not seeing a stimulus). Responses were coded as *blue*, *grey* or *other*. All *blue* or *grey* responses contained only one color term (i.e., *grey* but not *blue-grey*), but they may have been paired with an adjective describing lightness (e.g., *light-blue* or *dark-grey*). For Russian participants *goluboy* and *siniy* were both coded as blue because they

both fall under the term *blue* in English. Responses coded as *other* either contained one color term unrelated to blue or grey (e.g., *green*) or a combination of color terms (e.g., *bluish-green*). 89 percent of English-speaking responses were *blue*, 9 percent were *grey* and 1 percent were *other*. 50 percent of Russian responses were *blue*, 49 percent were *grey* and less than one percent were *other*. It should be noted that most Russian blue labels were the Russian word *goluboy*, meaning *light blue*.

Demographic Analyses

Demographic information on gender and age were collected from participants. To assess whether the proportion of males and females differed across languages, a 2 (language) x 2 (gender) chi square analysis was used, and no significant difference was found, $\chi^2(1, N=100)=0.38, p=.54$. Two 3 (condition) x 2 (language) chi square analyses were also used to determine whether the proportion of males and females was different across each level of the independent variables. The number of males did not significantly differ across conditions, $\chi^2(2, N=100) = 0.11, p=.94$, and neither did the number of females, $\chi^2(2, N=100) = 0.13, p=.94$. A 3 (condition) x 2 (language) ANOVA was used to analyze age differences across conditions. The main effect of language approached significance, $F(1,94) = 3.81, p=.05$. English speakers ($M=25.58, SD=10.83$) were younger than Russian speakers ($M=29.80$ years, $SD=10.35$). However, the correlation between age and mean blue tile rating was not significant, $r(98)=-0.10, p=.34$. There was no main effect of condition, $F(2,94) = 0.43, p=.65$, and there was no interaction, $F(2,94) = 0.79, p=.50$.

Tests of Predictions

In order to determine whether English speakers were more likely to describe an eye as blue than Russian speakers, the total number of blue responses was summed for each participant.

Table 4.
Means and Other Descriptive Statistics for Eyes and Tiles.
Higher means indicate more blue, while lower numbers indicate more grey.

Stimulus	Condition	English				Russian				Total			
		<i>n</i>	<i>M</i>	<i>SD</i>	Kurt	<i>n</i>	<i>M</i>	<i>SD</i>	Kurt	<i>n</i>	<i>M</i>	<i>SD</i>	Kurt
Eye	Perception	16	7.62	0.98	0.50	16	7.64	1.05	-1.53	32	7.62	1.00	-0.50
	Memory	17	7.13	1.28	-0.39	17	6.42	1.99	-0.42	34	6.78	1.69	0.43
	Label	17	6.95	1.38	-0.86	17	6.33	1.62	0.06	34	6.64	1.51	-0.20
	Total	50	7.23	1.24	0.04	50	6.78	1.68	0.19	100	7.00	1.49	0.64
Tile	Perception	16	6.06	0.75	0.30	16	6.05	0.83	-0.40	32	6.06	0.78	0.11
	Memory	17	5.46	0.66	-0.97	17	5.68	0.90	-0.59	34	5.57	0.79	0.13
	Label	17	5.22	1.08	0.58	17	5.9	1.11	-0.84	34	5.56	1.13	-0.25
	Total	50	5.57	0.91	0.70	50	5.87	0.95	-0.39	100	5.72	0.94	0.19

Overall, Russian speakers used blue as a label less ($M=5.00$, $SD=3.30$) than did English speakers ($M=8.71$, $SD=.91$). An independent samples t-test where language group was the independent variable (equal variance was not assumed) was used to analyze these data, and a significant difference for how the language groups categorized eyes was found, $t(18.92)=4.44$, $p<.001$.

In order to assess whether Russian and English speakers perceive or remember shades of grey or blue differently in general, a 3 (condition) x 2 (language) factorial ANOVA was used to analyze mean blueness ratings of the 10 tiles. There was no main effect of language, $F(1,94)=2.65$, $p=0.11$, nor was there an interaction, $F(2,94)=1.25$, $p=0.29$. There was a main effect of condition, $F(2,94)=3.19$, $p=0.04$. A Tukey HSD test was used to analyze this main effect. The largest mean difference was between the label and perception conditions. Participants in the perception condition rated the tiles as more blue ($M=6.06$, $SD = 0.78$) than did participants in the label condition ($M=5.56$, $SD = 0.94$), and the difference approached significance ($p=.07$). Refer to Figure 6 on page 35 for a graph of mean tile ratings.

A 3 (condition) x 2 (language) ANCOVA with participants' mean blue tile ratings of the participants as the covariate was used to analyze blueness ratings for the 10 blue/grey eye pictures. There were two reasons an ANCOVA was chosen. First, it increased power. It controlled for how participants perceived and remembered eye color by removing error variance for how they perceived and remembered color in general. Second, it more directly addressed the predictions than a 3 (condition) x 2 (language) x 2 (type of picture) repeated measures ANOVA. Overall, there was a main effect of language, $F(1,93)=4.71$, $p=0.03$, $\eta_p^2=.05$, $d=.44$. There was no main effect for condition, $F(2,93)=2.47$, $p=0.09$, $\eta_p^2=.05$, nor was there an interaction $F(2,93)=2.38$, $p=0.28$, $\eta_p^2=.03$. The ANCOVA can then be used to find adjusted cell means. Adjusted cell means are estimates of each group's blue eye mean if all participants had the same

mean blue tile rating. The adjusted means are shown in Figure 8 on page 37 (the means are evaluated at a mean blue tile rating of 5.72; they are estimates of what the mean of each condition would be if each participant rated the blue tiles as 5.72 on average). It should be noted that all the assumptions of ANCOVA were met. First, the covariate was correlated with the dependent variable ($r=.34, p<.01$). Second, the slopes did not significantly differ between conditions (see Figure 7 on page 36). This was tested by running a customized ANCOVA model examining all possible interactions between the covariate and dependent variables. No interactions differed between any combination of independent variables and the covariate (see Table 5 on page 36). Third, homogeneity of variance was not violated; a Levene's test of equality of variances was not significant, $F(5,94)=1.25, p=0.29$.

Exploratory Analyses

The data were input into a Bayesian model that used Gibbs sampling based on work by Kruschke (2010). Bayesian analyses have several advantages over traditional analytic methods. First, they calculate the probability of the model based on the data and the model's assumptions rather than the probability of obtaining data given the hypothesis. In other words, instead of testing the validity of the null hypothesis, they calculate how probable values of the parameter of interest are based on the data. Second, they are not dependent on how many statistical analyses are planned. In traditional null hypothesis significance testing, the family-wise error rate increases as the number of tests increase. This problem led to the development of post-hoc analyses (e.g., Tukey HSD). In Bayesian analyses, the inclusion of prior information and shrinkage of parameter estimates through the process of Gibbs sampling mitigates this problem. For example, the ANCOVA used to analyze the data obtained in this study yielded a significant main effect of language. It did not reveal a main effect of condition or interaction. It is not

recommended to do pairwise comparisons with null hypothesis significance testing when there are non-significant effects. This is not true of Bayesian analyses, in which one examines all data available because error rates are not inflated.

The Bayesian model used was analogous to ANCOVA and the general linear model insofar as it looked at how predictor variables influenced a dependent variable while controlling for a covariate. It was hierarchical and examined how each independent variable and combination of independent variables changed the grand mean of the data. The model assumed that an individual score, nested within a condition or combination of conditions, came from a normal distribution and standard deviation. The mean blue score of the eye pictures (y) of any level of any combination of the independent variables was allowed to vary based on the following model: $y = \beta_0 + \vec{\beta}_1\vec{x}_1 + \vec{\beta}_2\vec{x}_2 + \vec{\beta}_{1x2}\vec{x}_{1x2} + \vec{\beta}_3\vec{x}_3$, where \vec{x}_1 and \vec{x}_2 are the nominal predictors (language and condition respectively), \vec{x}_{1x2} is the interaction between those predictors and \vec{x}_3 is the covariate (mean blue tile score). β_0 is the grand mean, and $\vec{\beta}_1$, $\vec{\beta}_2$, $\vec{\beta}_{1x2}$ and $\vec{\beta}_3$ are deflections of the independent variables (i.e., estimates of how much an independent variable changes the grand mean β_0).

This Bayesian model required priors for each deflection (i.e., a prior belief as to how the deflections affect the grand mean). The priors on the deflections were based on normal distributions with a mean of zero. The standard deviations of those normal distributions were taken from gamma distributions where the shape and rate parameters were estimated based on a hierarchal prior (a gamma distribution with shape and rate parameters equal to one). In other words, the priors indicate a belief that there are no main or interaction effects. In order to estimate the values of the deflections, the model combines the probabilities provided by the priors and the probability of the actual data (i.e., the likelihood) to come up with a posterior

distribution for each level of each deflection. If most of the density (95%) of a posterior distribution (i.e., the highest density interval or HDI) of a deflection at any level of an independent variable does not include zero, then the deflection is deemed credible. It can be concluded that the corresponding independent variable had an effect on the dependent variable. It should be noted that, unlike null hypothesis significance testing, homogeneity of variance is not assumed (and didn't need to be). Overall, the grand mean was estimated to be 7.02. There was a credible deflection of the perception condition of 0.41, and a credible deflection due to language of 0.27 for English and -0.27 for Russian (see Figure 9 on page 37).

The above model was then used to estimate pairwise differences to see if the means of conditions are credibly different from each other. No adjustments needed to be made that would change power (i.e., Bonferroni corrections, Tukey HSD etc) because Bayesian analyses mitigate false alarms by incorporating prior beliefs. As a consequence there is shrinkage of parameter estimates (e.g., deflection estimates). In this model, it means the deflection estimates (and therefore mean differences) are closer to zero than they otherwise would have been had there been no priors (Kruschke, 2010). With this in mind, I was able to estimate the difference between means of the language groups (see Figure 10 on page 38), conditions (see Figure 11) and all combinations of language and condition (see Figure 12). A difference between means is significant if the 95% highest density interval does not include zero. According to the analysis, several means were credibly different. Russian-speakers rated the eyes 0.55 points greyer than English-speakers. Participants in the perception condition rated eyes 0.77 points bluer than the label condition. Russian-speakers in the label condition rated the eyes 0.94 points greyer than English-speakers in the memory condition, 1.16 points greyer than English speakers in the perception condition, and 1.14 points greyer than other Russian-speakers in the perception

condition. Finally, Russian speakers in the memory condition rated the eyes 0.97 and 0.95 points greyer than English and Russian speakers respectively in the perception condition. Within the label condition, there was almost a credible difference between Russian and English speakers (Russian-speakers rated the eyes 0.87 points greyer). There was virtually no difference between language groups in the perception condition

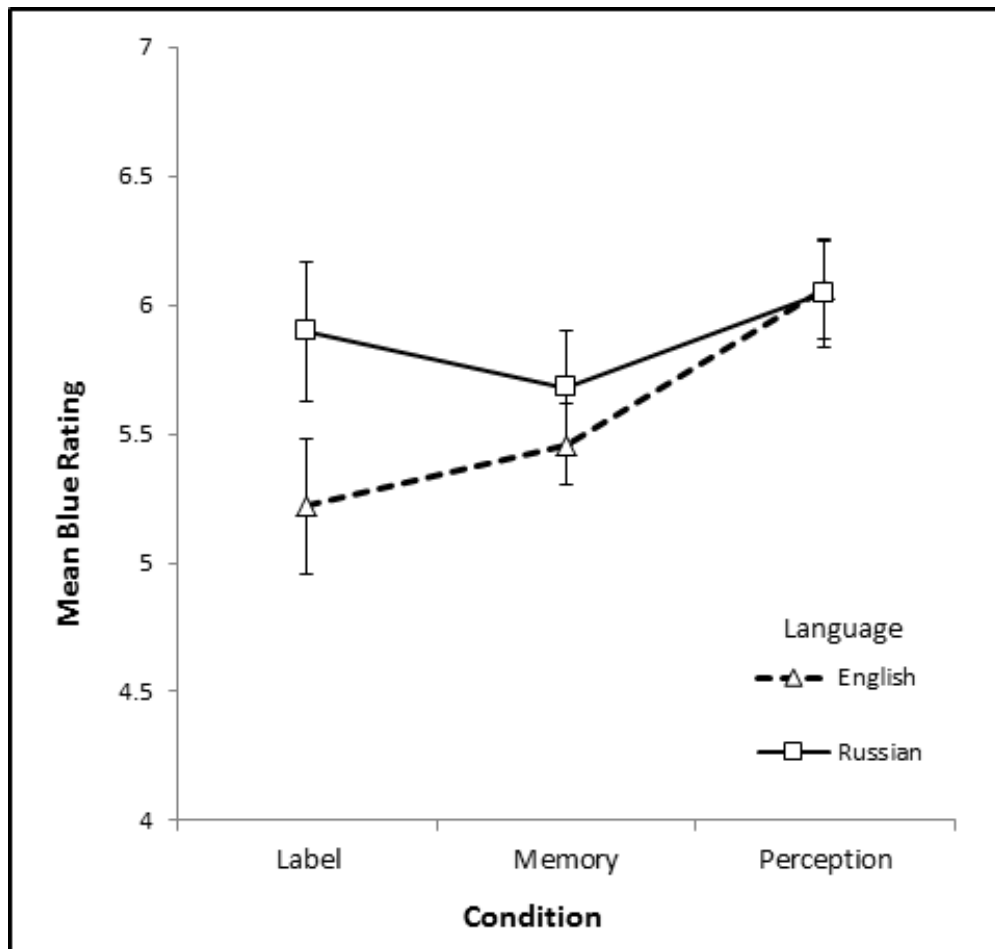


Figure 6. Means for blue tile ratings across conditions. Error bars represent standard errors

Table 5.

Model testing the relationship between the covariate (mean blue tile score) and dependent variable (mean blue eye score) at each level of the independent variables.

Source	SS	df	MS	F	p
Condition	1.74	2	.88	.46	.63
Language	2.39	1	2.39	1.28	.26
Mean Tile Score	19.25	1	19.25	10.29	<.01
Condition x Mean Tile Score	1.14	2	.57	.31	.74
Language x Mean Tile Score	1.19	1	1.19	.64	.43
Condition x Language x Mean Tile Score	3.23	2	1.62	.86	.43
Error	168.41	90	1.87		

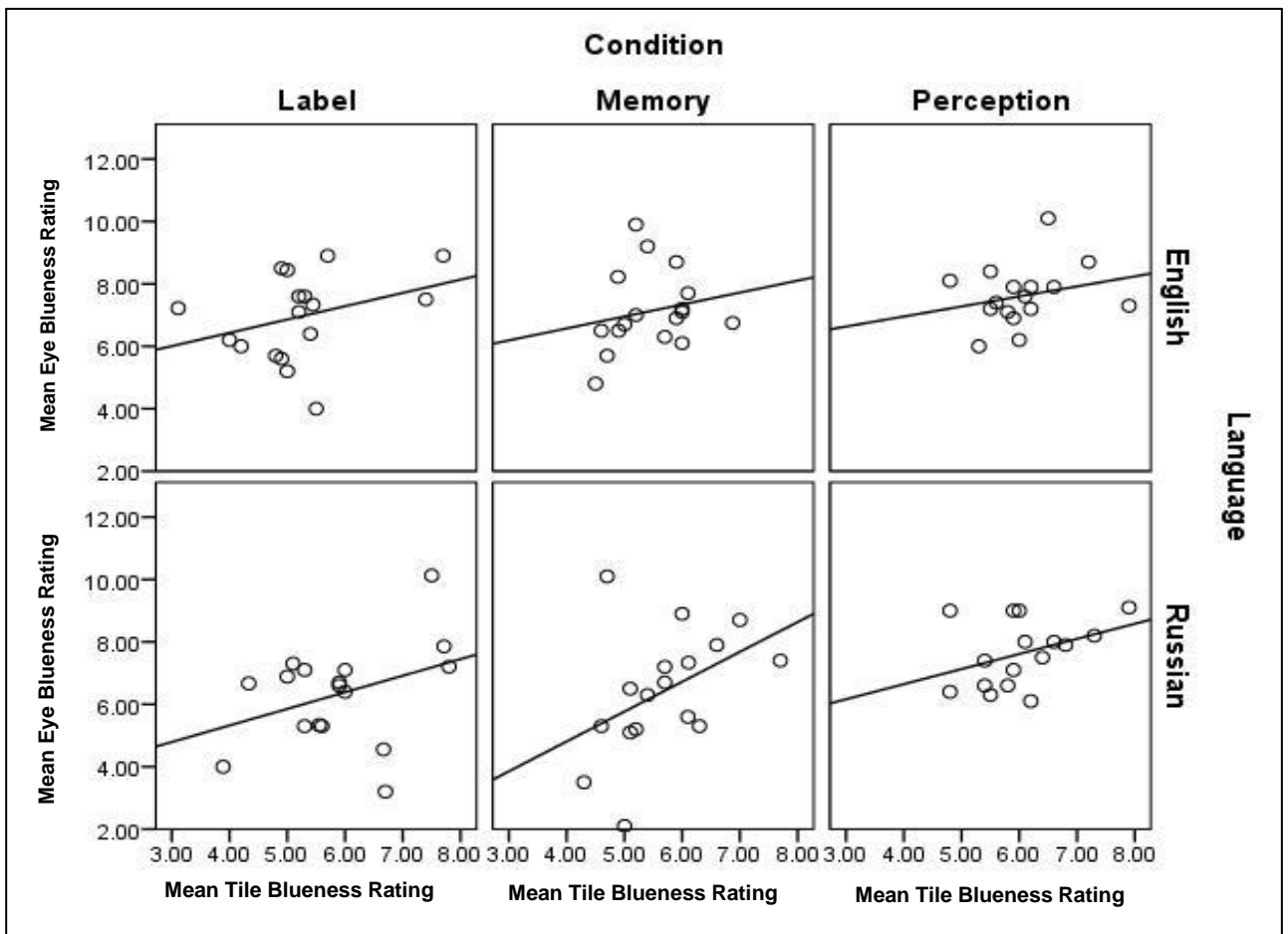


Figure 7. *Regression slopes of the dependent variable and covariate at all levels.*

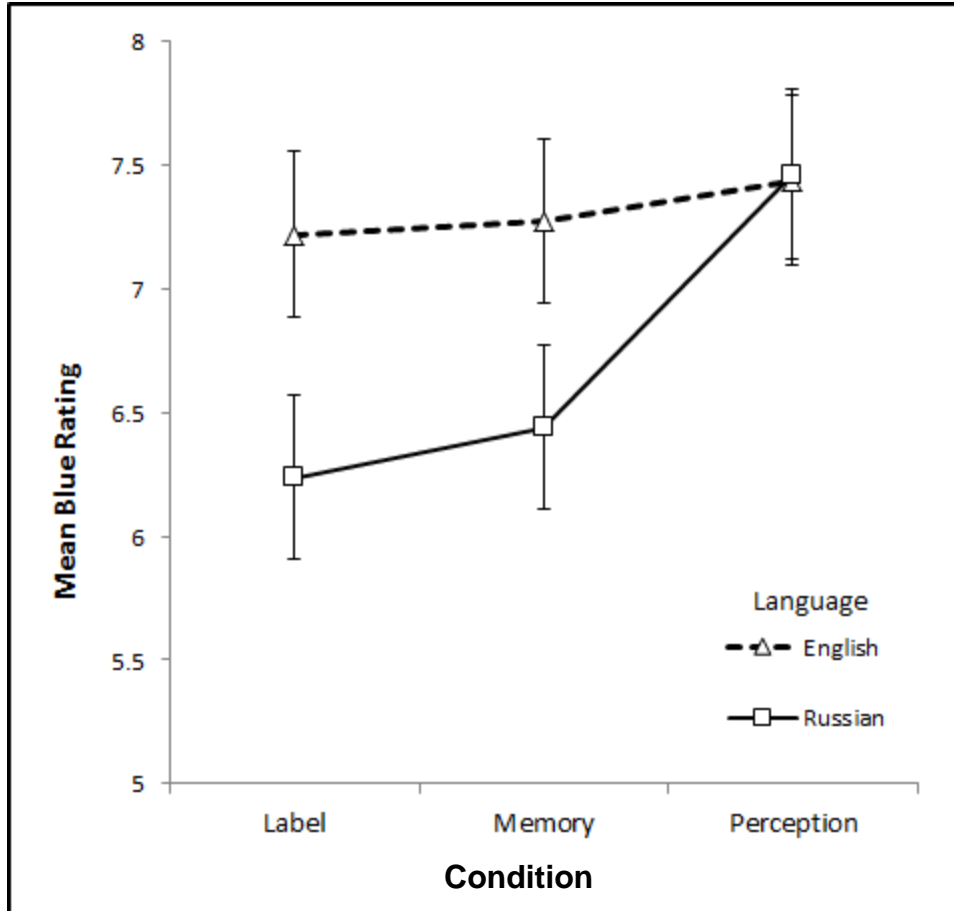


Figure 8. Adjusted mean estimates from the ANCOVA model. Error bars represent standard errors

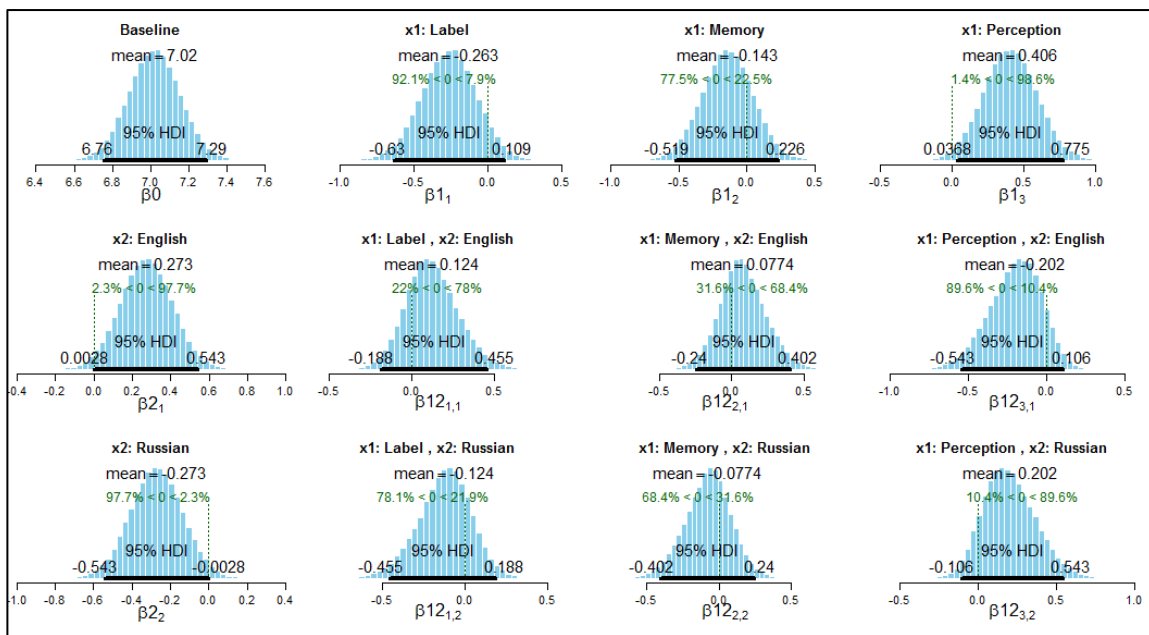


Figure 9. Estimates of deflections at each level of the independent variable.

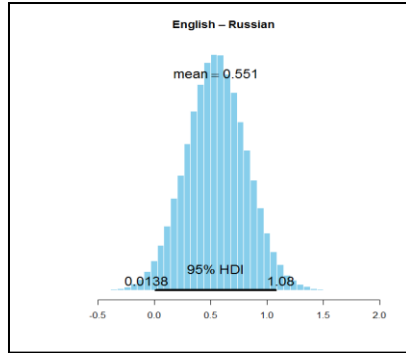


Figure 10. Mean difference of blue eye ratings between Russian and English speakers

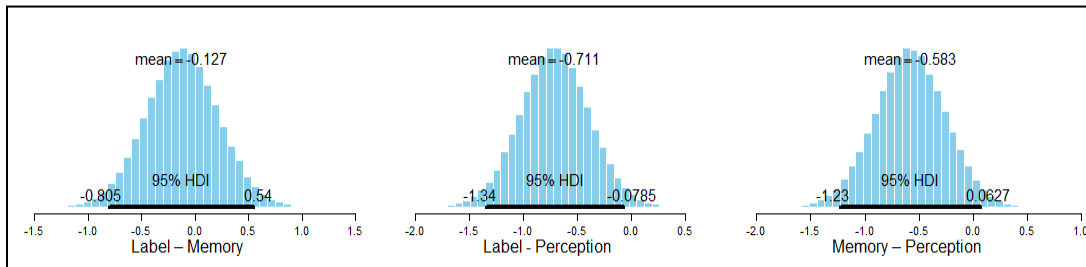


Figure 11. Mean differences of blue eye ratings between conditions

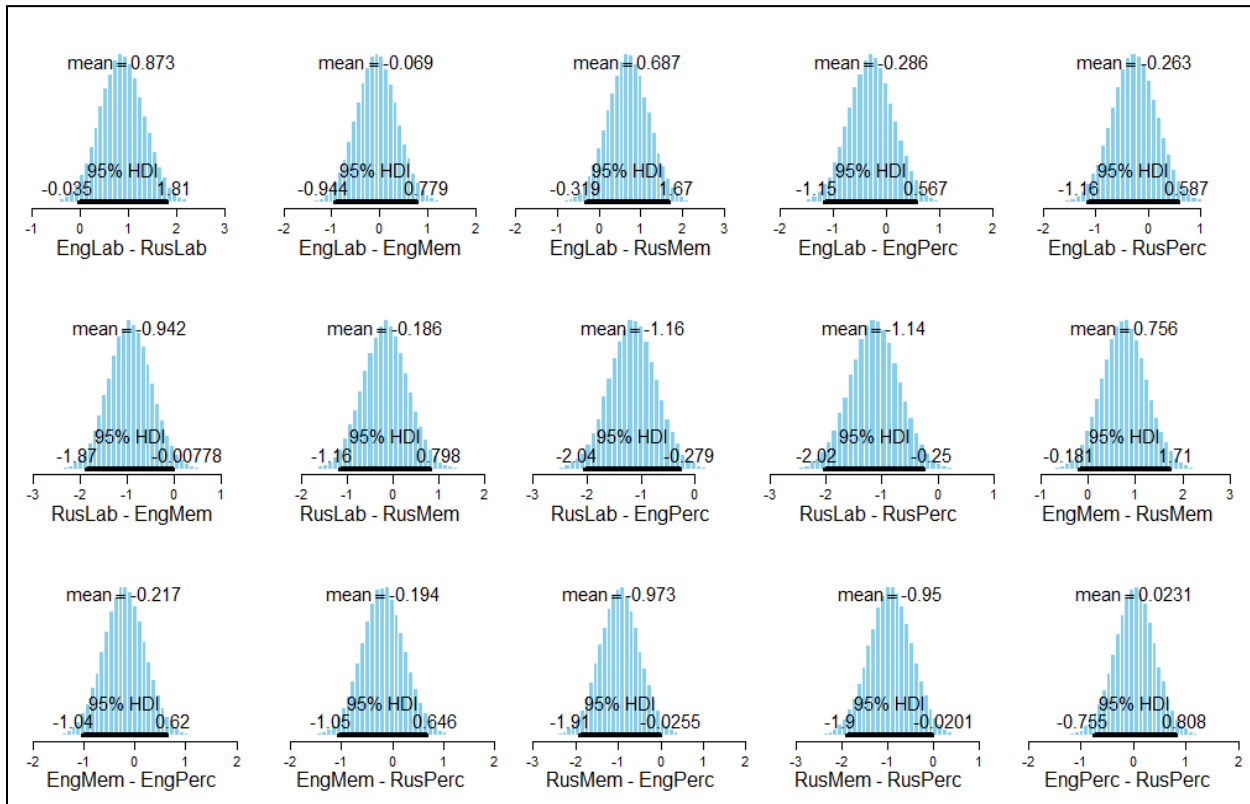


Figure 12. Mean differences of blue eye ratings for all levels of language and condition.

Chapter Four

Discussion

The overall goal of the study was to examine in an ecologically valid way how language influences thought. This was done by taking an object (i.e., a blue eye) that is described differently in two languages and seeing how that description affected participants' perception and memory. It was hypothesized that Russian-speaking participants would rate blue/grey eyes as being greyer than would English-speaking participants. If Russian speaking participants did so, then the linguistic relativity hypothesis would be supported. It was also hypothesized that the smallest difference between the English and Russian-speakers would be in the perception condition because color perception has such a strong universal biological component.

The underlying assumption of the study was that English speakers are more likely to describe an eye as blue, whereas Russian speakers are more likely to describe the same eye as grey. The data suggest this is the case. Overall, Russian speakers were less likely to label the eyes as blue. Participants were viewing the same stimuli, yet eye labeling was different between the two languages.

The results also suggest that there are no major differences in how Russian and English speakers perceive and remember shades of grey and blue when not associated with an eye. The fact that there was no main effect of language or an interaction between language and condition for mean blue tile ratings supports this conclusion. This also means that the results are not confounded by how the two language groups perceive and remember color in general.

Additionally, the fact that each tile's color was based on an eye's mean color suggests that colors found specifically in eyes don't affect eye ratings.

Interpreting how participants perceived and remembered eye color is more complicated. There are two approaches that are valid. The first is to analyze results from the ANCOVA, and the second is to analyze the Bayesian results.

As hypothesized, Russian speakers generally rated the 10 eyes as greyer than did English speakers. Based on the ANCOVA, there was only a main effect of language. The effect size was moderate. There was no main effect of condition or interaction between condition and language. Because there was only a main effect of language, I could conclude that language does influence thought, both perceptually and through memory for specific objects (e.g., eyes). In other words, one may be tempted to conclude that languages that categorize objects using different colors actually see the colors of those objects differently. This is based on the fact that there was no main effect of language for the tiles but there was for the eyes. Such a conclusion would be interesting if true, but it seems precarious at best. For example, one could argue that the study was underpowered, and instead point to the fact that there was virtually no difference between Russian and English speakers in the perception condition in how blue or grey they rated the eyes. A study with more participants may yield a significant interaction. Because of this possibility, the Bayesian analysis was conducted.

The Bayesian analysis offered more insight. In fact, it may be a more valid way of analyzing the data because it does not estimate the probability of the data given the hypothesis, but rather tests the hypothesis given the data. In addition, many of the assumptions needed in null hypothesis testing do not apply. Consequently, a more nuanced picture emerged by using the Bayesian analysis to calculate all possible pairwise differences. The results suggest there was

indeed a main effect of language. Russian speakers rated the blue/grey eyes as greyer than did English speakers. This finding supports the linguistic relativity hypothesis generally. The data also suggest that there was an interaction because there was no difference between Russian speakers and English speakers in the perception condition. This second finding supports universalism.

A more in-depth examination of mean differences between conditions among each language group makes interpreting the results more complicated. As expected, Russian-speaking participants rated the eyes greyer in the label and memory conditions than in the perception condition. This finding suggests that the way language codes a stimulus (i.e., the color of an eye) affects how the stimulus is recalled. In contrast, there was no difference among English speakers in how they rated the eyes across conditions. The pattern observed among English speakers is intriguing, and there are several conclusions that can be made, some more valid than others. First, it may be evidence that language does not influence thought. Because Russian speakers in the memory and label conditions remembered eyes as greyer than did Russian speakers in the perception condition, one might expect the English speakers to rate the eyes as more blue in the memory and label conditions than in the perception condition. Because this was not the case, one may be tempted to conclude that the differences found among Russians happened by chance, and that language and thought are unrelated. However, the overall main effect of language supports the linguistic relativity hypothesis, making this first possibility unlikely. Second, it may indicate that it was incorrect to conclude from the Bayesian analysis that there was an interaction. Perhaps there is only a main effect of language, and similarities between the Russian and English participants in the perception condition happened by chance. Third, there could be a ceiling effect because the bluest colors on the color scale were not believable to participants as possible

eye colors, regardless of language. If there was such a ceiling effect, then English speakers were limited by the functionally truncated scale, and therefore couldn't rate the eyes any bluer.

What general conclusions can then be made from the data? The first is that Russian and English speakers differ in how they describe blue/grey eyes. English-speakers tend to use blue terms, whereas Russian speakers use both blue and grey terms. The second is that Russians tend to rate blue/grey eyes greyer than do English-speaking participants. This is supported by both the ANCOVA and Bayesian model. There is disagreement in the models as to whether there is an interaction. The ANCOVA suggests there is not, whereas the Bayesian analysis indicates that there may be one. If there is an interaction, then there is evidence for both universalism and the linguistic relativity hypothesis in terms of color perception and color memory. Thus, language may have more influence in how colors of objects are retrieved from memory than how they are actually perceived. This difference could be explained by the fact that language directs attention to certain characteristics about a stimulus (e.g., the grey or blue colors in an eye). Russians may focus more on the grey colors in an eye than English speakers do. This makes Russian-speakers more likely to remember grey colors from the eye. However, when participants from both language groups are given the opportunity to match a color to an eye at the same time as they perceive the eye, input from the cones in the retina is the deciding factor when participants make their choice.

There are several limitations to this study. The first and foremost is that language is a grouping variable, not a true independent variable. Several uncontrolled factors could have influenced how the Russian participants remembered the eyes. For example it could be that Russians answer surveys differently than Americans because of cultural differences. However, language is a more parsimonious answer. In addition, the ANOVA testing how the independent

variables affect the tile ratings showed that Russian and American participants did not differ in their responses. If confounds affected how the language groups are reporting color, one would expect them to affect all stimuli equally.

The second limitation is sample size. Originally, I had aimed for 200 total participants. I found as many Russian participants as was feasible given the resources and time available. Nevertheless, more participants would have increased power. Because more data increase confidence in one's results, the fact that there were only 16-17 people in each condition per language group decreases certainty. Thus, it is possible that the main effect of language is spurious, or it is possible that the interaction would have been significant in the ANCOVA model. Further follow up studies are needed to see if the results are replicable.

The third main limitation is the fact that the laptop used to show stimuli was not ideal. The LED display changes contrast depending on the angle viewed. However, I was careful to monitor participants and make sure the viewing angle was similar and they were seated the same distance away from the monitor. Moreover, the LED display is not a confound. All participants were tested in a dark room on the same computer using the same settings. Any complications resulting from the display would have affected all participants. It is a limitation because it may have increased error variance, resulting in less sensitivity, especially in finding a significant interaction.

The last limitation is related to the tile stimuli. First, in retrospect, it would have been better to have sampled more pixels from each of the eyes in order to create a better average color for the tiles. This would also have increased the accuracy of each eye's estimated average color, made the scale more accurately reflect the average color of the eyes and may have increased the correlation between the eyes and tiles. Second, the tile stimuli were solid colors, whereas the

irises were not. There are multiple colors in each eye. Further research is needed to assess how Russian and English-speakers process multi-colored stimuli. It could be that Russian and English speakers remember multi-colored stimuli differently. Thus, the main effect found would not have been due to language, but would have been due to differences in how the two cultures perceive and remember multi-colored objects.

Conclusion

The legitimacy of the linguistic relativity hypothesis has been debated in psychology for decades. On the one hand, many cognitive scientists have argued against it (Berlin & Kay, 1969; Pinker, 1995). They have contended that cognitive processes, such as color perception, are unaffected by one's language. On the other hand, recent research suggests that color categories may affect color recognition (Goldstein, Davidoff & Roberson, 2009; Kwok et al., 2011; Winawer et al., 2007). The present study gives novel, ecologically valid evidence for the linguistic relativity hypothesis. The overall main effect of language, supported by both the ANCOVA and Bayesian analysis, suggests that language affects thought. A plausible explanation for why this happens is that language directs people to pay attention to certain characteristics of stimuli. Since Russian speakers are more likely than English speakers to describe eyes using grey terms, they may focus more on grey colors in eyes than English speakers do. This may make them more likely to remember eyes as being greyer than English speakers. Conversely, English speakers may focus on blue colors in eyes more than Russian speakers do. If language does direct attention, then it is not surprising that a main effect of language was found.

The Bayesian analysis offers further insight. Those results support the ideas of universalism because there was virtually no difference in how Russian and English speakers

rated eyes in the perception condition. This is what one would expect given the strong biological component of color perception. It would also suggest that there is a limit to how much language can affect thought. Language most affected how the color of the eyes was remembered and not how they were perceived.

In conclusion, this study suggests that there may be a compromise between universalism and the linguistic relativity hypothesis. Both ideas consider how thought and language interact. Universalism claims that cognition is independent of language. The linguistic relativity hypothesis states that thought and language are closely linked, and that the language one speaks influences how one thinks. Based on results from this study, the truth may lie somewhere in the middle. In terms of color, language influences memory the most and perception the least. Thus, blue may be in the eye of Russian speakers, but not their memory.

References

- Berlin, B., & Kay, P. (1969). *Basic color terms; their universality and evolution*. Berkeley, CA: University of California Press.
- Boaz, F. (1911). *Handbook of American Indian languages*. Washington, DC: Government Printing Office.
- Bornstein, M. H., Kessen, W., & Weiskopf, S. (1976). Color vision and hue categorization in young human infants. *Journal of Experimental Psychology: Human Perception and Performance*, 2(1), 115-129.
- Bunak, V.V.(1965). *Proiskhozhdeniye i Etnicheskaya Istoria Russkovo Naroda po Antropologicheskim Dannym* [The origin and ethnic history of Russian people according to Anthropological Data]. Moscow, Russia: Nauka.
- CIE 1976 L*a*b* colour space. (n.d.). 17-157 CIE 1976 L*a*b* colour space. Retrieved March 6, 2014, from <http://eilv.cie.co.at/term/157>
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: M.I.T. Press.
- Conklin H. (1955). Hanunoo color categories. *Southwestern Journal of Anthropology*, 11, 339-344.
- Cornsweet, T. (2012). *Visual perception*. Burlington, MA: Elsevier Science.
- Davenport, G.C., & Davenport, C.B. (1907) Heredity of eye-color in man. *Science*, 26, 590–592.
- Davies, M. (2008-2014). *The corpus of contemporary American English: 450 million words, 1990-present*. Available online at <http://corpus.byu.edu/coca/>.
- Ebner, M. (2007). *Color constancy*. Chichester, England: John Wiley.

- Goldstein, J., Davidoff, J., & Roberson, D. (2009). Knowing color terms enhances recognition: Further evidence from English and Himba. *Journal of Experimental Child Psychology*, *102*(2), 219-238.
- Heider, E. R. (1971). 'Focal' color areas and the development of color names. *Developmental Psychology*, *4*(3), 447-455.
- Heider, E. R. (1972). Universals in color naming and memory. *Journal of Experimental Psychology*, *93*(1), 10-20.
- Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist*, *86*, 65-79.
- Kuehni, R. G. (2012). *Color: an introduction to practice and principles* (3rd ed.). Hoboken, N.J.: Wiley.
- Kuehni, R. G. (2003). *Color space and its divisions : Color order from antiquity to the present*. Hoboken, NJ: Wiley-Interscience.
- Kwok, V., Niu, Z., Kay, P., Zhou, K., Mo, L., Jin, Z., & Tana, L. (2011). Learning new color names produces rapid increase in grey matter in the intact adult human cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(16), 6686-6688.
- Leavitt, J. H. (2010). *Linguistic relativities: language diversity and modern thought*. Cambridge, MA: Cambridge University Press.
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314-324.
doi:10.3758/s13428-011-0168-7

- Michaels, D. (1977). Linguistic relativity and colour terminology. *Language & Speech*, 20(4), 333-343.
- Pinker, S. (1995). Language acquisition. In L.R. Gleitman & M. Liberman (Eds.), *An invitation to cognitive science: Volume 1* (pp. 135-182). Cambridge, MA: MIT Press.
- RNC. (2003-2014). *Russian National Corpus*. Available online at <http://www.ruscorpora.ru/en/search-main.html>
- Spillmann, L. (1990). *Visual perception: the neurophysiological foundations*. San Diego, CA: Academic Press.
- Stone, J. V. (2012). *Vision and brain: How we perceive the world*. Cambridge, MA: MIT Press.
- Sturm, R. A., & Frudakis, T. N. (2004). Eye colour: portals into pigmentation genes and ancestry. *Trends in Genetics*, 20(8), 327-332.
- Sturm, R. A., & Larsson, M. (2009). Genetics of human iris colour and patterns. *Pigment Cell & Melanoma Research*, 22(5), 544-562. doi:10.1111/j.1755-148X.2009.00606.x
- Whorf, B., & Carroll, J. B. (1998). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. Cambridge, MA: MIT Press.
- Winawer, J., Witthoft, N., Frank, M., Wu, L., Wade, A., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences of the United States of America*, 104(19), 7780-7785.
- WotanJugend - INFO. (n.d.). VK. Retrieved March 6, 2014, from http://vk.com/wall-39340950_36403

Appendix

IRB Approval Letter



RESEARCH INTEGRITY AND COMPLIANCE
Institutional Review Boards, FWA No. 00001669
12901 Bruce B. Downs Blvd., MDC035 • Tampa, FL 33612-4799
(813) 974-5638 • FAX (813) 974-7091

6/7/2013

Mark Lowry, B.A.
Psychology
4202 East Fowler Ave.
Tampa, FL 33620

RE: **Exempt Certification**
IRB#: Pro00012758
Title: Color Perception

Study Approval Period: 6/7/2013 to 6/7/2018

Approved Items:
Protocol Document:
[Color Perception Protocol Version 1 04.17.13](#)

Consent Scripts:
[English Color Perception Script Version 1- May 13 2013](#)
[English Version of Informed Consent Given to Russian Participants Version 1- May 13 2013](#)
[Russian Color Perception Script Version 1- May 13 2013](#)

Dear Mr. Lowry:

On 6/7/2013, the Institutional Review Board (IRB) determined that your research meets USF requirements and Federal Exemption criteria as outlined in the federal regulations at 45CFR46.101(b):

(2) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless:
(i) information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation.

As the principal investigator for this study, it is your responsibility to ensure that this research is conducted as outlined in your application and consistent with the ethical principles outlined in the Belmont Report and with USF IRB policies and procedures. Please note that changes to this protocol may disqualify it from exempt status. Please note that you are responsible for notifying

the IRB prior to implementing any changes to the currently approved protocol.

The Institutional Review Board will maintain your exemption application for a period of five years from the date of this letter or for three years after a Final Progress Report is received, whichever is longer. If you wish to continue this protocol beyond five years, you will need to submit a new application at least 60 days prior to the end of your exemption approval period. Should you complete this study prior to the end of the five-year period, you must submit a request to close the study.

We appreciate your dedication to the ethical conduct of human subject research at the University of South Florida and your continued commitment to human research protections. If you have any questions regarding this matter, please call 813-974-5638.

Sincerely,

A handwritten signature in black ink, appearing to read 'Kristen Salomon', followed by a horizontal line.

Kristen Salomon, Ph.D., Vice Chairperson
USF Institutional Review Board