

The Impact of Red on Performance on Anagrams

IB Psychology SL

IB code: jwt991

Group Members: jwy162, jwx890, jwv020

Date Submitted: January 24, 2022

Word Count: 2200

## Table of Contents

Introduction.....	3
Exploration.....	5
Analysis.....	8
Evaluation.....	10
References.....	13
Appendix.....	14

## INTRODUCTION

Learned association is a process in classical conditioning where an unconditioned stimulus that causes an unconditioned response is paired repeatedly with a neutral stimulus so the neutral stimulus, now the conditioned stimulus, will also elicit the unconditioned response, now the conditioned response. Priming is when “prior exposure to a stimulus facilitate its subsequent identification and classification” (Horner et al. 2008). Bargh’s automotive model explains how goals inadvertently cause increased sensitivity to certain stimuli using unconscious priming (Chatrand et al. 1996). Automatically activated goals can govern behaviors without any purposeful involvement of the individual.

The study of *Color and Psychological Functioning: The Effect of Red on Performance Attainment* by Andrew Elliot uses learned association where “use of red to mark errors, warn of negative possibilities, and so on in society ... [causes] tendency to view red as a danger signal” (Elliot et al. 2007). The study aims to find the relation between red and psychological functioning. In Elliot’s study, the participants were split into 3 groups: red, green, or black. Each participant looked at their number which was written in their designated color before starting the anagram test. An anagram is a word formed by rearranging the letters of another. It resulted in the red condition performing worse than the green and black conditions, so Elliot concluded that viewing red led to decreased performance on the test.

Our investigation aimed to find how manipulating an anagram test text color affects academic performance and psychological functions like problem solving in high school students. This is relevant because if color can impact performance in an educational setting, testing environments should be adjusted accordingly so students perform their best, especially on important national or international assessments. Perhaps, this can be generalized to the workplace where changes in wallpaper can increase efficiency and productivity.

Our experiment is linked to classical conditioning by using the learned association between red and decreased performance. The unconditioned stimulus, danger, leads to the unconditioned response of decreased performance. Repeated pairing of red, neutral stimulus, with the unconditioned stimulus, red will become the conditioned stimulus and cause conditioned response of decreased performance. Linking to Bargh’s automotive model, the goal of completing the anagram test will cause the students to be more sensitive to red and its relation to failure which further decreases the participant’s performance.

Our null hypothesis is participants showed no significant difference in the number of anagrams solved correctly regardless of the color of the text the anagram test.

Our research hypothesis is if the anagram test text color is red then the number of anagrams solved correctly in ten minutes would decrease. Independent variable is color of text, red or black, and dependent variable is number of correctly solved anagrams.

## EXPLORATION

### Research Design

Our experiment uses independent design and randomly assigned the participants into two groups. The independent variable is the color of the text: experimental group is red and control group is black. The dependent variable is the number of anagrams solved correctly.

We chose independent design because giving the same anagrams test with Times New Roman, 28 sized font simultaneously, we limit confounding variables like different levels of difficulty and times of day, increasing internal validity. Participants are exposed to one version of test. We prevent practice effect where participants perform better by seeing the test twice and order effects as taking the black version first may increase performance or vice versa. Demand characteristics would arise if the participants are exposed to both version, where they may do better or worse to accommodate our hypothesis if they figured out the purpose of the experiment. Then results won't be a valid representation.

### Sampling

We used opportunity sampling by finding classes that are available during our class period, ensuring that there is no need to coordinate when and where the experiment will take place, making the process time efficient and convenient.

### Participants

Our target population is high school student and our sample of twenty-two participants with nine male, eleven female, and two non-binary students ranging from seventeen to eighteen years old is from a magnet high school with diverse population of sixteen hundred students in the United States. Most are in the International Baccalaureate or Advanced Placement programs, tracks with rigorous and challenging courses.

We chose these participants because they fit our target population and the effect of red on academic performance is most relevant in an educational setting. However, the overall results may be higher than the general population because this magnet school has more rigorous courses and higher

expectations. If participants take art classes, it may affect our results since they have exposure to red when painting things unrelated to danger, decreasing red's association with danger, so red text may have less effect on decreasing their performance.

We assigned each participants a number alphabetically. Using a random number wheel, we assigned half to the experimental group and the remaining half to the control group. This decreases participant variability as researcher bias is present if experimenters chose the groups.

### Controlled variables

The participants received the same anagram test within the same time frame, ten minutes, and same classroom. The text is in the same font, size, and boldness. The environment is controlled since some may work better during different times of day. The same instructions were read aloud so each knew what to do, and different talking speed or accents won't affect the participants.

If the anagrams were not the same, then participants with 5-letter anagrams have an easier test compared to participants with 6-letter anagrams causing one group to have lower scores caused not by the color of text, adding a third-variable problem. Likewise for other variables like time limit and text font. By limiting the confounding variables, internal validity increases.

### Materials

- consent form (Appendix A)
- debriefing statement (Appendix B)
- standardized instructions (Appendix C)
  - to control confounding variables like different wordings
- timer
  - increasing our control over experiment by ensuring the same time limit for everyone
- class roster
- excel
  - keep track of data and calculating mean and standard deviation
- anagrams ("Random Word Generator") (Appendix D and E)

- a simple and easy to score activity and so experiment is as close to the study replicated as possible
- number picker wheel ("Picker Wheel", 2021 Version 76)
  - prevent researcher bias and to randomly assign the participants so that participant variability decreases like people with more exposure to anagrams are not in the same group
- raw data sheet (Appendix G)

### Procedures

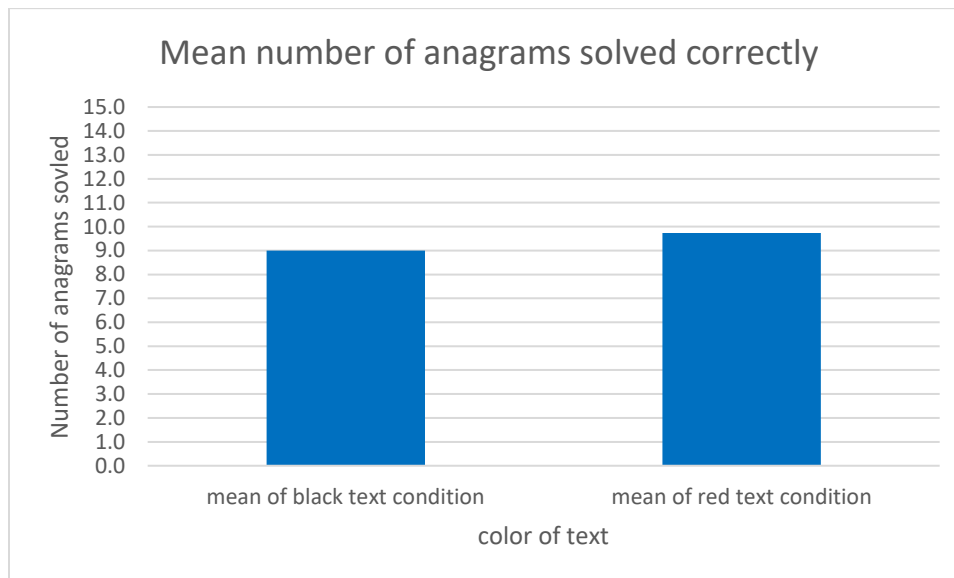
1. get signed consent form
2. assign number alphabetically; randomly assign into two conditions using random number wheel
3. read standardized instructions
4. distribute corresponding tests for each group
5. start timer for 10 minutes
6. collect tests
7. debrief

### Ethical considerations

Our group followed all APA ethical guidelines, including giving informed consent, anonymity, right to withdraw, debriefing and causing no harm, physically or psychologically.

## ANALYSIS

Mean number of anagrams solved correctly		
	black text condition	red text condition
Mean	9.000	9.727
Standard Deviation	3.795	2.370



Our experiment had ratio data and both data sets have normal distribution, so we chose to use the mean and standard deviation. The mean of black condition is nine correctly solved anagrams while the red condition is 9.72. The standard deviation of the red condition was 2.370 while the black condition was 3.975. So, the control group had a wider range of scores while scores for the experimental group were more condensed, thus more reliable. On average the red condition performed better than black condition, which conflicted with our hypothesis that the black condition will perform better.

We used the Mann Whitney U Test to determine the significance of our data because our experiment used independent measures and had a small sample size with less than 25 participants. The Mann Whitney U test indicated no significant difference between the number of anagrams solved correctly for the black text group (Median = 9) and the red text group tested (Median = 10),  $U = 69$ ,  $p = 0.298$ . Based on my inferential statistics, I reject my research hypothesis and accept the null hypothesis since the p-value (0.298) is greater than 0.05.



As a result, we can conclude at a  $p = 0.298$  level of significance that changing the text color of the anagrams from black to red had no effect on the number of anagrams solved and academic performance of high school students.

## EVALUATION

In our results the red condition had higher scores on average than the black condition. The red group solved 9.727 anagrams correctly on average while black group solved 9. We concluded there was no significant results at  $p = 0.298$  level of significance.

These results contrast with what the learned association model predicted. In the model, the unconditioned stimulus, danger, led to the unconditioned response, decreased performance. Various association of red with danger and failure like stop signs and red pens for grading, cause red to become a conditioned stimulus causing the now conditioned response of decreased performance. In Bargh's automotive model, unconscious priming explains how participants are more sensitive to association of red with danger causing further decrease in performance when having goal of excelling on the test (Chatrand et al. 1996).

In Elliot's study, participants were exposed to the color for a short period of time before the anagram test, and on average the red group scored lower than the black group. However, participants were constantly exposed to red in our experiment causing difference between Elliot's and our results. The effectiveness of red in eliciting anxiety and worse performance may decrease once the participants were accustomed to the color.

The original study used college undergraduates while we used high school seniors. The difference in age and mental ability may account for the differences in the results. Our participants from a magnet school are accustomed to working under pressure and stress may increase their performance instead. Other confounding variables like vocabulary between the two conditions all contribute to why the experimental group performed better.

### Research Design

One strength of implementing independent measures was eliminating practice effects. Participants won't do better on the second test due to prior exposure or extra time if exposed to only one version. Another strength is limiting demand characteristics. It's harder to decipher the experiment's purpose if participants are unaware that the test has two versions. Otherwise, participants may purposefully do worse on the red test through speculation of experimenter's expectancies.

One limitation was more confounding variables existed among the participants. Experimental group may contain participants with better vocabulary or experience with anagrams causing higher scores and increasing probability that results were due to participant variability than manipulation of the text color. We can modify the experiment by using a matched-pairs design. A pre-test can determine their vocabulary level to match them into pairs and then randomly assign into groups. This increases internal validity and decreases possibility that variables like experience and vocabulary affect our results.

### Sample

We used opportunity sampling to find 22 non colorblind students from an ethnically diverse magnet high school in United States. A strength of using opportunity sampling was increased efficiency. Coordinating with participants is easier without need to match the schedules of everyone. Another strength is that the sample consisted of seniors. Seniors may take the test more seriously than freshmen, so results are more reliable.

A limitation is the characteristics specific to the sample. Our sample is in English class so recent exposure to vocabulary or being in the English classroom may lead to higher scores compared to what the population may score, since our test involves spelling and vocabulary. Another characteristic is that our sample is all seniors, so we can't generalize our results to high schoolers. Seniors may have more education and higher vocabulary than underclassmen. We can modify by using a volunteer sample to get participants from different grades and subjects. Increased variety among participants would increase representativeness of the sample to the population by considering all grades subject areas.

### Procedures

A strength is including a standardized instructions with an example of an anagram, so everyone was clear about expectations and how to answer the test. Our clear instruction limited confusion. Everyone was read the same instructions, so confounding variables like different wordings were controlled. Another strength was operationalizing the independent and dependent variables which increases the construct validity of our experiment. Each correct answers earns one point (total of fifteen), making it easier to grade the results without confusion between right or wrong answers.

A limitation is standing the front and watched the participants as proctors causing higher scores due to Hawthorne effect where participants modify their behavior because they are being watched. We stood closer to the experimental group and increased attention may increase productivity and effort. We can modify by standing at the back. Our presence is less pronounced while still monitoring the exam, so Hawthorne effect decreases. Another limitation is passing out the red text exam first. The red ink shows through the paper, so the experimental group had longer exposure to the test. A longer or earlier look can enhance performance. To modify this, we can create a cover page to ensure that no one can get a head start. This decreases any confounding variables to increase internal validity.

### Conclusion

As a result of our study, we conclude at a  $p = 0.298$  level of significance that changing the text color of anagrams from black to red has no effect on the number of anagrams solved and academic performance of high school students.

## REFERENCES

- Elliot, A.J., Maier, M.A., Moller, A.C., Friedman, R. & Meinhardt, J. (2007). Color and Psychological Functioning: Effect of Red on Performance Attainment. *Journal of Experimental Psychology*, 136 (1), 154-168. 10.1037/0096-3445.136.1.154
- Chartrand, T.L. & Bargh, J.A. (1996). Automatic Activation of Impression Formation and Memorization Goals: Nonconscious Goal Priming Reproduces Effects of Explicit Task Instructions. *Journal of Personality and Social Psychology*, 71 (3), 464-478. 10.1037/0022-3514.71.3.469.
- Horner, A.J. & Henson, R.H. (2008). Priming, Response Learning and Repetition Suppression. *Neuropsychologia*, 46 (7), 1979-1991. <https://doi.org/10.1016/j.neuropsychologia.2008.01.018>.
- Worksheet generator*. Maker | Education.com. (n.d.). Retrieved December 14, 2021, from <https://www.education.com/worksheet-generator/reading/word-scramble/>.
- Spin the wheel to decide a random choice*. Picker Wheel. (n.d.). Retrieved December 14, 2021, from <https://pickerwheel.com/>.
- Random word generator*. Random Word Generator. (n.d.). Retrieved December 14, 2021, from <https://randomwordgenerator.com/>.

## APPENDIX

## Appendix A: CONSENT FORM

**CONSENT FORM**

We are looking for 20-25 participants for an IB Psychology experiment. This study will replicate experiment 1 in *Color and Psychological Functioning: The Effect of Red on Performance Attainment* designed by Andrew J. Elliot, Markus A. Meier, Alen C. Moller, Ron Friedman, and Jörg Meinhardt. We aim to investigate number of anagrams solved under five minutes.

This experiment will take approximately thirty minutes.

If you choose to participate:

- Any information you give out during this experiment will be kept confidential and anonymous
- You may withdraw at any point in the experiment for any reason
- You will receive all the information about the nature of our experiment and the results after the study
- If you have any questions, contact us anytime at:
  - [Experimenter phone number]
  - [Experimenter email]

If you agree, please sign the form below and fill in the following questions relevant to our experiment. If your age is less than 16 at time of experiment, also include signature of parent or guardian.

Gender: M F Non-Binary

Age: \_\_\_\_\_

Ethnicity: Hispanic/Latino American Indian/Alaskan Native Asian  
Black/African American Native Hawaiian/Other Pacific Islander White

Color Blindness: Y N If yes, please indicate which colors: \_\_\_\_\_

Language Impaired Disability: Y N

Fluent in English: Y N

I, \_\_\_\_\_, understand this experiment has been approved by the IB psychology teacher and follows the International Baccalaureate Psychology Ethics Guidelines, and agree to participate voluntarily. I give the researchers the permission to use my data as part of their experimental study.

Student Signature: \_\_\_\_\_

Date: \_\_\_\_\_

Parent/Guardian signature: \_\_\_\_\_

Date: \_\_\_\_\_

Appendix B: DEBRIEFING STATEMENT

## DEBRIEFING STATEMENT

Thank you for taking part in this experiment. While your papers are being collected, I will share the aim of our investigation. Our investigation was to test the effect of color on performance. There were two groups, the Red Group and the Control Group (Black Group). The anagram tests were identical, except that it was written in red font for the Red Group and black font for the Control Group. Elliott-Maier in 2007 found that exposure to the color red decreases performance. We will share the results of this experiment with you by emailing the results to your teacher. Please contact us if you need to ask any questions or want to withdraw yourself or your data.

---

Appendix C: STANDARDIZED INSTRUCTIONS

## STANDARDIZED INSTRUCTIONS

Hello everyone, we are conducting an experiment that is going to test you by solving anagrams. Before starting the experiment, we are going to rearrange you all into 2 groups. When we call your name please go to the designated side of the room.

Before we start, an anagram is a puzzle where a word, phrase or name is formed by rearranging the letters of another. For example, from the word iceman, you could form the word cinema.

You have 10 minutes to solve 15 anagrams.

Please write your answer legibly next to the lines provided. If you find that there is more than one answer to the anagram, please only write down one answer for each question.

If you do not know the answer, skip it, and move on to the next question.  
Please do not look at another person's paper, this is an independent test.

Do you have any questions?  
(Pass out papers)

Please answer to the best of your abilities and we will begin working now. (Press timer)

(time's up) Time is up, please stop working. (Collect papers)

---

Appendix D: TEST VERSION BLACK

1. PALEP \_\_\_\_\_
  2. TMSAP \_\_\_\_\_
  3. RNWBO \_\_\_\_\_
  4. ORUHG \_\_\_\_\_
  5. ELNIN \_\_\_\_\_
  6. SEKMO \_\_\_\_\_
  7. RTGNA \_\_\_\_\_
  8. HSRAE \_\_\_\_\_
  9. DARIY \_\_\_\_\_
  10. UAFLT \_\_\_\_\_
  11. DSLIE \_\_\_\_\_
  12. WFRON \_\_\_\_\_
  13. KNSAC \_\_\_\_\_
  14. GNEFI \_\_\_\_\_
  15. KCARC \_\_\_\_\_
- 

Appendix E: TEST VERSION RED

1. PALEP \_\_\_\_\_
  2. TMSAP \_\_\_\_\_
  3. RNWBO \_\_\_\_\_
  4. ORUHG \_\_\_\_\_
  5. ELNIN \_\_\_\_\_
  6. SEKMO \_\_\_\_\_
  7. RTGNA \_\_\_\_\_
  8. HSRAE \_\_\_\_\_
  9. DARIY \_\_\_\_\_
  10. UAFLT \_\_\_\_\_
  11. DSLIE \_\_\_\_\_
  12. WFRON \_\_\_\_\_
  13. KNSAC \_\_\_\_\_
  14. GNEFI \_\_\_\_\_
  15. KCARC \_\_\_\_\_
-



APPENDIX F: ANSWER KEY

1. apple

2. stamp

3. brown

4. rough

5. linen

6. smoke

7. grant

8. share

9. dairy

10. fault

11. slide

12. frown

13. snack

14. feign

15. Crack

---

## APPENDIX G: RAW DATA

control group: black text		experimental group: red text	
number	score	number	score
1	7	1	8
2	14	2	12
3	3	3	6
4	15	4	11
5	6	5	9
6	7	6	13
7	10	7	7
8	9	8	8
9	12	9	13
10	11	10	10
11	5	11	10

## Appendix H: VASSER STATS CALCULATIONS

Data Entry:

count	Ranks for		Raw Data for	
	Sample A	Sample B	Sample A	Sample B
1	6	8.5	7	8
2	21	17.5	14	12
3	1	3.5	3	6
4	22	15.5	15	11
5	3.5	10.5	6	9
6	6	19.5	7	13
7	13	6	10	7
8	10.5	8.5	9	8
9	17.5	19.5	12	13
10	15.5	13	11	10
11	2	13	5	10
Reset	Calculate from Ranks		Calculate from Raw Data	

Mean Ranks for				
Sample A	Sample B	$U_A =$	$P_{(1)}$	$P_{(2)}$
10.7	12.3	$z = -0.53$	0.2981	0.5961

Note that mean ranks are provided only for descriptive purposes. They are not part of the Mann-Whitney test. ~ Note also that the z-ratio is calculated only if  $n_A$  and  $n_B$  are both equal to or greater than 5.

Critical Intervals of  $U_A$  for  $n_A=11$ ;  $n_B=11$ 

	Level of Significance for a		
	Directional Test		
	.05	.025	.01
	Non-Directional Test		
	--	.05	.02
lower limit	34	30	25
upper limit	87	91	96

The adjacent critical intervals are calculated only if  $n_A$  and  $n_B$  both fall between 5 and 21, inclusive. For sample sizes smaller than 5, you can refer your results to a standard table of Mann-Whitney critical values, such as the following, provided by the Department of Mathematics & Statistics at the University of Saskatchewan:

<http://math.usask.ca/~laverty/S245/Tables/wmw.pdf>

with  $n_A=11$ ,  $n_B=11$ ,  $U_A=69$ , and  $U_B=52$