Triad Investigations of Color Appearance Similarity Relations in Korean-English Bilinguals

Erin H. Moon
University of California, Irvine

Author Note

I would like to thank my mentors Dr. Kimberly A. Jameson and Professor Louis Narens for providing me the opportunity to conduct research under their guidance as well as for overseeing the project and providing the necessary resources. A big “Thank You” to Tim Satalich (IMBS, UC Irvine) for providing the code for the MDS scaling procedures. Thanks also to David Bimler (Univeristy of Massey, NZ) for the loan of the color triad stimuli. This research was supported by an award from the Undergraduate Research Opportunities Program, at the University of California, Irvine. Additional support was provided by The University of California Pacific Rim Research Program, 2010-2015 (K.A. Jameson, PI) and The National Science Foundation 2014-2017 (#SMA-1416907, K.A. Jameson, PI). IRB Approval for Human Subjects Research: HS#2014-1644.

Abstract

Various research efforts had been dedicated to understand better of the possible effects that culture may have on its member's way of perceiving and grouping colors. To study this effect in a narrower context, the present study sought to observe and measure the way the individual categorized colors in a Korean-English population. There were 30 bilingual Korean-English native speaking participants that consisted of both male and female. Three tasks were given to each subject in which the first and third followed the same guidelines except in different languages. The first and third tasks consisted of color triad tasks that required the individual to identify the colored shade that he or she thought did not belong with the other two shades. The subject was then asked to rate his or her confidence level of the choice from a scale of 1 to 10. The second task was the Farnsworth-Munsell hue 100 Hue Test as standard color vision diagnostic which required participants to arrange color samples in a smooth color gradient order. The main objective of this task was to test any possible anomaly found in the participants' standard perception of color such as whether he or she could not see a certain color. Triad choice data from the two language conditions was analyzed using Multidimensional Scaling Procedures to examine the shared similarity structure observed in participants' triad choices and how they varied across the two language modes assessed. Results were interpreted to examine variation of observed similarity structure as a function of language mode. Findings were informative and raise important questions about the independence of perceptual color similarity relations from linguistic influences contributed from culturally specific language mode cognitive processing.

Keywords: triad similarity, language, categorical perception

Triad Investigations of Color Appearance Similarity Relations in Korean-English Bilinguals

1.0 Introduction

Much interest has been garnered toward the topic of color cognition and its relation to the bilingual population. One of the early assumptions on the subject had been that color terms are "isomorphic" or correspond with the colors that are perceived. As evident in the study conducted by Berlin and Kay (1969), there seemed to be "basic" color terms that are shared across different languages. However, this finding should not dismiss the variability also observed across these different semantic structures. Further research has been carried out to find a definite answer of whether language does indeed determine what the individual perceives as a particular color. Yet, findings suggest that the debate of the language effect is rather inconclusive. As also observed in the present study, it is difficult to determine whether color categorization is solely dictated by the individual's semantic structure.

Much of the direction of the research was provided by the Sapir-Whorf hypothesis. A key aspect of this hypothesis was that language "determines the world-view [the individual] will acquire as he learns the language" (Kay and Kempton, 1984). The language of the individual was suggested to be an essential determinant of shaping the individual's perception of his or her surroundings. This then implied that due to linguistic relativity, it would not be surprising to have individual differences in perspective. Expanding this notion to the field of color perception, the individual's language could possibly affect how he or she processed color.

Further research had been conducted to see if this notion held true in particularly bilingual populations. In the study of Romney, two groups of participants were gathered to explore the relation between color cognition and bilingual populations. One group consisted of Mandarin Chinese speaking participants in Taiwan and the other of monolingual English speaking participants in California. Two tasks were given to the participants, one being a paired comparison rating task and the second being a triadic comparison. The first task consisted of 28 pairs of 8 colors while the second task included 56 triads of 8 colors.

In another study conducted by Roberson, the participation pool shifted to focus on the Korean participants just as the present study will do. Based on the triad tests that were given to these participants, findings revealed that Korean monolinguals were faster in response rate when the triads chose colors in "cross categories" than in "within categories". For example, Korean native speakers treated what bilingual Korean-English speakers would consider varying shades of green as two separate color categories of green and yellow. This discrepancy was due to the lack of naming for these varying shades as would be labeled in English as "lime green" or "yellow-green" (Robinson Pak et al, 2007). This detail concerning the color group of green resurfaced in the present study. After limiting the data pool to three high-quality participants, the results revealed that there was a discrepancy between one of the green color stimuli of the English trial and the same one in the Korean trial. The difference in locations of that particular green color stimuli further supported the lack of color names regarding this color group in the Korean language observed in Roberson's study.

Another notable finding was that of the study conducted by Caskey-Simmons Hickerson (1997). Bilinguals seemed to present a broader list of color names compared to monolinguals. The former showed choices that varied in comparison to one another more so than those of the monolinguals. This variability may be accounted for the cognitive process of using two languages instead of one. While monolinguals have one language system to base off their color choices, bilinguals have two systems to rely on. This supports the notion that language can act a factor in affecting the individual's cognitive processing of colors.

In the present study, we will explore the same question from the previous studies in the Korean-English population. Contrary to the studies prior to the current one, the language that the directions are given, may possibly trigger the semantic structure that dictates how the individual perceives and categorizes the colors. By using the triad similarity judgment task, we will measure the color appearance similarity found in the bilingual population when tested in different language modes.

1.1 Empirical Question

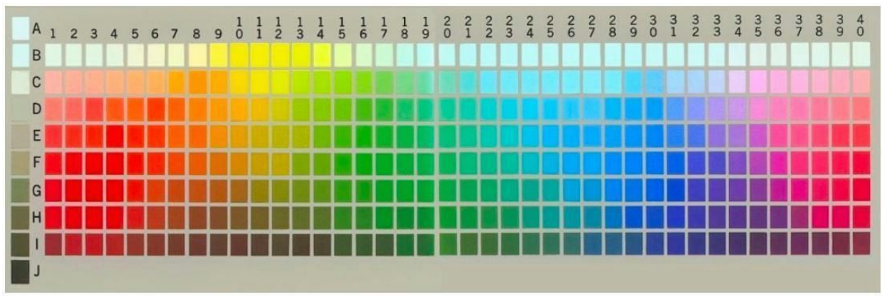
Several interesting color similarity questions can be raised for Korean-English bilinguals, and these relate to recognized differences across the two languages and how they differ in the ways speakers of those languages categorize color space. For example, Figure 1(a) shows a standard color stimulus that has been widely used by researchers, across many empirical investigations, to elicit color names from a wide range of ethnolinguistic groups. This color structure is what is named categories shown in Figures 1(b) and 1(c), for English and Korean speakers respectively. The important point to notice is ways the two languages differ in how they name the colors.

One notable aspect of this comparison is the emphasis of green and blue colors found in the Korean naming system. In the English naming system, there is a distinction between each color group. However, this is shown to be the contrary when using the Korean naming system. The color groups of green and blue seem to heavily overlap. The color group for green seems to also overlap with that of yellow. In English, there are various names for the same color group. For example, a color that falls under both blue and green would be called as “green-blue”. However, as shown in Figure 1c, the brightness of this sort of combination of either place the color in an entirely green color group or blue color group. Another example is the “red-purple”. In the English naming system, this would fall under the purple color group. However, depending on the brightness of the shade, the color would either classify as purple or red in the Korean naming system. It seems that the Korean naming system limits or restricts its color naming based on the brightness of the color. The same shade of color can fall into an entirely different color group based on the darkness or lightness of it. This distinction in the color naming systems was also seen similarly in the study conducted by Caskey-Simmons Hickerson (1997). In the study’s findings, monolinguals had fewer overlapping areas in colors than bilinguals. For instance, there were more color combinations with blue, red, and orange (Caskey-Simmons Hickerson, 1997). These features of the two languages, and their differences just described in how the two languages map color, are some of the features we will be looking for in our results from the present triad similarity studies.

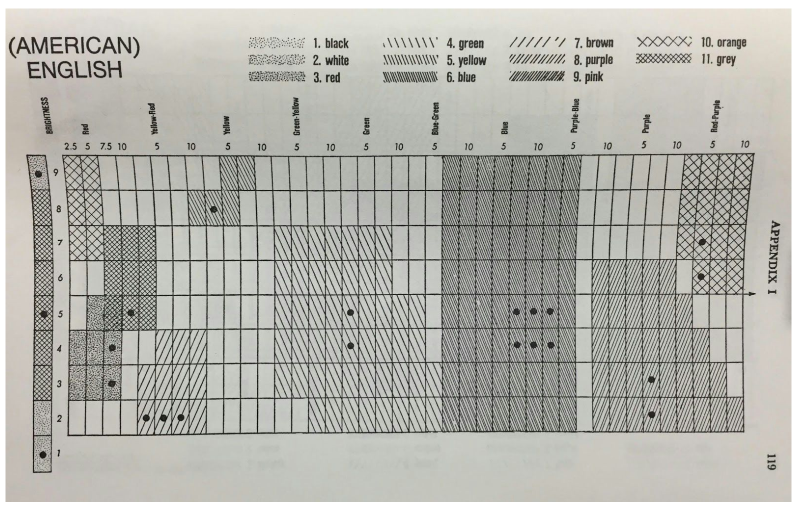
Among the empirical questions we will examine in the present study are:

1. Are the color naming differences found in previous studies (e.g., Figure 1 (a-c)) consistent with the color similarity structures we observe for Korean-English bilinguals triad choices. For example, in the green region of the color space where Koreans and English have been shown to differ, is it the case that the color similar found in the triad data of the two language conditions?

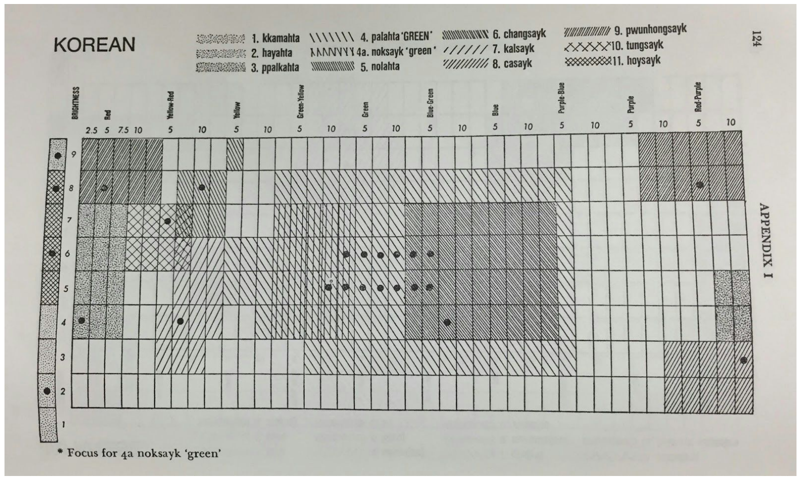
2. Or, do we find that there is no effect of language mode and Korean-English bilingual color similarity structures do not reveal differences across the two language conditions.
3. Is there also a third variable that accounts for the possible difference or similarity across the triad choices such as prior exposure to the task?



(a.) Color stimulus



(b.) English color naming



(c.) Korean color naming

Figure 1. Color categorization differences known to exist across Korean and English language speakers. Panel (a) shows the Berlin & Kay (1969) stimulus used to map color categories across different languages. Panel (b) show a mapping result from English speakers. Panel (c) shows a different mapping result from Korean speakers. (Adapted from Berlin & Kay 1969).

2.0 Methods

2.1 Participants. The present study recruited voluntary participants that were Korean and English native speakers. The participants were undergraduate students that studied at the University of California, Irvine. Their ages fell in the range from 18 to 21 years old. In addition, the participation pool did not have a gender bias, including both male and female subjects.

2.2 Materials. The apparatuses that were employed in the present study were Farnsworth-Munsell 100 Hue Test (Farnsworth, 1957) and 75 triad similarity cards (Bimler et al. 2014). For the fm hue 100, there were four wooden cases of color dots that were to form a smooth gradient color spectrum. And, the 75 triad similarity cards consisted of three colored circles printed on a gray background of each card. A letter of a, b, and c were printed under each colored circle, and the cards themselves were labeled from 1 to 75.

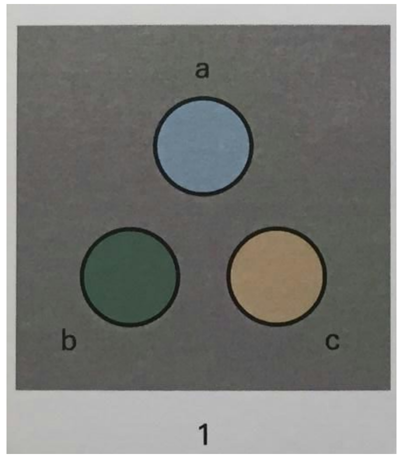


Figure 2. Triad stimulus card. As shown in the illustration above, all 75 cards had three circles colored in. Each circle had a color listed in the range below. Under each color was a letter a, b, and c. The gray background allows the colors to be viewed in constant adaptation via the light booth.

Figure 3. Approximation of color appearances of the 16 individual color stimuli employed in the Lambda1 Design for the 75 experimental triads. See Appendix C for the list of 75 triads.

In addition, a D65 daylight approximate illuminant was provided via a calibrated light source housed in a viewing booth. This was used to provide a consistent lighting for each testing subject. The entire light booth was a standard Munsell gray approximate (value = 5). No other light source or chromatic appearances were present in the testing area. Experimenter wore a black lab coat. Black cloth covered the table where the light booth was placed to eliminate chromatic contrast effects on the color stimuli.

2.3 Procedure. The participant of the study was to sit in front of the light booth while the experimenter sat next to the table. The subject was asked his or her email and name as the experimenter jot down the information. Then, the experimenter explained in English the directions of the first task. After I elaborated the directions, I showed the subject two examples of the cards that would be used in the first task. When the subject completed each example, I then asked if the subject understood what was to be done during the first task. If no further questions were asked, I continued on to execute the first task. The first task consisted of 75 triad similarity cards that had been randomized. Each card would be placed on the surface under the light portal for the subject to see and decide which colored circle did not belong with the other two. The subject would then give the letter written underneath the colored circle that he or she identified as the one that did not belong. Meanwhile, I would measure the subject's response time by counting to myself. After identifying which colored circle did not belong, the subject was asked to rate his or her confidence level regarding the choice. The subject would rate based on a Likert-scale from 1 being the least confident to 10 being the most confident. After the subject gave his or her rating, the current card would be put to the side and the next card would be flipped over and placed on the surface. After all 75 cards were shown, the subject then was given the FM 100 Hue Test. I would explain the directions regarding this test in English as well. The subject was told that there were four wooden cases that contained color dots. The task for the participant was that he or she would arrange the color dots shown into a smooth color gradient order. On each end of the case was a color dot that acted as reference of starting and ending points. After each wooden case was arranged, the subject was then informed of the last task. The third task was essentially a repetition of the first except the directions were explained in Korean. Then, the subject was to complete the last task as he or she had done so similar to the first task. After all 75 cards were shown, the subject had completed the whole experiment.

3.0 Results

Analyses of the three forms of collected data are presented below.

3.1 Color Vision Assessment:

All 20 participants were assessed for normal color vision through the FM100 Hue test. By calculating the total error scores or the number of wrong choices that participants had made, the normality of their color perception was measured. By assessing the participant's color perception, the possibility of having skewed results can be taken into account. With the exception of a few, most of the 20 participants were found to have normal color perception. Out of the 20 participants, six were shown to have "pathologic" or abnormal color perception (see the FM100 Hue chart in Appendix A). The results fell under the range from 4 being the lowest total error score to 177 being the highest total error score. The average of the total error scores was 53.450, and the standard deviation was shown to be 42.226. Comparing the performance of the participants on the FM100 Hue Test to their correlation coefficient, there seems to be a link between the two. Those who were particularly diagnosed as having pathologic color perception scored the highest in the total error scores. This particular group of participants also showed to have lower correlation coefficients regarding the consistency of their triad choices across both their English and Korean trials.

3.2 Color Triad Similarity Analyses:

The aim of all the color triad analyses was to investigate color similarity relations across two language conditions (English and Korean) and explore the possibility of between-language mode influences seen on within- and across-subject similarity and within-subject performance consistency across the two language modes. For this we examined (1) between-subject group color similarity scalings across two language conditions using Metric Multidimensional Scaling (see MDS description in Appendix B), (2) within-subject individual consistency in triad-choice consistency across language conditions (refer to Table 1). The first analysis presented aimed to plot the distance between the color stimuli and assess color similarity across participants' triad choices. The second analysis helped identify the high-quality participants whose data would be reevaluated for a more reliable assessment.

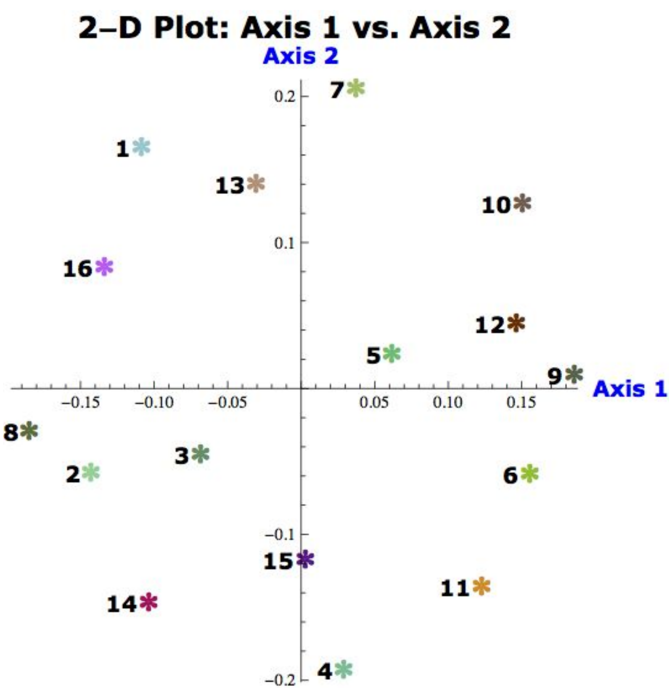
3.2.1 Analysis 1: Group MDS scaling results.

A standard way of examining color similarity relations shared across participants was to scale the similarity distances in a two-dimensional MDS plot. This method was widely used on color triads (e.g., Moore, Romney & Hsai 2002) and typically would reveal a solution of hue circle similarity relations with a circumplex similarity structure for the color stimuli. That was, a structure with color similarity ranging from reds to purples to blues to turquoise to green to chartreuse to yellow to orange and back to red -- forming a circular organization of color similarity that follows the hues associated with the spectrum. This was referred to in the literature as a "color circumplex" and had 2 dimensions that could be described reddish-to-greenish and bluish-to-yellowish. Figure 2 showed such 2-dimensional MDS solution obtained from the two language conditions tested for all 20 participants.

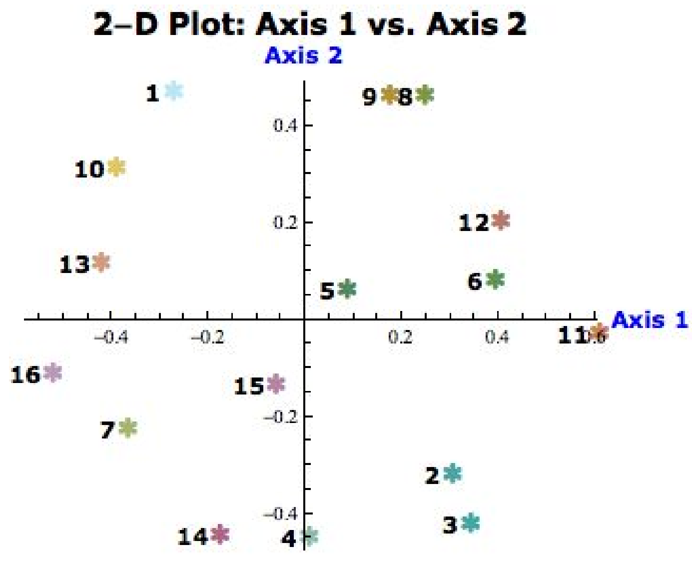
Note that the aforementioned isolation of the color circumplex in 2-D MDS scalings of these data Figure 2 results was somewhat disappointing. For example, neither Figure 2(a) (for English condition) nor Figure 2(b) (for Korean condition) showed the characteristically ordered color circle expected. Moreover, the stress levels observed for these MDS solutions was much higher (English MDS_Stress= 0.305124 and Korean MDS_Stress= 0.291437) than one would expect from a well modeling solution (MDS_Stress = 0.08 or lower is acceptable), and neither scree plot of the solutions (not shown) exhibited the characteristic "elbow" typically found in robustly estimated similarity structures. Albeit, there were signs that the solutions were tracking distances among similar items in a reasonable manner. For example, stimuli 2 and 3 (and stimuli 5 & 6, to a degree also) were very similar chromatically, and in the MDS plots they appeared relatively more adjacent to one another, suggesting a closer estimated distance that coincides with their closer color similarity. So, this kind of result suggested that while distances seemed to be properly estimated, there were perhaps inconsistencies or different strategies for selecting the odd-one-out in a triad across participants' response patterns.

Therefore, the results of these two MDS solutions were not as one would expect and suggest that either (a) there may have been some problems with too few participants (N=20) contributing to the solutions, or (b) or there were unreliabilities or inconsistencies in the participant data. To address possibility (a) we aggregated the English and Korean datasets to effectively produce a dataset of 40 participants, and recomputed the MDS analysis to determine if improvements in MDS fit, Stress and Skree could be observed. This alternative analysis using 40 participants is presented in Figure 2(c).

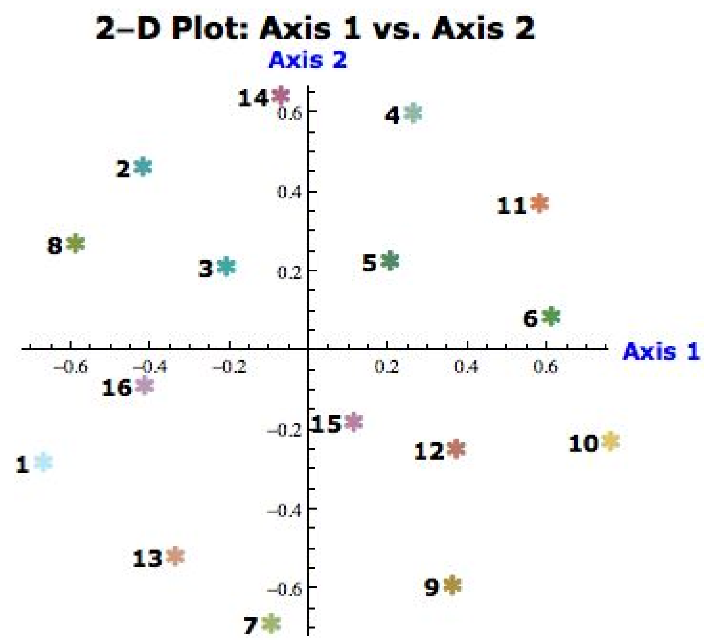
Figure 2(c)'s MDS solution on 40 participants did not dramatically improve the derivation of a color circumplex for the color similarity structure. Moreover, the MDS Stress remained high and unacceptable (MDS_Stress =0.328524) and the Scree plot also lacked the desired "elbow" formation (not pictured). There were, however, some indications that increasing participants to 40 does help the normalized dissimilarity relations to model distances (plot not shown), but the other indicators (lack of circumplex, Stress and Scree) suggest that problems are due to participant data, or alternative "(b) or there are unreliabilities or inconsistencies in the participant data." To analyze this possibility we examined some individual participant measures.



(a.) English



(b.) Korean



(c.) English + Korean

Figure 2. Two-dimensional color similarity MDS structures. (a.) English language condition (20 participants). (b.) Korean language condition (20 participants). (c.) MDS structure for both English and Korean language conditions together from a group of 40 participants.

3.2.2. Analysis 2: Assessing individual participant data.

To investigate if problematic participants could be identified and eliminated from further MDS analyses we first considered “Repeated measure correlations” based on each participant’s choices for the 75 triads run in the English condition compared with their choices for the same 75 triads when run under the Korean condition. Table 1 shows those Repeated measure correlations.

Rank-ordered "r" Correlation coefficient Subject #	
18	0.924
16	0.882
20	0.832
13	0.746
7	0.738
19	0.734
10	0.711
12	0.709
17	0.662
15	0.644
9	0.634
11	0.601
4	0.599
1	0.595
6	0.593
14	0.579
8	0.545
5	0.514
2	0.489
3	0.350

Table 1. Rank-Ordered “r” Correlation Coefficient. The subjects were ranked based on the numerical values for the correlation coefficient found between their triad choices in both the English and Korean trial. The 18th subject had the highest value for the correlation coefficient, suggesting high consistency between the triad choices in both the English and Korean trials. The 3rd subject was ranked last for having the lowest correlation value.

As noted in the table above, some subjects scored a higher value for their correlation coefficient than others. In general, correlation coefficients refer to the strength of the relationship between the two items being compared. In this case, the subjects’ triad choices in the English trial and their choices in the Korean trial were being matched to assess any similarity found in the two. Higher correlation coefficient indicates stronger overlapping of their color choices in both of the English and Korean trials. The range of these correlation coefficients can be considered one of an extreme, seeing that the highest is 0.924 and the lowest being 0.350. As previously discussed (see 3.1 Color Vision Assessment), some of the lower correlation coefficients matched with the lower performances of the participants. Those who had the highest total error scores from FM100 Hue Test (that is, those who performed poorly on the FM100 test) also seemed to display lower values for their correlation coefficients. The ranking of subjects based on their correlation values will later on allow us to further assess the present data.

Another way to assess whether participants were consistent in their triad choices across the two language conditions is to count up the overall percent matches in response for each participant when comparing English triad choices to Korean triad choices for all 75 triads. Figure 4 displays the possible consistency found between the participants' triad choices during the

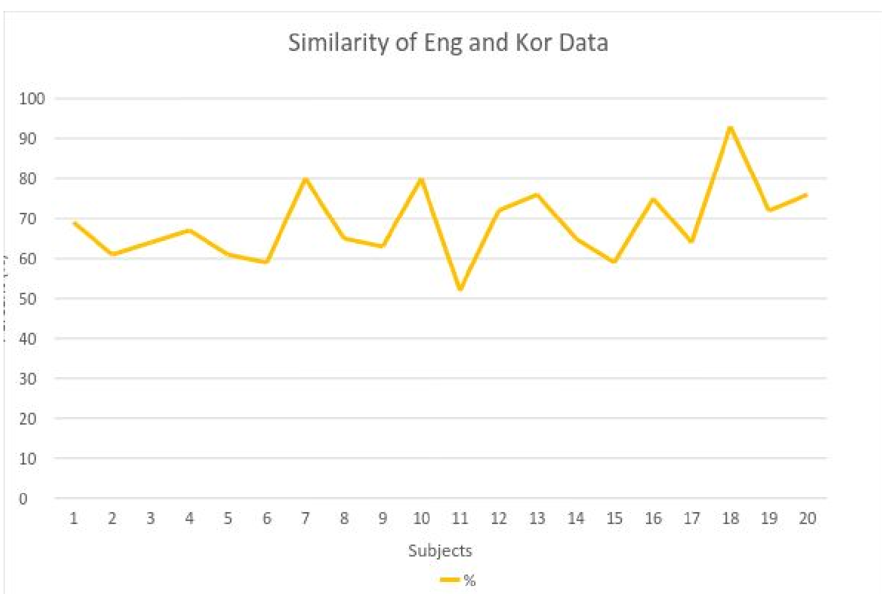


Figure 4: Percent consistency for within-subject triad choices across two language conditions. This gold line shows that participants were not all consistent across the two conditions completed.

English trial and the Korean trial.

As displayed in the graph listed above, there appeared to be moderate level of consistency across all twenty subjects. Despite a few fluctuations among the subjects, the general pattern seemed to suggest that many of the subjects had overlapping color choices in both trials. Especially notable are the extreme peaks such as that of the 18th subject. For the case of the 18th subject, there was a very high level of consistency or overlapping of color choices between both the English and Korean trial. As later noted in the ranking of subjects in Table 1, this particular subject also had the highest correlation value. Another notable detail of the graph is the extreme drop found for the eleventh subject. This subject had the lowest consistency found across the data collected in the English and Korean trial. Interestingly enough, the same subject did not have the lowest correlation value noted in Table 1. This discrepancy brings about the question of the reliability of the data collected of this particular subject.

3.3. Analysis 3: Assessing variation in participant "confidence" and "time to respond to task" measures.

English **Korean**

Subject ID	Conf AVG	Conf SD	RT AVG	RT SD
001H	6.474	2.312	1.353	0.812
002S	6.408	1.328	2.4	1.185
003J	6.368	1.924	1.2	0.465
004A	8.039	2.023	1.02	0.128
005Y	6.684	2.483	1.087	0.322
006E	7	1.818	1.153	0.394
007J2	7.303	1.819	1.147	0.492
008J3	6.947	1.818	1.32	0.661
009K	7.711	2.513	2.007	1.229
010C	8.947	1.118	1.02	0.554
011Y2	5.526	1.872	1.287	0.459
012P	8.474	1.205	1.06	0.232
013HJ	5.237	1.704	2.013	1.075
014I	4.829	2.849	1.0667	0.251
015B	6.737	1.586	1.013	0.081
016EJ	6.974	1.833	1.487	0.809
017A2	6.961	2.1	1.073	0.228
018H2	7.684	1.618	1.667	1.206
019J3	6.013	2.023	1.033	0.171
020EM	5.803	1.876	1	0

Subject ID	Conf AVG	Conf SD	RT AVG	RT SD
001H	7.4	2.15	1.707	0.912
002S	6.667	1.018	2	0.915
003J	5.827	2.029	1.467	0.644
004A	7.133	2.094	1.04	0.197
005Y	7.16	2.563	1.014	0.117
006E	8.053	2.111	1.127	0.386
007J2	7.8	2	1.16	0.396
008J3	6.853	1.608	1.687	0.951
009K	8.52	2.379	1.953	1.252
010C	9.24	0.819	1.36	0.225
011Y2	5.347	1.842	1.74	0.589
012P	8.627	1.25	1.347	0.604
013HJ	5.413	1.508	2.187	1.023
014I	3.187	1.872	1.007	0.058
015B	6.627	1.063	1.007	0.058
016EJ	6.693	1.15	1.207	0.436
017A2	6.88	1.497	1.093	0.256
018H2	6.917	2.285	1.405	0.915
019J3	6.931	2.233	1.382	0.889
020EM	5.013	0.647	1	0

Table 2. Confidence and Response Time. Above are listed two tables, the left pertaining to the results from the English trial and the right, the Korean trial. The average and standard deviation were calculated for the confidence level and the response time of all 20 subjects. The confidence level was measured on a rating from 1 to 10, 1 being the least confident and 10 being the most confident. The response time accounted for how long the subject took in making his or her triad choice.

Listed above are two tables, one summarizing the average and standard deviation of confidence level and response time in the English trial while the other in the Korean trial. The average and standard deviation were calculated as a method of assessing possible outliers such as subjects who may had extremely inflated confidence or unusually deflated confidence. Those calculations also would account possible abnormality of response time such as subjects who took too long or rushed through the study. The first table, showing the results from the English trial seems to suggest a stable and consistent pattern. The second table on the right for the Korean trial, displays more fluctuations in the numerical values between subjects. In comparison between the two tables, the general pattern for most of the subjects was a decrease in their average confidence level. However, the decrease is not extreme and rather slight for many. For example, for the 15th subject, the average confidence level during the English trial was a 6.737 which decreased to 6.627 during the Korean trial. Further calculations showed that the overall average confidence level of the 20 subjects was 6.757 during the English trial. This average then slightly increased during the Korean trial to 6.790. Contrary to the general decrease in the average confidence level across subjects, the slight increase in the overall average of all 20 subjects suggests the possibility of a third variable. Possibly, the prior exposure of the task may have biased the subjects in unconsciously selecting similar color choices and display an increase in confidence level toward those choices.

Another interesting finding to note is that the top three participants with the highest correlation value did not necessarily have the highest average confidence level or fastest response time. Quite the contrary, for example, the sixth subject appeared to have a high average confidence level of 7 in the English trial and 8.053 in the Korean trial, an approximate increase of 1. However, this same subject later was found to be one of the six outliers who had the highest total error scores on the FM100 Hue Test. The fourteenth subject also appeared to display a similar discrepancy found with the sixth subject. This subject had the lowest average confidence level in both the English and Korean trial. The average confidence level for the English trial was 4.829, which then decreased to 3.187 for the Korean trial. This particular subject also was part of the six noted subjects to have the highest total error scores. Hence, results from specific participants such as those exemplified suggest the possibility of the overall data being skewed and even flawed.

Additional information has been collected pertaining to those particular subjects that may have given skewed results to the overall data.

3.4. Analysis 4: Constraining analyses from twenty to three high-quality participants.

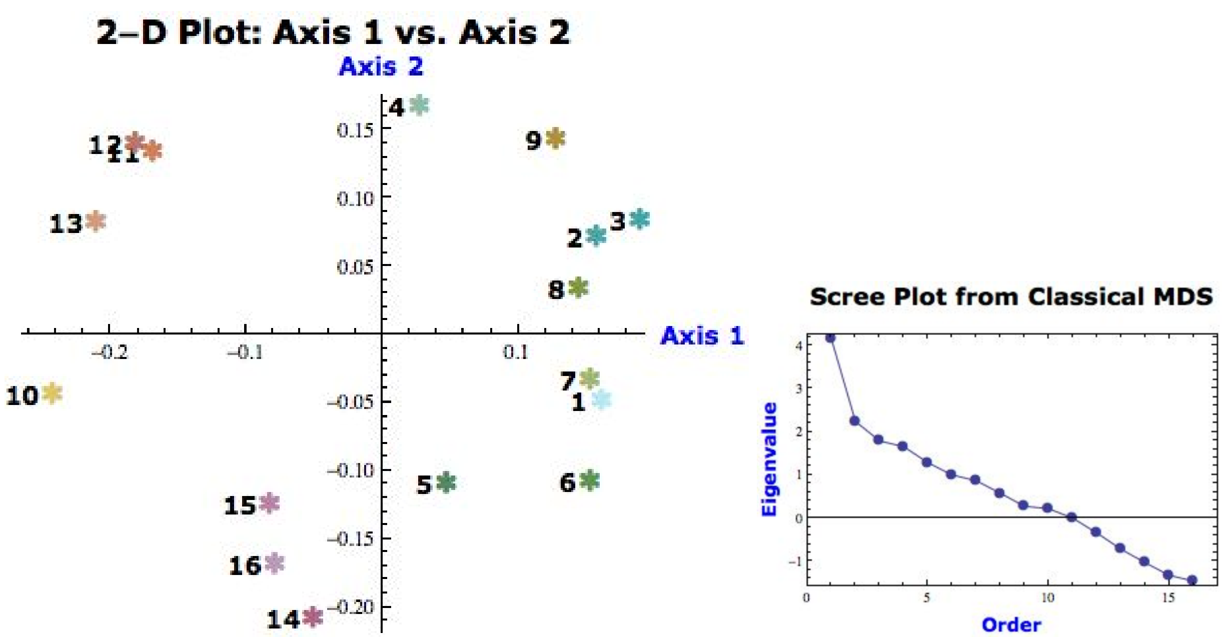
In analyses presented in Sections 3.2.2 and 3.3 above a detailed assessment was done to identify participants through assessing individual participant data. The net result of these analyses was to isolate desireable participants that had consistent choice patterns across the two language conditions, and to identify subjects that might possibly be excluded from group analyses (based on various indicators like confidence, response time, color vision error scores, and repeated measure correlations) for the purposes of improving the MDS results seen earlier in Figure 2(a), (b) & (c), and for assessing the original empirical questions discussed earlier in this paper, including whether language mode is found to have an effect on the judged perceptual similarity of the stimulus triads.

The individual analyses above led us to a process of testing subgroups of subjects to determine if a color similarity circumplex relation could be observed in smaller groups of participants' data when high quality participants were used. Some subgroup analyses used individuals that satisfied a criterion level of *Pearson's r* values equal or exceeding 0.7 seen in Table 1's repeated measure correlations. Eight participants satisfied this criterion and were run on the same kinds of MDS procedures shown earlier. The results found that even with this group of eight higher-quality participants, there remained a poor MDS fit as represented by a lack of a color circumplex solutions (not shown here), unacceptable levels of MDS solution *Stress*, and poor scree plot configurations. Other subgroup analyses attempted also did not provide satisfactory solutions for the data (results not detailed here).

Finding satisfactory solutions ....

Despite the failures seen in the abovementioned analyses, the results began to improve when we constrained analysis of participant data to the top ranked 3 participants in Table 1. That is, those individuals exhibiting a Table 1 repeated measure *Pearson's r* values equal or exceeding 0.83. Of these 3 participants (Subjects 16, 18, and 20) none showed any signs of color vision anomalies. All three showed normal levels of confidence: English mode found the 3 subjects had average confidence values measured at 6.9, 7.7, and 5.8, (group average = 6.76); in Korean mode they showed 6.7, 6.9, and 5.0, respectively (group average = 6.79). And all three showed average triad-completion response times similar to the observed group averages. The feature that distinguishes these 3 participants from the other 17 in the sample is there shared color similarity relations as shown in the MDS solutions of Figure 5 below.

(a.) English language condition 2-D MDS plot. Scree plot for English solution at right.



(b.) Korean language condition 3-D MDS plot. Scree plot for English solution at right.

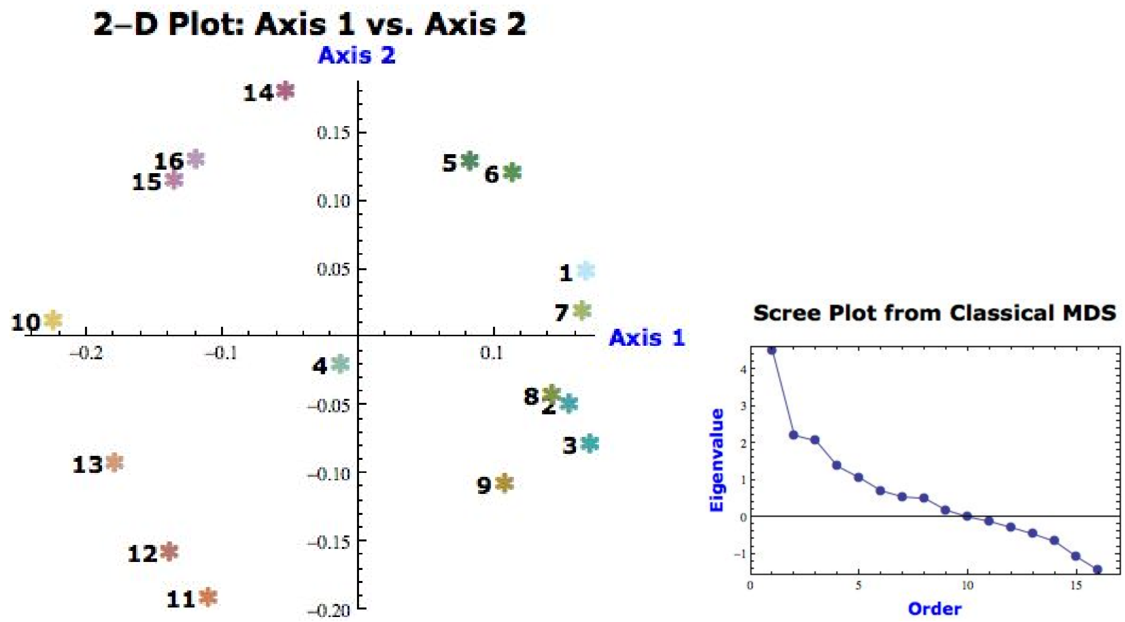
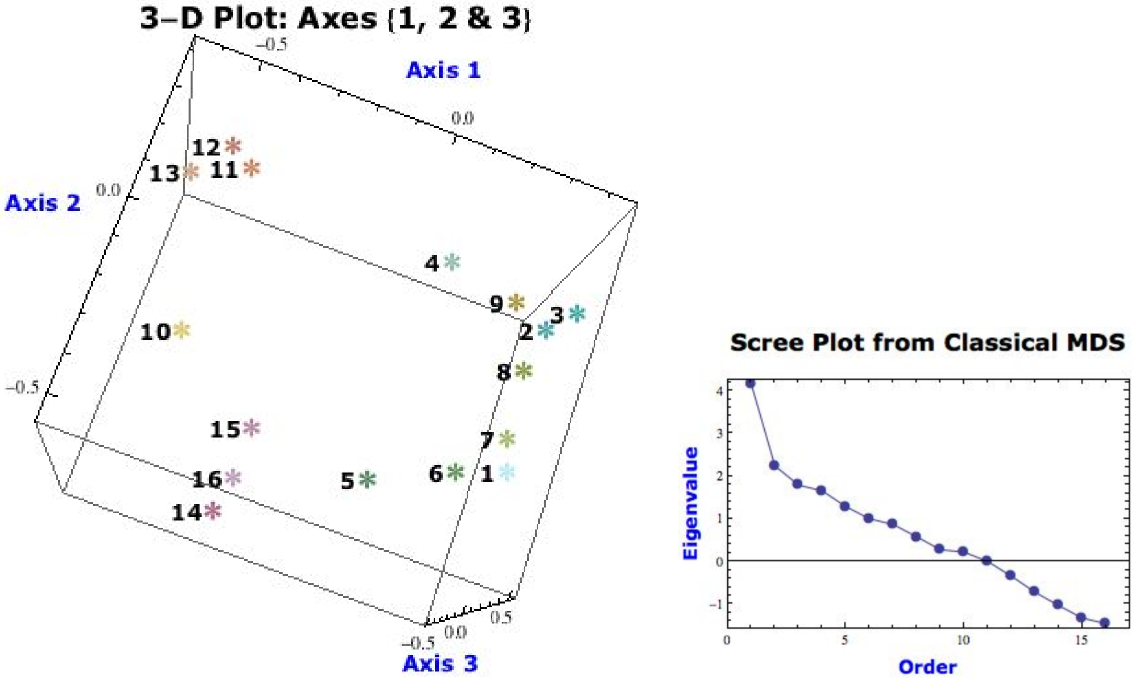


Figure 5. Two-Dimensional MDS solutions for (a.) English language condition triads. *Stress*= 0.29, (incl. Scree plot shown), and (b.) for Korean language condition triads. *Stress*= 0.25.

3.5. Analysis 5: Extending MDS scaling from 2-D to 3-D solutions.

Given improvements found in our analyses that reduced participants from 20 to 3 in group MDS analyses, we chose to further explore analysis of these 3 high-quality participants using a higher-dimensional MDS analyses of the data to see if a difference in the Korean and English MDS solutions obtained would suggest a potential language influence on this perceptual triads task. The rationale for considering a 3-dimensional solution is that the addition of the extra dimension may allow for improved estimation of the similarity distances, and solutions occuring at lower stress levels, producing a more robust color similarity circumplex. This involved increasing the dimensionality of the multidimensional procedure used from 2-dimensions to 3-dimensions. The results for these analyses are shown in Figure 6 below.

(a.) English language condition 3-D MDS plot. Scree plot for English solution at right.



(b.) Korean language condition 3-D MDS plot. Scree plot for English solution at right.

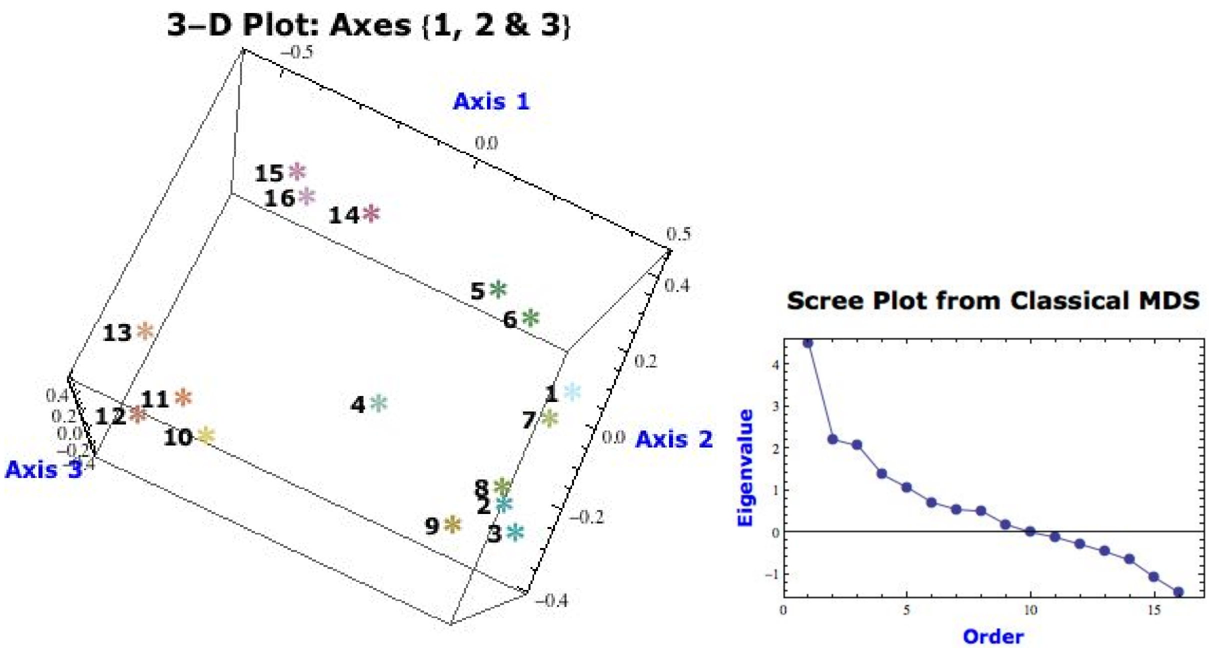


Figure 6. Three-Dimensional MDS solutions for (a.) English language condition triads. *Stress*= 0.23, (incl. Scree plot shown), and (b.) for Korean language condition triads. *Stress*= 0.19.

The figures just shown above appeared to demonstrate a language specific effect on color similarity relations. This can be understood by looking again at Figure 1. Figure 1 B which displays the English version of color categorization shows “green” to be the main color group depicting its shades. The “green” is also found to be distinct from the blue group set next it. However, this distinction of the categorization of green seems to differ in the Korean version, shown in Figure C. In Figure C, the color “green” seems to split into two subgroups one named “palahta” or green and the other “noksayk” or yellow green. Unlike how in Figure B, all shades of green were categorized under the overarching group of “green”, Figure C shows a variation between green and yellow and even an overlap of blue. Depending on the brightness of the color, the green can be classified as “green” as also shown in the English version or “yellow green”, which seems to be the prevailing categorization in the middle of the green color group.

As evident above in Figure 6’s 3D MDS plot, the color stimuli 4 seemed to be found in the middle of both green and yellow in the Korean triad trial (Figure 6 b). In the English triad test (Figure 6 a), the same color stimuli 4 appeared to relate more so with the green group. This then supported the distinct categorization of the green color group found in the English color naming system. The wavering of the color stimuli 4 then coincided with the subtypes of the green color group in the Korean color naming system. Just as shown in Figure C, there exist yellow green and green. For the Korean triad test to display the color stimuli 4 somewhat in the middle of yellow and green, it suggested that language may indeed have an influence on the individual’s color perception. The suggested notion that language may affect the perceiver’s cognitive processing of color may seem to hold true at least for this bilingual population. Another notable aspect of the above figures was that it took the results of the top three participants with the highest correlation coefficients. The results of the rest of the participants were dismissed, seeing that the previous Figure 2 showed a rather chaotic and inconclusive assessment.

As mentioned before, the shape of the “elbow” was necessary in noting whether the data is conclusive and reliable. Here, in Figure 6, the desired “elbow” can be pinpointed, further strengthening the reliability and consistency in those specified participants’ data.

Summary of Findings

The present study had yielded interesting findings along with essential factors to be taken into consideration while assessing the data at hand. The study explored the possible effect of language on color perception based on the assumption proposed by the Whorfian hypothesis. This assumption suggested that a linguistic factor plays a vital role in determining how the individual perceives his or her surroundings. This concept expanded to include whether a linguistic factor truly had a significant impact on shaping an individual’s color perception.

To have a stronger sample with less likelihood of bias, 20 subjects were recruited voluntarily. After data had been collected from the study, the results were assessed into MDS or multidimensional scaling plots. Contrary to what was expected from the MDS solutions, the plots did not accurately display the specific sets of solutions. Instead, a rather chaotic array of color markings resulted from the MDS analyses. Thus, the data collected from all 20 subjects were reevaluated by the usage of other measurements.

The consistency between the subjects’ triad choices during the English trial and the Korean trial was then measured in hopes of eliminating any possible outliers in the participation pool. The similarity of the triad choices found between the English trial and the Korean trial for each participant was measured. As shown in Table 1, the correlation coefficient was then calculated for each of the 20 subjects based on the similarity of their triad choices during both language trials. After each correlation value was calculated, the subjects were ranked based on those who had the weakest correlation to those with the strongest correlation value. The higher the correlation coefficient was, the higher the consistency was between the subject's triad choices during both trials. This would help eliminate any extreme result from the rest of the data pool as well as identify high-quality participants that would provide more reliable data to draw conclusions from.

Based on the ranking of the subjects, the correlation values from the top three participants were taken. This indicated that these three were considered as the high-quality participants of the study; therefore, their data would be more reliable and useful for the purpose of the study. Just as the initial data from all 20 subjects were analyzed into the MDS plots, similar measures were done with the new set of data. And, the desired MDS color circumplex was able to appear. Not only were the plots give a more reliable and sensible display of the color markings, but also the Scree plots showed the desired "elbow". Previously, the MDS plots showed the color stimuli to be everywhere without any direction. However, in the current MDS plots, the color circumplex was created with the color stimuli marked in the proper corresponding axis. This indicated that despite the reliance of fewer data, more reliable assessments were made.

The empirical questions that were posed at the beginning seemed to be supported by the results yield in Figure 5 and 6. Questions 1 and 2 had focused on whether language did affect the individual's cognitive processing of color. And, to these inquiries, Figure 5 and 6 suggest that in some cases, language did seem to play a role in the individual's color perception while in other cases, language may not the sole determinant in the processing of color. This; however, seems to rely on individual differences as discussed before of the errors found in the 17 subjects who did not seem to present strong data. The third question is also taken into consideration with the given results. For the participants who may have showed lower-quality results, factors such as prior exposure of the task may have led the subjects to make flaws in giving their triad choices. Perhaps not all participants had taken the task with the same level of significance as others.

One notable finding of the present study was the assessment of green. Based on the data collected from the three high-quality participants, the color stimuli 4 which was a lighter shade of green seemed to be marked in different areas in the MDS plots. In the plot of the English trial, the color stimuli was found near the other green shades. However, in the plot of the Korean trial, the same color stimuli was located towards the middle of the plot near both the yellow color group and the green color group. This difference in axis location suggested that language may to a certain degree influence how an individual processes color. The first empirical question was then revisited and answered. In a task that was not meant for a linguistic purpose, such a finding displayed the possible effects of language.

Conclusions

Overall, the findings of the present study emphasized the importance of having good participants in order to obtain reliable set of data. Otherwise, the data would become problematic in finding conclusive results for the purpose of the designated study. The present study also observed the possible effect of language on some of our participants. This finding further supported the literature that the study’s empirical questions were based on. And so, some individuals did appear to exhibit a language dependent influence on triad choices, coinciding the literature's own findings of English color categories being distinct from those of Korean. However, these findings do not ultimately conclude that there truly exists a bilingual effect on color triad choices. Yet, the present preliminary study does provide a promising outlook on further exploring the bilingual effect as well as possible future research.

References

Bimler, D. L., Paramei, G. V., Feitosa-Santana, C., Oiwa, N. N. and Ventura, D. F. (2014), Saturation-specific pattern of acquired colour vision deficiency in two clinical populations revealed by the method of triads. Color Res. Appl., 39: 125–135. doi: 10.1002/col.21794

Caskey-Sirmons, L. A., & Hickerson, N. P. (1977). Semantic shift and bilingualism: Variation in the color terms of five languages. *Anthropological linguistics*, 358-367.

Farnsworth, D. (1957). *The Farnsworth-Munsell 100-hue test for the examination of color discrimination*. Munsell Color Company.

Jameson, K. A., & Alvarado, N. (2003a). Differences in color naming and color salience in Vietnamese and English. Color Research & Application, 28(2), 113−138.

Jameson, K. A. & Alvarado, N. (2003b). The Relational Correspondence between Category Exemplars and Naming. Philosophical Psychology, 16(1), 26-49.

Jameson, K. A. (2005). Guest Editor. "Studies of color cognition, categorization and naming" Journal of Cognition & Culture, vol. 5(3/4).

Jameson, K. A. & Alvarado, N. (2005a). Guest Editors. "Cross-cultural studies of color cognition, categorization and naming." February 2005. Cross-Cultural Research: The Journal of Comparative Social Science, vol. 39(1).

Jameson, K. A. & Alvarado, N. (2005b). Guest Editors. "Cross-cultural studies of color cognition, categorization and naming." May 2005. Cross-Cultural Research: The Journal of Comparative Social Science, vol. 39(2).

Kay, P., & Kempton, W. (1984). What Is the Sapir-Whorf Hypothesis? American Anthropologist, 86(1), 65–79.

Kay, P., Berlin, B., Maffi, L., Merrifield, W. R., & Cook, R. (2011). The World Color Survey (1 edition.). Stanford, California: Center for the Study of Language and Inf.

Kay, P., & Regier, T. (2003). Resolving the question of color naming universals. Proceedings of the National Academy of Sciences, 100(15), 9085–9089.

Kay, P. & Regier, T. (2006). Language, thought, and color: Recent developments. Trends in Cognitive Sciences, 10, 51–54.

J. B. Kruskal, Psychometrika, 1964a,b

Lindsey, D. T., & Brown, A. M. (2014). The color lexicon of American English. Journal of Vision, 14(2):17, 1–25.

Moore, Carmella C., A. Kimball Romney, and Ti-Lien Hsia. "Cultural, Gender, and Individual Diferences in Perceptual and Semantic Structures of Basic Colors in Chinese and English." *Journal of Cognition and Culture* 2.1 (2002): 1-28. Web.

Moore, C. C., Romney, A. K., & Hsia, T. L. (2000). Shared cognitive representations of perceptual and semantic structures of basic colors in Chinese and English. *Proceedings of the National Academy of Sciences*, *97*(9), 5007-5010.

Roberson, D., Pak, H., & Hanley, J. R. (2008). Categorical perception of colour in the left and right visual field is verbally mediated: Evidence from Korean.*Cognition*, *107*(2), 752-762.

Romney, A. K., Boyd, J. P., Moore, C. C., Batchelder, W. H., & Brazill, T. J. (1996). Culture as shared cognitive representations. *Proceedings of the National Academy of Sciences*, *93*(10), 4699-4705.

Sayim, Bilge, Kimberly A. Jameson, Nancy Alvarado, and Monika K. Szeszel. "Semantic and Perceptual Representations of Color: Evidence of a Shared Color-Naming Function." *Journal of Cognition and Culture* 5.3 (2005): 427-86. Web.

APPENDICES

APPENDIX A: Twenty Participant's Farnsworth-Munsell 100 Hue Test Results:

Subject ID	FM 100 TES	Diagnosis
001H	51	Normal
002S	47	Normal
003J	95	Pathologic
004A	76	Pathologic
005Y	37	Normal
006E	94	Pathologic
007J2	23	Normal
008J3	177	Pathologic
009K	27	Normal
010C	33	Normal
011Y2	39	Normal
012P	39	Normal
013HJ	19	Normal
014I	102	Pathologic
015B	100	Pathologic
016EJ	36	Normal
017A2	12	Normal
018H2	12	Normal
019J3	46	Normal
020EM	4	Normal
AVG	**53.45**	
SD	**42.225554**	

Table 3. Subjects fell under the age group of 18 to 21. For this specified age group, a numerical range from 75 to 76 would be deemed as normal color perception. Any total error score above that given range would be noted as pathologic, referring to the possibly abnormal color perception of the subject.

APPENDIX B: Metric Multidimensional Scaling Procedure

Metric Multidimensional Scaling: Steepest descent method to optimize Kruskal's "Stress" goodness of fit criterion. Algorithm written by: Timothy A. Satalich 3/2015, IMBS UC Irvine. Based on procedures in: J. B. Kruskal, Psychometrika, 1964a,b (with modifications).

Triads to Dissimilarity Matrix

Timothy A. Satalich 3/2015 IMBS, University of California, Irvine

A triad consists of three stimuli {Subscript[s, i], Subscript[s, j], Subscript[s, k]} where i<>j<>k. The subject chooses one of the three that is most different from the other two. For the purposes of this program the Subscript[s, i] are integers representing stimuli in the experiment numbered 1 through n. The two not chosen as the most different are considered to be the most similar and it is this pair, say {j, k} that is given a similarity count of 1(one) and added to the count in the j-th row an k-th column of a n by n similarity matrix. This is done for all triads and subjects which then produces an nxn similarity count matrix. The count in each cell of the matrix is divided by the number of times that pair was possible from the set of triads presented multiplied by the number of subjects. This results in the proportion of times that pair was deemed to be most similar across all possible times that pair could have possibly been chosen for all subjects. This is a similarity matrix. If all pairs j,k are not possible because of an unbalanced triad design, the missing pairs are considered "missing". In this case, they are "estimated" by taking the mean of the proportions in the column of the missing cell. The mean excludes the diagonal cell of the column or any other "missing" cells. The MDS procedure I wrote assumes a dissimilarity matrix so the similarity proportions are subtracted from 1 (one) to produce dissimilarity like data. The final output of this program is a symmetric matrix of dissimilarites averaged across all subjects that can then be used in procedures that assume this type of data.

Each stimulus is assigned a unique integer starting at 1 and proceeding consecutively to n (n = the number of stimuli).

This program takes as input a matrix of triads data structured like this:

Col 1: Subject Number

Col 2: An integer representing the stimulus in the triad considered most different

Col 3: Subscript[s, i] an integer representing a stimulus i in the triad

Col 4: Subscript[s, j] an integer representing a stimulus j in the triad

Col 5: Subscript[s, k] an integer representing a stimulus k in the triad

When the program is run, a dialog box will appear that requires you to supply an input file and an output file location and name. The default output file name is "Dissimilarity data.dat". Change it to any name you want following the rules of your operating system. You needn't supply a file extension. The output file name will include the ".dat" file extension automatically. If the number of triads per subject is not the default of 75 change it to reflect the correct number contained the dataset. There is only one consistency check on the data. The number of triads per subject is checked. The number of triads per subject should be the same for each subject and equal to the value supplied in the dialog box. If it is not, a warning is printed and the program completes the calculations.

APPENDIX C: Seventy-five Triads used in the Experiments

1	6	10
6	10	8
12	7	9
11	13	16
2	16	5
13	8	1
16	12	7
6	15	11
13	11	15
15	9	4
2	13	8
10	4	15
16	9	5
8	11	9
3	7	5
7	11	9
7	2	13
10	8	7

2	4	6
13	2	6
10	11	13
16	1	3
3	8	14
6	12	1
12	15	13
9	14	4
16	2	14
16	4	2
2	12	7
2	9	14
5	7	4
16	12	14
8	10	13
15	10	5
5	11	16
15	4	8
13	9	11

14	12	11
1	11	6
10	14	5
9	6	4
12	5	1
8	6	5
1	3	6
6	11	8
9	11	14
2	15	16
16	5	10
4	9	13
3	10	15
6	4	8
14	13	16
14	9	3
1	12	14
5	7	9
11	16	4

3	12	8
6	16	11
2	5	3
2	15	13
1	7	12
13	1	15
15	4	1
5	1	3
4	1	2
7	10	5
9	7	6
8	10	12
5	3	8
2	7	11
13	3	8
7	3	14
3	15	1
10	14	12
3	14	16

APPENDIX D: Subjects Description

Subject ID	Status	Proficiency in Eng/Kor	Behavior during Task
001H	tired, hungry		
002S			
003J			
004A			appeared more confident during the English trial
005Y		Korean	
006E			
007J2			seemed hesitant more during the English trial than during Korean trial
008J3			seemed to not really take the task seriously; kind of gave up in the fm100
009K		English	took the task very seriously
010C		Korean	
011Y2	appeared tired	Both	took long on the triad
012P		English	lower confidence rating when taking long to rate her answer
013HJ		Korean	took longer and showed precision during triad & fm100
014I		Korean	answered on cue; steady & consistent answering (questionable if she answered on the whim or not)
015B			

016EJ		English	
017A2		English	distinguished some colors by Earthy tones
018H2			
019J3	was sick		took moderate time to rate and select color choices
020EM		Both	

Table 4. Subject Description. The following information had been recorded about each subject. These details allowed the experimenter to further evaluate the reliability of the subject and his or her performance during the tasks. The subject's physical status, proficiency in English and Korean, and behavior during the task were noted.The yellow highlighted rows indicated that the subjects were noted as having "pathologic" color perception in the FM100 Hue Test.