

# Science of Decisions

## Interactive Stroop Effect Experiment

### Analysis Results for Udacity P1 Study

Draft for Review

Prepared for: General Publication

Prepared by:

Laurie S. Reynolds

San Jose, California

[laurie\\_reynolds@yahoo.com](mailto:laurie_reynolds@yahoo.com)

Reviewed by: Udacity P1 Evaluator

# Table of Contents

[Abstract](#)

[Introduction](#)

[QFI #1: Independent and Dependent Variables](#)

[QFI #2: Set of Hypotheses](#)

[Methods](#)

[QFI #3: Descriptive Statistics](#)

[Table 1: Descriptive Statistics](#)

[QFI #4: Supporting Visualizations](#)

[Figure 1: Raw data ordered by Congruent condition](#)

[Figure 2: Congruent and Incongruent plotted together](#)

[Figure 3: Congruent condition histogram](#)

[Figure 4: Incongruent condition histogram](#)

[Figure 5: Timing Differences histogram](#)

[Results](#)

[QFI #5: Report Results](#)

[Figure 6: t-critical region](#)

[Discussion](#)

[QFI # 6: Digging Deeper](#)

[Observations on the study design](#)

[References](#)

[Appendix A: Study Instructions and Rubric](#)

[Appendix B: Experiment Protocol](#)

[Appendix C: Poll results for all samples to date](#)

[Appendix D: Stroop Effect - Raw Data](#)

[Appendix E: Data Preparation and Analysis Workflow](#)

## Abstract

In 1935, John Ridley Stroop published the results of an experimental study on the interference in verbal reactions [Ref 1]. Stroop found that conflicting word stimuli had a considerable interference effect. In the study, subjects are asked to say out loud the name of a color that is displayed under two different conditions. The time to say out loud a series of names was measured in seconds. In one condition, the color is displayed in the color name., ( e.g the word Red is displayed in red.) In the other condition, the color is displayed in a name of a different color. (e.g. the word Green is displayed in red.)

In this study, the Stroop experiment was recreated using an online interactive version. Timing data from participants were analyzed using a repeated measures right tailed t-test. The analysis results corroborated Stroop's original findings that the interference due to word stimuli is significant.

## Introduction

Dr. Eric Chudler, University of Washington, has made available an online interactive version of the Stroop experiment. [Ref. 3] For the purposes of this study, we were provided the timing data for 24 participants under each condition. Appendix A provides background for the study. Appendix B outlines the study protocol.

This investigation aims to determine if the word interference is significant in Chudler's interactive version of the experiment.

### QFI #1: Independent and Dependent Variables

What is the independent variable?
<p>The independent variable is the display condition (Congruent/Incongruent) for the colors. The congruent set contains colors that are displayed with the matching name. The incongruent set contains colors that are displayed with the text of colors that do not match.</p> <p><b>Independent Variable:</b> Display condition: Congruent/Incongruent display of colors with color names</p>

**What is the dependent variable?**

The dependent variable is the reaction time for a participant to read all the colors presented. This is measured in seconds.

**Dependent Variable:** Reaction time in seconds

## QFI #2: Set of Hypotheses

**What is the appropriate set of hypotheses for this task?**

In the original study, the Incongruent condition was found to be significantly more than the Congruent condition. This will serve as the basis for our hypotheses. Our null hypothesis is that the timing of the incongruent condition is equal to or less than the timings for the congruent condition. If the results follow the original study, the null hypothesis will be rejected in favor of the alternative hypothesis which is that the timing of the incongruent condition is significantly greater than the congruent condition.

**Null Hypothesis:**  $\mu_{incongruent} \leq \mu_{congruent}$

**Alternative Hypothesis:**  $\mu_{incongruent} > \mu_{congruent}$

**What kind of statistical test do you expect to perform?**

In this study, timing data is from 24 participants that took the test under the two conditions. We do not have any population statistics. Given these conditions, we will use a repeated measure right tailed t-test

**Statistical Test:** Repeated measures right tailed t-test

## Methods

### QFI #3: Descriptive Statistics

**Report some descriptive statistics regarding this dataset. Include at least one measure of central tendency and at least one measure of variability**

Table 1 shows measures of central tendency and variability for the data along with the calculated difference and percent difference.

As you can see the means and medians for each set are fairly close to each other which indicate that the data is relatively normally distributed. Additionally, for all the data except the Incongruent, the min and max are each within one to two standard deviations from the mean.

As we are estimating the population standard deviation, we apply Bessel's correction when calculating the sample standard deviation

In addition, we examine what percentage of the original time the difference represents in the Percent Diff column. We see that the average percentage difference is 63.42%. In other words, it takes an average of 63.42% longer to say out loud the names of colors in the Incongruent condition than under the Congruent condition.

[Ref 4 is a link to the Github repository which contains the IPython notebook that was used to perform the analysis. It also contains an excel spreadsheet which was used to double check the calculations.

Table 1: Descriptive Statistics

	Congruent	Incongruent	Difference	Percent Diff
count	24	24	24	24
mean	14.05	22.02	7.96	63.42
median	14.36	21.02	7.67	63.42
min	8.63	15.69	1.95	9.83
max	22.33	35.26	21.92	177.21
sstd	3.56	4.80	4.86	42.32
se	0.73	0.98	0.99	8.64

## QFI #4: Supporting Visualizations

**Provide one or two visualizations that show the distribution of the sample data. Write one or two sentences noting what you observe about the plot or plots**

Figure 1 represents the raw data which is listed in table form in Appendix D. The data is sorted by Congruent timings. As we can see, the Incongruent timings are all greater than the congruent timings. The differences between the Congruent and corresponding Incongruent times vary in the size of the difference with two distinct outliers.

Figure 2 shows the Congruent and Incongruent conditions overlaid on the same graph. This shows the differences between the means as well as the difference between the spread of the distribution of the values.

Figures 3 - 5 show that the Congruent, Incongruent, and Differences are somewhat normally distributed.

Figure 1: Raw data ordered by Congruent condition

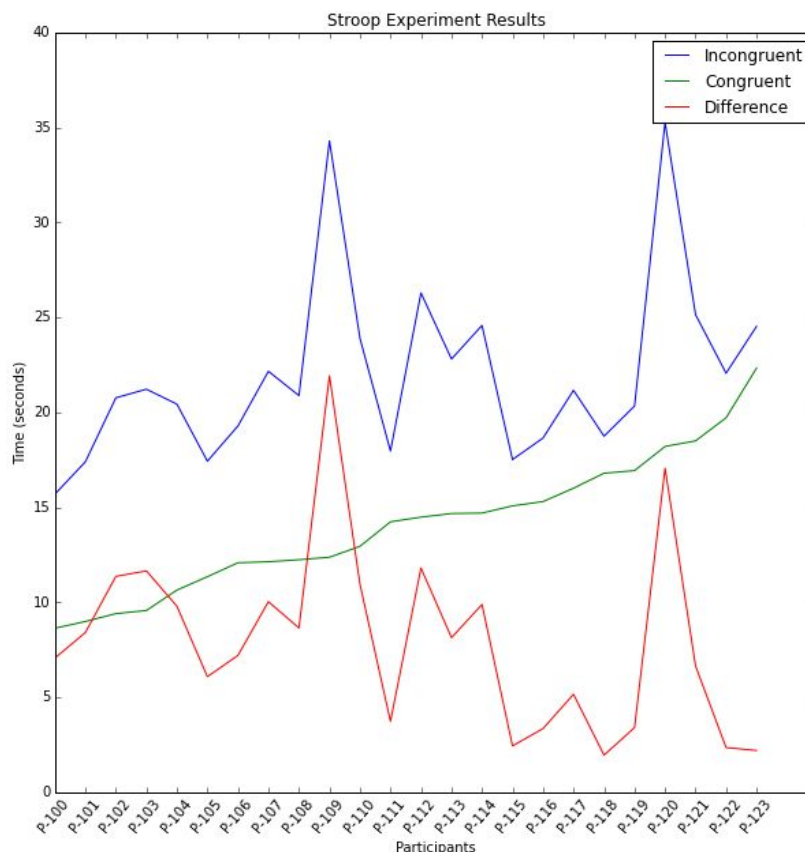


Figure 2: Congruent and Incongruent plotted together

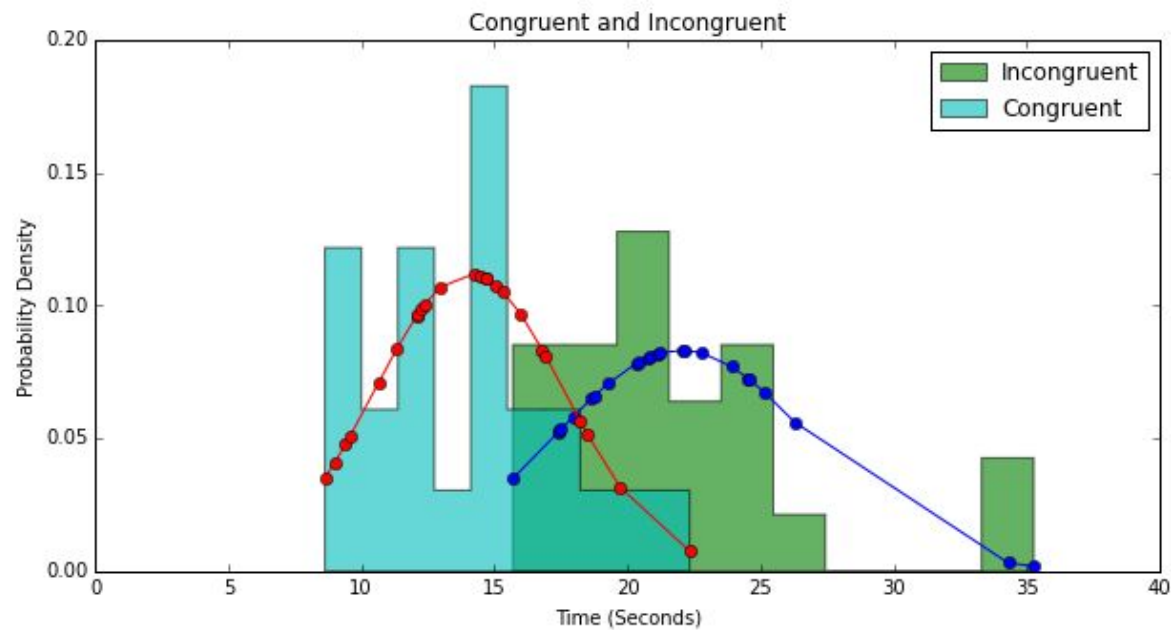


Figure 3: Congruent condition histogram

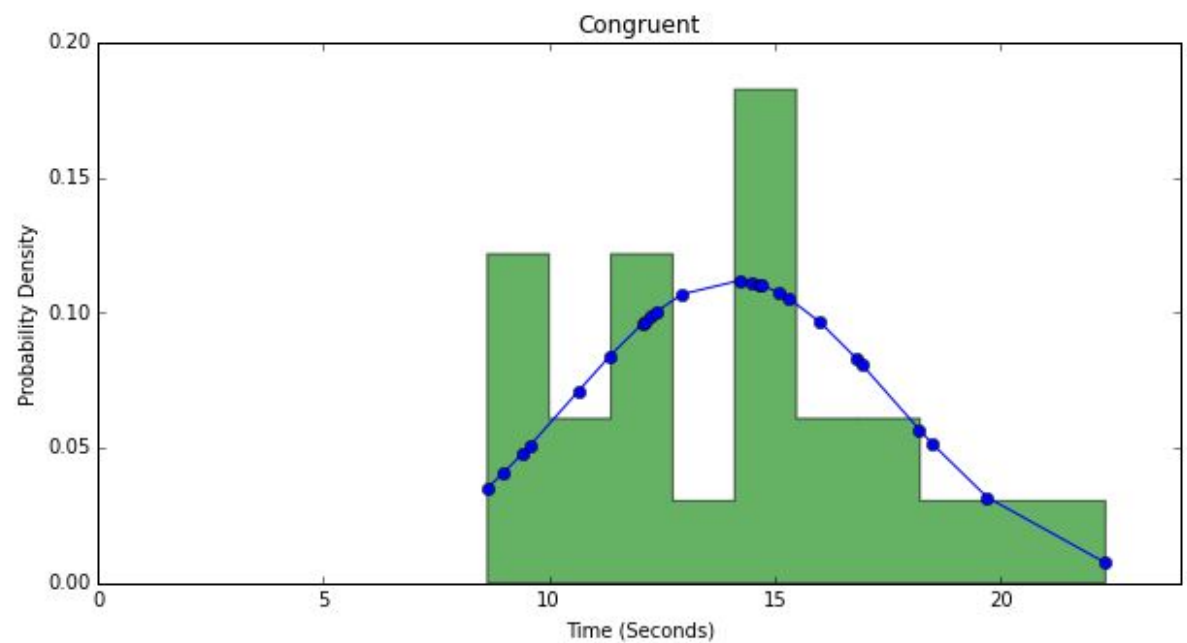


Figure 4: Incongruent condition histogram

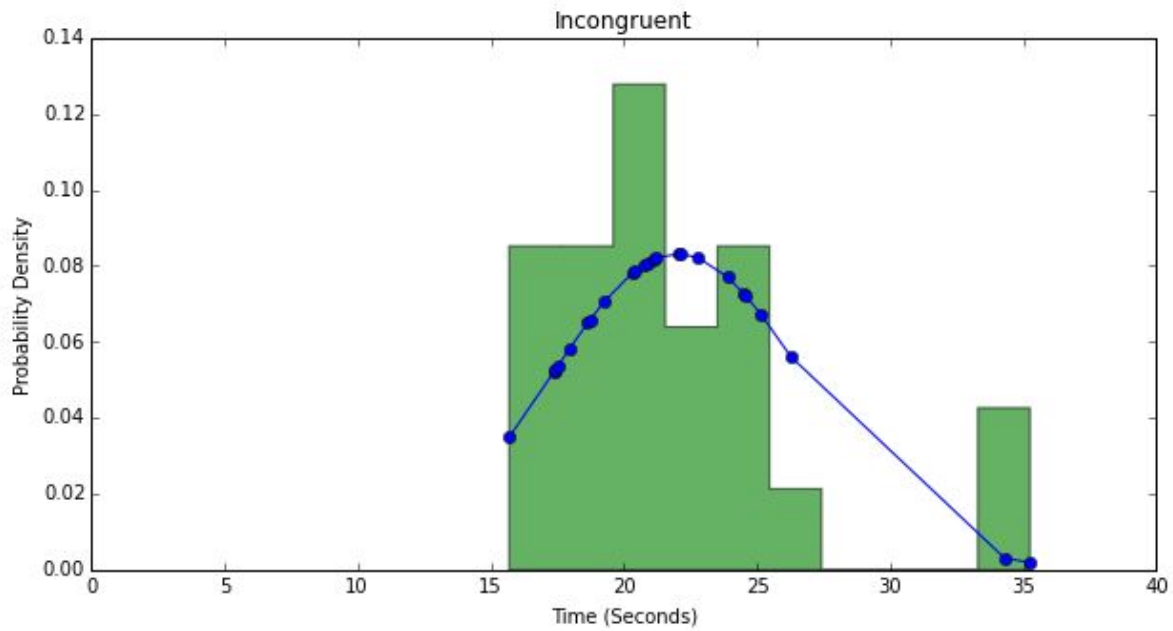
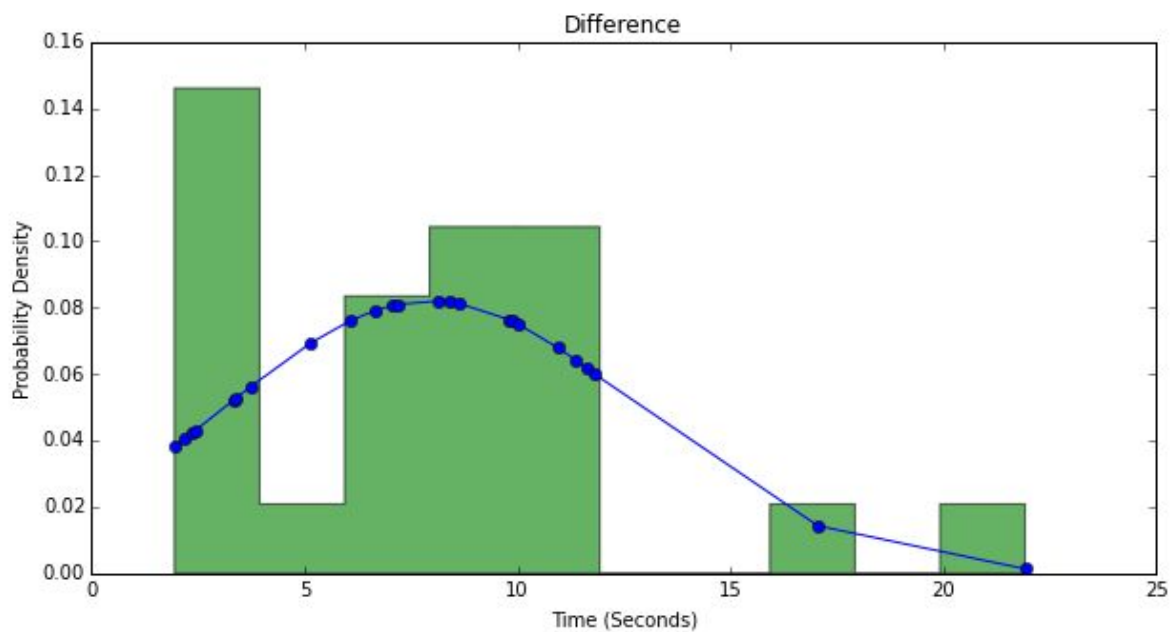


Figure 5: Timing Differences histogram





## Results

### QFI #5: Report Results

#### Now, perform the statistical test and report your results

Figure 6 shows the t critical region for  $\alpha = 0.05$ ,  $df = 23$ . This illustrates that the t -statistic of 8.02 is well beyond the t-critical point on the plot. The effect size measures of Cohen's  $d = 1.64$  and  $r^2 = .74$  further demonstrate the significance of the results.

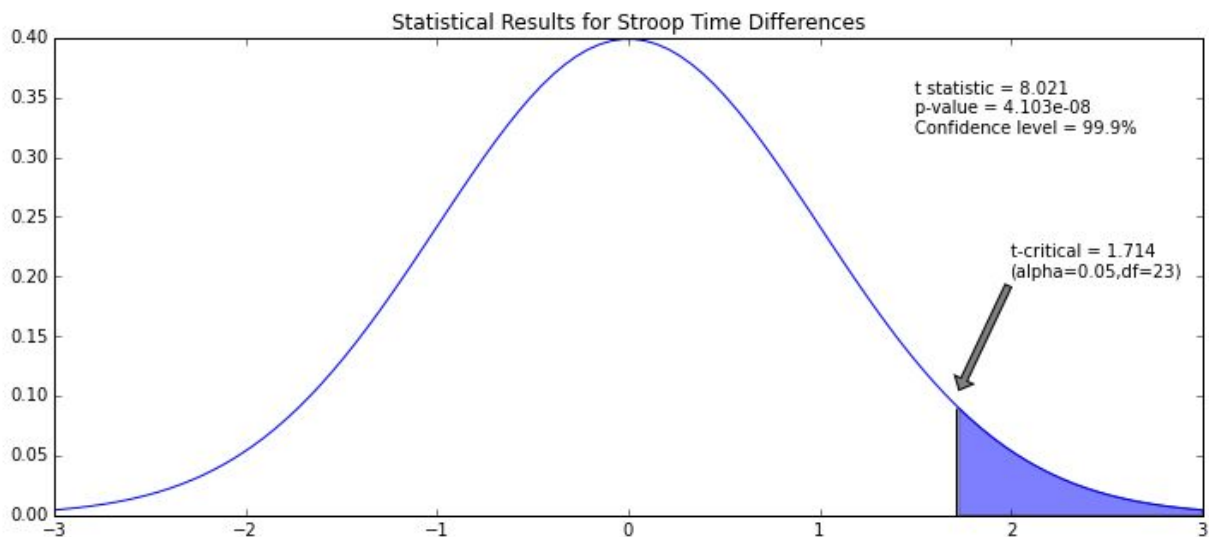
$t(23) = 8.02$ , p-value =  $4.10\text{e-}08$ , right-tailed

Confidence interval on the mean difference; 99.9%, CI = (6.262741 to 9.666843)

Cohen's  $d = 1.64$

$r^2 = .74$

Figure 6: t-critical region



#### What is your confidence level and your critical statistical value?

**Confidence Level:** 99.9%

**Critical Statistic Value:** 1.714,  $\alpha = 0.05$ ,  $df = 23$

**Do you reject the null hypotheses or fail to reject it?**

Given the strong statistical effect size measures, we reject the null hypothesis that  $\mu_{incongruent} \leq \mu_{congruent}$  and accept our alternative hypothesis that  $\mu_{incongruent} > \mu_{congruent}$

**Come to a conclusion in terms of the experiment or task.**

Word stimuli has a significant effect on the timing of saying out loud color names.

**Did the results match up with your expectations**

This has corroborated Stroop's findings from the original study and matches our expectations.

## Discussion

### QFI # 6: Digging Deeper

**What do you think is responsible for the effects observed?**

It appears that our verbal reading abilities are more dominant and faster than the recall abilities for the names of colors. Hoi-Chung Leung, et al in "An Event-related Functional MRI Study of the Stroop Color Word Interference Task" [Ref 2] measured the brain activity while a participant was performing the Stroop experiment. They found different areas of the cortex were activated when performing the test under different conditions.

**Can you think of an alternatives or similar task that would result in a similar effect?**

For an alternative to be effective, it would need to be universally recognized and the choices should be similar enough so that there is room for interference if labeled differently. The following are alternatives to consider

- Fruits ( apple, oranges, pears, grape, melon)

- Emotions ( happy, mad, sad, none)

The following might not be good candidates as the association with the shape may be stronger than the name.

- Common signs (stop, man, woman, yield)
- Transportation (car, plane, boat, bus, truck)

## Observations on the study design

The following factors of the study design may have an impact on the study results.

1. Self reporting of results. The study relies on participants saving and recording the timing results. Potential errors that could occur are as follows:
  - Incomplete reporting. This is seen by the difference in the number or results for each set.
  - Swapping of values for the incongruent and congruent cases.
  - Estimation of values. In cases where the participant did not remember to record the value presented, some may enter an estimate for completeness.
  - Adjustment of values. There is the potential that a participant could adjust the values for what they feel may be expected for the study.
2. Manual calculation of differences: The study relies on participants calculating the difference between the two timings. Potential errors that could occur are as follows:
  - Miscalculation. There are two types of miscalculations that could occur.
    - i. Incorrect difference value
    - ii. Swapping the operands during the subtraction.
3. Entry of values. The recording of timings is done by selecting from a set of choices that describe time ranges. It is possible that a participant could accidentally select a choice that does not correspond to the actual values.
4. Multiple results in one session. The study does not detect whether or not a participant takes the test multiple times in one session which could impact the results

## References

1. Stroop, John Ridley (1935). "Studies of Interference in Serial Verbal Reactions", *Journal of Experimental Psychology*, Vol 18(6), Dec 1935, 643-662. Retrieved from: York University Classics in the History of Psychology - <http://psychclassics.yorku.ca/Stroop/>
2. Hoi-Chung Leung, Pawel Skudlarski, James C. Gatenby, Bradley S. Peterson, and John C. Gore. "An Event-related Functional MRI Study of the Stroop Color Word Interference Task". *Cerebral Cortex* (2000) 10 (6): 552-560.doi: 10.1093/cercor/10.6.552 Retrieved from Cerebral Corted - <http://cercor.oxfordjournals.org/content/10/6/552.full>
3. Chudler, Eric H. Executive Director, [Center for Sensorimotor Neural Engineering](#)

### Sources

- Interactive Stroop Effect Experiment: <https://faculty.washington.edu/chudler/java/ready.html>
  - Stroop Effect: <https://faculty.washington.edu/chudler/words.html#seffect>
4. Github repository for analysis artifacts: <https://github.com/golgistudio/udacity-data-analyst-stroop>

## Appendix A: Study Instructions and Rubric

These instructions were provided by Udacity

### Background Information

In a Stroop task, participants are presented with a list of words, with each word displayed in a color of ink. The participant's task is to say out loud the *color of the ink* in which the word is printed. The task has two conditions: a congruent words condition, and an incongruent words condition. In the *congruent words* condition, the words being displayed are color words whose names match the colors in which they are printed: for example **RED**, **BLUE**. In the *incongruent words* condition, the words displayed are color words whose names do not match the colors in which they are printed: for example **PURPLE**, **ORANGE**. In each case, we measure the time it takes to name the ink colors in equally-sized lists. Each participant will go through and record a time from each condition.

Table: Study Rubric provided by Udacity

Study Rubric	
Criteria	Meets Specifications
<b>Question 1: Identify variables in the experiment</b>	Question response correctly identifies the independent and dependent variables in the experiment.
<b>Question 2: Establish a hypothesis and statistical test</b>	An appropriate hypothesis test has been stated along with an appropriate statistical test to apply to collected data, with appropriate justification.
<b>Question 3: Report descriptive statistics</b>	Descriptive statistics, including at least one measure of centrality and one measure of variability, have been computed for the dataset's groups.
<b>Question 4: Plot the data</b>	One or two visualizations have been created that show off the data, including comments on what can be observed in the plot or plots.
<b>Question 5: Perform the statistical test and interpret your results</b>	A statistical test has been correctly performed and reported, including test statistic, critical test statistic or p-value, and test result. The test results are interpreted in terms of the experimental task performed.
<b>Question 6: Digging deeper and extending the investigation</b>	Hypotheses regarding the reasons for the effect observed are presented. An extension or related experiment to the performed Stroop task is provided, that may produce similar effects. <b><i>This question is optional and does not need to be answered in order to meet project specifications</i></b>

## Appendix B: Experiment Protocol

The following outlines the Stroop Effect protocol used for this experiment.

1. Word set #1 is displayed - Congruent - 25 color names shaded with the same color as the name.
2. Participant reads each color out loud and selects [Finish] when complete.
3. The application displays the amount of time the page with word set #1 has been displayed.
4. Participant records the time.
5. Participant selects [Continue Experiment] to continue
6. Word set # 2 is displayed - Incongruent - The same color names are displayed. This time, they are shaded a color that does not match their names.
7. Participant says each color (not the text name) out loud and selects [Finish] when complete.
8. The application displays the amount of time the page with word set #2 has been displayed.
9. Participant records the time.
10. Participant selects [Enter and compare your results] to continue
11. Participant is asked to select the range in which each time falls and select [Submit] after each time is entered.
  - Word set #1
  - Word set #2
  - Difference - Participant is asked to calculate the difference by subtracting the first word set time from the second word set time.
12. After each time is submitted, the results for the current set of participants is displayed in tabular and pie chart form.

### Screenshot from example run

1) Instructions: Participants are directed to navigate to the following URL to start the test: <https://faculty.washington.edu/chudler/java/ready.html>

## Interactive Stroop Effect Experiment

In this experiment you are required to say the color of the word, not what the word says. For example, for the word, **RED**, you should say "Blue."

As soon as the words appear on your screen, read the list as fast as you can. When you have finished, click on the "Finish" button. The time it took you to read all of the words will be shown. If you want to try the same set of words, click on the "reload" button of your browser. If you want to continue with the experiment, click on "Continue Experiment."

[Go to the first test.](#)

2) Word Set # 1 - <https://faculty.washington.edu/chudler/java/timesc.html>

### Word Set #1

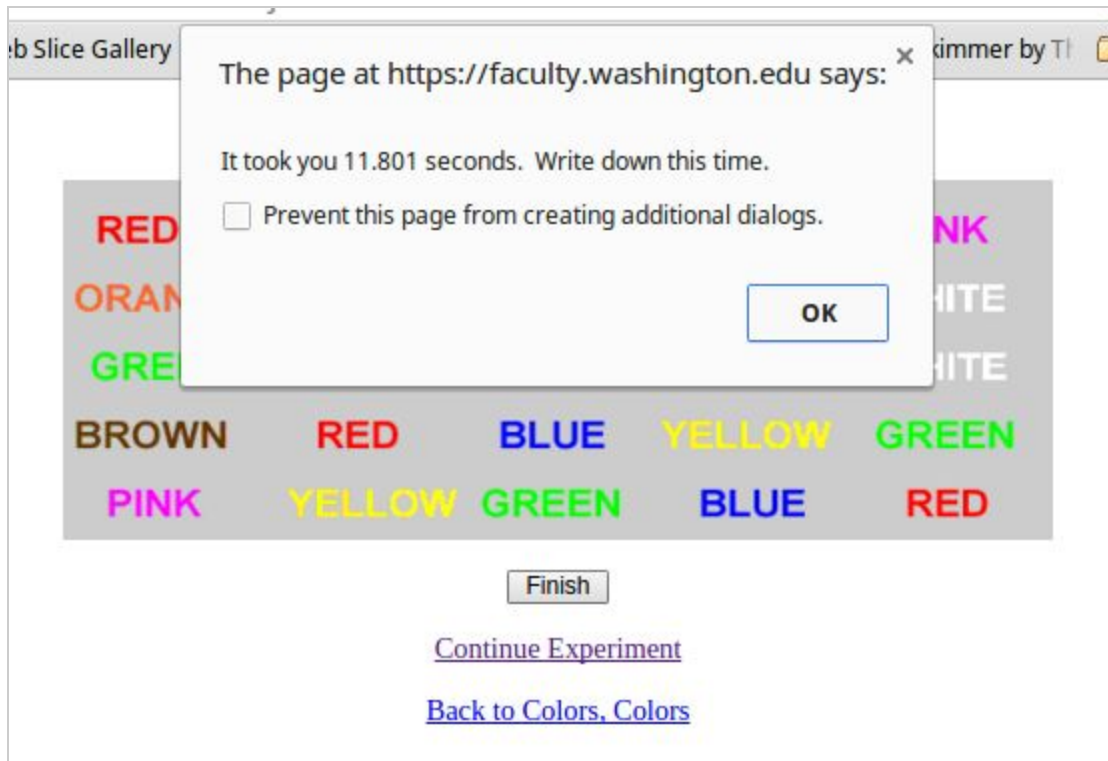
RED	GREEN	BLUE	YELLOW	PINK
ORANGE	BLUE	GREEN	BLUE	WHITE
GREEN	YELLOW	ORANGE	BLUE	WHITE
BROWN	RED	BLUE	YELLOW	GREEN
PINK	YELLOW	GREEN	BLUE	RED

Finish

[Continue Experiment](#)

[Back to Colors, Colors](#)

3) Timing results for Word Set #1.

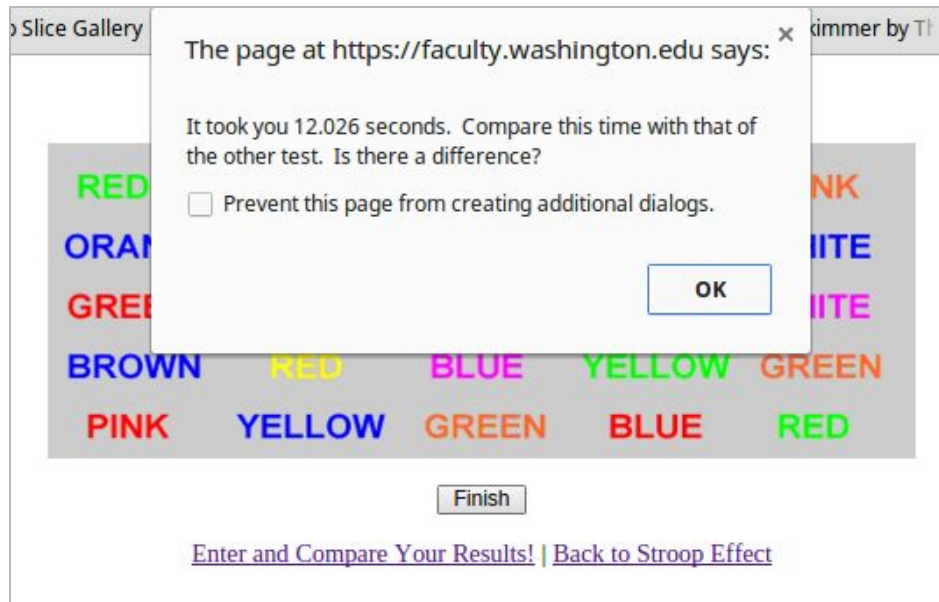


4) Word Set #2 - <https://faculty.washington.edu/chudler/java/timestc.html>



5) Timing results for Word Set #2





6) Record Results - <https://faculty.washington.edu/chudler/java/strvote.html>

**Stroop Effect**

Enter your times for the first and second word sets.

Time to Complete Word Set 1

- ☐ 5-10 sec
- ☐ 10-15 sec
- ☐ 15-20 sec
- ☐ 20-25 sec
- ☐ 25-30 sec
- ☐ 30-35 sec
- ☐ 35-40 sec

Time to Complete Word Set 2

- ☐ 5-10 sec
- ☐ 10-15 sec
- ☐ 15-20 sec
- ☐ 20-25 sec
- ☐ 25-30 sec
- ☐ 30-35 sec
- ☐ 35-40 sec

What is the difference between your Word Set 2 time and Word Set 1 time (Difference Time)

- ☐ Word set 2 time is LESS than word set 1 time
- ☐ 0-5 sec
- ☐ 5-10 sec
- ☐ 10-15 sec
- ☐ 15-20 sec
- ☐ 20-25 sec

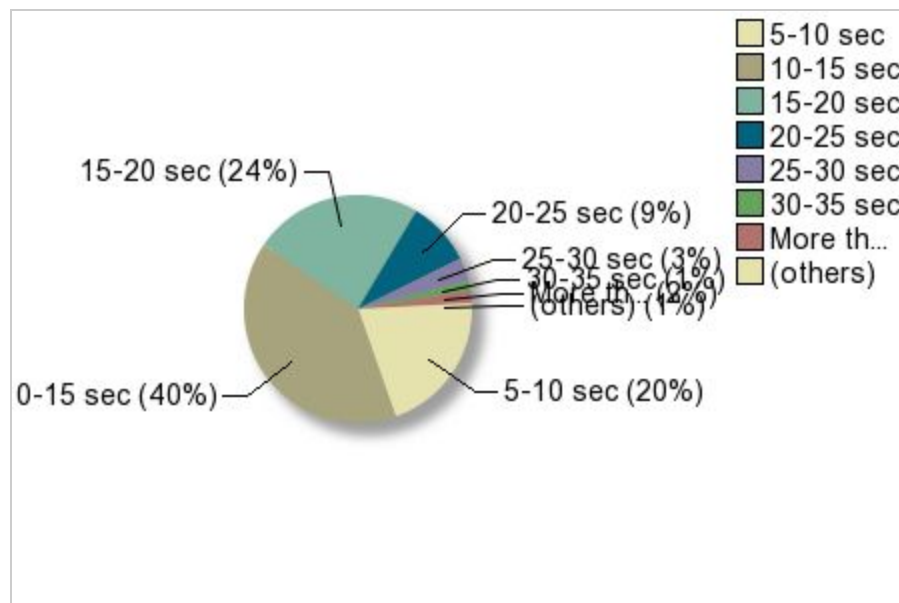
7) After selecting [Submit] for each time entered, the cumulative results are displayed

## Appendix C: Poll results for all samples to date

(Recorded November, 2015 - 2:30 pm PST)

### Time to Complete Word Set 1

Poll Results		
Total submissions: 105300		
<i>Answer choice</i>	<i>Frequency</i>	<i>Percentage</i>
5-10 sec	20651	19.61%
10-15 sec	41991	39.88%
15-20 sec	25363	24.09%
20-25 sec	9640	9.15%
25-30 sec	3461	3.29%
30-35 sec	1457	1.38%
35-40 sec	849	0.81%
More than 40 sec	1888	1.79%

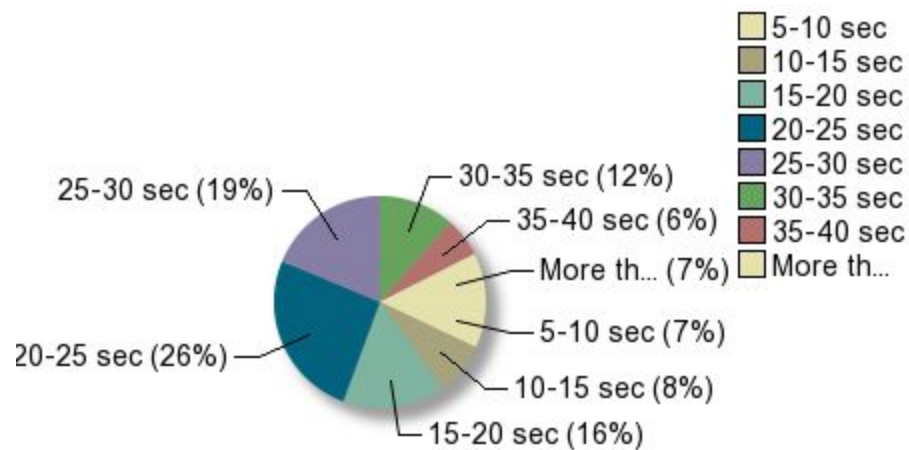


### Time to Complete Word Set 2

## Poll Results

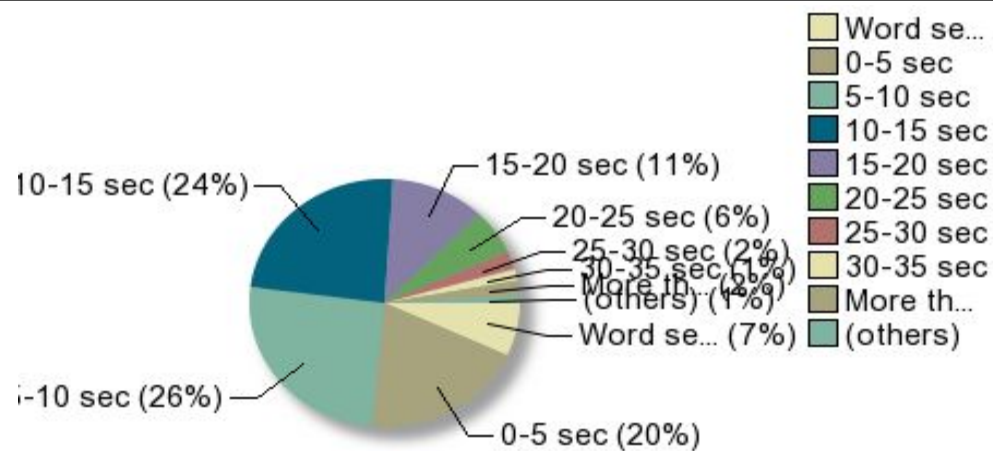
Total submissions: 105610

Answer choice	Frequency	Percentage
5-10 sec	7442	7.05%
10-15 sec	8432	7.98%
15-20 sec	16448	15.57%
20-25 sec	27117	25.68%
25-30 sec	19771	18.72%
30-35 sec	12207	11.56%
35-40 sec	6303	5.97%
More than 40 sec	7890	7.47%



## Poll Results for the differences

Poll Results		
Total submissions: 181097		
Answer choice	Frequency	Percentage
Word set 2 time is LESS than word set 1 time	12571	6.94%
0-5 sec	35421	19.56%
5-10 sec	46261	25.54%
10-15 sec	43427	23.98%
15-20 sec	20550	11.35%
20-25 sec	10617	5.86%
25-30 sec	4329	2.39%
30-35 sec	2683	1.48%
35-40 sec	1690	0.93%
More than 40 sec	3548	1.96%



## Appendix D: Stroop Effect - Raw Data

Sample Results were provided by Udacity

Congruent	Incongruent
12.079	19.278
16.791	18.741
9.564	21.214
8.63	15.687
14.669	22.803
12.238	20.878
14.692	24.572
8.987	17.394
9.401	20.762
14.48	26.282
22.328	24.524
15.298	18.644
15.073	17.51
16.929	20.33
18.2	35.255
12.13	22.158
18.495	25.139
10.639	20.429
11.344	17.425
12.369	34.288
12.944	23.894
14.233	17.96
19.71	22.058
16.004	21.157

## Appendix E: Data Preparation and Analysis Workflow

### Stroop Sample Result Data

The sample result data of 23 samples - two columns (Congruent/Incongruent) was provided in the form of a comma separated values. The following steps were taken to prepare the data for analysis.

1. Assigned sequential participant ids starting with P-100
2. Calculated the difference between the Congruent and Incongruent timings using the built in function tools in Google Sheets
3. Saved data as a comma separated file
4. Converted data to JSON format using Convert CSV to JSON (see table Tools for tools used)
5. Analysis and visualizations done in Jupyter, a python development environemnt.
6. Duplicate calculations were performed in Google Sheets.

Table: Data Preparation and Analysis Tools

Data Preparation and Analysis Tools		
Tool name	Source	Description
CSV to JSON	<a href="http://www.convertcsv.com/csv-to-json.htm">http://www.convertcsv.com/csv-to-json.htm</a>	
Google Sheets	<a href="https://www.google.com/sheets/about/">https://www.google.com/sheets/about/</a>	
Jupyter	<a href="http://jupyter.org/">http://jupyter.org/</a>	Python development environment