

Name the Color

Mechanical Turk Behavioral Experiment

MIT Conway Lab — Deployed March 14-15 and May 5-14, 2016

Experiment created by: Katherine Hermann
Experiment edited by: Christine Vonder Haar
Experiment deployed by: Christine Vonder Haar
Analysis created by: Christine Vonder Haar

1 Overview

Before conducting the fMRI and MEG scans and collecting data in the lab, we are deploying a pair of behavioral experiments on Mechanical Turk. The results from these experiments will help us decide what stimuli are likely to produce the best results in the real-life scans.

2 Goals

The goal of this behavioral experiment is to determine which color terms are closely associated with which color patches. Ideally, we will be able to determine what word a person would likely use to identify a certain patch of color; vice versa, we also will be able to determine what color patch most people would choose to describe a given color term.

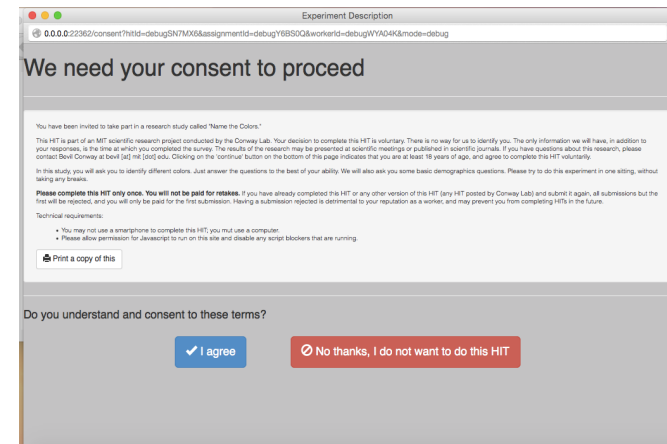
3 Important Note

Two versions of this experiment were deployed. The difference lies in the first half of the experiment. In the first version, 16 color spirals are shown (DKL Cardinal and Intermediate hues at two luminance levels); in the second version, only 8 color spirals are shown (DKL Intermediate hues at two luminance levels). **For this point forward in this paper, the first version with 16 color spirals will be referred to as Experiment 1 while the second version with 8 spirals will be referred to as Experiment 2.** This paper will only analyze the DKL Intermediate Spiral Data from Experiment 1. For more information on the entirety of Experiment 1, please refer to the previous experiment overview.

3 Experiment Implementation

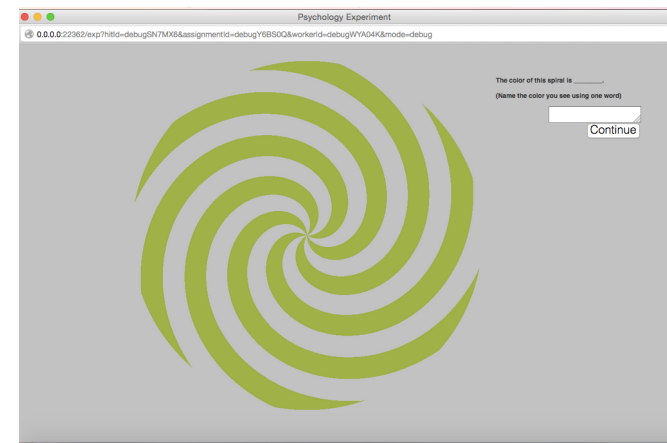
3.1 Terms and Agreement

When the subject accepts a HIT, they are taken to the page shown. It summarizes the task and outlines the requirements to complete in the experiment. For example, the subject must complete the HIT on a desktop or laptop and cannot complete the experiment multiple times. When the subject agrees to these terms, the experiment begins.



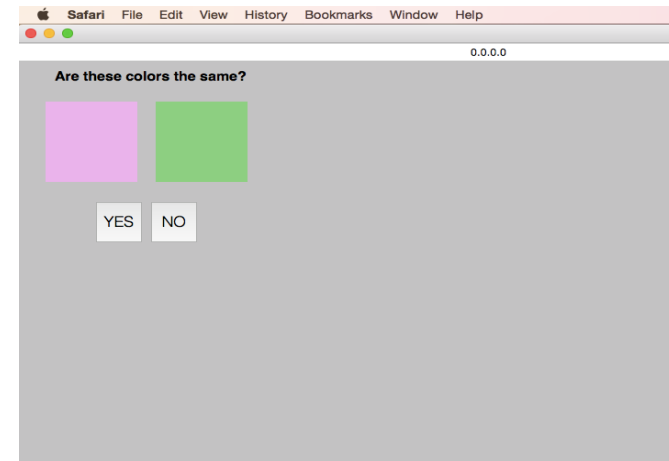
3.2 Part 1 — Given Color, Write Term (Color Spirals)

The subject is shown a color spiral and must type in the name of the color they identify the spiral as into a text box. The subject is limited to a one word response, which should cut down on result variability. Eight/four hues are presented, each at two levels of luminance, for a total of 16/8 unique color spirals. All color spirals exhibit the same level of saturation. Each color spiral is then shown to the subject three times, for a total of 48/24 colors terms to collect. This repetition of stimuli was implemented in order to check for within-subject reliability.



3.5 Attention Questions

The experiment was coded and deployed with the three main sections as described above. One HIT was deployed, and the subject finished the experiment in about 8 minutes; it was expected to take about 20 minutes. After looking at his/her results, “attention questions” were added in order to make sure the subject was focused on the experiment and their answers. There are 8 attention questions — 4 in part one and 4 in part 2, dispersed randomly throughout each section. The subject is presented with two color squares and has to answer the prompt “are these colors the same?” with a yes or no button. Half of the attention questions present the same colors, and half present different colors (same saturation and luminance, different hues).



4 Results

The results of the experiment were collected and stored in a database. Python scripts were used to reject subjects and store the data of the accepted subjects in a cleaner way. The spiral section (Part 1) data was then used to create meaningful charts and graphs to analyze the data. The grid section (Part 2) of the experiment has yet to be analyzed.

4.1 Rejected Subjects

Subjects were rejected on many criteria: failure to complete the experiment, colorblindness, non-English first language, attention questions, and failure to answer questions within specified constraints.¹ In Experiment 1, in total, 76 subjects participated in the study; 26 of these responses were rejected and 50 of these responses were analyzed. In Experiment 2, in total, 70 subjects participated in the study; 18 of these responses were rejected and 52 of these responses were analyzed.

4.2 Demographic Statistics of Subject

A few main demographic details were calculated across the subjects:

Experiment 1:

- Average age = 33.10204
- Minimum age = 20
- Maximum age = 60
- Female to male ratio = 28:32
- Total accepted subjects = 50

Experiment 2:

- Average age = 32.2
- Minimum age = 19
- Maximum age = 63
- Female to male ratio = 29:33
- Total accepted subjects = 52

¹ Note: subjects may have been rejected on more than one criteria. The criteria stated is the first criteria that the subject did not pass.

4.3 Within-Subjects Analysis

The goal of the within-subjects analysis was to determine how consistent each individual subject was with their color term usage. Each of the subjects' responses were analyzed to determine how many color terms a subject used to describe the same spiral. Because there were three trials for each spiral, a subject could respond with "pink" each time it was shown, or with multiple terms, such as "pink" and "purple". A subject could use 1, 2, or 3 terms to describe the same spiral. The color spirals were split into two categories, Cardinal and Intermediate colors, based on hue. The percent of subjects who used 1, 2, and 3 color terms was plotted as a histogram. A bar graph of average number of terms used per subject per spiral was also generated. See Graphs Section 5.1 and 5.2.

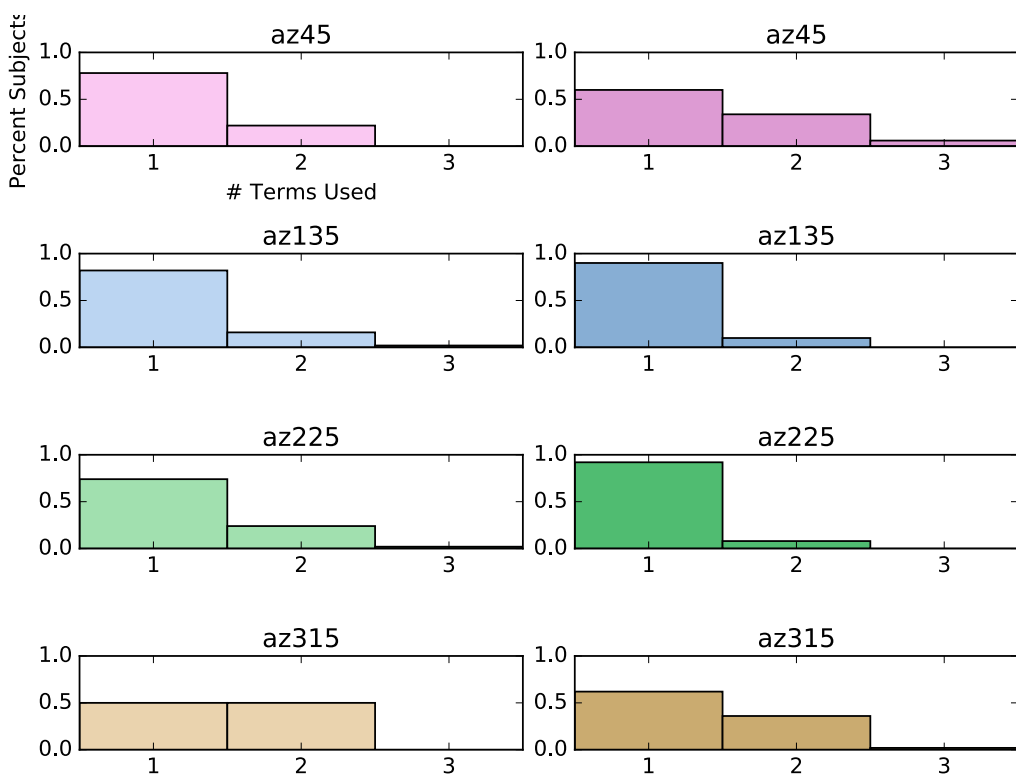
4.4 Across-Subjects Analysis

The goal of the across-subjects analysis was to determine how much variability in color term usage there was across all subjects. One trial was randomly chosen from each of the 50 subjects for use in the across subjects analysis. For each spiral, a bar graph of the percentage of subjects who used a certain color term versus all the color terms used to describe the spiral was created. A bar graph of total terms used per spiral was also generated. See Graphs Section 5.3 and 5.4.

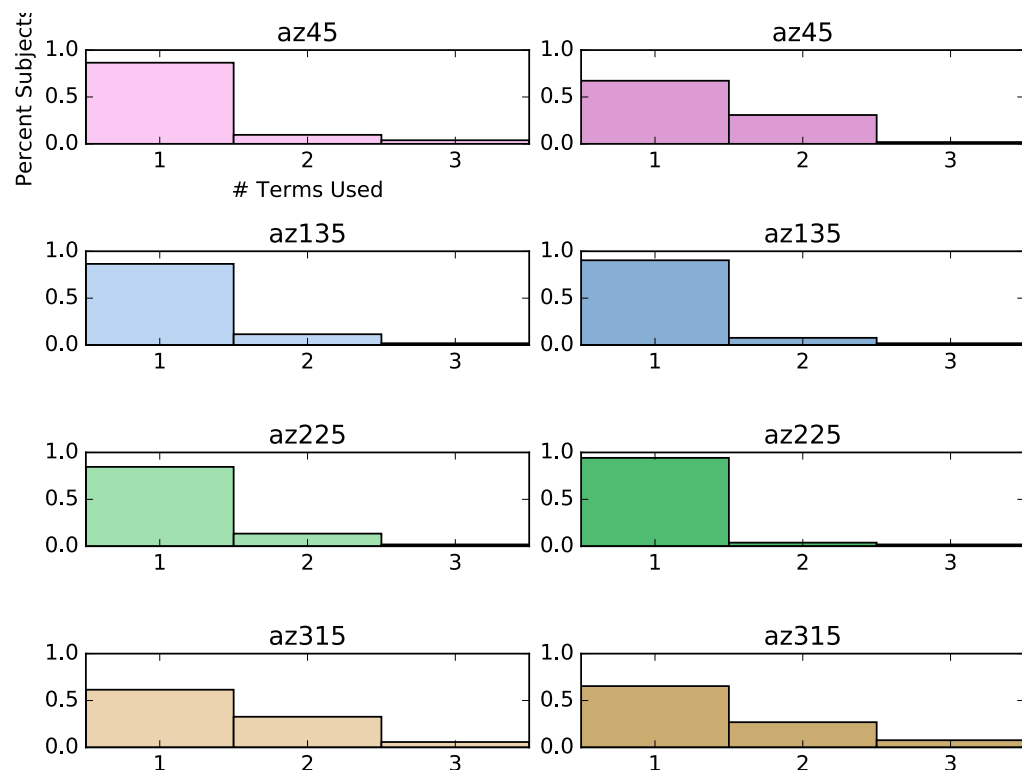
5 Graphs

5.1 Within-Subjects — Percent of Subjects who used 1, 2, or 3 color terms for each spiral

Experiment 1

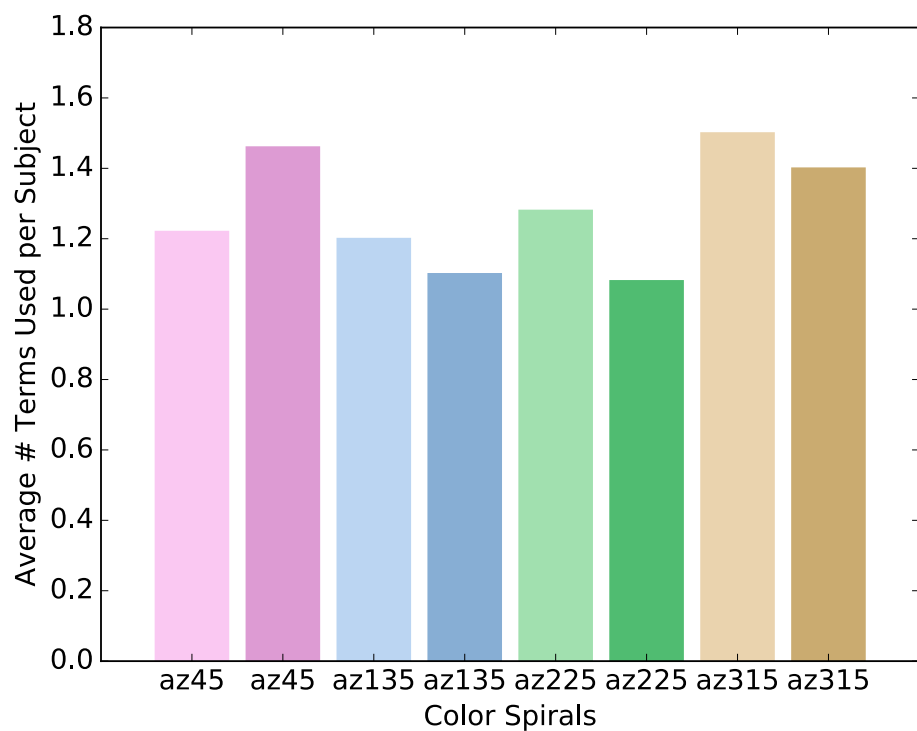


Experiment 2

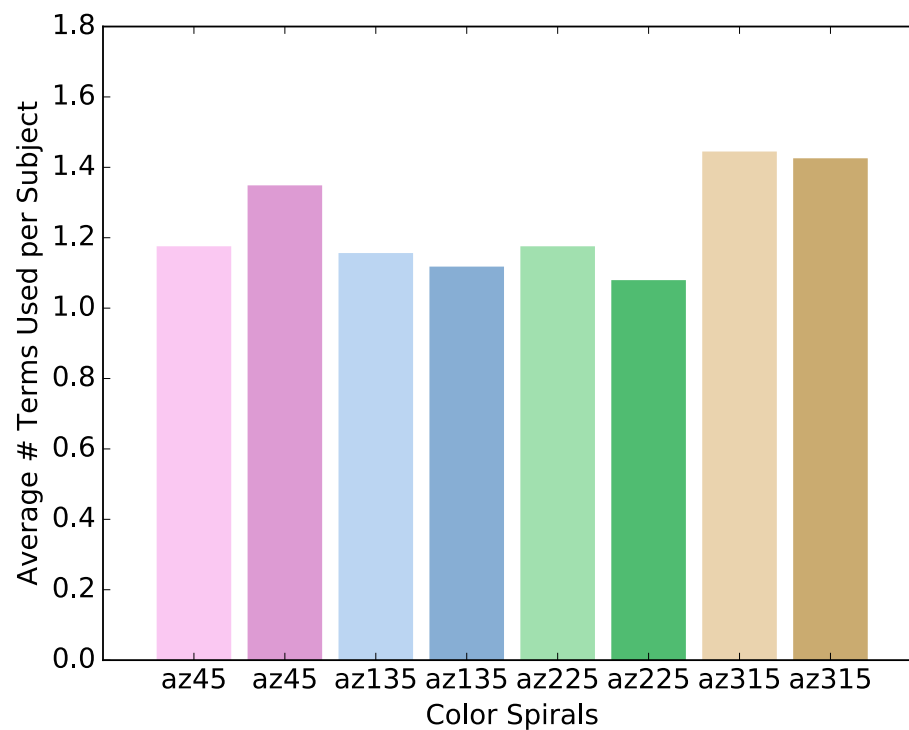


5.2 Within-Subjects — Average Number of Terms Used per Subject per Spiral

Experiment 1



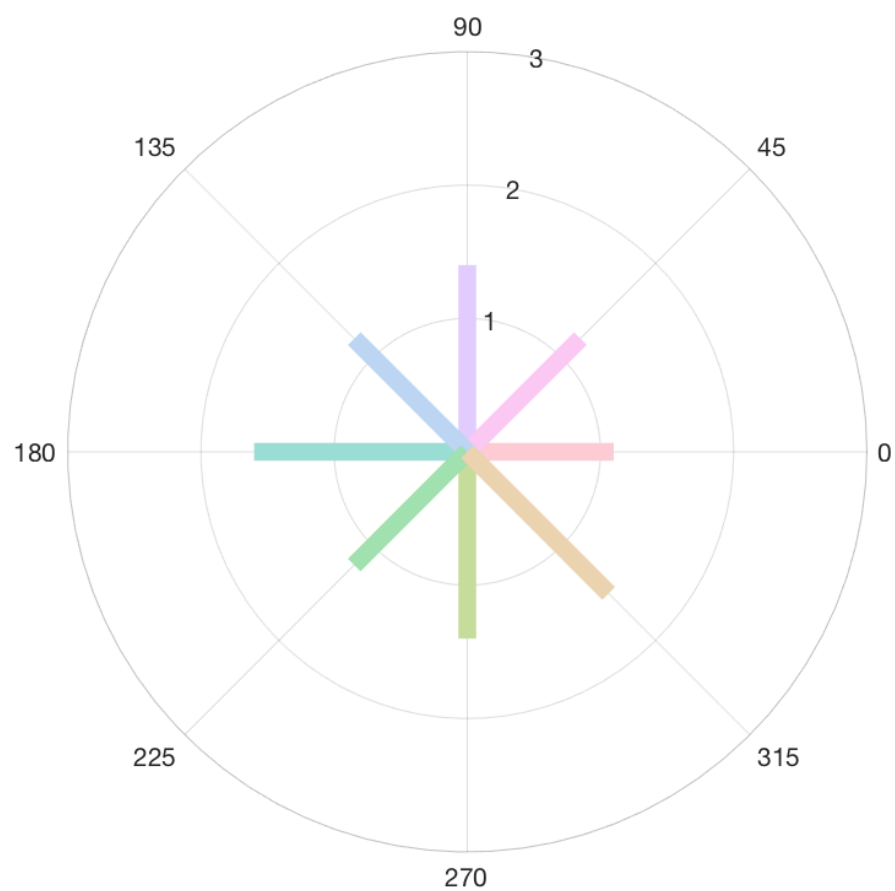
Experiment 2



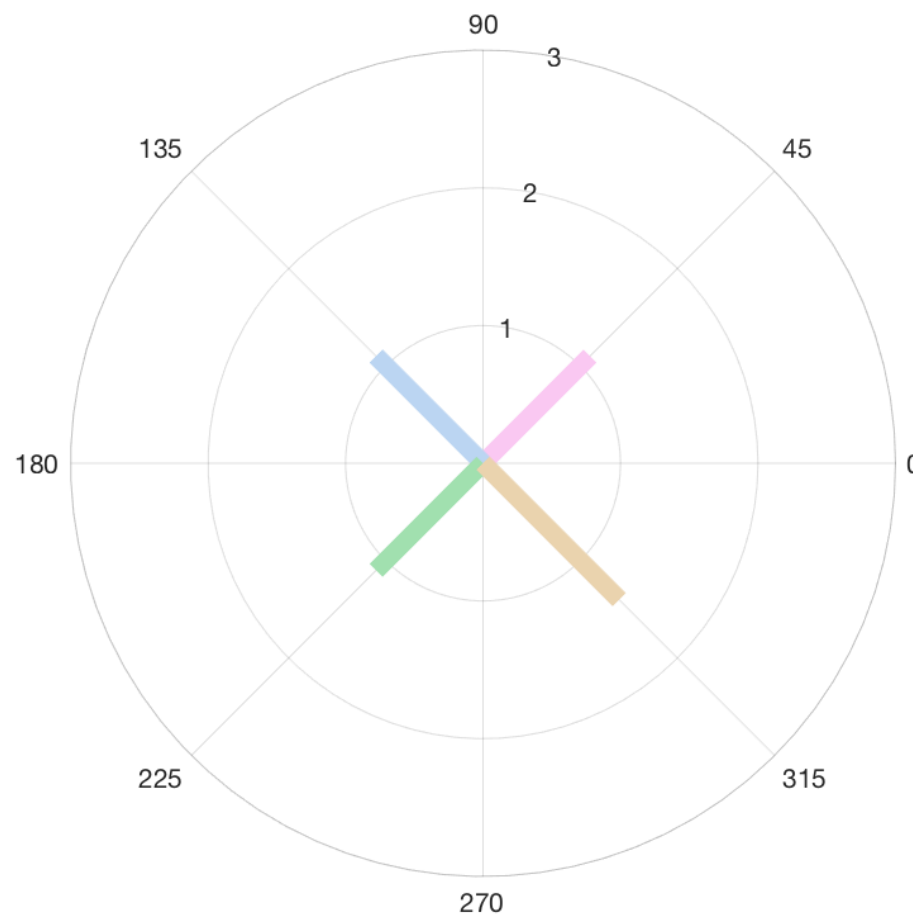
Light Luminance (Polar Graph Representation of Average Terms Used)

This is the same data as above, just presented in a different way

Experiment 1

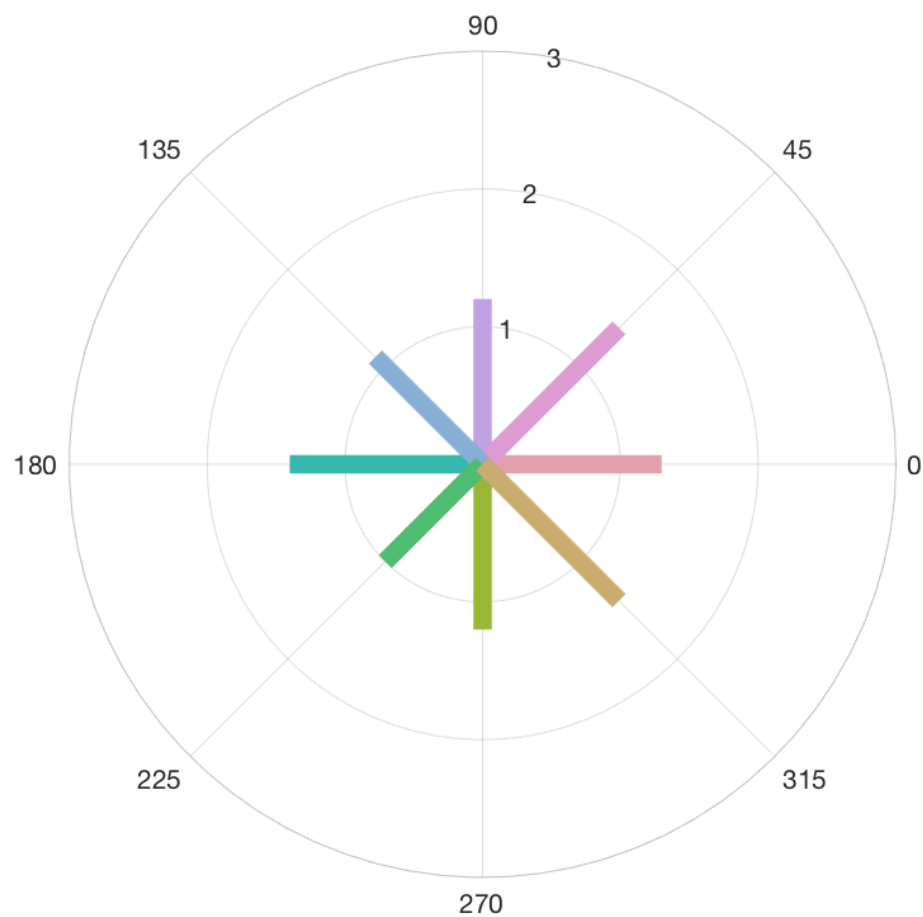


Experiment 2

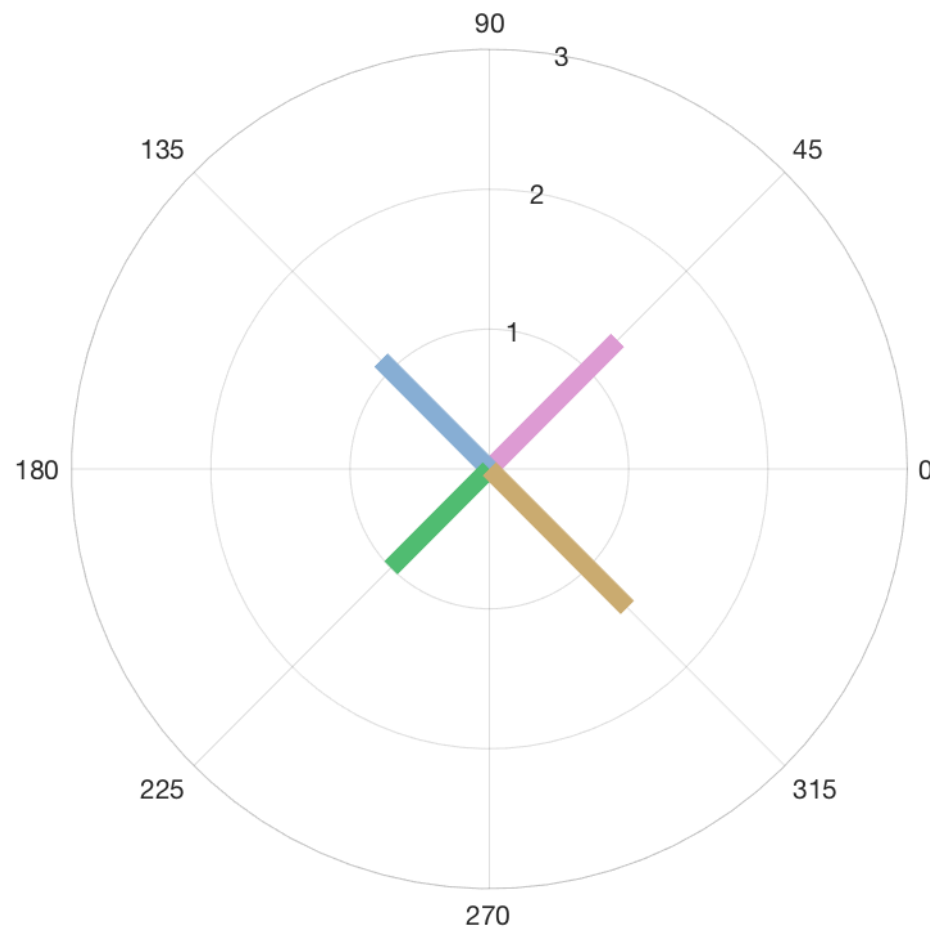


Dark Luminance (Polar Graph Representation of Average Terms Used)

Experiment 1

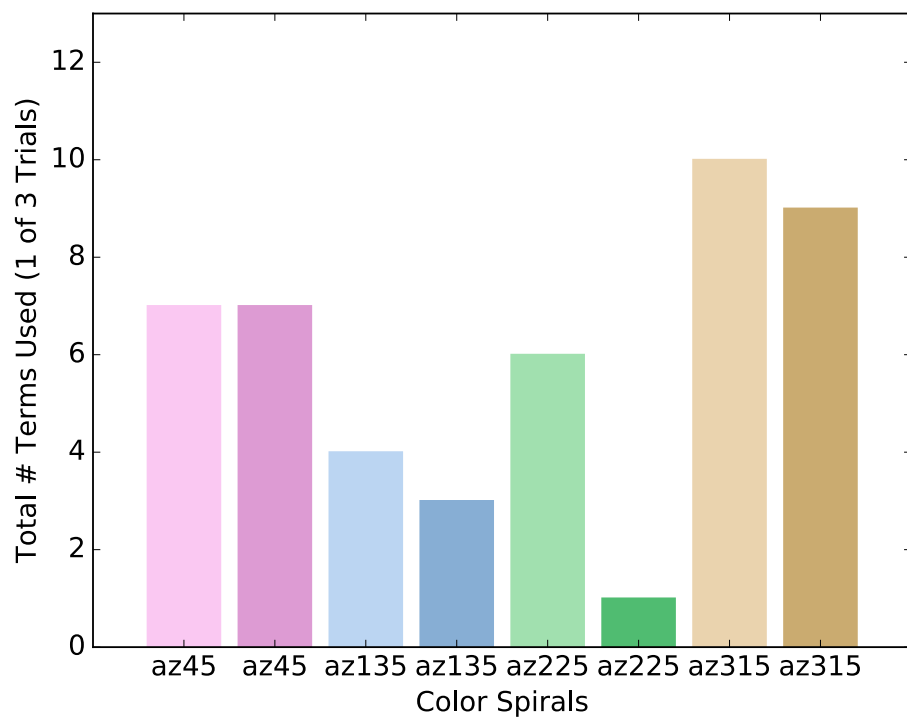


Experiment 2

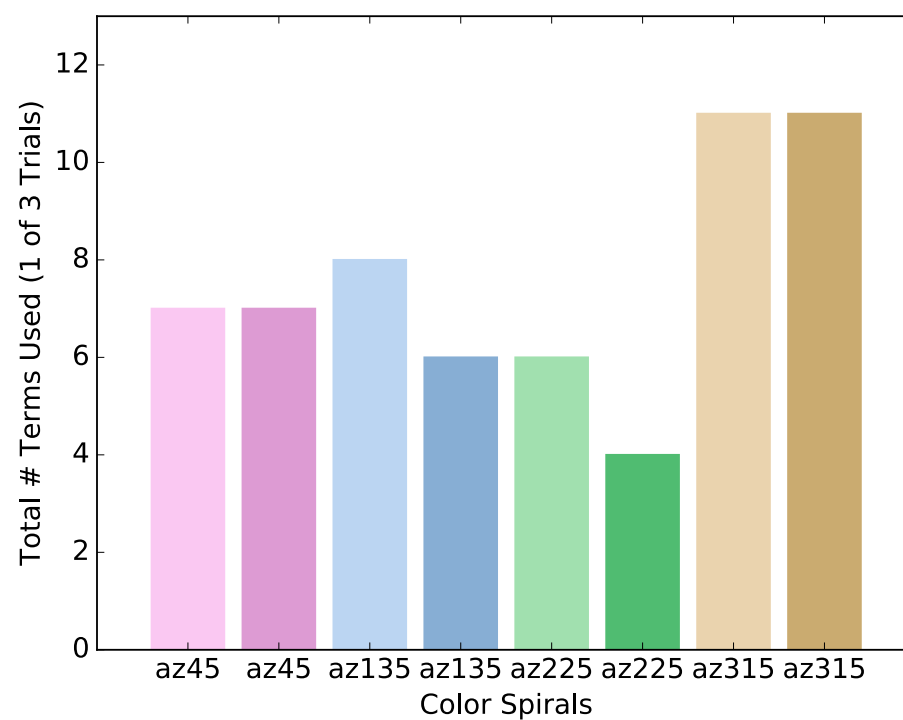


5.3 Across-Subjects — Total Number of Terms Used Across Subjects Per Spiral

Experiment 1

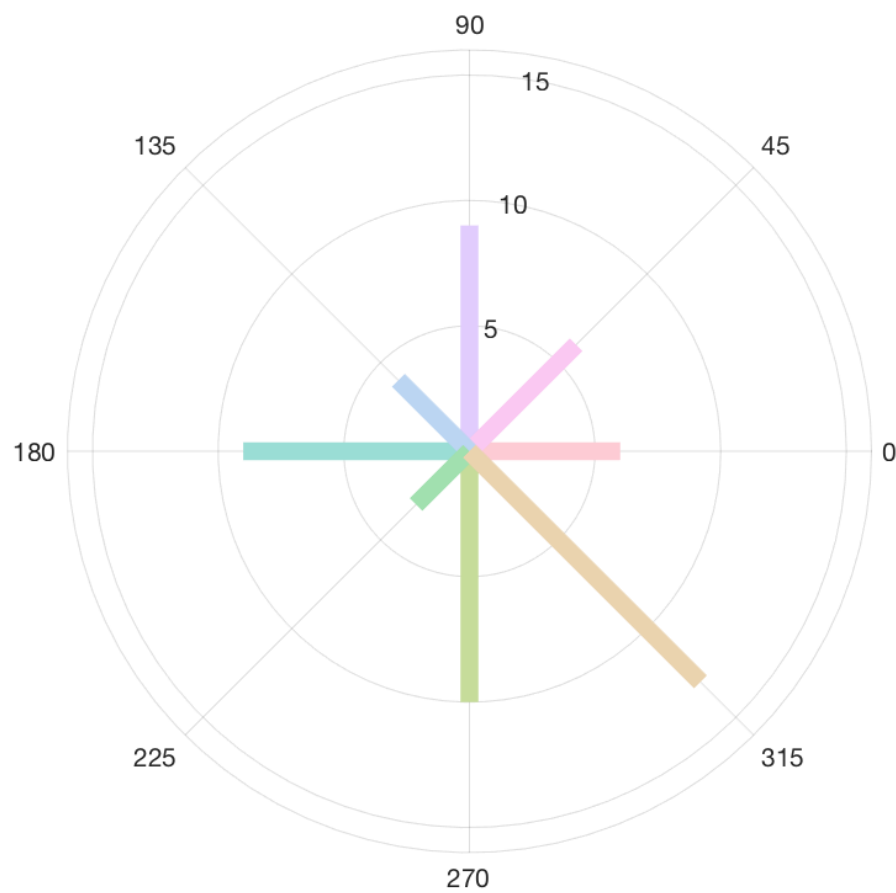


Experiment 2

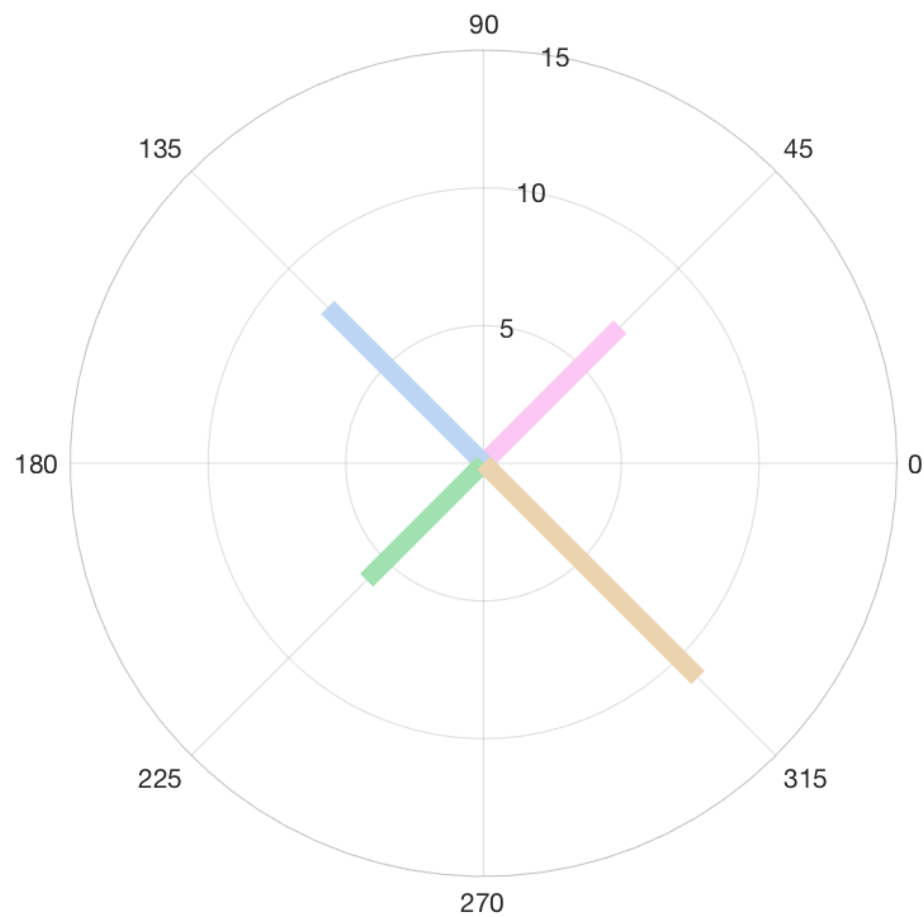


Light Luminance (Polar Graph Representation)

Experiment 1

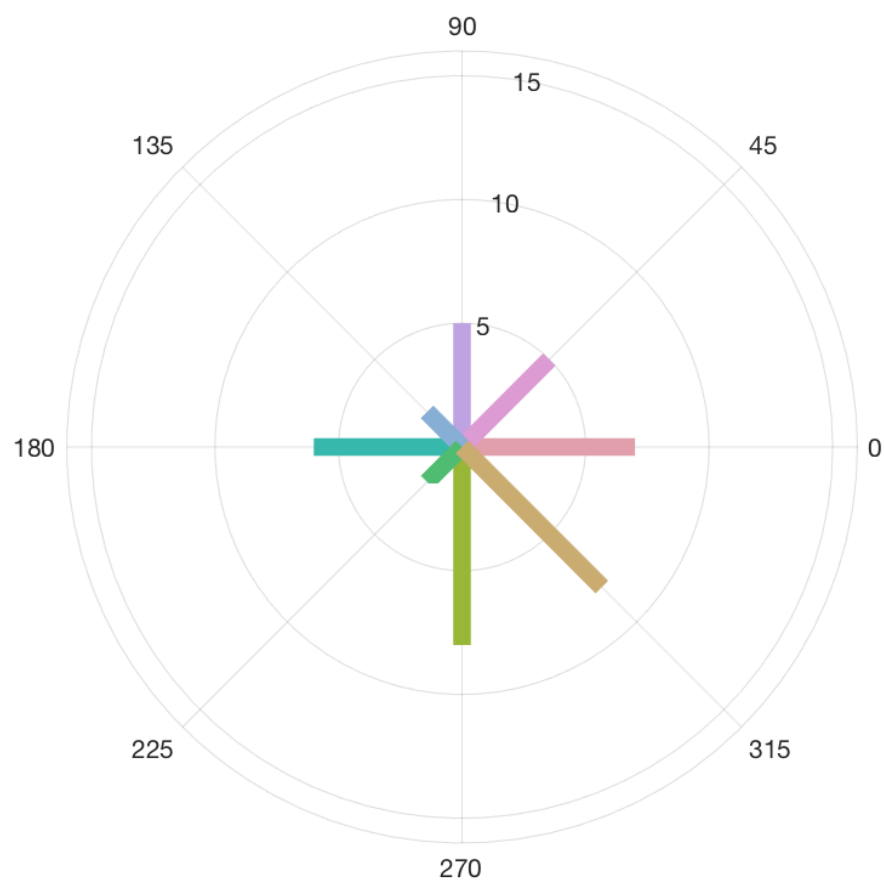


Experiment 2

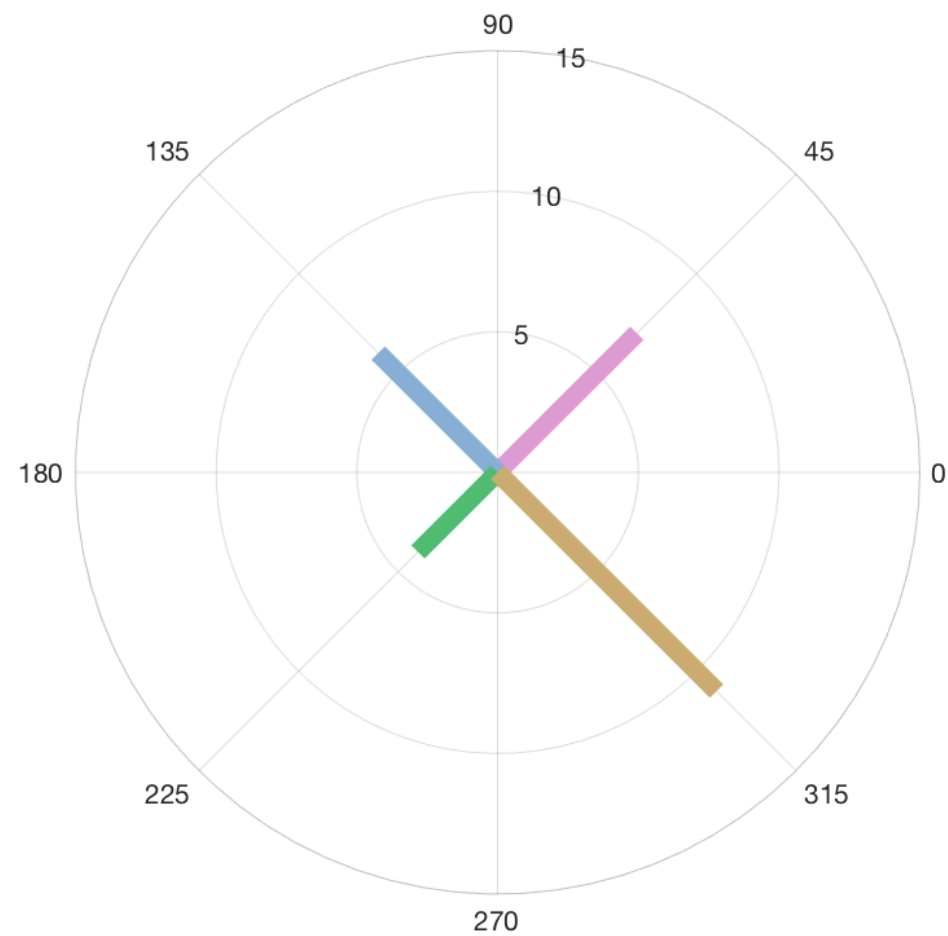


Dark Luminance (Polar Graph Representation)

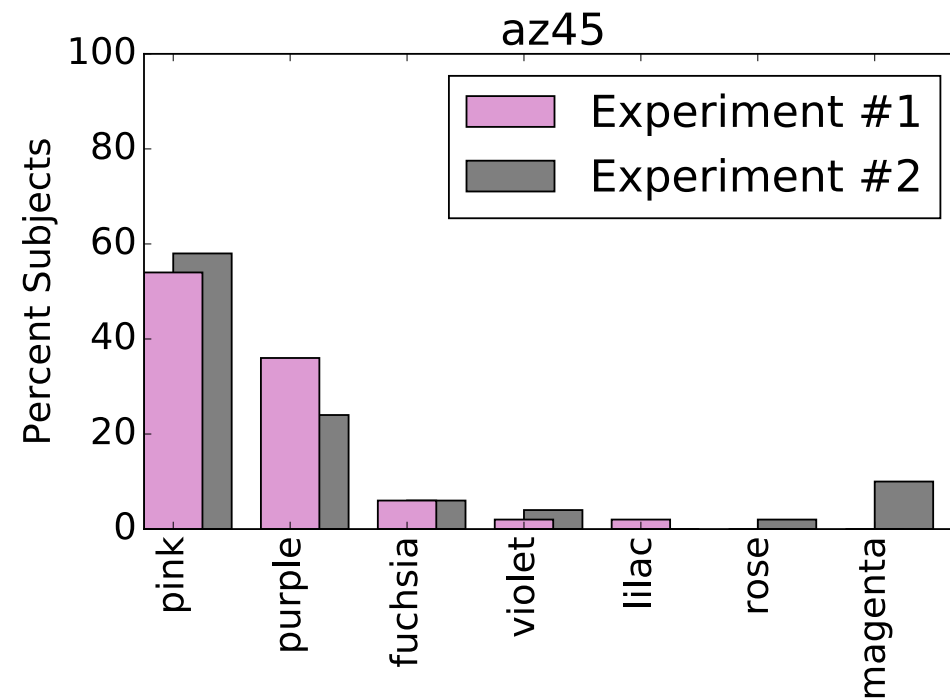
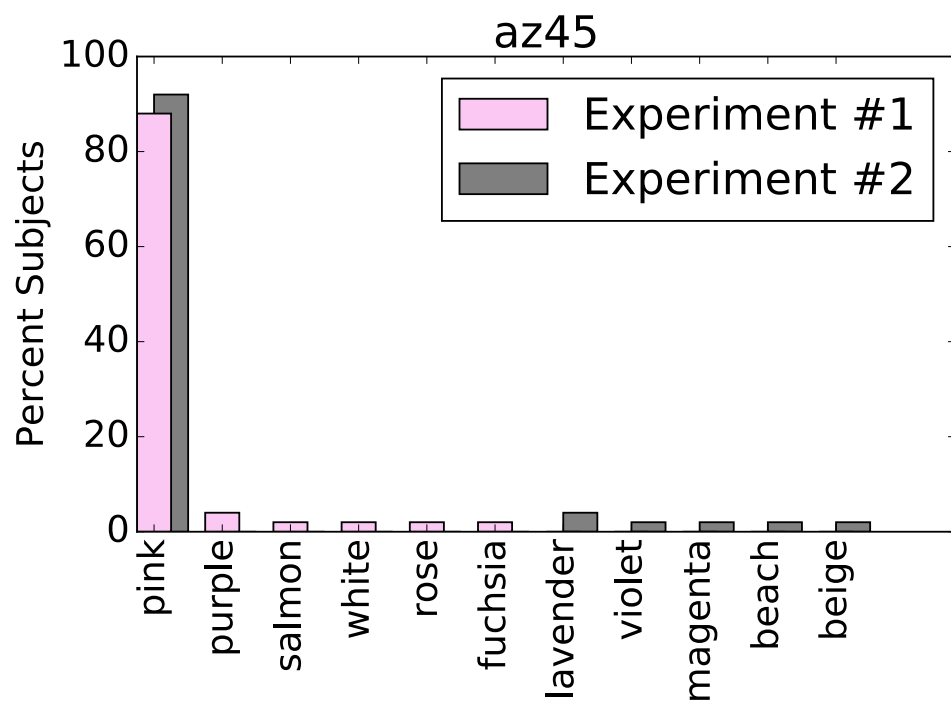
Experiment 1

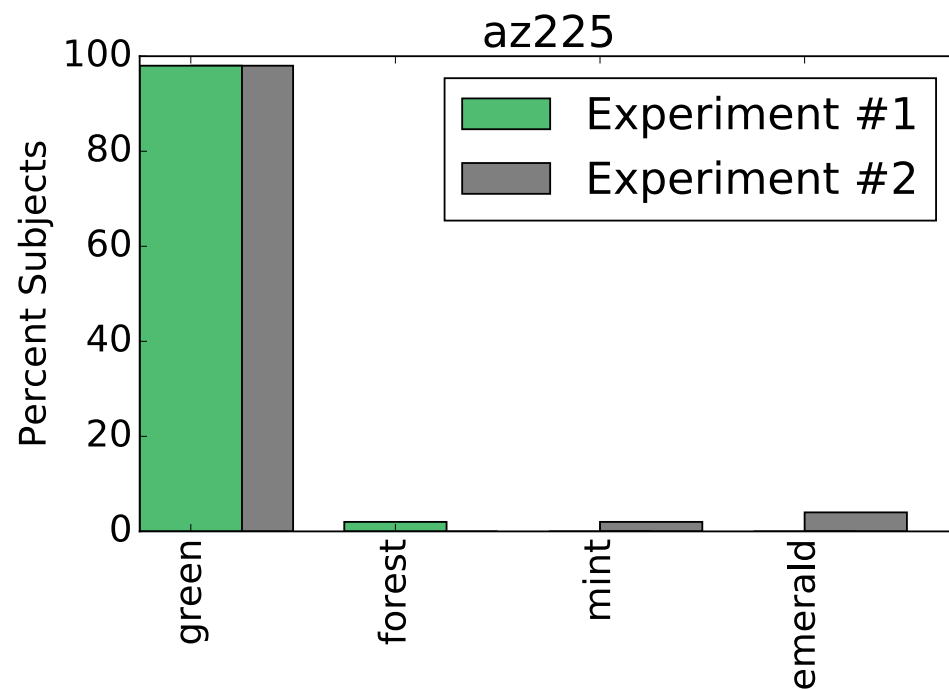
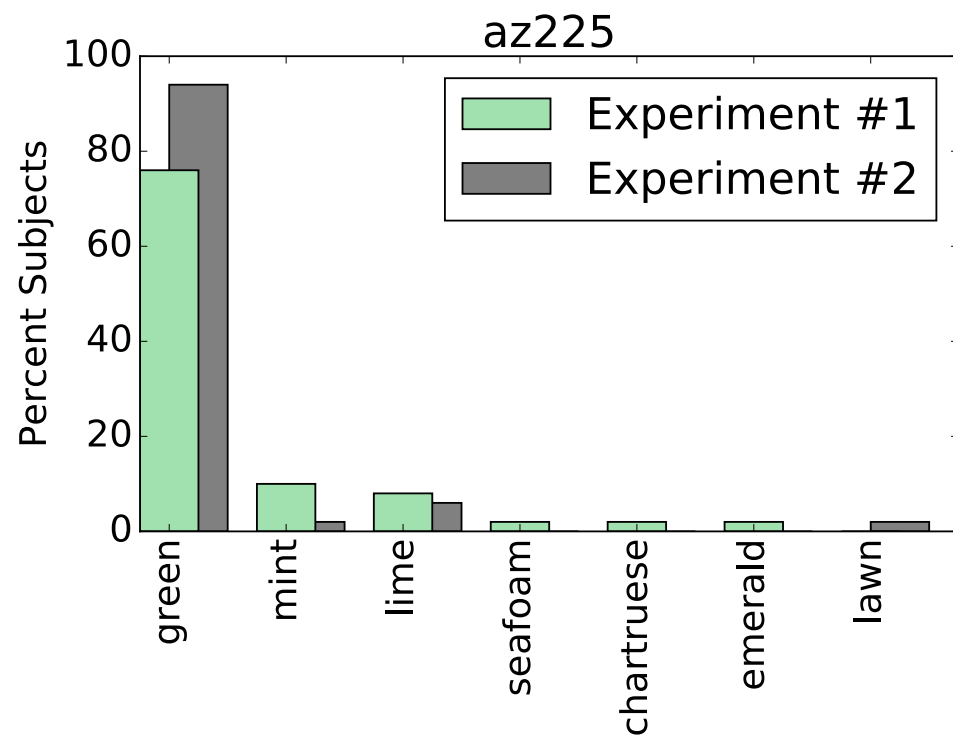
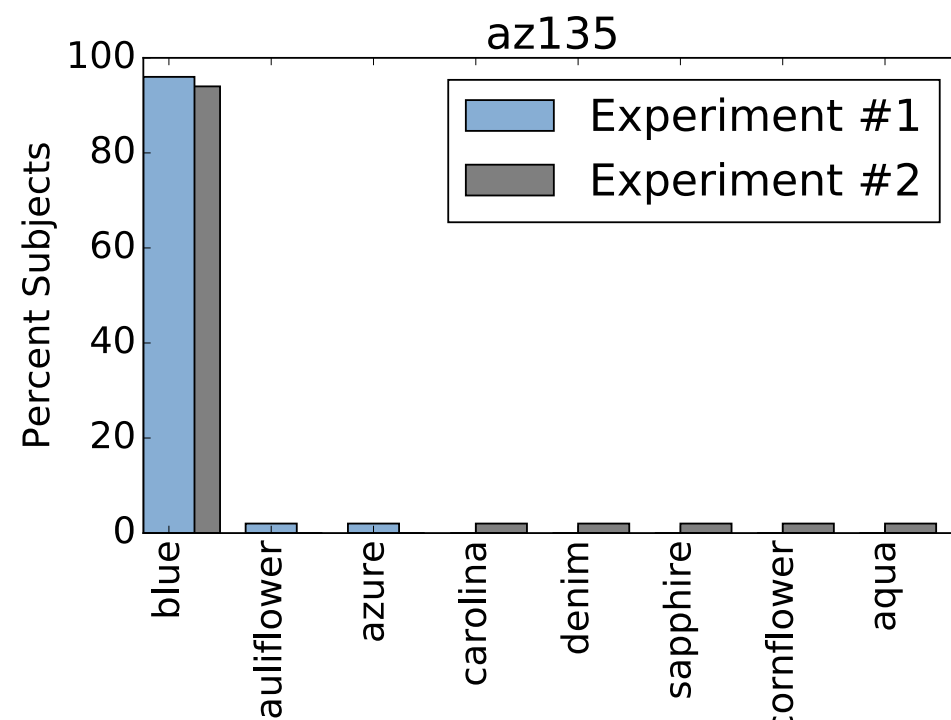
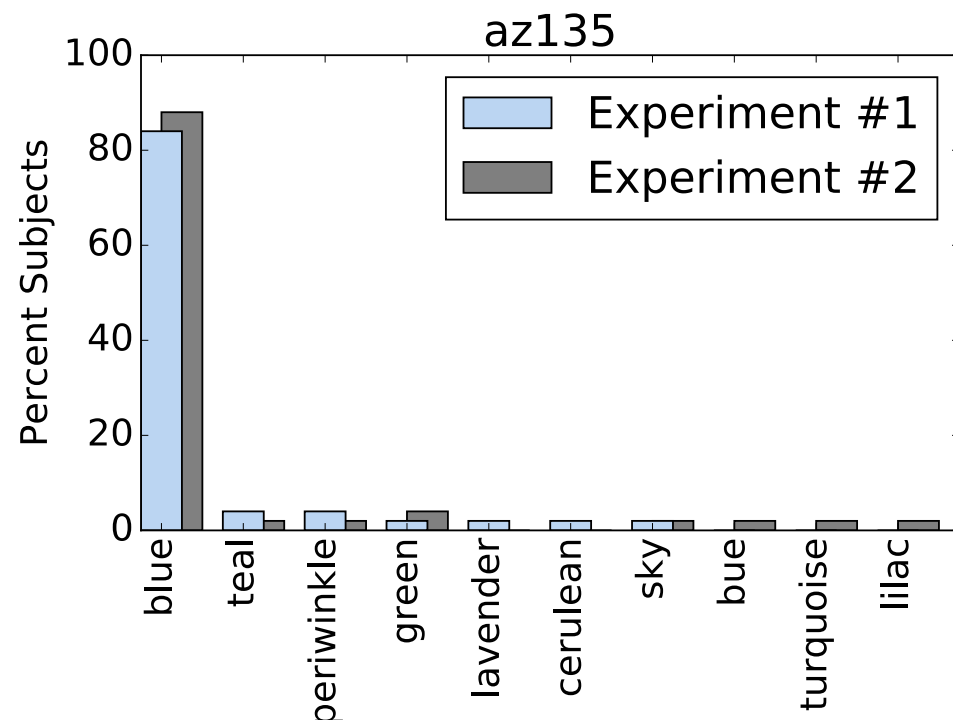


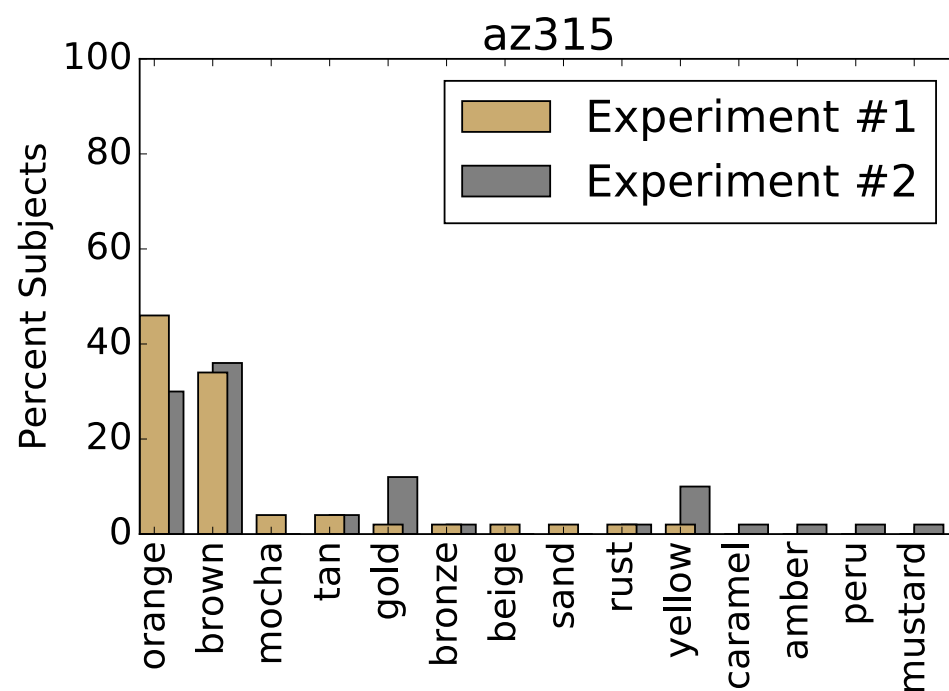
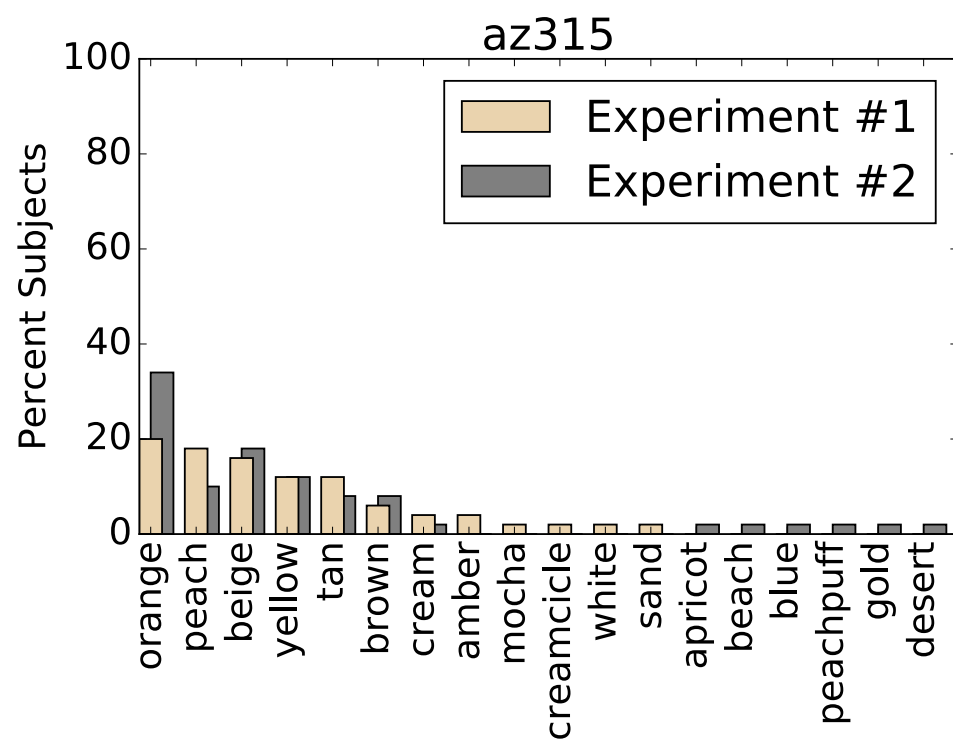
Experiment 2



5.4 Across-Subjects — Percent of Subjects Using a Specific Term for Each Color Spiral







6 Correlation Calculations

6.1 Spearman's Rank Correlation

The correlations between the data — both within and across subjects and conditions — were calculated using MATLAB's corr function, specifically using Spearman's Rho.

6.2 Results

Comparison	Correlation
Average Number of Terms Used (Experiment 1, within-subjects) and Total Number of Terms Used (Experiment 1, across-subjects)	0.9037
Average Number of Terms Used (Experiment 2, within-subjects) and Total Number of Terms Used (Experiment 2, across-subjects)	0.8049
Average Number of Terms Used (Experiment 1, within-subjects) and Average Number of Terms Used (Experiment 2, within-subjects)	0.9701
Total Number of Terms Used (Experiment 1, across-subjects) and Total Number of Terms Used (Experiment 2, across-subjects)	0.8773

7 Concluding Remarks

Based on the data and graphs above, it's clear that certain color terms are more closely associated with certain color spirals. Our ultimate goal was finding closely related color terms and colors to present as stimuli in our MEG experiment; to realize this goal, certain color spirals should be eliminated.

7.1 Chosen Color Spirals

The blue and green (az135 and az225) from the Intermediate DKL Color Hues were ultimately the spirals chosen for stimuli in the MEG scans. These were the spirals with the least variability in responses. These two colors were also attractive options because both luminances produced encouraging results, which makes it simple to test MEG subjects for both luminance and hue data.

7.2 Other Comments

Looking at the data, several generalizations and observations can be made:

- There was not much difference in the within-subjects data analysis between Experiment 1 and 2.
 - The polar representations of the average number of terms used for both conditions are almost identical.
- All the values calculated in 6.2 suggest strong correlation both within and across subjects for each experiment, as well as strong correlation between each experiment.
 - The correlation of the within-subject data for Experiment 1 and 2 was extremely high.
- There was little difference in response between subjects who were tested with both DKL Cardinals and Intermediates and subjects who were tested with only DKL Intermediates.
 - For the az315 spirals, the majority of people use the color term “orange” in both Experiment 1 and 2.