the Berkeley Carroll School

Proceedings of the Berkeley Carroll

9TH INDEPENDENT RESEARCH CONFERENCE

SCIENCE RESEARCH PARTICIPANTS

Seniors

Zachary Ashen
Luca Broder
Phoebe Degn
Donald Minerva
Zoe Rabinowitz
Isabella Solimine
Sophia Sondey
Hope Swetow

Juniors

Alma-Lilas Anzalone
Shreya Balaji
Jonathan Barley
Lila Blank
Antonio Domonkos
Luke Ehrenfreund
Kai Ellis
Nichole Fernandez
Lucy Graham
Harry Justice
Michael Khoury
Christopher Kretschman
Gabi Prieto
Natasha Rasnick
Adeleen Rosner
Daniel Safian
Margaret Schaffzin
Ava Sonies
Andreas Welch

Sophomores

Daniel Chian
Louisa Egolf
Grace Freeman
Joseph Gordon
Leonardo Intelisano
Oluwakemi Iyageh
Ruby Kass
Simon Korotzer-Mariani
Zoë Krashes
Zoe Pyne
Anya Renzi
Carly Saba
Nika Sadeghi
Margot Schiller
Annika Sheridan
Nathaniel Snow
August Swetow
Carter Swetow
Michael Warshaw
Weezie Wilson

SRD Coordinator

Dr. Danny Gentile

SRD Facilitators

Vanessa Prescott
Essy Levy Sefchovich
Scott Rubin

Adjunct Facilitator

Jason Gaines

In addition, we would like to acknowledge the support and assistance of the following members of the Berkeley Carroll administration, without whom this program would not have been possible:

Jane Moore
Director of the Upper School

Dr. Lisa Waller
Head of School

Welcome from the Editor

In his premature eulogy in the closing minutes of the film *The Wrath of Khan*, James T. Kirk says of his dear friend Spock "of all the souls I have encountered in my travels, his was the most human." The fascination with why humans do the things they do is universal—and it's a theme almost all of the student scientists featured in this Journal have built on this year.

Any scientist can tell you that when you work with humans, you're asking for trouble. You have to recruit subjects, monitor their safety, make sure they don't figure out the question you're trying to ask, and try to keep them from dropping out of your study.

Those difficulties didn't stop our researchers from investigating human subjects in many different ways.

Sophia Sondey and Phoebe Degn, for example, looked at how humans process emotion through the senses. Sophia investigated whether gender-targeted fragrances elicit a corresponding gender-based response, while Phoebe asked whether listening to music can reduce stress.

Zoe Rabinowitz and Bella Solimine looked at the human memory. Bella wanted to find out why people sometimes remember things that didn't happen, and Zoe wanted to see if musical training improves verbal memory.

Hope Swetow and Luca Broder looked at how humans sleep. Hope investigated the reasons why we dream, and Luca made his subjects sleep better by getting them to cut down on their screen time.

Zack Ashen directly probed the meaning of humanity by writing a computer program that could function as both a judge and a poet. As you'll see when you read his paper, his computer's poetry was good enough to fool a fair number of Berkeley Carroll English teachers!

The one exception to this year's trend of studying aspects of human behavior was Donald Minerva, who compared the benefits of hydroponics to aeroponics in a study that has applications to big issues like food security in an increasingly populous world. The advantage of using plant subjects is that they never forget appointments, they work around your schedule, and if you lose a subject or two, no one really misses them. Plants are a lot easier to work with than people.

That said, I think all of our student scientists would argue that despite the trouble, working with humans has some appeal. Since we're all human, conducting research on people allows you to look inward and outward at the same time. I encourage you to do the same as you read the pages of this year's journal.

Goodbye for now,
Scott Rubin
Upper School Science Chair

Science Research and Design
SENIOR RESEARCHERS

Zack Ashen '20

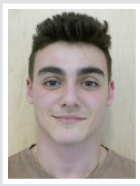
Zack is presenting his research on machine learning and, more specifically, intelligent computing. Zack has been interested in artificial intelligence and computer science since middle school and was inspired by a paper written by Alan Turing. He was fascinated by the paper which theorized that a computer could be indistinguishable from a human. Over the course of three years, Zack has attempted to develop an "intelligent" computer that could be considered of human-level intelligence.

Luca Broder '20

Luca is presenting his research on the relationship between screen usage and sleep quantity. As someone who has struggled to balance school work, screen usage, and sleep, he wanted to dig into the relationship between two of those factors. In 11th grade, Luca worked with Dr. John Mariani to determine whether there was a correlation between screen usage and sleep quantity among high schoolers. As a senior he has worked to determine whether this correlation is causal. He hopes to continue this research and to begin testing the effect of screen usage on adult sleep quantity.

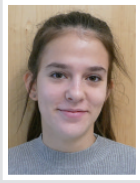
Phoebe Degn '20

Phoebe is presenting her research on the relationship between music and stress. She worked closely with Dr. Karen Kauffmann to figure out an experimental design that would best answer her scientific questions. In 11th grade, she began by focusing on self-reported stress and experimented with different music types. In 12th grade, she decided to focus more specifically on physiological measures of stress and used a heart rate monitor to determine students' heart rate and heart rate variation during a stress-inducing task. Phoebe hopes to continue to research psychological and physiological measures of stress to determine if music benefits adolescents who face stressors daily.



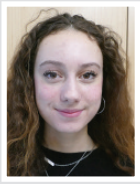
Donald Minerva '20

For the past three years, Donald has been interested in sustainability, specifically investigating how to develop agricultural techniques to allow food yields to keep pace with a growing global population. He began in 10th grade testing how the difference in nutrient content in feces from different species of fish would affect the growth rate of lettuce plants grown without soil (hydroponically). In the following year, Donald conducted a study in which he varied the levels of nitrates, phosphates, and ammonia in a hydroponic system. With the data he gathered from 10th and 11th grade, he began investigating the advantages of different types of soilless growing systems for lettuce plants. Ultimately Donald plans on incorporating all the different studies he has done since 10th grade and continuing his hydroponic investigations outside the classroom with the help of family members who also use hydroponics.



Zoe Rabinowitz '20

Zoe is presenting her research on the effect of musical training on verbal memory. Her interest in the relationship between music and memory started in eleventh grade when she conducted her first study, asking students to complete a memory test while listening to different types of mood-inducing songs. Her research then evolved to musical training, as she became more interested in the long term effects of music on the brain. This year, Zoe used subjects in middle school and high school, creating her own version of a verbal memory test to examine the different verbal memory skills of students enrolled in Berkeley Carroll's music program. In the future, she hopes to continue to examine this relationship and others, to determine the potential ways in which music can be used to help developing brains.

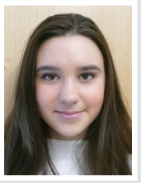


Isabella Solimine '20

Bella is presenting her research on the accuracy of memory recall and the formation of false memories. She began her research on the accuracy of memory recall in simulated eye-witness testimonies in the 11th grade using 5th graders as participants. The conclusions of this study prompted her to continue testing the influence of biased information on memory recall in a simulated eyewitness testimony scenario. Bella's 12th grade research was intended to test whether there is a difference in accuracy of memory recall when accuracy of co-witness testimonies and language in follow-up questions is varied. She used 9th and 10th grade subjects who were asked to watch a video, answer a set of comprehension questions, read either an accurate or inaccurate testimony, then answer a set of follow-up questions containing either correct or misleading language. Bella wishes to continue to research the concept of false memory formation in the future, as it combines her interest in law and her fascination with memory and psychology.

Sophia Sondey '20

For the past three years, Sophia has been attempting to determine the reason for the gender divide in the fragrance industry. Why is perfume for women and cologne for men? Is there something in the biological makeup of men and women that causes them to react differently to perfume and cologne, or are we just responding to what the marketing information of the products tells us we should like better? In order to answer these questions, Sophia has been examining differences in the way males and females emotionally respond to perfume and cologne. She has also been attempting to determine how this emotional response is affected by the marketing information of perfume and cologne. During her junior year, she was mentored by two consumer insight specialists at *Bustle*, a women's lifestyle and insights magazine company, who worked with her to host a focus group on fragrance marketing strategies. This year, Sophia has incorporated heart rate as a tool to measure emotional response into her experiment. In the future, she wishes to advance her research to continue to uncover the reason for the gender divide in the fragrance industry.



Hope Swetow '20

Hope is presenting her research on the significance of dreaming in adolescents. In 11th grade she focused on the difference in dream recall frequency between male and female participants, which prompted her 12th grade case study. Hope decided that in order to truly understand dreams on a deeper level, she should read and analyze the dreams of several Berkeley Carroll students. Instead of using a survey to collect data as in previous years, Hope gave participants a dream diary with the hope that she would receive in-depth accounts of the participants' dreams. Working with her mentor, Ms. Sarah Moore, she employed several qualitative data analysis techniques to come to her conclusions. Because many aspects of dreaming remain a mystery to scientists, Hope continues to be fascinated by them. She wishes to further her understanding of dreaming in the future.

Table of Contents

Computational Methods in Intelligent Computing

by Zack Ashen

Abstract

First theorized by Alan Turing in 1950, machine learning and artificial intelligence allow computers to mimic human decision making and pattern recognition abilities. Recent advancements of machine learning programs have given more attention to Alan Turing's original question of whether a computer can fully mimic human-level intelligence. The goal of this study was to develop an intelligent computer. Initially, a computer program (a Computer Judge) was developed to detect consciousness through written dialogue and was tested using a modified Turing Test. The Computer Judge in the modified Turing Test showed no significant difference between its assessment accuracy of humans and computer chatbots (p-value: 0.33). A Lovelace Test was used after the modified Turing Test to test the creativity of the program. The developed computer program was able to create poetry that was not significantly different from human poetry (p-value: 0.0929). Furthermore, younger participants were much more likely to mistake a computer poem for a human poem (p-value: 0.1214) whereas older participants were less likely to make a mistake (p-value: <.0001). It was concluded anecdotally that the developed computer program does not represent human level intelligence.

Introduction

Machine learning has emerged in the field of computer science as a tool to augment human processes (Ghahrami, 2015). Most common computer programs are only able to perform basic quantitative tasks and must follow explicit instructions written by a programmer (Géron, 2017). When a computer is explicitly programmed, the program is written in a stepwise fashion similar to how a standard laboratory procedure may be written. In other words, the computer completes a series of tasks written by the programmer. However, with the emergence of machine learning, computers can recognize patterns without being explicitly programmed (Samuel, 1959). In machine learning, the programmer writes instructions for a computer to find patterns in data and then output a prediction or conclusion from that data, without explicitly writing the process which the computer should follow to get to the end conclusion or prediction (Géron, 2017). Therefore, a computer programmed effectively with machine learning should be able to function independently from the code explicitly written (Samuel, 1959). Thus, computers can appear to function autonomously, but, in reality, they follow a mathematical process that only appears autonomous. For example, if a human were to multiply two numbers, they may follow a stepwise process in their heads (e.g multiply the ones by the ones place, multiply the tens place, etc.). Thus, the process for humans seems like it could be explicitly written. However, a computer appears to skip all of these steps when it multiplies and can come to a mathematical answer without explicit instructions (Weizenbaum, 1976). To many humans, multiplication is a multi-step process, but to a computer, the calculation is completed in one step. This is analogous to machine learning. To humans,

computing large sums of data and creating conclusions is a multi-step process, but to a computer, this process is just a mathematical algorithm (Geron, 2017).

Furthermore, a machine learning program can be trained to repeat favorable actions in order to achieve a goal. This method of training a computer using machine learning is referred to as reinforcement learning (Silver, 2017). Reinforcement learning algorithms are often utilized in machine learning programs that play games such as chess, checkers or Go (Silver, 2017). These programs are trained much like children—with positive or negative reinforcement. If a computer system programmed with a reinforcement algorithm reaches a negative outcome, the process that the program took to achieve that outcome will be weighted less heavily. However, if a computer program reaches a positive outcome, the process that the program took to achieve that outcome will be weighted more heavily. In other words, the more a computer reaches the positive outcome the better it becomes at arriving at that outcome. Another type of machine learning is supervised learning. Supervised learning is a type of machine learning in which the programmer labels the entered data such that the computer learns what the desirable data looks like prior to being deployed (Kotsiantis, 2007). Supervised learning makes training the machine learning program faster and easier due to the fact that the computer already knows what the desired outcome looks like and therefore has a much better ability at matching new data to the already labeled data. For example, in a spam email detector that uses a supervised machine learning algorithm, the user labels emails as spam or not spam and then the computer learns based on previous decisions what type of text in the emails is spam and looks at new emails to compare to the labeled ones. Supervised learning is faster and more successful than unsupervised learning (when the computer is not trained prior to deployment) because the computer is being assisted and taught the pattern instead of learning the pattern independently. In addition, a supervised learning program is less likely to make a mistake than a program that does not have labeled data sets. This is due to the fact that it has been manually trained, so the computer has more precision. In fact, reinforcement learning algorithms in tandem with supervised learning methods create a program that can be trained and goal-oriented (Lin, 1992).

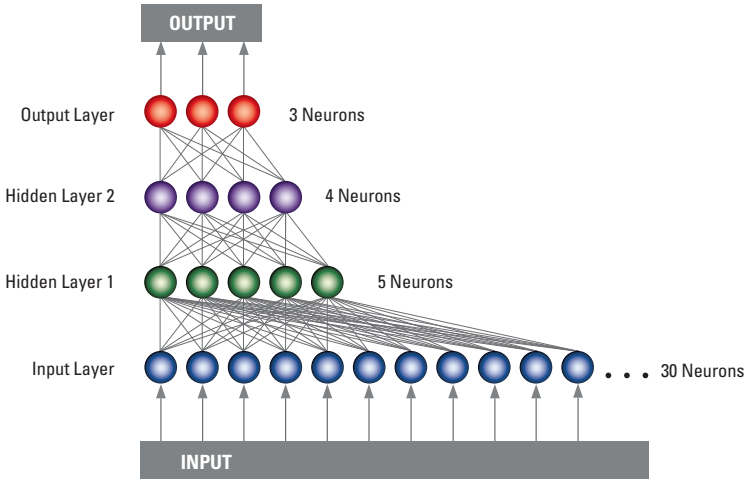
Furthermore, with the emergence of machine learning, computers are able to follow a mathematical process to analyze data and learn patterns from them (Géron, 2017). The most simple machine learning algorithm is a regression analysis (Géron, 2017). These simpler prediction algorithms have been abstracted to create a modeled prediction system known as a neural network (Schmidhuber, 2015). Neural networks are sets of algorithms that use nodes and layers to model human cognition (LeCun, 2015). Moreover, in a network, there is an input layer that takes in data, hidden layers that take into consideration different weights and biases, and an output layer which is the final prediction (Dietterich, 2000). To an outsider, the result of a neural network may appear like the program is making a conscious decision. However, in a neural network, a computer uses algorithms to take input data and make predictions or conclusions (Silver, 2017).

A more advanced type of machine learning architecture that is used for language processing is a Recurrent Neural Network (RNN) (Schmidhuber, 2015). The piece of the RNN that makes it advanced is the inclusion of deep learning. Deep learning is a type of machine learning that uses nodes in a layered pattern to mimic the design of a human brain (LeCun, 2015). These nodes are meant to mimic the biological neurons in humans. In this layout *(Figure 1),* the first layer of neurons take the input data and then that data is processed through hidden layers using specific deep learning algorithms (LeCun, 2015). This allows for mass amounts of data to be analyzed and considered. RNN's use the same neural network structure but are designed such that each node in the different layers follow a sequential path (Schmidhuber, 2015). In other words, data from one node goes to another node which goes to another

node and so on. This is effective in text analysis because the text is read in a sequential manner and thus translates well to the sequential nature of the neural network. Furthermore, for text analysis RNNs are able to be trained off of large amounts of texts known as corpuses (Li, 2016). Corpuses have thousands of lines of text that the RNN processes and learns from. After the RNN has processed the data, seed text can be used to prompt the RNN to output related text from the corpus or to weigh the seeded text more heavily in its recreation (Géron, 2017).

Figure 1

A Visual Representation of a Neural Network.



Machine learning programs can be built using programming packages, which include previously built code. Python programming packages, such as Scikit-learn (Pedregosa, 2011), make basic machine learning easier due to the fact that they automate much of the learning process. However, in addition to that, more complex tools are necessary for more advanced machine learning algorithms. For example, the most effective way to implement reinforcement learning is with a program called TensorFlow (Abadi, 2016). TensorFlow is a free software library for machine learning; it uses the cloud to simplify challenging computational tasks that are too complex for a desktop computer. In addition, TensorFlow creates a framework for setting up deep learning such that the programmer does not have to start from scratch.

As machine learning techniques and neural networks have improved, they have also become more effective at resembling human-level cognition (Moscovitch, 1995; LeCun, 2015). This improvement has evolved from basic supervised learning algorithms that could only recognize patterns in simple data to complex neural networks that are able to recognize patterns in images that build large data sets (Kotsiantis, 2007). Therefore the question at hand arises: can a computer be deemed of human-level intelligence?

Chatbots are programs that conduct written conversations and replicate human dialogue (Gamez, 2008). Due to this, they are the most effective exhibit of computers that demonstrate human level cognition. (Radziwill, 2017). Furthermore, through the use of neural networks these programs are able to replicate human sentiment (Gamez, 2008), dialogue (Li, 2016), and include a conversational context (Vinyals, 2015). Chatbots exhibit the human-level intelligence that machine learning is able to create.

The Turing Test (Turing, 1950), also called the Imitation Game, is a method that systematically assesses the intelligence of chatbots. In the Turing Test, two confederates, a computer and a human, answer questions from an interrogator whose goal is to ascertain their true identity. The interrogator, through a series of questions, identifies the subjects as human or computer. According to Turing, if an interrogator can not accurately distinguish the computer from the human subject then the computer must be intelligent. The Imitation Game is played annually in the Turing Test which assesses new computer programs' ability to play the Imitation Game (Livingstone, 2006). Computers entered into the Turing Test are tested on their ability to fool an interrogator into thinking that they are human, thus showing they have human intelligence and can mimic human decisions. This is the standard method to test chatbots (Livingstone, 2006).

However, the Turing Test is an ineffective means of assessing artificial intelligence (LaCurts, 2011). The problem with the Turing Test is that chatbots utilize linguistic deception as a manner for deceiving interrogators. In other words, computer subjects make intentional grammatical errors or design their personalities in an attempt to deceive the judges. It is also not possible to tell judges to be cautious of linguistic deception because if a human truly does have poor grammar they may be more susceptible to consider them as a computer after being warned. Linguistic deception serves solely to deceive the interrogator and not to show computational intelligence and is, therefore, a flaw in the Turing Test (Livingstone, 2006). This makes it impossible to detect whether a chatbot is, in fact, intelligent. Moreover, in the Turing Test, it is impossible for chatbots to make decisions which is a key characteristic of artificial intelligence (McCarthy, 1979). Chatbots in this test are only able to respond to questions and do not come to conclusions or make decisions. Therefore, the Turing Test is considered ineffective (LaCurts, 2011).

In the current study, The Turing Test was reworked such that the chatbot being tested for intelligence was inserted into the Turing Test as a judge such that there was a human judge, computer judge, a human confederate, and a computer confederate. This modification to the Turing Test was based on a study that showed that a defining characteristic of humanity is being able to discern what has consciousness and what does not (Cheesman, 1986). A successful computer judge would demonstrate an equal ability to assess consciousness to a human by being able to determine the difference between a human and a computer through dialogue. Therefore, the human judge's assessment accuracy was compared to the Computer Judge's assessment accuracy to see if there was a difference between their abilities at assessing consciousness. If the chatbot can discern what has consciousness, between humans and computers, then it fits another criteria of an intelligent chatbot and is even more human-like. This new configuration also removes the problem of linguistic deception due to the fact that the Computer Judge does not need to fool the confederate to be deemed intelligent. The Computer Judge will use supervised reinforcement learning to come to accurate decisions and make the chatbot more versatile such that it is able to do two things: ask questions and come to accurate conclusions. By fixing the Turing Test and changing the standard programming scheme of intelligent chatbots, this new design should be able to take the first step in creating an intelligent chatbot.

Beyond simply detecting consciousness an additional ability to exhibit creative thought needs to be included for the computer program to be considered intelligent (Bringsjord, 2003). This facet of intelligence is characterized by the Lovelace Test which tests a computer's creative intelligence (Bringsjord, 2003). In the test, the Computer Judge should be able to create a work of art and the programmer should not be able to explain exactly how the program was created, demonstrating the lack of explicit programming. Poetry was subjectively chosen as the artistic medium for the computer program to illustrate its creativity. In other words, the Computer Judge was turned into a computer poet.

There should be no difference in the proportion of poems deemed to have been written by humans. This would suggest that people identify human poems as frequently as they identify computer poems, indicating creativity. After both the creative aspects of the Computer Poet and the decision making aspect of the Computer Judge are combined, the computer would take another step towards human-level intelligence. (Gamez, 2008).

Methods

Computer Judge Program Development

A supervised reinforcement learning program was developed to serve as a Computer Judge in the Turing Test. A list of 25 preset questions was imported into an array to serve as a random question bank for the confederates. Answers to questions were paired together and stored in a dictionary. The dictionary was then rewarded and reinforced contingent on a moderator verifying the correct assessment of the confederate. A reinforcement learning software agent was imported from the TensorFlow module to monitor specific word usage in the responses to questions. Words were classified as emotional or unemotional using TensorFlow's pretrained text embedding modules. In addition to emotion, the program measured repeated words, character length, and grammatical accuracy. Grammatical accuracy was monitored based on the number of commas, semicolons, colons, m-dashes, and quotations. Grammar was monitored in such a way that the computer judge would not be subject to linguistic deception. The reinforcement learning network associates a correct response with the weighting value of multiple measures. In this case one of those is grammar, and, therefore, if computers make more grammar mistakes the Computer Judge would still be effective in detecting it because it would learn that pattern and change its weighting. In other words, at one point during its learning process, it may correlate more grammar mistakes to a chatbot, but as it learns more, it may correlate grammar mistakes to a human. Therefore, the computer judge, unlike human judges, would not categorically associate grammar mistakes with one confederate identity. Similarly, the computer judge did not associate any of the other measures (emotional word count, repeated word count, and character length) with a confederate identity. Rather the program, after learning if its assessment was correct, would adjust its weighting of the measures to make more accurate predictions as to the identity of the confederate.

Reverse Turing Test

Four groups of trials were conducted: Human Judge (HJ) – Human Confederate (HC), HJ – Computer Confederate (CC), Computer Judge (CJ) – HC, CJ – CC.

There was no specific recruitment method for acquiring the participants due to the fact that the only criterion for participation was being human. Participants were sampled from the Berkeley Carroll high school's student body and faculty using a variety of methods including emails to the school, planned sampling during school meetings, and school surveys. 97 assessments were conducted by the CJ and HJ. The CJ made 55 assessments while the HJ made 42 assessments. There were 79 participants in the study. 5 participants were older than 18. Having 79 humans allowed both judges to assess a variety of HCs.

Participants in the HJ trials were assigned roles as either an intermediary, judge, or confederate using simple randomization. Both trials could be conducted using a standard Turing Test. Participants were not blinded to the rules and goal of the adjusted Turing Test. However, due to the rules of the Turing Test they were blinded to the roles of the other participants. The judge's job was to ascertain the identity of the confederate by emailing questions to an intermediary. The intermediary was either

assigned to a HC or to a CC. If the intermediary was assigned to a human, they would transcribe the HC's answers into their response to the judge through email. The intermediaries were used to maintain anonymity of the confederate, such that the judge did not know who they were talking to. HCs were told to answer the questions in any manner they choose. Judges were restricted from asking questions specific to the location in which the test occurred and any other information that may be personally known by participants. After 20 minutes, judges would return their assessed identity of the confederate.

The trials with the CJ were completed independently from each other. The group with the CJ – CC was completed having a moderator transcribe questions from the CJ into four chatbots. The moderator would then transcribe the chatbot's response to the questions back into the CJ. This would go on for 25 questions until the CJ arrived at its final assessment, and was repeated for each chatbot multiple times to allow the CJ to make numerous assessments. For the assessments with the CJ – HC, the HC was instructed to answer the CJ's questions in any manner they desired. All the HCs throughout the study were permitted to lie.

Poetry Generation Program Development

After developing the CJ program, an additional piece of code – a Computer Poetry Generation program – was developed to add creativity to pass the Lovelace Test. Furthermore, a multi-layer recurrent neural network (RNN) was developed to train a model based on children's poetry. Children's poetry was chosen due to its popularity and ease of recognition. Commonly known poets like Dr. Seuss, Shel Silverstein, and E.B. White were included. To develop the corpus used to train the model, a 31,829 line text file was created. The RNN developed was a two-layer network and the size of the RNN's internal layers was 128. Throughout the study, the corpus was edited, removing non-English words, multiple spaces, and reinforcing positive sentence structure. The reinforcement of positive sentence structure was done through the addition of multiple lines of identical sentences that had correct grammar and rhyming schemes. This was done to encourage the neural network to learn more off of these sentences. The resulting poetry was deemed satisfactory for an initial test based off of anecdotal feedback from peers and subjective comparison to other poetry *(Figure 2).*

Figure 2

Sample Computer Generated Poem before Revisions.

A singer cannot forget to see and say,
Find a thousand ways on his own
Blow mine own love in a fire.
Mind, my home to scent and spray
When you're dead on
He drowned his cat,
O me think's I lived to see
What a time

He thinks of you in a moon: never are you to be done,
It's lost in a starry night.
I do not love when he is home,
No horry hope in the sea, but if he shall go on
And can he go all on the air with me?
I can't get to the song.

Initial Lovelace Test

An initial Lovelace Test was conducted using 5 of the computer-generated poems and 4 poems from three selected authors: John Keats, Michel Foucault, Roberto Unger. These authors were chosen based on the feedback supplied by a member of the Berkeley Carroll English department. A survey was created to get initial feedback so that the program could be revised and improved upon before a final survey was distributed. The initial survey was structured such that there were three sets of three poems. In set one (S1) there were two computer written poems and one human written poem. In set two (S2) there were three computer written poems and zero human-written poems. In set three (S3) there were zero computer written poems and three human-written poems. In addition to the Lovelace Test, the participants were asked to gauge how similar the computer poems were to the human-written poems and determine the extent to which their choice was a guess. Furthermore, the participants were asked to include feedback about how they were able to tell the difference between human poems and computer poems. 11 English teachers in the Berkeley Carroll High school English department were surveyed and provided data and anecdotal feedback to aid in revisions.

Poetry Generation Program Revisions

Based on the feedback of the English teachers, the program was revised to include necessary features to increase the quality of the poetry. For example, one English teacher indicated: "Basically I was looking for sustained (and sophisticated) syntax and imagery. I was also looking for regular structure on the page. The more it had those things, the more likely I was to think it was a human." The program was revised to include more thematic structure and repetition. The RNN was seeded using strings of related text to improve the repetition of words to induce the sensation of a theme. The result *(Figure 3)* was deemed satisfactory to move on to the Final Lovelace Test trial based on subjective approval made by the Berkeley Carroll English department.

Figure 3

Computer Generated Poem after Revisions Suggested by the English Department.

The strange stars and sands that come at dawn
In the morning we feel distant to love.
Although, it seems to the beat of death that the sun and stars shine brightest,
And the coarseness of the sand and the coldness of the mountains distance us from ourselves.
Love, I say, are good stars who are born in the wall when capees cover the light of the dead.
And the dark flowers are straightened and strangely with a stream of the sea
and sights and shade.

Final Lovelace Test

The final Lovelace test was conducted in the same manner as the initial Lovelace Test. However, instead of only being conducted on one group of English teachers, the test was conducted amongst 70 participants. The computer poems from the original Lovelace Test were replaced with the revised poems from the program revisions. However, the human poems and the format of the survey remained the same. The Berkeley Carroll Upper School Science, Language, Math, History, and Arts department faculty were all sampled to increase participants over the age of 18 as well as to gauge the effect of age on the assessment accuracy (n = 24). Furthermore, participants under the age of 18 were sampled in the Berkeley Carroll Upper School and Middle School (n = 51).

Results

In the modified Turing Test, there was no difference in the assessment accuracy of the HJ and the CJ *(Figure 4).* Therefore, the CJ is of comparable ability to the HJ at discerning intelligence. Furthermore, there was no significant difference in the assessment accuracy of the HJ between CCs and HCs *(Figure 5)* or in the assessment accuracy of the CJ between CCs and HCs *(Figure 6).* Therefore, neither the CJ nor the HJ was significantly better at determining the identity of HCs or CCs.

In the Lovelace test, the proportion of poems chosen as human, given the author of the poem was a computer, was significantly greater for the trial compared to the pretrial *(Figure 7).* In other words, the poetry was significantly better at fooling participants in the trial group. For the main test *(Figure 8),* there was no difference in the proportion of human poems chosen as human vs. computer poems chosen as human in the Total Sample and the 18 Under group. However, for the 18+ group the proportion chosen as human was larger for human poets than for computer poets.

Figure 4

HJ Assessment Accuracy vs. CJ Assessment Accuracy.

HJ and CJ confederate assessment accuracy (mean % ± SEM). Both HJ assessment accuracy (M = 69.05%, SEM = 0.07) and CJ assessment accuracy (M = 58.18%, SEM = 0.07) was determined based on the proportion of accurately identified confederates. A comparison of proportions test showed no significant difference (p-value: 0.33).

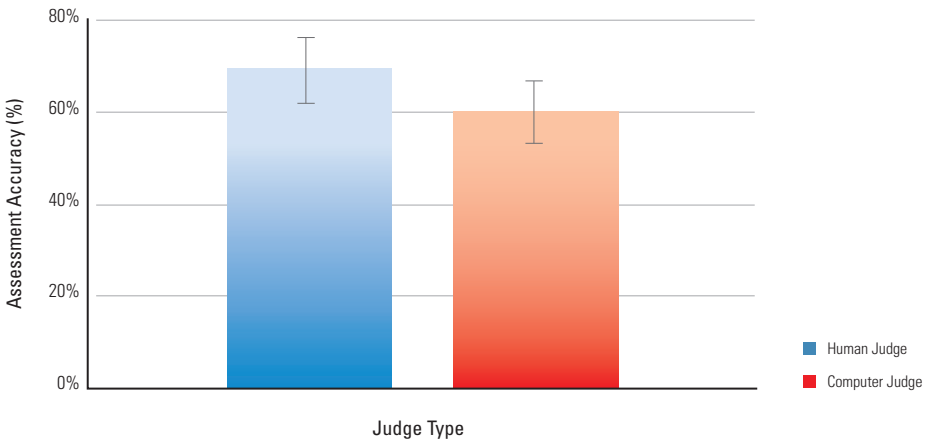


Figure 5

HJ Assessment Accuracy with CC vs. HC.

HJ's assessment accuracy (mean % ± SEM) was compared between the CCs (M = 69.23%, SEM = 0.09) and the HCs (M = 68.75%, SEM = 0.12). A comparison of proportions test showed no significant difference (p-value: 0.4868).

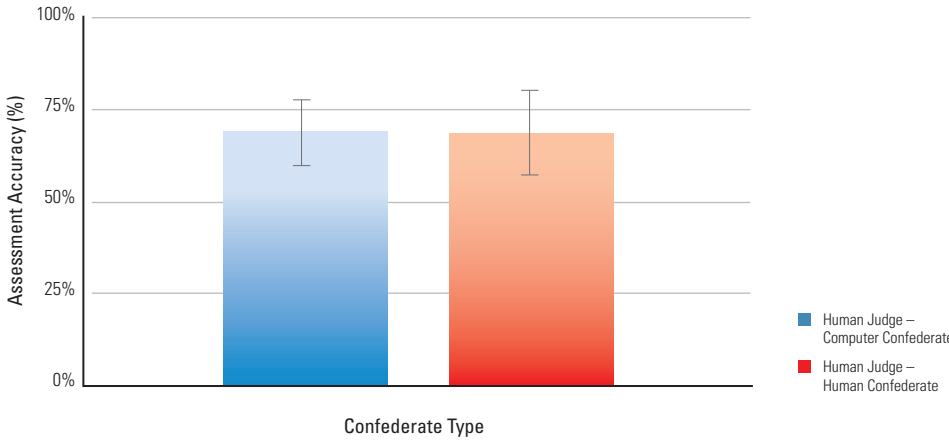


Figure 6

CJ Assessment Accuracy with CC vs. HC.

CJ's assessment accuracy (mean % ± SEM) was compared between the CCs (M = 55.88%, SEM = 0.09) and the HCs (M = 61.90%, SEM = 0.11). A comparison of proportions test showed no significant difference (p-value: 0.33).

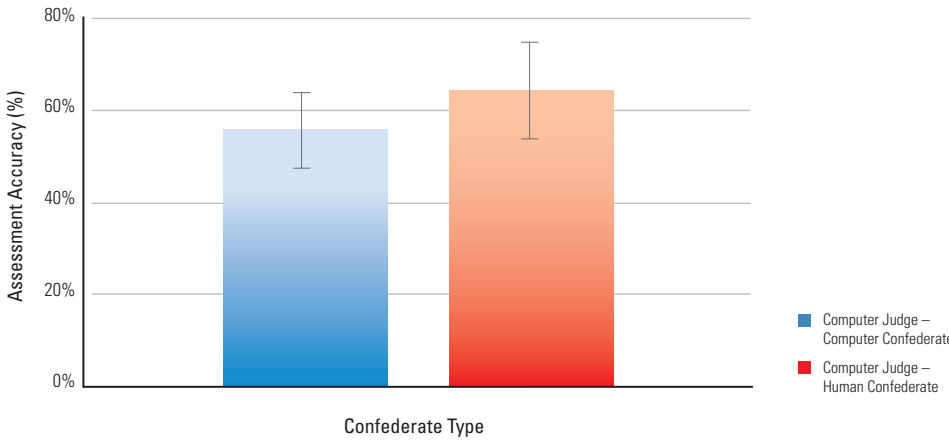


Figure 7

Pretrial Computer Poems Selected Human vs. Trial Computer Poems Selected Human.

The proportion of computer poems chosen as human for the Pretrial computer poems was compared to the proportion of computer poems chosen as human for the Trial group. A comparison of proportions test showed a significant difference (p-value: 0.0059)

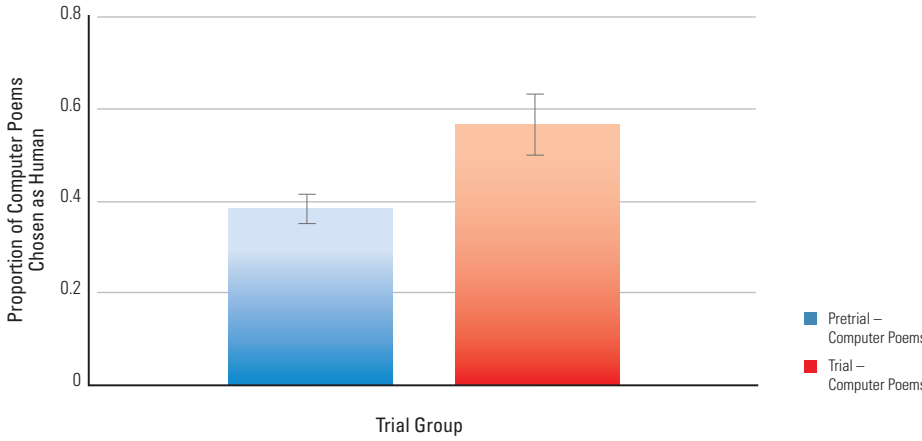
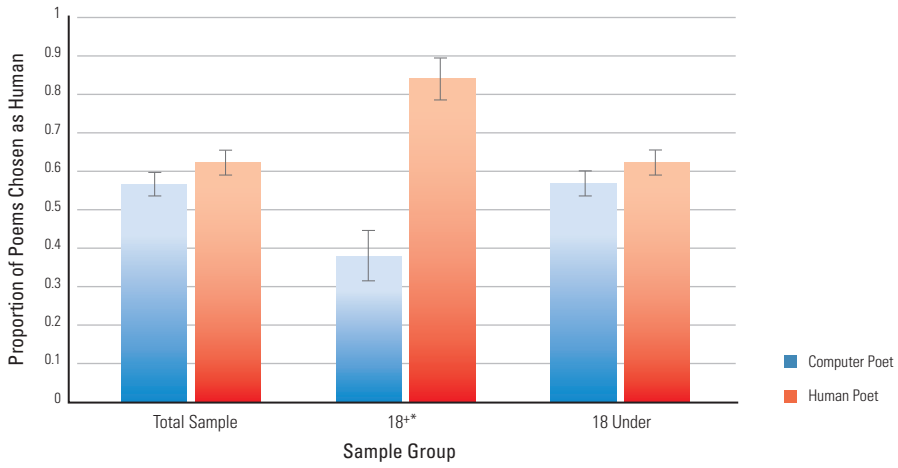


Figure 8

Human poems Selected Human vs. Computer Poems Selected Human.

The proportion of poems chosen as human was compared with two types of authors: human poets and the developed computer poet. Three groups were designated: Total Sample, 18+, and 18 Under. Total sample is the entire sample. 18+ includes all of the participants who are 18 and older, and 18 Under is all of the participants under the age of 18. Each group was tested independently using a comparison of proportions test. There was no significant difference in the Total Sample (p-value: 0.0929) or in the 18 under (p-value: 0.1214). There was a significant difference in the 18+ (p-value: <.0001).



Discussion

Throughout the study, the CJ faced several limitations. It was not possible to implement into the CJ a sequential learning algorithm such that the CJ would be able to take an answer from the previous question and formulate a new question based on that answer. This is due to the fact that TensorFlow does not have a sequential learning algorithm and rewriting a new one in the code conflicted with the supervised reinforcement learning algorithm and caused a fatal error. It was also not possible to statistically track the CJ learning. This was due to the fact that data was lost from each individual trial and therefore it was not possible to see the weights increase and then plateau as they should have if the CJ was learning. Another challenge that the CJ may have encountered is machine learning bias. The CJ only asked questions to 5 participants over 18. Therefore, if the CJ only learned how participants younger than 18 act, the CJ may have developed bias to only be able to accurately detect humans under 18. In order to account for this in a future study, the chatbot should be reset to its weighting parameters where it has had no experience and then the trial should be rerun while tracking the weights. After a diverse subject pool has been tested through the chatbot the weighting values from that trial should be compared to the younger subject pool. If they are the same there is no bias. If they are different there is bias.

While conducting the poetry test with the computer program that could generate poetry there were several limitations. First of all, the poetry that the computer-generated poems could be compared to was limited to a specific array of human authors which could have skewed results. These specific poems were selected because of the suggestions made by the English department at Berkeley Carroll who believed these authors would be most similar to the poems developed by the computer. A different genre of poetry could have been noticeably different from the computer-generated poetry making it easier to detect. Furthermore, although the computer-generated poetry was able to generate subtle themes in its poems it was not able to coherently write about a single topic for an extended period of time, rather it would just repeat words and similar words from the seed text *(Figure 3).*

The computer was more likely to fool younger participants in the study and make them believe the computer poetry was human. More specifically, it was clear that the 18^{+} participants were easily able to detect computer poems. Whereas, participants under the age of 18 struggled significantly with detecting the computer-generated poems. However, the computer poet is still a competent poetry writer since it was able to fool the entire sample including the 18^{+} group. I speculate that the younger group struggled to identify the human poems simply due to a lack of intelligence and an inability to analyze texts or simply a lack of experience with poetry. However, I would like to look further into why younger participants struggled significantly more than the older group.

Even if the CJ and the computer poetry generator were combined into one program, or, even better, more aspects of human intelligence were added on (e.g. chatbot, voice capability, personality) in my opinion the developed program would never be able to be considered of human-level intelligence. Furthermore, in order to design an intelligent computer one would have to design multiple programs that complete human tasks, like making jokes, writing stories, or speaking. These programs would then all have to be combined in such a way that the end product would not be efficiently designed nor effective.

Therefore, the next step in intelligent computing should be to design machine learning programs that use specific elements of human intelligence to accomplish human tasks. I think this would alleviate the main struggle with designing an holistically intelligent computer. With this, the computer would simply have to demonstrate one specific element of human intelligence.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Kudlur, M. (2016). TensorFlow: A system for large-scale machine learning. In OSDI (Vol. 16, pp. 265-283).

Berrar, D. P., & Schuster, A. (2014). Computing machinery and creativity: Lessons learned from the Turing test. *Kybernetes*.

Bringsjord S., Bello P., & Ferrucci D. (2003) Creativity, the Turing Test, and the (Better) Lovelace Test. In: Moor J.H. (Eds.), The Turing test. Studies in Cognitive Systems, (vol. 30, pp. 201-213) Springer, Dordrecht

Dietterich, T. G. (2000). Ensemble methods in machine learning. In International workshop on multiple classifier systems (1-15). Springer, Berlin, Heidelberg.

Gamez, D. (2008). Progress in machine consciousness. *Consciousness and cognition,* 17(3), 887-910.

Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. *Nature, 521*(7553), 452-459.

Geron, A. (2017). Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems. Sebastopol, CA: O'Reilly Media.

Koch, C., & Tononi, G. (2011). A test for consciousness. *Scientific American, 304*(6), 44-47.

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering, 160,* 3-24.

Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: A review of classification and combining techniques. *Artificial Intelligence Review, 26*(3), 159-190.

LaCurts, K. (2011). Criticisms of the turing test and why you should ignore (most of) them. Official Blog of MIT's Course: Philosophy and Theoretical Computer Science.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436.

Li, Jiwei, ... (2016). Deep reinforcement learning for dialogue generation. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. *Association for Computational Linguistics, (1192-1202)* (2016).

Lin, L. J. (1992). Self-improving reactive agents based on reinforcement learning, planning and teaching. *Machine learning, 8*(3-4), 293-321.

Livingstone, D. (2006). Turing's test and believable AI in games. *Computers in Entertainment (CIE), 4*(1), 6.

Makridakis, S. (2017). The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. Futures, 90, 46-60.

McCarthy, J. (1979). Ascribing mental qualities to machines (No. STAN-CS-79-725). Stanford University Department of Computer Science..

Moscovitch, M. (1995). Models of consciousness and memory. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (1341-1356). The MIT Press.

Özesmi, S. L., & Özesmi, U. (1999). An artificial neural network approach to spatial habitat modelling with interspecific interaction. *Ecological modelling, 116*(1), 15-31.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. j (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12* (Oct), 2825-2830.

Radziwill, N. M., & Benton, M. C. (2017). Evaluating quality of chatbots and intelligent conversational agents.

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of research and development, 3*(3), 210-229.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., ... & Lillicrap, T. (2017). Mastering chess and shogi by self-play with a general reinforcement learning algorithm.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks, 61,* 85-117.

Schulz, P. J., & Nakamoto, K. (2013). Patient behavior and the benefits of artificial intelligence: The perils of "dangerous" literacy and illusory patient empowerment. *Patient education and counseling, 92*(2), 223-228.

Serban, I. V., Sankar, C., Germain, M., Zhang, S., Lin, Z., Subramanian, S., ... & Mudumba, S. (2017). A deep reinforcement learning chatbot..

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology, 61*(12), 2544-2558.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, LIX,* Issue (236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Vinyals, O., & Le, Q. (2015). A neural conversational model.

Weizenbaum, J. (1976). Computer power and human reason: From judgment to calculation. Oxford, England: W. H. Freeman & Co.

The Effect of Screen Usage on Sleep Quantity of Adolescents

by Luca Broder

Abstract

This study focuses on the impact of screen usage on sleep quantity among high school students. I found that higher screen usage was associated with decreased sleep quantity when I conducted a between-subject comparison of 10 participants' average screen usage (collected via the iPhone 'Screen Time' Setting) and sleep quantity (collected with the Fitbit Alta HR). Then, in order to determine whether screen usage was the cause of this decreased sleep, I conducted a second trial on new participants. These individuals used their phones a typical amount for one week, and then reduced their screen usage by 20% for the next one to two weeks (depending on the trial).

Background

Sleep deprivation has many adverse effects. It has been shown to negatively affect academic performance, as seen in decreased Grade Point Averages. (Smaldone, Honig & Byrne, 2007; Zeek, et al., 2015; Chee & Chuah, 2008; Taylor, et al, 2013; Amin, et al., 2015; Gilbert & Weaver, 2010). It also leads to a variety of physical ailments, including obesity (Wong, et al, 2013), depression (Wolfson, & Carskadon, 1998), and decreased reaction time and coordination (Patrick, Lee et al., 2017).

Unfortunately, high school students (Mary, Jodi & Christopher, 2006) in both private and public schools (Wheaton Jones, Cooper, & Croft 2018) across the country don't get enough sleep (Kathryn Orzech, David, Salafsky & Leem 2011). Sleep deprivation is both common, with 72.7% percent of high school students reporting insufficient sleep, and extreme, as about 20% of those students reported sleeping fewer than 6 hours on school nights (Wheaton, Jones, Cooper, & Croft, 2018).

Workload (Britz & Pappas, 2010), sedentary behavior (Kakinami, et al., 2017), and mood altering substances such as caffeine (Davies & Sue, 2010), marijuana (Schierenbeck, Riemann, Berger, & Hornyak 2008), and nicotine (Jaehne, Loessl, Bárkai, Riemann, & Hornyak, 2009) have all been found to decrease an individual's sleep quantity.

There is also overwhelming evidence of a link between screen usage and sleep quantity among adolescents (Carter, Rees, Hale, Bhattacharjee, & Paradkar, 2016; Harbard, Allen, Trinder, & Bei, 2016; Neralie & Gradisar, 2010). In one study, researchers examined the screen usage behavior of 9,846 adolescents ages 16–19 (Hysing, Pallesen, Stormark, Jakobsen, Lundervold & Sivertsen, 2015). They measured "frequency of electronic devices at bedtime and hours of screen time during leisure time" among participants, and found that participants who used screens more tended to sleep less, concluding "a negative relation between use of technology and sleep." In another study researchers administered

questionnaires to 738 adolescent participants and found that "Frequent use of all technology types was significantly inversely associated with weekday sleep duration." (Arora, Broglia, Thomas & Taheri, 2014).

The limitation of all of these studies is that since they are observational, rather than experimental, they do not establish causality. One can't determine causality without controlling for all the other factors that could affect sleep, something that you can only accomplish by using the same subjects throughout. In a systematic literature review of twenty studies concerning the "association between portable screen-based media device access or use and sleep outcomes," with data from 125,198 children with a mean age of 14.5, researchers acknowledged that "the limitations of research in this area include: measurement error of self-reported data;" and most importantly "difficulty in ascertaining causality," and "Therefore, a degree of caution is needed when interpreting these findings." (Carter, Rees, Hale, Bhattacharjee & Paradkar, 2016). Ultimately, they concluded that "media device access and use at bedtime is significantly associated with detrimental sleep outcomes, and lead to poor health outcomes," but were unable to determine whether screen usage was causing a decrease in sleep quantity. In another literature review, researchers concluded "that screen time is adversely associated with sleep outcomes (primarily shortened duration and delayed timing)" but that due to the consistent between-subject experimental designs of the studies, "causal association [could] not [be] confirmed" (Hale & Guan, 2015).

In this study, I examine this same correlation but by using a within-subject design, and by experimentally manipulating screen usage I determined if screen usage is truly inhibiting sleep among high school students, or if it is just occupying time that would have been spent doing other activities. This within-subject design allows me to control for all of the factors that differ between individuals that could affect sleep quantity.

Methods

This study is composed of two trials. In Trial #1, I monitored participants' sleep and screen usage, and compared them to each other to determine whether people who used their phones more tended to sleep less. In Trial #2, the experimental portion of my study, participants reduced their screen usage, and I monitored how their sleep was affected.

Trial #1 (2019)

To collect subjects I reached out to the Berkeley Carroll 10th Grade Dean, who forwarded my email to the entire grade seeking study participants. Ten individuals (all of whom had an iPhone) either responded to the email or came up to me in person asking to take part in the study. The choice to have ten subjects was based on limited resources (Fitbits are expensive) and I felt that ten subjects would provide enough data to yield statistically significant results and meaningful conclusions. After I acquired subjects, I distributed Fitbits and instructed subjects to wear them all day, every day, for the next week. To track phone usage, participants turned on "Screen Time" in the iPhone settings application. Each week, participants shared screenshots of their screen time as well as their Fitbit logins which allowed me to access their sleep data. Of my original ten subjects, three individuals did not wear their Fitbit for the duration of the study.

Trial #2 (2a and 2b, 2019)

Sampling

Subjects of ages 16-18 from the Berkeley Carroll Upper School participated in Trial 2 of the study. Subjects were recruited via an email to the 11th grade (2a), an email to the 12th grade (2b), and an announcement during 12th grade morning meeting. A total of 29 individuals agreed to participate, (10 in Trial 2a and 19 in Trial 2b), of which 21 reduced their screen usage by at least 20% and completed all other tasks required for their data to be used in my analysis. 11th graders (2a) and 12th graders (2b) were chosen to increase the diversity of the subject pool (I sampled from the 10th grade previously in Trial 1) and increase the number of potential participants. The subject count was limited by the number of Fitbits, and lack of interest in the study (after sending out the initial email, several requests for participation were made on a one-on-one basis to no avail). For both trials, anonymity was impossible as it was necessary for me to personally message participants with their decreased screen usage goal. However, once the data was collected and organized for analysis, names were removed.

Materials

Fitbit Alta HRs (2a) and Inspire HRs (2b) were used to monitor sleep quantity (minutes) of subjects. These devices were chosen for their affordability ($99 2a and $89 2b) and ability to track sleep quantity with accuracy (Zambotti, et al. 2016; Ferguson, Rowlands, Olds, & Maher 2015). The decision to purchase Fitbit Inspire HRs for Trial 2b was based on their longer battery life and their waterproof property, allowing participants to wear them throughout the entire day. For all trials, the iPhone "Screen Time" setting was used to monitor participant screen usage (minutes). This software was ideal as it is pre-installed into all iPhones and provided automatic phone usage data. A daily form was sent out to participants (2a) that provided them with a place to disclose any extenuating circumstances that could have affected their sleep quantity and a Feedback form was provided to subjects (2b) where they could articulate any issues they encountered with the apparatus or provide any suggestions for how to revise the experiment.

Procedure

Participants were provided with a Fitbit and instructed on how to access the iPhone "Screen Time" software. Participants spent a week using their phones as normal and recording their sleep with their Fitbits. Daily forms (2a) or a Feedback Form (2b) were sent out to participants where they could disclose any extenuating circumstances that could have affected sleep quantity. After the first week, participants shared their screen usage and their sleep quantity. Participants were instructed to limit their screen usage by 20% for the next one[1] (2b) or two weeks (2a). To help ensure this, they enabled a Screen Time Limit in their settings app, which reminded them when they were getting close to reaching their maximum screen usage for the day. After the period of reduced screen usage, participants shared their new screen usage Fitbit sleep quantity data. I compared their average sleep quantity before and after reduced screen usage.

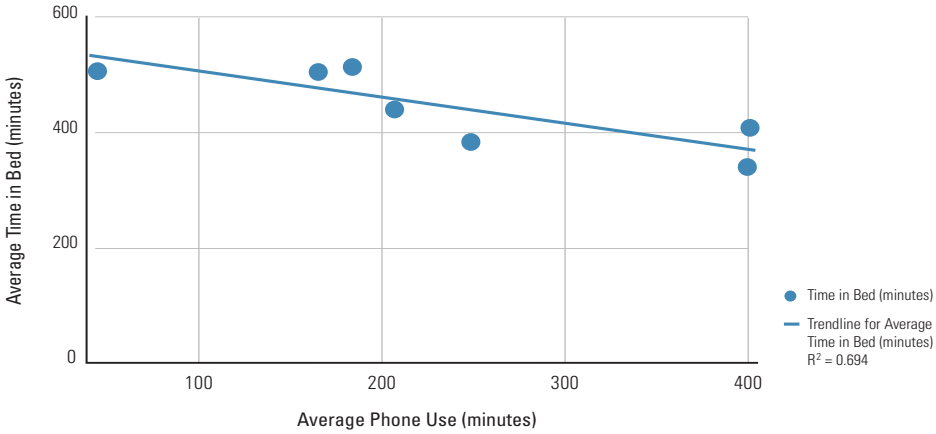
[1] Trial 2b occurred over two, rather than three, weeks so that one full round could be completed before Thanksgiving break, and another round could be completed before winter break.

Trial #1 Results

Figure 1

The Average Sleep Quantity and Screen Usage of Participants.

Participants' average sleep quantity and screen usage were plotted to determine whether participants with higher rates of screen usage tended to have a lower sleep quantity (r2 = .694). Ages ranged from 15-17 (N=7). A 2-tailed linear correlation and regression t-test indicated that there is a statistically significant negative correlation between screen usage and sleep quantity (p = .01989).

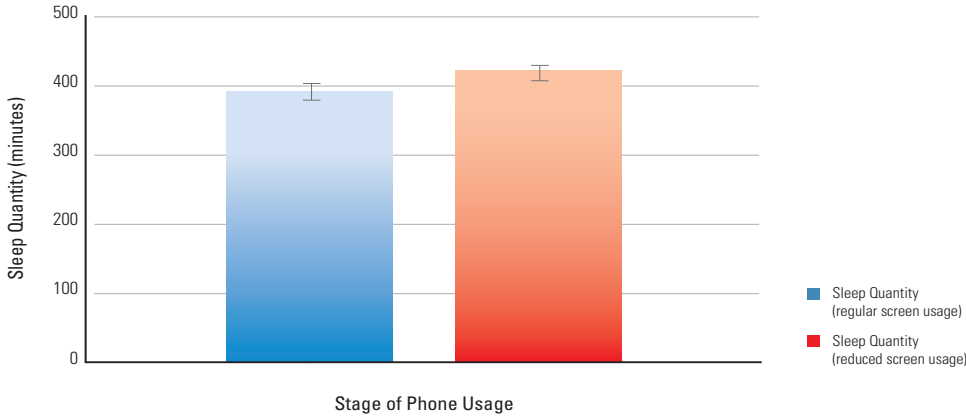


Trial #2 Results

Figure 2

Average Participant Sleep Quantity during Period of Regular Screen Usage vs. Period of Reduced Screen Usage.

The results above compare the average sleep quantity of participants during a period of regular screen usage and a period of reduced screen usage. Ages ranged from 15-17. Error bars represent the standard error of mean (average ± SEM). A 2-tailed correlated t-test showed that there was a statistically significant difference (P = .008658) between sleep quantity during the period of regular screen usage and the period of reduced screen usage.



Results and Discussion

For the observational part of the study (Trial #1) a statistically significant negative correlation was found between screen usage and sleep (P = .02) *(Figure 1).* Once screen usage and sleep quantity data were documented, the data was anonymized. For the experimental part of the study (Trial #2), there was a statistically significant increase in sleep (p = 0.009) when participants reduced their screen usage *(Figure 2).*

The two trials worked together to demonstrate that screen usage inhibits sleep quantity. In the first trial, I was able to find a correlation between screen usage and sleep quantity. The link I observed supported results from previous studies (Carter, Rees, Hale, Bhattacharjee, & Paradkar, 2016; Harbard, Allen, Trinder, & Bei, 2016; Neralie & Gradisar. 2010). Like those studies, because I was comparing screen usage between-subjects, the link was not necessarily causal. There are so many variables to control for in this between-subject design that it is impossible to attribute the lower levels of sleep quantity among some participants to higher levels of screen usage. The only variable that I could feasibly account for was workload, which I controlled for by selecting all participants from the same grade. However, there was no effective way to account for other factors such as different living situations, work habits or varying workloads, among tenth graders.

The focus of Trial #2 was to advance my research into unexplored territory. The link that I had discovered in Trial #1 was necessary to progress into researching the causality of screen usage and sleep quantity. The experimental design I developed and implemented in Trial #2 allowed me to determine causality between screen usage and sleep quantity. When participants used their phones 20% less, they slept almost half an hour more. This shows that screen usage is truly inhibiting sleep among high school students, rather than just occupying time that would have been spent doing other activities.

There were numerous limitations to my study. First, the subject pool wasn't completely random because I believe members of our community that were familiar and comfortable with me were more likely to volunteer and participate in the study. Additionally, the subject pool was not very diverse as it consisted solely of Berkeley Carroll high school students. It is also worth noting that Fitbits are not medical devices specifically designed for monitoring sleep quantity, so there is room to improve the accuracy of the sleep measurements. In all trials, I did not account for other types of screen usage. Going forward, I would like to use software that would allow me to monitor participants' screen usage across all devices. Additionally, in further iterations of my study, I may investigate whether the decreased levels of sleep result in higher self-reported rates of sleepiness and mental and physical impairment. This might be interesting to examine, as different individuals require different quantities of sleep to function at their best. Lastly, I would like to examine how screen usage affects adults. I think a strong place to start would be among teachers at the high school where I sampled students for my trials. The discussion surrounding screen usage and its effect on sleep quantity is nowhere near finished, but I believe this study is a good first step.

Works Cited

2008 sleep in America poll—pleep, performance and workplace. (2015). *Sleep Health, 1*(2). doi:10.1016/j.sleh.2015.04.007

2011 sleep in America poll—technology use and sleep. (2015). *Sleep Health, 1*(2). doi:10.1016/j.sleh.2015.04.010

2015 sleep in America poll. (2015). *Sleep Health, 1*(2). doi:10.1016/j.sleh.2015.02.005

Amin, H., Almazroua, I., Alsahlan, A., Alrishan, M., Elmourad, H., Alotaibi, M., & Almohaisin, A. (2016). Effect of sleep deprivation on the attitude and performance of medical students, Riyadh, Saudi Arabia. *International Journal of Medical Science and Public Health, 5*(3), 575. doi:10.5455/ijmsph.2016.17112015207

Arora, T., Broglia, E., Thomas, G. N., & Taheri, S. (2014). Associations between specific technologies and adolescent sleep quantity, sleep quality, and parasomnias. *Sleep Medicine, 15*(2), 240-247. doi: 10.1016/j.sleep.2013.08.799

Britz J & Pappas E. (2010) Sources and outlets of stress among university students: Correlations between Stress and Unhealthy Habits. URJHS, 9, 1

Carskadon, M., Acebo, A., Seifer, C., Extended nights sleep loss and recovery sleep in adolescents. *Arch Ital Biol* 2001;139 (3) 301-312.

Carter, B., Rees, P., Hale, L., Bhattacharjee, D., & Paradkar, M. S. (2016). Association between portable screen-based media device access or use and sleep outcomes: A systematic review and Meta-analysis. *JAMA pediatrics, 170*(12), 1202-1208. doi:10.1001/jamapediatrics.2016.2341

Chee, M. W., & Chuah, L. Y. (2008). Functional neuroimaging insights into how sleep and sleep deprivation affect memory and cognition. *Current Opinion in Neurology, 24*(4), 417-423. doi:10.1097/wco.0b013e3283052cf7

Davies, Sue. Adolescents living the 24/7 lifestyle: Effects of caffeine and technology on sleep duration and daytime functioning. *Primary Health Care,* vol. 20, no. 3, 2010, pp. 14-14., doi:10.7748/phc.20.3.14.s23.

Fuller, Caitlyn, et al. Bedtime Use of Technology: Effects on sleep in children and adolescents. *Journal of Adolescent Health,* vol. 62, no. 2, 2018, doi:10.1016/j.jadohealth.2017.11.253.

Gilbert, S. P., & Weaver, C. C. (2010). Sleep quality and academic performance in university students: A wake-up call for college psychologists. *Journal of College Student Psychotherapy, 24*(4), 295-306. doi:10.1080/87568225.2010.509245

Hale, L., & Guan, S. (2015). Screen time and sleep among school-aged children and adolescents: A systematic literature review. *Sleep medicine reviews,* 21, 50-58. doi:10.1016/j.smrv.2014.07.007

Harbard, E., Allen, N. B., Trinder, J., & Bei, B. (2016). Whats keeping teenagers up? pre-bedtime behaviors and actigraphy-assessed asleep over school and vacation. *Journal of Adolescent Health, 58*(4), 426-432. doi:10.1016/j.jadohealth.2015.12.011

Hysing, M., Pallesen, S., Stormark, K. M., Jakobsen, R., Lundervold, A. J., & Sivertsen, B. (2015). Sleep and use of electronic devices in adolescence: Results from a large population-based study. *BMJ Open, 5*(1), e006748. doi:10.1136/bmjopen-2014-006748

Jaehne, A., Loessl, B., Bárkai, Z., Riemann, D., & Hornyak, M. (2009). Effects of nicotine on sleep during consumption, withdrawal and replacement therapy. *Sleep Medicine Reviews, 13*(5), 363—377. doi: 10.1016/j.smrv.2008.12.003

Kathryn M. Orzech PhD, David B. Salafsky MPH & Lee Ann Hamilton MA, CHES (2011) The state of sleep among college students at a large public university. *Journal of American College Health,* 59:7, 612-619, doi: 10.1080/07448481.2010.520051).

Kakinami, L., Oloughlin, E. K., Brunet, J., Dugas, E. N., Constantin, E., Sabiston, C. M., & Oloughlin, J. (2017). Associations between physical activity and sedentary behavior with sleep quality and quantity in young adults. *Sleep Health, 3*(1), 56-61. doi: 10.1016/j.sleh.2016.11.001

Lowry, Megan & Dean, Kayla & Manders, Keith. (2010). The link between sleep quantity and academic performance for the college student. *The University of Minnesota Undergraduate Journal of Psychology. 3.*

Mary AC, Jodi AM, Christopher D. 2006 sleep in America poll – teens and sleep. *Sleep Health: Journal of the National Sleep Foundation, Volume 1, Issue 2.*

Paruthi S, Brooks LJ, D'Ambrosio C, Hall WA, Kotagal S, Lloyd RM, Malow BA, Maski K, Nichols C, Quan SF, Rosen CL, Troester MM, Wise MS. Recommended amount of sleep for pediatric populations: A consensus statement of the American Academy of Sleep Medicine. *J Clin Sleep Med* 2016;12(6):785-786.

Patrick, Y., Lee, A., Raha, O., Pillai, K., Gupta, S., Sethi, S., Mukeshimana, F., Gerard, L., Moghal, M. U., Saleh, S. N., Smith, S. F., Morrell, M. J., … Moss, J. (2017). Effects of sleep deprivation on cognitive and physical performance in university students. *Sleep and Biological Rhythms, 15*(3), 217-225.

Schierenbeck, T., Riemann, D., Berger, M., & Hornyak, M. (2008). Effect of illicit recreational drugs upon sleep: Cocaine, ecstasy and marijuana. *Sleep Medicine Reviews, 12*(5), 381-389. doi: 10.1016/j.smrv.2007.12.004

Smaldone, A., Honig, J., Byrne, M (2007). Sleepless in America: Inadequate sleep and relationships to health and well-being of our nation's children. *Pediatrics, 119*(Supplement 1). doi:10.1542/peds.2006-2089f

Taylor, D. J., Vatthauer, K. E., Bramoweth, A. D., Ruggero, C., & Roane, B.(2013). The role of sleep in predicting college academic performance: Is it a unique predictor?. *Behavioral Sleep Medicine, 11*(3), 159-72.

Wheaton, A. G., Jones, S. E., Cooper, A. C., & Croft, J. B. (2018). Short sleep duration among middle school and high school students - United States, 2015. *MMWR. Morbidity and mortality weekly report, 67*(3), 85-90. doi:10.15585/mmwr.mm6703a1

Wolfson, A. R., & Carskadon, M. A. (1998). Sleep schedules and daytime functioning in Adolescents. *Child Development, 69*(4), 875. doi:10.2307/1132351

Wong, W. W., Ortiz, C. L., Lathan, D., Moore, L. A., Konzelmann, K. L., Adolph, A. L., Smith, E. O., Butte, N. F. (2013). Sleep duration of underserved minority children in a cross-sectional study. *BMC Public Health, 13,* 648. doi:10.1186/1471-2458-13-648

Zeek, M. L., Savoie, M. J., Song, M., Kennemur, L. M., Qian, J., Jungnickel, P. W., & Westrick, S. C. (2015). Sleep Duration and Academic Performance Among Student Pharmacists. *American Journal of Pharmaceutical Education, 79*(5), 63. doi:10.5688/ajpe79563

The Effects of Music on Stress in Adolescents

by Phoebe Degn

Abstract

The goal of this study was to determine how listening to music during the school day affected the stress of Berkeley Carroll students. The first trial investigated whether taking time to listen to self-selected music for 10 minutes during class reduces the perceived stress levels of ninth-grade students on a modified version of the Perceived Stress Scale. The difference in stress between experimental groups (music) and control groups (no music) demonstrated no significant difference suggesting that listening to music prior to taking a test did not alter perceived stress. The aim of the second trial was to compare the average differences in heart rate and heart rate variation readings of eighth grade students before and during an online math quiz for experimental and control groups ($n = 24$). The results showed that there was a significant decrease in heart rate ($p = 0.03$) and increase in heart rate variation for experimental groups compared to control groups ($p = 0.04$).

Background

Stress in Adolescents

The adolescence period in a person's life is known to be particularly stressful (Seiffge-Krenke, Aunola, & Nurmi, 2009). A study conducted at Tel Aviv University in Israel with the aim of systematically identifying the most relevant stressors perceived by undergraduate students in their school lives used Lazarus' definition of stress, which characterizes it as: "A person's perception of an environmental demand, addressed to her/him, which taxes or exceeds that person's adaptive resources" (Shirom, 1986). In other words, stress arises when someone is not able to meet the demands set for them by someone or something else. The results from this study supported this definition by indicating that examination-related stress was the most prominent among all the self-reported stressors (Shirom, 1986). Research on everyday stressors of adolescents shows that they are constantly faced with stress-inducing experiences relating to friends, parents, and school life (Seiffge-Krenke, et al. 2009). Physical and emotional development that occurs during adolescence is another stress-inducing factor. The imbalance of the development of the prefrontal and subcortical regions of the brain leads to a wide range of emotions adolescents experience (Casey, et al., 2010). Psychological changes heighten emotional reactions, which can give rise to anxiety disorders and other mental illnesses (Casey, et al., 2010). The evidence found about increased stress and anxiety in adolescents highlights the importance of finding effective coping methods.

Coping Methods for Stress

Compas (1987) explains that coping is divided into two categories: problem-focused coping, which is intended to act on the stressor, and emotion-focused coping, which attempts to regulate the emotions that arise as a response to the stressor. In his research on the problem and emotion-focused coping methods that were previously identified, Shahmohammadi (2001) found that adolescents tend to withdraw from their stressors rather than face them, showing that they are not properly coping. Planning and organization strategies (Leonard, et al., 2015), turning to friends and family (Leonard, et al., 2015), and seeking support in school (Armacost, 1990; Hains, 1994) have been identified as methods that effectively reduce stress. While these coping mechanisms do decrease stress and anxiety, not every student is able to participate in a training or counseling session. Therefore, it is important to identify other possible coping strategies that are available to a wider range of students that are equally effective.

Music and Stress

Music has played an influential role in civilizations for thousands of years. Yehuda (2011) states that researchers looking into music can "bridge between the clinical, therapeutical and basic research in music and its effects on stress and stress emotions." Khalfa (2003), Knight & Rickard (2001), and Thoma (2013) have all conducted studies on young adults using the Trier Social Stress Test (public speaking task followed by a mental arithmetic task in front of an audience) to determine whether music affects physiological responses to stress. Their results all support the idea that listening to music prior to completing a stressful task has a significant effect on cortisol levels and blood pressure. However, none of these studies have focused on the self-reported stress of participants, specifically in children and adolescents. Following Lazarus' definition of stress as being self-perceived, it is important to understand how subjects respond to the presence of music in order to find a coping method that best suits their needs. It would also provide a psychological aspect to the effects of music on stress rather than solely the physiological aspect.

Music in a School Environment

Campbell and Beegle's study on the effects of music and performance arts on middle and high school students demonstrated that adolescents develop a very personal relationship with music and the practice of it (2007). Similarly, music has been shown to improve elementary, high school, and college students' academic performance in math and reading (Schlichting, et al., 1970; Hall, 1952). However, these studies fail to look at the subjects' responses to the music. By allowing students to report how they respond to the presence of music during an exam, researchers could gain a better understanding as to why music leads to this rise in exam grades.

Music and Heart Rate Variability

Music therapy has been used as a way to increase parasympathetic nervous system activity, a measurement of stress that slows your heart rate and relaxes your muscles when an otherwise stressful situation would increase heart rate and lower heart rate variation (Latha, et. al., 2015; Kachanathu, et. al., S. 3013; Chuang, 2010; Roque, 2013). In other words, when subjects are in a more stress-inducing situation, such as the time before a sports competition or medical treatment, listening to different types or genres of music has been shown to increase heart rate variation, or the variation in time between each heartbeat. These studies have used classical and rock music, as well as self-selected music, which have been shown to be effective in terms of reducing stress and allowing the subject to be in a more

relaxed state (Latha, et. al., 2015; Kachanathu, et. al., S. 3013; Chuang, 2010; Roque, 2013). However, all of these studies have provided music therapy before the stress-inducing task, rather than at the same time. Listening to music while completing a stress-inducing task could have a different and potentially greater benefit on a subject's stress. This is why the current study will determine if listening to relaxing classical music while performing a math quiz affects the heart rate and heart rate variation of subjects.

Current Study

Trial 1

The first trial of the study looked generally at stress in a school environment to determine whether music is an effective coping mechanism. Each student chose their own music to listen to. There has not been much research done on the effects of self-selected music on stress, which is why this trial incorporated it into the study. It is important to investigate self-reported stress because as Lazarus says in his definition, self-perceived stress appears to be the only kind of stress (Shirom, 1986). Listening to self-selected music will ensure that students enjoy the music they are listening to. This trial used a modified version of the Perceived Stress Scale (a scale that measures self-reported stress over the past month) to measure psychological stress, which has been shown to be an effective way to measure psychological stress in previous studies (Masa'Deh, Rami, et. al., 2015; Ngyen-Rodriguez, 2007; Schiffrin, 2008).

Trial 2

The goal of the second trial was to determine how relaxing music affects heart rate variation during a stress-inducing task. This trial had all subjects listen to the same music rather than self-selected music. Classical music has been used in previous studies and has shown to be an effective means of reducing stress, which is why it was chosen for this study (Latha, et. al., 2015; Roque, et., al., 2013; Chuang, et. al., 2010). Previous studies I've seen have used music therapy as a way of reducing physiological responses prior to a stress-inducing task (Khalfa, 2003; Knight, & Rickard, N. S., 2001; Thoma, et al., 2013; Latha, et. al., 2015; Kachanathu, et. al., S. 3013; Chuang, 2010; Roque, 2013). However, this study had subjects listen to music while completing the stress-inducing task to see if it significantly affected their heart rate variation. As the first trial investigates psychological measures of stress, determining the physiological responses to stress in this trial will further demonstrate how music can affect a person in different ways. Much of the research that has been done on students has used university students as their subjects, while this study will use middle schoolers (ages 12-13). This trial used an online math quiz with a time limit as a way to create a stress-inducing environment that remains the same for each subject.

Methods: Trial 1

Sampling

Subjects of ages 14-15 from the Berkeley Carroll Upper School participated in the study during their history class. Two history classes were randomly assigned to participate in the experimental group and the other two in the control group. The study was conducted while students were writing research papers during class time, so participating wasn't a large inconvenience to teachers or students. Teachers were asked to select one of their classes to act as the experimental group and the other as the control group, while still allowing each student to opt out of the experiment (four classes in total, n = 63). This trial received IRB approval before it was conducted and each subject was given an information sheet and provided informed consent.

Materials

The Perceived Stress Scale

The PSS-10 is the most commonly used scale for determining a person's self-reported stress (Cohn, 1994). The 10 items in the scale ask how "unpredictable, uncontrollable, and overloaded respondents find their lives" over the course of the past month. Scores on each question range from 0 to 4, with 0 being never and 4 very often. Four of the questions are categorized as "positively stated items," such as "In the last month, how often have you felt confident about your ability to handle your personal problems?" while the remainder are categorized as "negatively stated items", such as "In the last month, how often have you been upset because of something that happened unexpectedly?" Scores on those questions are reversed to calculate the overall measure of stress (0=4, 1=3, etc.). The PSS-10 has been used in many studies to determine subjects' perceived stress as a result of variables such as parenting and school (Masa'Deh, Rami, et. al., 2015; Ngyen-Rodriguez, 2007; Schiffrin, 2008). For the purpose of the current study, the PSS-10 was modified to better suit the experimental design. Rather than having 10 items that subjects would fill out at once, the scale was split into two parts so subjects wouldn't remember the questions when they completed the scale a second time. The items were evenly split so there would be 8 positively phrased questions and 8 negatively phrased questions. There were also 6 items added to the scale that asked questions similar to those on the PSS-10 except worded differently so rather than asking how an individual had been feeling over the past month, it asked how they were feeling at that specific moment. The scaling of the modified stress scale is the same with each score ranging from 0-4 per item.

Procedure

For the experimental groups, participants completed the modified version of half of the PSS-10 right at the beginning of class. Then, they were asked to silently sit and listen to their own music for 10 minutes with headphones. While listening, they were asked to work on their research paper and then complete the other half of the modified version of the PSS-10. Students in the control group completed half of the modified version of the PSS-10, silently did school-related work for 10 minutes, and then completed the other half of the modified version of the Perceived Stress Scale.

Summary of Results: Trial 1

For the two classes that acted as the experimental groups (who listened to music for 10 minutes between completing the stress scale), a two-sample correlated t-test determined that the average difference between scores wasn't statistically significant ($p = 0.82$). This indicates that the music didn't lower their stress levels compared to before they listened to the music. A two-sample correlated t-test also determined that the average difference between stress levels of control groups also wasn't statistically significant ($p = 0.12$). This suggests that neither the experimental or control groups had a reduction in stress before and after the 10 minutes of either listening to music and doing schoolwork or just doing schoolwork. When comparing the differences in stress levels before and after the 10 minutes between experimental and control groups, a two-sample independent t-test demonstrated that the scores weren't significantly different ($p = 0.17$). This suggests that while the change in stress wasn't the same for experimental and control groups, the differences in stress weren't significantly greater for experimental groups. Overall, the results from this trial of the study weren't significant, which indicates that music doesn't have an effect on stress.

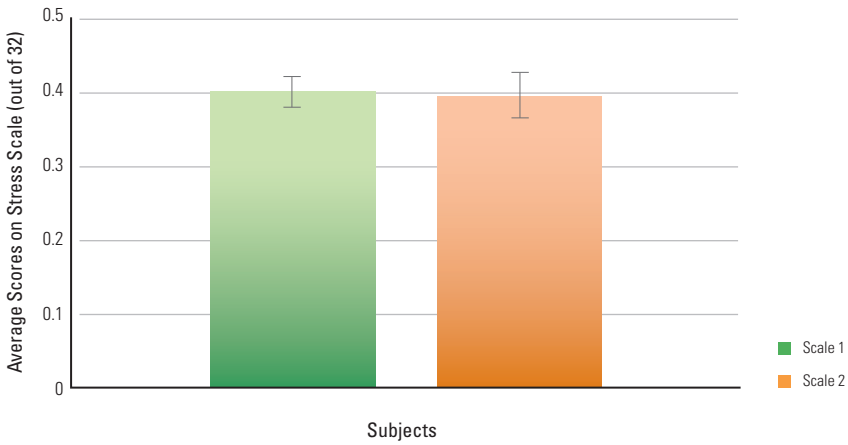
Results: Trial 1

Figure 1

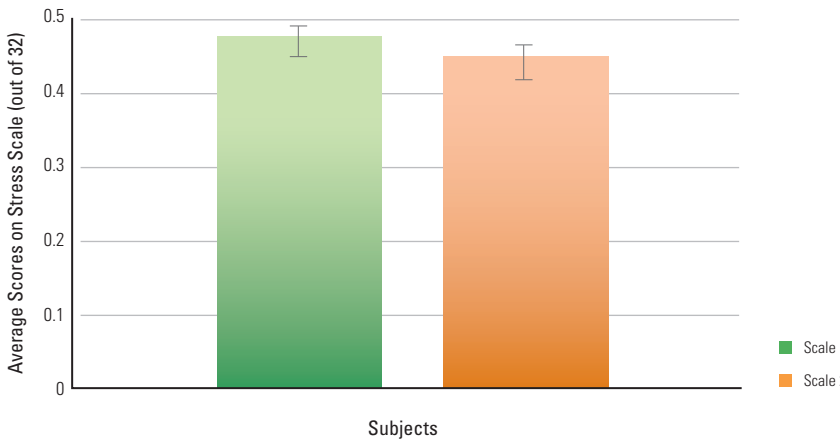
The Differences in Stress Levels for Experimental and Control Groups.

The modified version of the PSS-10 (Perceived Stress Scale), scored out of 32 points, was used to quantify students' level of perceived stress before and after listening to self-selected music for 10 minutes (n = 29 for experimental groups, n = 34 for control groups). Ages range from 14-15. The results below depict the self-reported stress levels of students before and after listening to the music. Error bars represent the standard error of mean (average ± SEM). A two-sample correlated t-test showed that the difference in stress levels for experimental groups and control groups is not significant (p = 0.82 for experimental groups, p = 0.12 for control groups).

Experimental Groups



Control Groups



Methods: Trial 2

Sampling

The dean of the Berkeley Carroll Middle School provided permission to allow the study to be conducted during 8th grade study hall periods, with ages of the subjects ranging from 13-14. 8th graders were preferred for this study because I felt that they were going to be more comfortable putting the heart rate monitors under their shirts than younger students. I also didn't use older students because I had used them in previous studies and wanted a new subject pool. In each study hall, students were given a basic overview of the procedure and volunteered to participate. Members of the experimental and control groups alternated, with the first volunteer being in the experimental group and the second in the control group. The number of students participating in the study hall depended on how many expressed interest (four study halls, n = 24).

Materials

Heart Rate Variation Monitor

The Mindfield eSense Pulse is a pulse sensor connected to an app that measures heart rate variation. It comes with a chest strap that is worn below the chest, as well as an electrode contact spray. This measurement tool was chosen because it is small, easy to strap on, and less expensive in comparison to other heart rate variation monitors considered for this study. The app provides a wide range of physiological measurements, including heart rate variation score (calculated through a formula developed by Mindfield), average heart rate, minimum and maximum heart rate, stress index, RMSSD (Root Mean Square of Successive Differences), etc. For the purpose of this study, only RMSSD and average heart rate values were used to determine results because those were the only two measurements that addressed the experimental question and have been used as a measurement of stress in previous studies (Latha, et. al., 2015; Kachanathu, et. al., 2013; Chuang, 2010; Roque, 2013). RMSSD, or the Root Mean Square of Successive Differences, describes how strongly the heart rate changes from one beat to the next. Heart rate is measured in beats per minute, and a lower heart rate indicates that an individual is in a more relaxed state (Mindfield eSense Pulse, 2019).

Preliminary Questions

Each student was given a list of 10 questions to answer before completing the math quiz in order to determine pre-experimental RMSSD and Heart Rate readings. The questions were non-stress inducing so they could be compared to the RMSSD and Heart Rate readings during the quiz.

Math Quiz

The online math quiz contained 30 questions which were evenly split into 5 pages. The problems involved either addition, subtraction, multiplication, division, or a combination of these operations. Each page had a 45 second time limit and the time would count down in the top right corner of the screen. Once the time ran out, the next page of the quiz would appear even if the student hadn't answered all the questions on the previous quiz.

Relaxing Music

A classical music piece by Beethoven was chosen for this experiment. Beethoven was selected due to his popularity and because there are no lyrics in his work, since they could potentially distract the subject while taking the quiz. Classical music has been used in previous studies as a type of relaxing music for reducing stress (Latha, et. al., 2015; Roque, et. al., 2013; Chuang, et. al., 2010).

Procedure

Within each of the four study halls, 8th graders were given a brief overview of the study and they were asked if they wanted to participate. The students in the experimental group were first asked to put on the heart rate monitor under their shirt in the bathroom following a set of instructions. Once the chest strap was working, they answered a set of preliminary questions so a pre-experimental heart rate and RMSSD reading could be taken. They were then asked to put headphones on to listen to a YouTube video playing a classical piece of music by Beethoven. They started the online math quiz immediately after beginning the music. The online math quiz was created on FlexiQuiz (FlexiQuiz) as a way to induce stress in the 8th graders. The 30 problems were separated into six different pages with five problems per page. The problems contained addition, subtraction, multiplication, and simple algebra. This level of difficulty was chosen because all 8th graders know how to complete it. Each page had a 45 second time limit and the next page would appear if a student hadn't answered all the questions in time. Once the quiz ended, subjects stopped the music and returned the chest strap. They were then sent a Google Form asking questions about the math quiz and its level of difficulty, the music and how much they enjoyed it, if they felt comfortable putting the chest strap on under their clothes, and if they had any suggestions for the study. The procedure for the control group was very similar, with subjects first completing the preliminary questions to determine their pre-experimental readings. However, after they completed the math quiz, they did not listen to music and just sat in a silent room. They were also sent a Google Form later that day, but didn't answer questions about the music.

Summary of Results: Trial 2

The goal of this trial was to determine whether the music that the students listen to during the quiz could significantly decrease their heart rate and significantly increase their RMSSD. The difference in heart rate and RMSSD was calculated for each student in the music and no music groups and the average differences between the two groups were compared. A two-sample independent t-test demonstrated that for heart rate, there was a significant difference between the music and no music groups ($p = 0.03$), where the average heart rate for the music group decreased during the quiz and increased for the no music group. A two-sample independent t-test showed that for RMSSD, there was also a significant difference between the music and no music groups ($p = 0.04$), where the average RMSSD for the music group increased during the quiz and decreased for the no music group.

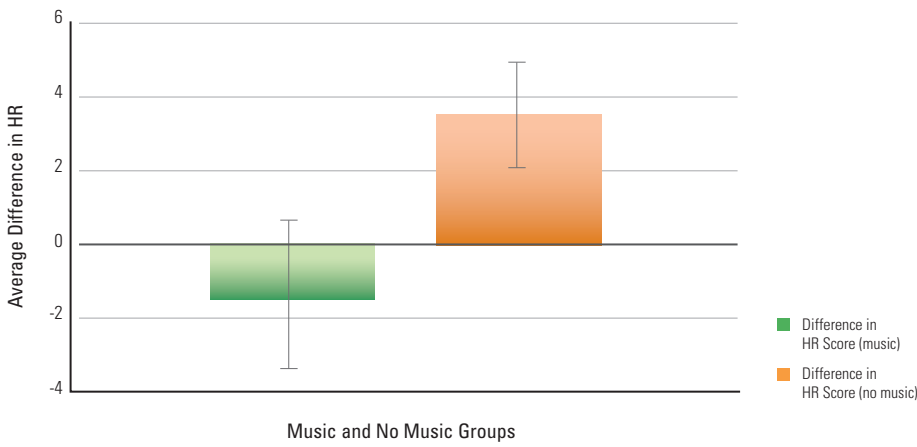
Results: Trial 2

Figure 2

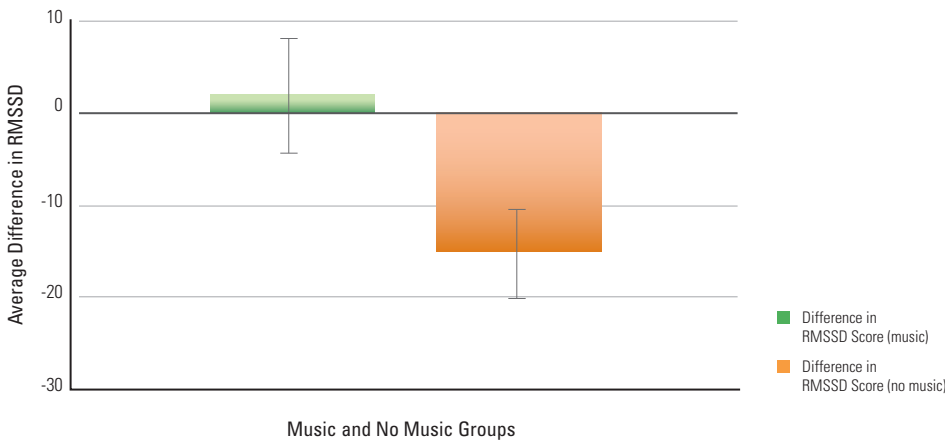
The Average Differences in Heart Rate and RMSSD for Experimental and Control Groups.

The Mindfield eSense Pulse was used to quantify students' heart rate and RMSSD before and during the online math quiz (n = 24). Ages ranged from 13-14. The results below compare the average difference in heart rate and RMSSD before and during the math quiz for experimental and control groups. Error bars represent the standard error of mean (average ± SEM). A two-sample independent t-test showed that the difference in average heart rate change between experimental and control groups is significant (p = 0.03) and the difference in average RMSSD change between experimental and control groups is significant (p = 0.04).

Heart Rate Difference for Music and No Music Groups



RMSSD Difference for Music and No Music Groups



General Discussion

The two trials in this study worked together to determine the role of music in different aspects of adolescents' stress. The first trial demonstrated that listening to self-selected music while completing schoolwork does not lower the perceived stress of students. This allows me to conclude that it's possible that perceived stress might not change or there is no relationship between music and stress. However, there were also limitations present in this study. It is possible that the 10 minute period of listening to music was not long enough to significantly affect their perceived stress. There could have also been distractions in the classroom that made it difficult to concentrate on the music and schoolwork. Future studies should have students participate individually in a quiet room so there are no distractions and have them listen to music for longer to determine if there is in fact an effect of music on stress.

The second trial took a different approach by examining physiological measures of stress rather than self-reported stress. Unlike the first trial, there was also an induced stressor to ensure that students would feel stressed, which would therefore increase heart rate and decrease heart rate variation. Students in the experimental group all listened to the same piece by Beethoven to act as a control. The results from this study were significant. The average heart rate decreased for experimental groups and increased for control groups, while the average heart rate variation increased for experimental groups and decreased for control groups. However, there were still limitations that arose in this study. Each student was emailed the post-experiment survey right after they completed the experiment, but the majority of students did not complete it. This made it difficult to get proper feedback on the experimental design. The survey asked whether they felt comfortable putting the chest strap on under their shirt and a few people responded that they didn't. Going forward, it was made clear that students did not have to participate in the study if they did not feel comfortable.

While I have not found evidence that music affects self-reported stress, the second trial demonstrated that listening to music while completing a stress-inducing task can lower students' heart rate and RMSSD. These contradicting results raise an important question, which is whether or not students are aware of their body's response to music. The results from the first trial demonstrate that students don't in fact feel that music relieves their stress, but their physiological responses to music say otherwise. However, the results from the second trial support the findings in previous studies that music, classical music in particular, can decrease heart rate and therefore increase heart rate variability when an individual is under stress (Latha, et. al., 2015; Kachanathu, et. al., S. 3013; Chuang, 2010; Roque, 2013). However, unlike these trials, the trial in the current study had students listen to music while completing the stress-inducing task rather than before, which was shown to be effective.

References

Armacost, R. (1990). High school student stress and the role of counselors. *The School Counselor, 38*(2), 105-112.

Campbell, P. S., Connell, C., & Beegle, A. (2007). Adolescents expressed meanings of music in and out of school. *Journal of Research in Music Education, 55*(3), 220-236. doi:10.1177/002242940705500304

Casey, B., Jones, R. M., Levita, L., Libby, V., Pattwell, S. S., Ruberry, E. J., & Somerville, L. H. (2010). The storm and stress of adolescence: Insights from human imaging and mouse genetics. *Developmental Psychobiology*. doi:10.1002/dev.20447

Chuang, Chih-Yuan & Han, Wei-Ru & Li, Pc & Young, Shuenn-Tsong. (2010). Effects of music therapy on subjective sensations and heart rate variability in treated cancer survivors: A pilot study. *Complementary Therapies in Medicine 18,* 224-6. 10.1016/j.ctim.2010.08.003.

Cohen, S. (1994). Perceived Stress Scale. Retrieved from www.mindgarden.com/documents/PerceivedStressScale.pdf

Compas, B. E. (1987). Coping with stress during childhood and adolescence. *Psychological Bulletin, 101*(3), 393-403. doi:10.1037//0033-2909.101.3.393

FlexiQuiz. Retrieved from www.flexiquiz.com/

Friedman, Isaac & Bendas-Jacob, Orit. (1997). Measuring Perceived Test Anxiety in Adolescents: A Self-Report Scale. Educational and Psychological Measurement - EDUC PSYCHOL MEAS. 57. 1035-1046. 10.1177/0013164497057006012.

Hains, A. (1994). The Effectiveness of a School-Based, Cognitive-Behavioral Stress Management Program With Adolescents Reporting High and Low Levels of Emotional Arousal. *The School Counselor, 42*(2), 114-125.

Hall, J. C. (1952). The Effect of Background Music on the Reading Comprehension of 278 Eighth and Ninth Grade Students. *The Journal of Educational Research, 45*(6), 451-458. doi:10.1080/00220671.1952.10881962

Kachanathu, S. J., Verma, S. K., & Khanna, G. L. (2013). Effect of Music Therapy on Heart Rate Variability: A Reliable Marker to Pre-competition Stress in Sports Performance. *Journal of Medical Sciences (Faisalabad),13*(6), 418-424. doi: 10.3923/jms.2013.418.424

Khalfa, S., Bella, S. D., Roy, M., Peretz, I., & Lupien, S. J. (2003). Effects of Relaxing Music on Salivary Cortisol Level after Psychological Stress. *Annals of the New York Academy of Sciences, 999*(1), 374-376. doi:10.1196/annals.1284.045

Knight, W. E., & Rickard, N. S. (2001). Relaxing Music Prevents Stress-Induced Increases in Subjective Anxiety, Systolic Blood Pressure, and Heart Rate in Healthy Males and Females. *Journal of Music Therapy, 38*(4), 254-272. doi:10.1093/jmt/38.4.254

Latha, R., Tamilselvan, K., Susiganeshkumar, E., & Sairaman, H. (2015). Effect of Classical music on heart rate variability between genders. *International Journal of Biomedical Research, 6*(3), 192. doi: 10.7439/ijbr.v6i3.1841

Leonard, N. R., Gwadz, M. V., Ritchie, A., Linick, J. L., Cleland, C. M., Elliott, L., & Grethel, M. (2015). A multi-method exploratory study of stress, coping, and substance use among high school youth in private schools. *Frontiers in Psychology, 6.* doi:10.3389/fpsyg.2015.01028

Masa'Deh, Rami., Bawadi, Hala., Saifan, A.hmad, & Aburuz, Mohannad. (2015). Perceived stress of Jordanian parents: A comparative study between mothers and fathers. *Journal of Nursing Education and Practice, 5*(11). doi:10.5430/jnep.v5n11p89

Mindfield eSense Pulse (2019). eSense Pulse Manual. *Mindfield Biosystems, 4*(2).

Ngyen-Rodriguez, S. T., Chou, C., Unger, J. B., & Spruijt-Metz, D. (2007). BMI as a Moderator of Perceived Stress and Emotional Eating in Adolescents. *PMC, 9*(2), 238-246. doi:10.1016/j.eatbeh.2007.09.001

Roque, A., Valenti, V., Guida, H., Campos, M., Knap, A., Vanderlei, L., … Abreu, L. (2013). The effects of auditory stimulation with music on heart rate variability in healthy women. *Clinics, 68*(7), 960-967. doi: 10.6061/clinics/2013(07)12

Schiffrin, H. H., Rezendes, D., & Nelson, S. K. (2008). Stressed and Happy? Investigating the Relationship Between Happiness and Perceived Stress. *Journal of Happiness Studies, 10*(4), 503-503. doi:10.1007/s10902-008-9118-1

Schlichting, H. E., & Brown, R. V. (1970). Effect of Background Music On student Performance Effect of Background Music On Student Performance. *The American Biology Teacher, 32*(7), 427-429. doi:10.2307/4443158

Seiffge-Krenke, I., Aunola, K., & Nurmi, J. (2009). Changes in Stress Perception and Coping during Adolescence: The Role of Situational and Personal Factors. *Child Development, 80*(1), 259-279.

Shahmohammadi, N. (2011). Students' coping with Stress at high school level particularly at 11th & 12th grade. *Procedia - Social and Behavioral Sciences, 30,* 395-401. doi:10.1016/j.sbspro.2011.10.078

Shirom, A. (1986). Students' Stress. *Higher Education, 15*(6), 667-676.

Sills, D. & Todd, A. (2015). Does Music Directly Affect a Person's Heart Rate?. *Journal of Emerging Investigators.*

Southgate, D. E., & Roscigno, V. J. (2009). The Impact of Music on Childhood and Adolescent Achievement. *Social Science Quarterly, 90*(1), 4-21. doi:10.1111/j.1540-6237.2009.00598.

Thoma, M. V., Marca, R. L., Brönnimann, R., Finkel, L., Ehlert, U., & Nater, U. M. (2013). The Effect of Music on the Human Stress Response. *Plos One. 8*(8). doi: 10.1371/journal.pone.0070156

Yehuda, Nechama. (2011). Music and Stress. *Journal of Adult Development.* 18. 85-94. doi:10.1007/s10804-010-9117-4.

Difference in Growth of Romaine Lettuce Plants: Hydroponics vs. Aeroponics

by Donald Minerva

Abstract

The goal of this experiment was to replicate a previous study (Gashgari, Alharbi, Mughrbil, Jan, & Glolam, 2018) to investigate if lettuce growth differs between hydroponics, a method of growing plants without using soil, and aeroponics, a method of growing plants suspended in the air. Lettuce samples grown in both hydroponics and aeroponics systems were measured using four dependent variables: change in root length (cm), change in stem length (cm), change in mass (grams), and change in the number of leaves (number). The setup consisted of two hydroponic kits, two aeroponics kits, grow lights (placed above all four kits; Sylvania 20W fluorescent tube light), and seven lettuce plant seedlings. Data was recorded Monday and Friday for three weeks. A two tailed independent t-test was conducted and demonstrated a significantly greater change in mass for the hydroponics kits compared to the aeroponics kits ($p= 0.000112$). However, there were no statistical differences for change in stem length, root length, and number of leaves.

Background

Introduction

Hydroponics is the process of growing plants in a soilless medium, with the help of nutrient rich solution that contains phosphates, nitrates, ammonia, and other chemical components of soil. With the help of clay beads, which act as a growing medium for the roots of the plants, the solution is absorbed and transferred to the plants. Aeroponics is a similar form of soilless agriculture, except there are no clay beads and the roots hang freely, getting sprayed periodically with a nutrient solution (Raviv, Lieth, & Raviv, 2008).

The use of soilless agriculture is revolutionizing the way in which humans grow food. Both hydroponics and aeroponics are fairly new to commercial use. They have improved the ways in which food is grown (Sheikh, 2006; Raviv, Lieth, & Raviv, 2008; Lakkireddy, et al., Apr. 2012.)

Advantages of Soilless Agriculture

Hydroponics is versatile. In one study, researchers talked about how soil is completely eliminated from this type of farming (Lakkireddy, et al., Apr. 2012.), causing a decrease in reliance on clean soil. This decrease is very important looking into the future, when there is a constant increase in population and

industrialization and less accessible farmland. Studies have shown that soilless agriculture, specifically hydroponics and aeroponics, is a very powerful tool for growing crops and is better than traditional methods (Carver & Wasserman, 2012; (Martin-Laurent, Lee, Tham, Jie, & Diem, 1999). A study comparing the growth of *Acacia mangium* seedlings, using aeroponics vs. soil, concluded that the plants grown aeroponically grew three times larger than those grown in soil (Martin-Laurent, Lee, Tham, Jie, & Diem, 1999).

Hydroponics also has great potential to compete with soil (Gashgari, Alharbi, Mughrbil Jan, Trial 1& Glolam, 2018). The researchers mentioned in the discussion section that plants grown hydroponically grew faster than plants grown using soil (Gashgari, Alharbi, Mughrbil, Jan, & Glolam, 2018).

In a study conducted by Martin-Laurent et al. researchers compared the difference in growth of various cucumber types and found that the lengths of the cucumbers were greater in hydroponics by almost double compared to the same variety of cucumber grown in soil (Martin-Laurent, Lee, Tham, Jie, & Diem, 1999).

Research has shown that hydroponics and aeroponics can still be adapted to fit even in conditions where the farmland is unsuitable for growing plants, such as when the land is too dry (Yoshitani, Norimichi, & Merabtene, 2007) or there is constant flooding (Treftz, & Omaye, 2016). These soilless methods are becoming popular due two main issues facing the world today: rapid increase in population and change in climate (Sengupta, Amrita, & Banerjee, 2012).

Hydroponics vs. Aeroponics

However, it now comes down to determining which of the two is better. In a study conducted by Gashgari and their team, they investigated whether hydroponics or aeroponics was better to cultivate lettuce plants in different categories (height of plants and length of leaves) (Gashgari, Alharbi, Mughrbil, Jan, & Glolam, 2018). They planted lettuce seeds in a hydroponic system, aeroponic system, and traditional soil beds as a control. From this experimental setup, they concluded that plants grown hydroponically and in soil weighed more than plants grown aeroponically. Additionally, the roots were more mature (root volume, length, yield) than plants grown aeroponically. Overall, lettuce grown hydroponically matured more efficiently in the same amount of time compared to those grown aeroponically (Gashgari, Alharbi, Mughrbil, Jan, & Glolam, 2018).

Justification

For my experiment, I investigated which soilless approach, hydroponics or aeroponics, allows for increased growth of Romaine lettuce plants. My experimental question was: Is there a difference in growth rate of Romaine lettuce between hydroponics and aeroponics using a manufactured solution? I measured change in number of leaves, mass, and root/stem length for plants grown in hydroponics and aeropinics kits. I controlled the solution type (FloraPro 9-11-19), light exposure (Sylvania 20W fluorescent tube lights), and plant type (Romaine lettuce). The sample size was 28 Romaine lettuce plants. This was a replication study similar to the one Gashgari and their team conducted. I wanted to understand firsthand whether hydroponics or aeroponics is better in terms of how mature the lettuce plants become.

Methods

Overview

For this experiment, I started germinating lettuce seeds using Sylvania 20W fluorescent tube lights and traditional soil from Miracle Grow. Seeds were planted in a 27 x 12 in. growing trough and watered with 150 mL of water daily. After one week, I transferred seedlings into smaller growing containers (2 in nursery pots) so they had more room to grow independently. The plants were exposed to the same amount of light, but less water (~50 mL) than used during the germination phase because they were in smaller growing cups. After another week, the plants were then tagged. Tags were made out of tape and were numbered 1-7 for each plant. Initial measurements were taken (see *Taking Initial and Final Measurements* for details). Samples were marked with a sharpie to act as a reference/control as the plants grew and they were separated evenly by size into hydroponic/aeroponic growing kits. Stem, root, and number of leaves were measured every Monday and Friday for three weeks. This time difference between data collection days was chosen to give enough time to see a noticeable difference. The mass and root length were only calculated at the beginning and end of the experiment. After three weeks, all plant samples had final measurements taken (the same as the initial) and statistical tests were run.

Growing Plants

First, I planted the lettuce seeds within a large grow bed *(Figure A).* Seeds were placed about a knuckle length deep and 3-5 seeds were placed in each hole in order to increase the yield of plants needed for the experiment. All plants were placed under grow lights and grown in the same type of soil. After a week, plants were transferred to smaller growing containers *(Figure B)* to allow better growing conditions. After another week, plants were transferred into the growing systems.

Figure A

Growing Plants in Large Container.

Figure B

Transferring Plants into Smaller Container to Increase Growth.

Taking Initial/Final Measurements

Plants were individually weighed using a weight boat and a scale. Additionally, the root lengths and meristem apical lengths (from where the stem began to where the first two leaves bloomed) were measured. All measurements were recorded in centimeters because of how small they were. Finally, the number of leaves was recorded (only live ones – if plants were brown, limp or losing leaves, they were considered dead). All data was recorded in a *Google Spreadsheet.* The mass and root length were recorded before and after the plants were growing – stem length and number of leaves were recorded every data collection day.

Figure C

Final Measurements on Last Day of Experiment.

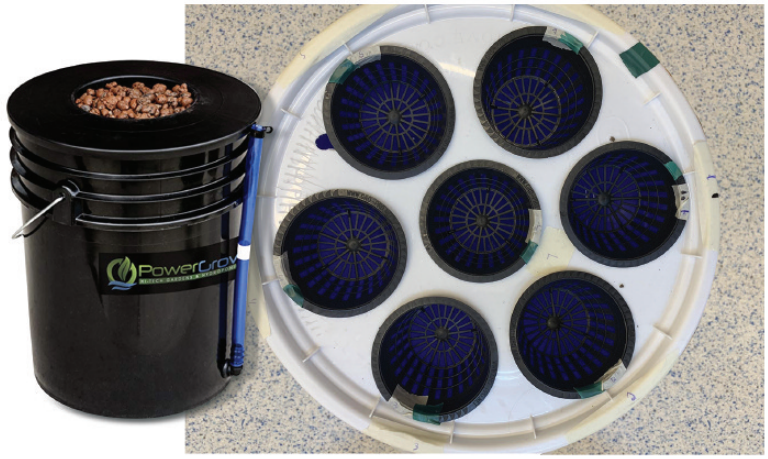
Setting Up Buckets

Both kits required different parts all provided by *Garden Pool,* the manufacturer for the aeroponics kits, as well as *step by step instructions* on how to assemble each piece (Model Number: 0fbd32471aa9). PowerGrow Systems, the manufacturer for the hydroponics kits, gave the instructions within the package (Model number: B00CHEIO6Y)

Figure D

This is a Picture of One of the Aeroponics Kits.

The cups were set inside the lid of the bucket. There were holes placed for the roots to go through and have the nutrient solution be sprayed onto them via a pump inside of the bucket. The difference between this bucket and the hydroponic bucket (left) is that the hydroponic buckets had clay beads that sat in a strainer and absorbed the nutrient solution. The solution would then be absorbed through the roots of the lettuce plants.)



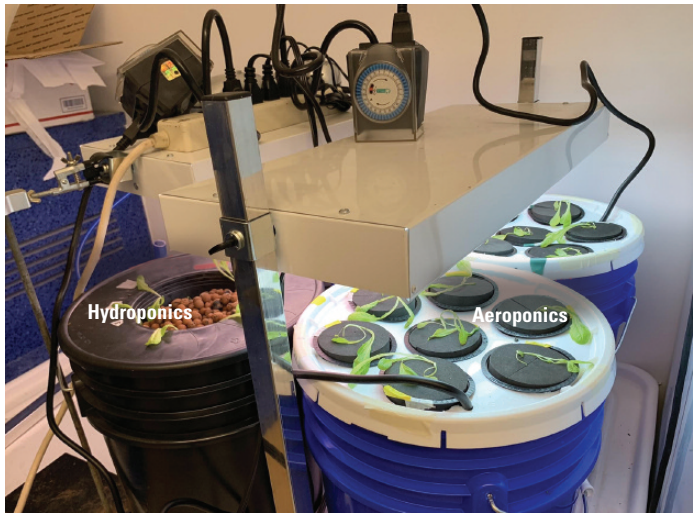
Transferring Plants

Plants were removed from their small growing containers and gently washed off with water to remove soil from roots. To prevent them from going completely dry as measurements were being taken, each plant was dried off evenly and placed on a damp towel. Measurements were then taken (see *Taking Initial/Final Measurements*). To place the plants in the hydroponics kits, grow beads were removed and the plants went in root first, with the beads placed back on top. For the aeroponics kits, the roots hung freely through a grow cup with a foam cover placed around the stem of the plant to prevent it from falling through.

Figure E

Experimental Design.

Hydroponics on the left and aeroponics on the right.



Making Solution

Reference chart was found online by the manufacturer (General Hydroponics® FloraPro™ Grow Soluble 9-11-19). The solution was made out of FloraPro 9-11-19, a soluble powder that contained all the nitrates and phosphates that were similar to soil. It was then calculated, via the reference chart, how much should be added to water to create an optimal nutrient solution. I modified the amount of water in my buckets since I had five gallons *(Figures F and G*). My solution was measured in grams per liter. Water was measured out using a 1000 mL graduated cylinder and the premade powder was massed out on a scale, measured in grams. Both the powder and water were then added to the bucket and mixed.

Figure F

Reference Sheet for how much Nutrient Powder should be Mixed.

General Hydroponics® FloraPro™ Grow Soluble 9-11-19.

Ounces of Dry Fertilizer needed **per Gallon** of Water to obtain desired Nitrogen Concentration. EC Value assumes using RO Water.

Desired N Feed Rate	Injector Setting				EC value mS/cm
	1:15	1:50	1:100	1:200	
50 ppm	1.2	3.9	8.0	16.8	0.6
100 ppm	2.4	8.0	16.9	3.85	1.1
150 ppm	3.7	12.4	27.1	***	1.7
200 ppm	4.9	17.1	38.7	***	2.2

Gallons of Water needed to dissolve one **25 lb. bag of fertilizer.**

Desired N Feed Rate	Injector Setting			
	1:15	1:50	1:100	1:200
50 ppm	333.4	102.5	50.3	23.8
100 ppm	165.2	49.8	23.7	10.4
150 ppm	109.1	32.2	14.8	***
200 ppm	81.1	23.4	10.3	***

*** *Exceeds maximum solubility of 3.5 lbs/gallon*

Figure G

Reference Sheet for how much Nutrient Powder should be Mixed.

General Hydroponics® FloraPro™ Grow Soluble 9-11-9.

Final Nutrient Solution Qualities	Grow (18h photoperiod)		Bloom (12h photoperiod)	
	Early Growth	Late Growth	Early Bloom	Mid Bloom
Total Nitrogen (ppm)	120	180	150	130
EC Range (mS/cm)	1.1-1.4	1.7-2.0	1.6-1.9	1.5-1.8
PPM Range (500 ppm scale)	500-700	800-1000	750-950	700-900

Data Analysis

Overview

For this trial of my experiment, I set up four kits (two hydroponic and two aeroponic). Plants were divided into four groups (one for each kit) based off of similar height so there would be a consistent starting point for all buckets. Data was taken for three weeks and all plant samples were exposed to the same type/amount of light (Sylvania 20W fluorescent tube light). Additionally, the solution for each kit was the same (FloraPro 9-11-19). After three weeks, final measurements were taken and two tailed independent t-tests were run for each graph. Out of the graphs below, the only one to have a significant difference was the change in mass *(Figure 1),* demonstrating a p-value of .000112. The rest of the graphs had no significant difference (*Figure 2* = 0.450003; *Figure 3* = 0.639117; *Figure 4* = 0.251112). The error bars on each of the graphs were in place to represent the standard error of mean (average ± SEM). Samples size for the experiment was n = 28.

Figure 1

The Average Difference in Mass (grams) of Lettuce Plants for Hydroponics vs. Aeroponics.

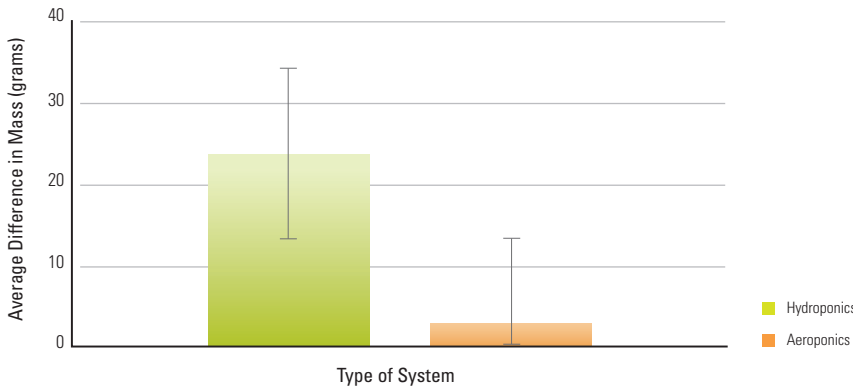


Figure 2

The Average Difference in Stem Length (cm) for Hydroponics vs. Aeroponics.

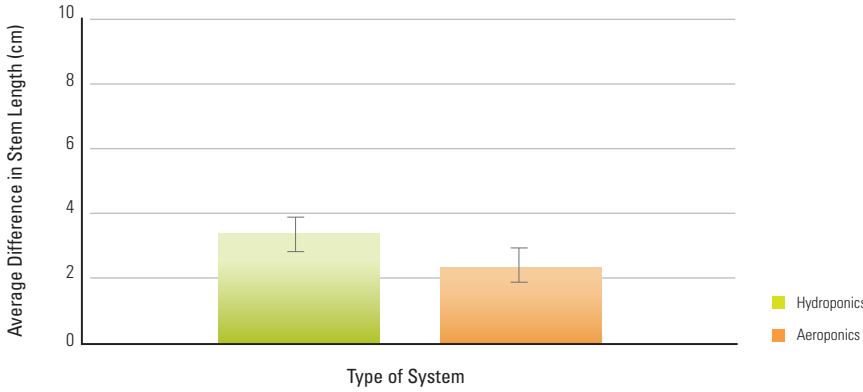


Figure 3

The Average Difference in Root Length (cm) between Hydroponics and Aeroponics.

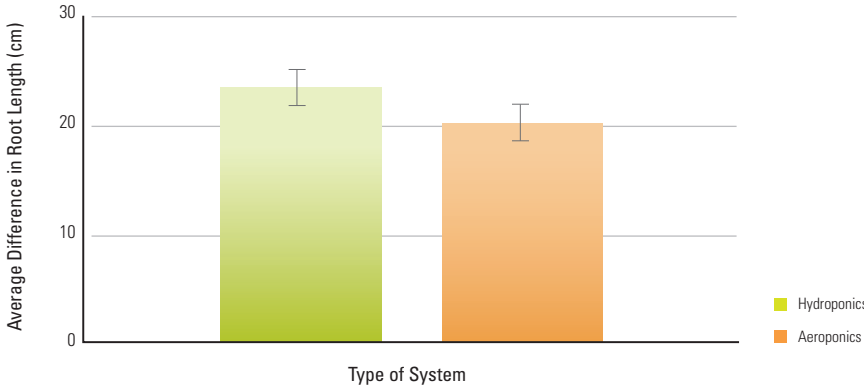
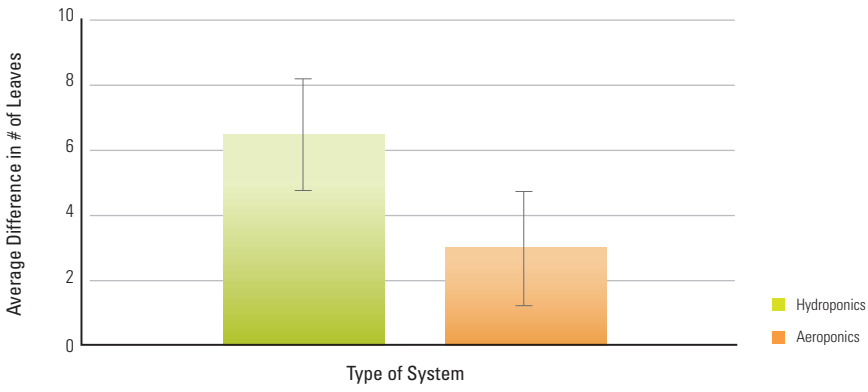


Figure 4

The Average Difference in Number of Leaves between Hydroponics and Aeroponics.



Conclusion

From the data above, plants that were grown hydroponically displayed a greater change in growth, due to their masses changing a larger amount than plants that were grown aeroponically. Going forth, because this was a replication study, I now have the last piece of the puzzle to continue with my second trial in Spring 2020 where I plan to test the growth of Romaine lettuce plants while varying the concentration of nitrates and phosphates.

There were some errors/limitations in the data collection that involved various factors. For example, some of the plants died early in the study when I was measuring and transitioning them from the soil to their kits. This caused there to be fewer subjects in each bucket and less data taken. Another error came from inconsistent data collection because of holidays and weekends. Additionally, there is some uncertainty in the data because when I measured plants roots and stem length, it wasn't always exact to the marking on the ruler. This is human error, but it is also hard to be exact due to the hash markings on the plants getting washed away over time, which caused the measurements to either be over/under exaggerated depending on the situation. Finally, another form of error that skewed the data was the plants tearing/losing leaves when measurements were being taken. The plants were very delicate and handling them took a good amount of precision to avoidt damaging them. This caused leaves to break off, ultimately resulting in data points lost.

Going forth, now that I have this information, I plan on running another trial in the spring using hydroponics, but testing different nitrate and phosphate levels in different solutions. This is essentially a shift in focus from the different types of systems that would affect growth to the different solutions that would affect growth. My main goal of this next experiment is to now see how varying the nutrient type will affect the growth rate of Romaine lettuce, which will allow me to determine whether or not hydroponics is really a viable growing medium, as it will be compared against samples being grown in soil.

References

Carver, J., & Wasserman, B. (2012). Hands-on hydroponics. *The Science Teacher, 79*(4), 44-48. Retrieved from https://search.proquest.com/docview/1017897397?accountid=35114

Gashgari, R., Alharbi, K., Mughrbil, K., Jan, A., & Glolam, A. (2018). Comparison between growing plants in hydroponic system and soil based system. *Proceedings of the 4th World Congress on Mechanical, Chemical, and Material Engineering.* doi:j10.11159/icmie18.131

General Hydroponics® FloraPro™ Grow Soluble 9-11-19: Hawthorne Gardening Company. (n.d.). Retrieved February 28, 2020, from www.hawthornegc.com/shop/product/general-hydroponics-florapro-grow-soluble-9-11-19

Lakkireddy, K. K. R., Kasturi, K., & Sambasiva, R. K. R. S. (2012). Role of hydroponics and aeroponics in soilless culture in commercial food production. *Journal of Agricultural Science & Technology, 1*(1), 26-35.

M. D. (2013). A review on plants without soil – Hydroponics. *International Journal of Research in Engineering and Technology, 02*(03), 299-304. doi:10.15623/ijret.2013.0203013 http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.514.4323&rep=rep1&type=pdf

Martin-Laurent, F., Lee, S.-K., Tham, F.-Y., Jie, H., & Diem, H. G. (1999). Aeroponic jproduction of acacia mangium saplings inoculated with AM fungi for reforestation in the jtropics. *Forest Ecology and Management, 122*(3), 199-207. doi: 10.1016/s0378-1127(99)00006-7

Raviv, M., Lieth, J. H., & Raviv, M. (2008). Significance of soilless culture in agriculture. *Soilless culture theory and practice,* 1-11.

Sengupta , Amrita, and Hirak Banerjee. Soil-Less culture in modern agriculture. *World Journal of Science and Technology,* vol. 2, no. 7, 2012, pp. 103-108., s3.amazonaws.com/academia.edu.documents/34706897/as.pdf?

Sheikh, B. A. (2006). Hydroponics: key to sustain agriculture in water stressed and urban environment. *Pak. J. Agric., Agril. Eng., Vet. Sci, 22*(2), 53-57.

Treftz, C. and Omaye, S. (2016), Hydroponics: potential for augmenting sustainable food production in non-arable regions, *Nutrition & Food Science,* Vol. 46 No. 5, pp. 672-684. https://doi.org/10.1108/NFS-10-2015-0118

Yoshitani, Junichi, Norimichi Takemoto and Tarek Merabtene. 2007. *Factor analysis of water related disasters in angladesh.* Tsukuba: International Centre for Water Hazard and Risk Management (ICWHRM) and Public Works Research Institute (PWRI).

The Effect of Musical Training on Verbal Memory in Seventh and Eleventh Graders at Berkeley Carroll

by Zoe Rabinowitz

Abstract

This experiment explored the effect of musical training on verbal memory in seventh and eleventh graders at Berkeley Carroll. 37 eleventh grade and 54 seventh grade musicians and nonmusicians received the Rey Auditory Verbal Learning Test to assess their verbal memory skills. Participants completed tasks that involved recalling a word list immediatly after hearing it, recalling it after hearing an interfering list, recalling the list after a delay period, and recognizing the words from a larger list. No significant difference was found between the scores of the musicians and nonmusicians in the seventh or eleventh graders. Therefore, it was concluded based on the data that musical training does not affect verbal memory in both seventh and eleventh grade students at Berkeley Carroll.

Background

Brain plasticity is defined as the "brain's ability to change and adapt as a result of experience" (Kolb & Whishaw, 1998). Multiple studies conclude that the effects of musical training on cognitive function are evidence that musical training helps induce brain plasticity. These studies have suggested that there are numerous non-musical benefits to long-term musical training. Some of these benefits lie in phonological awareness, the ability to recognize sounds in language (Anvari, Trainor & Woodside, 2002), early reading ability (Lamb & Gregory, 1993), and multiple subtests of the general IQ test (Schellenberg, 2004). However, when examining if brain plasticity can be induced through musical training, perhaps it is most useful to look at verbal memory as an adaptation, for there is a vast amount of research in this area. (Franklin et al., 2008; Costa-Giomi, 1990; Helmbold et al., 2005; Ho et al., 2003).

Neuroimaging research has indicated a cognitive link between verbal memory skills and musical training, with studies finding that musical training affects brain function in the left-hemisphere. For example, Ohnishi et al. (2001) studied both nonmusicians and musicians under functional Magnetic Resonance imaging (fMRI) while participating in a passive musical listening task. The researchers found differences in the amount to which the planum temporale (PT) and the left dorso-lateral prefrontal cortex (DLPFC) were activated, with musicians showing greater activation. This activation increased in

correlation with the amount of training each musician received. This is significant due to the fact that the left PT, also referred to as Wernicke's Area, is correlated with language comprehension (Poldrack & Gabrieli, 1998). Schlaug, Jancke, Huang, Y. and Steinmetz (1995) found more asymmetry of the PT of musicians than nonmusicians, with it being more lateralized to the left hemisphere, and noted that this asymmetry has been accepted as a cause of left-hemisphere dominance in auditory language comprehension tasks, such as verbal memory tasks. Verbal memory refers to memory for presented language terms, and involves recalling these terms either immediately after presentation or several minutes later. Both these studies have suggested that musicians retain better verbal memory skills due to brain plasticity.

Researchers have also investigated the relationship between musical training and verbal memory. Chan, Ho, and Cheung (1998) examined the verbal memory skills of female college students, half of them having at least six years of musical training and the other half having none. They found that after being presented with a word list three times, the musicians remembered significantly more words than nonmusicians. This study provided evidence that long-term musical training affects verbal memory skills, but did not speak to the effects of short term training. Ho et al. (2003) complicated these results, examining children who received 0-5 years of training, and found a correlation between the number of years a student has studied music and their verbal memory skills. Although I have found no neuroimaging research on the brains of adolescent musicians, I believe that that the structural changes in these musicians may still exist, given that adolescence is the age in which brain plasticity is most likely to occur (Pauwels, Chalavi, & Swinnen, 2018). I am in a situation where I have access to groups of students experiencing musical training all in the same setting. I am also able to examine students of different ages, which I have done by using both seventh grade and eleventh grade musicians and nonmusicians and subjects. The goal of this study is to investigate the effects of musical training on different aspects of verbal memory, and to examine if this effect correlates to or is affected by years of musical training. I am hoping that I can add to the overall understanding of this topic in science research and medical research, to see if musical training could aid people with language difficulties.

I chose to use the Rey Auditory Verbal Learning Test (RAVLT) which has been shown to be a reliable method of collecting verbal memory data in other studies (Rosenberg et al., 1984; Franklin et al, 2008; Rezvanfard et al., 2011; Khosravi et al., 2016) The RAVLT is a notable measure of a person's capacity to encode, store, and recover verbal information. It is able to evaluate immediate recall, the effect of an interfering verbal stimulus, delayed recall skills, and recognition. Compared to similar studies that use other tests to operationalize verbal memory, this test allows me to acquire data on multiple aspects of verbal memory.

Methods

Subjects

Seventh grade participants were recruited through an email sent to faculty study hall advisors, and the study was completed during their study hall period on different days. Eleventh grade participants completed the study during their advising period on a single day. Ultimately, 37 eleventh graders and 54 seventh graders participated in this study. While participation in the study was optional, more eleventh graders chose not to participate, leading to a lesser number of eleventh grade than seventh grade subjects. Seventh grade musicians participated in at least a full year of musical training in the Berkeley Carroll music program, and at most two years. Eleventh grade musicians had received at least four years

of musical training at Berkley Carroll, and at the most eight. In order to keep the study blind, participants were not asked directly if they were in the instrumental music program at Berkeley Carroll, and this information was collected by the music directors.

Rey Auditory Verbal Memory Test (RAVLT)

Participants were given a packet of paper consisting of blank sheets labeled number 1-5 and told to write their initials on the front cover. Before the experiment began, participants were instructed to completely fold over each page after completing a stage of the test, ensuring that they could not see any previous words they recalled while completing the subsequent page. They also were asked to only use pencils, preventing them from seeing through the pages in the packet. Following the RAVLT, participants were read a 15 noun-word list (List A) and given 45 seconds to record these words on page "1" and then fold the page. They were told that the order of the words were not of importance. Participants repeated these steps on pages "2" and "3." These three trials evaluated their immediate recall skills. Participants were then read another 15-noun word list (List B), which consisted of words semantically different from List A, and were given 45 seconds to record these words on page "4." Then, participants were instructed to record the words from List A on page "5," measuring their immediate recall skills after hearing an interfering list of words. After a 15 minute delay period, participants were given a sheet of paper, one side having a numbered list of 1-30, and the other being blank. They were instructed to write their initials on the blank side, and then given 45 seconds to write down the words they remembered of list A, measuring their delayed recall skills. On the other side of the sheet, students were read a 30-noun word list, which consisted of the words from List A, as well as 15 other semantically similar words. After hearing each word, participants were instructed to write down "Y" (yes) meaning they recognized the word from List A, or "N" (no) meaning they didn't. They were given two seconds in between each word to do this, and this aspect of the test measured their recognition abilities.

Experimental Procedure

Seventh grade participants and eleventh grade participants were administered the RAVLT in different ways. Seventh graders were tested on different days at 2:45 pm. I went into their study hall to administer the test. The eleventh grade participants were all tested on the same day. The instructions were read to them by their advisors and the words were read to them by a recording. This recording was used to ensure that all the subjects were hearing the words at the same tone and pace. While it would've been ideal for the seventh grade participants to also have been tested on the same day using this recording, the Middle School dean requested that the responsibility to administer this test not be put on the study hall advisors. During the delay period, both eleventh graders and seventh graders were asked to complete any school work as they normally would. This instruction was given solely to prevent subjects from doing drastically different activities during that time.

Results

Seventh Grade Immediate Word Recall and Delayed Recall Scores

Two-tailed t-tests were performed using a statistical significance of $p < 0.05$ to calculate the p-value for both the sum of the three immediate recall scores and the delayed recall scores. There was no significant difference in the immediate recall scores between the musicians and nonmusicians (Figure 1a), with a p-value of 0.24. Similarly, for the delayed recall scores (Figure 2a), no significant difference was found between the scores of musicians and nonmusicians, with a p-value of 0.10. These results indicate that musical training does not improve verbal memory in seventh graders at Berkeley Carroll. This study specifically focused on the data from these two sections because in other studies exploring this topic, they also only focused on these aspects of verbal memory (Lamb & Gregory, 1993; Chan et al., 1998; Anvari et al., 2002; Ho et al., 2003). The data from the post-interference and recognition section of the RAVLT was still evaluated, and for both data sets no significant difference was found in the scores between the musicians and nonmusicians.

Eleventh Grade Immediate Recall and Delayed Recall Scores

To calculate the p-value for both the sum of the three immediate recall scores and the delayed recall scores, two-tailed t-tests were performed using a statistical significance of $p < 0.05$. For the immediate recall scores *(Figure 1b),* no significant difference was found between the sum of the musician and nonmusican scores, with a p-value of 0.98. Similarly, for the delayed recall scores *(Figure 2b),* no significance was found between the scores of musicians and non musicians, with a p-value of 0.38. These results indicate that musical training does not improve verbal memory in eleventh graders at Berkeley Carroll. It should be noted while drawing conclusions from this data that a low number of eleventh grade musicians was used (n=10), which might have produced less reliable results.

Figure 1a

Immediate Recall Section of the RAVLT for Seventh Graders.

Shows no significant difference between the mean sum of the three recall trials for seventh grade musicians and nonmusicians (p=0.24).

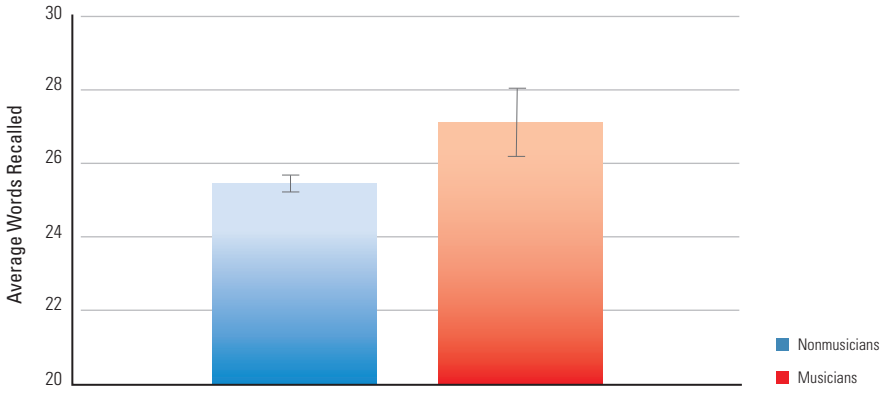


Figure 2a

Delayed Recall Section of the RAVLT for Seventh Graders.

Shows no significant difference between the mean delayed recall scores of the seventh grade musicians and no musicians (p = 0.10).

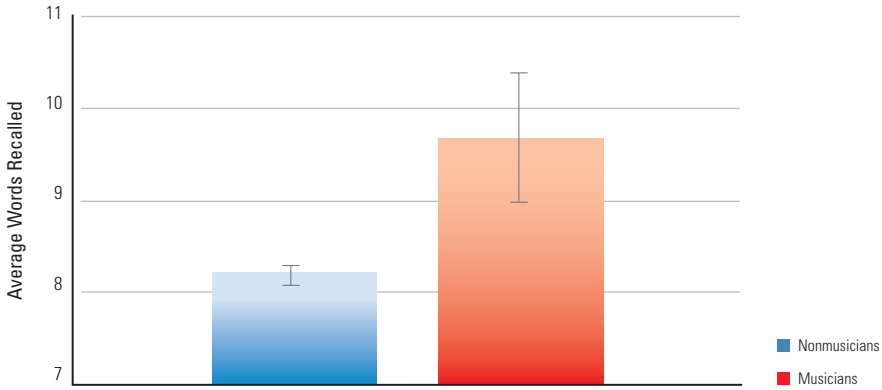


Figure 1b

Immediate Recall Section of the RAVLT for Eleventh Graders.

Shows no significant difference between the mean sum of the three recall trials for eleventh grade musicians and nonmusicians (p=0.98).

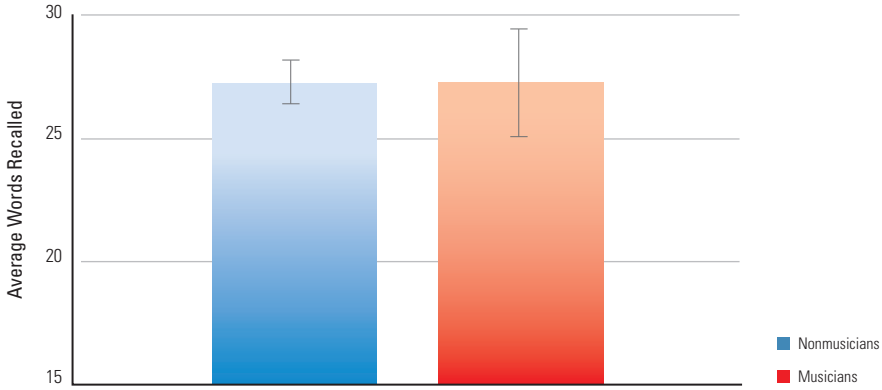
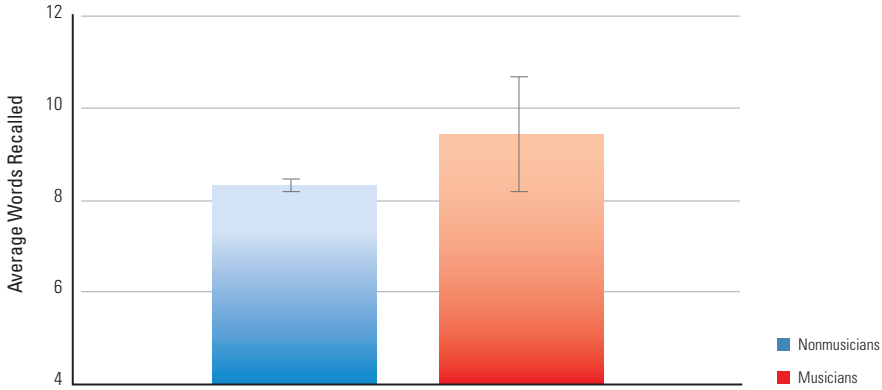


Figure 2b

Delayed Recall Section of the RAVLT for Eleventh Graders.

Shows no significant difference between the mean delayed recall scores of the eleventh grade musicians and nonmusicians (p = 0.38).



Discussion

Due to the lack of significance when comparing the results between the musicians and nonmusicians in both the seventh and eleventh grade, the data supports the conclusion that musical training does not improve verbal memory for both grades of students. No significant difference was found between the results of the musicians in the seventh and eleventh grade. There are multiple explanations for the conclusions drawn from this data. One explanation is that the benefits from long-term musical training in verbal memory would only appear in adulthood, when the brain is more fully developed. This idea comes from the study of Schlaug et al. (1995), in which the physiological brain structures that appeared to result from musical training were observed in the brain of adults. No studies have been done to observe these brain structures in adolescents, so it is possible these brain structures have not finished forming during adolescence, and only exist in adults. Another explanation lies in the limitations of the experimental procedure. In both the eleventh and seventh grade, many students chose to opt out, leading to a smaller number of subjects. Having a higher number of subjects for this experiment is ideal because it creates more data and therefore more reliable conclusions. In addition, having more participants would help to control for other factors that could affect verbal memory. There is also a limitation in the type of musical training that is provided by Berkeley Carroll. Although some students in both the seventh and eleventh grade music programs already have been given introductions to music theory and musical skills prior to the program, this cannot be assumed. The program itself mostly involves practicing pieces as a group. Studies that have completed similar research, also using young participants enrolled in a music program, have used programs that are introducing students to instruments and musical theory (Gardiner et al., 1996; Rauscher & Zupan., 2000; Ho et al., 2003; Rickard et al., 2010; Roden et al. 2012). For example, Roden et al. (2012) introduced students to instrumental training in first grade and noted the importance of teaching them skills relating to rhythm and pitch identification. Skills involving rhythm and pitch learning seem to be the most related to verbal memory, and therefore the length and level into which

participants practiced these skills would most likely affect the results. This experiment did not control for these skills, which could explain why no significance was found in the verbal memory ability of musicians at Berkely Carroll.

References

Anvari, S., L.J. Trainor, J. Woodside, et al. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of Experimental Child Psychology. 83,* 111-130.

Chan A. S., Ho Y. C., Cheung M. C. (1998). Music training improves verbal memory. *Nature 396:* 128.

Gardiner M. F., Fox A., Knowles F., Jeffrey D. (1996). Learning improved by arts training. *Nature 381,* 284-284.

Ho, Y.-C., Cheung, M.-C. and Chan, A.S. (2003). Music training improves verbal but not visual memory: cross-sectional and longitudinal explorations in children', *Neuropsychology 17*(3): 439-50.

Khosravi Fard, E., L Keelor, J., Akbarzadeh Bagheban, A., & W Keith, R. (2016). Comparison of the Rey Auditory Verbal Learning Test (RAVLT) and Digit Test among typically achieving and gifted students. *Iranian Journal of Child Neurology, 10*(2), 26-37.

Kolb B, Whishaw IQ. (1998) Brain plasticity and behavior. *Annual Review of Psychology; 49:* 43-64.

Lamb, S. J. & A.H. Gregory (1993) The relationship between music and reading in beginning readers. *Educational Psychology 13:* 19-27.

Ohnishi, T., Matsuda, H., Asada, T., Aruga, M., Hirakata, M., Nishikawa, M., Katoh, A. and Imabayashi, E. (2001) Functional anatomy of musical perception in musicians. *Cerebral Cortex, 11,* 754-60.

Pauwels, L., Chalavi, S., & Swinnen, S. P. (2018). Aging and brain plasticity. *Aging, 10*(8), 1789-1790.

Poldrack RA, Gabrieli JD. (1998). Memory and the brain: what's right and what's left? *Cell Press 93:* 1091-1093.

Rauscher F. H., Zupan M. A. (2000). Classroom keyboard instruction improves kindergarten children's spatial-temporal performance: a field experiment. *Early Child Research Quarterly.* 15, 215-228

Rezvanfard M., Ekhtiari, H, Noroozian M, Rezvanifar A, Nilipour R, & Javan GK. (2011). The Rey Auditory Verbal Learning Test: alternate forms equivalency and reliability for the Iranian adult population (Persian version). *Archives of Iranian Medicine,14*(2), 104-109.

Rickard, N. S., Vasquez, J. T., Murphy, F., Gill, A., and Toukhsati, S. R. (2010). Benefits of a classroom based music program on verbal memory of primary school children: a longitudinal study. *The Australian Journal of Music Education.* 1, 36-47.

Roden, I., Kreutz, G., Bongard, S. (2012). Effects of a school-based instrumental music program on verbal and visual memory in primary school children: a longitudinal study. *Frontiers in Psychology.* 3:572.

Rosenberg SJ, Ryan JJ, & Prifitera A. (1984). Rey auditory-verbal learning test performance of patients with and without memory impairment. *Journal of Clinical Psychology, 40*(3):785-7.

Schellenberg, E.G (2004)Music lessons enhance IQ. *Psychological Science. 15:* 511-514.

Schlaug, G., Jancke, L., Huang, Y. and Steinmetz, H. (1995) In vivo evidence of structural brain asymmetry in musicians, *Science. 267:* 699-701.

Accuracy of Memory Recall and Formation of False Memories in Eyewitness Testimonies

by Bella Solimine

Abstract

My experiment is intended to test whether there is a difference in accuracy of memory recall when accuracy of co-witness testimonies and language in follow-up questions is varied. Subjects in the preliminary iteration included 143 9th and 10th grade Berkeley Carroll students. Subjects were asked to watch a video, answer a set of comprehension questions, read either an accurate or inaccurate testimony, then answer a set of follow-up questions containing either correct or misleading language. A follow-up iteration on the study was run on 34 9th grade subjects using the same procedure, yet using a different video and only inaccurate testimonies. Data from the preliminary iteration showed statistically significant results for all four trial groups. Data from the follow-up study was not statistically significant. Further studies with a more sensitive measurement tool must be run in the future to draw concrete conclusions regarding the experimental question.

Background

In the legal sphere, false memories often hinder the accuracy of eyewitness testimonies. It is extremely common for those who have witnessed a crime to recall false details when asked to testify in a courtroom setting or elsewhere (Shaw & Garven, 1997). Finding the causative factors that play into the formation of false memories would be extremely valuable in improving the accuracy of eyewitness testimonies.

In many instances the human memory can be manipulated to reflect events and facts that are not true to reality (Frenda, Nichols, & Loftus, 2011). In an experimental setting, this manipulation of the human memory is titled the "misinformation effect" (Frenda, Nichols, & Loftus, 2011). The misinformation effect has been thoroughly discussed in past research (Godfrey-Smith, 1989; Kasprzyk, Montano, & Loftus, 1975), but one study (Frenda, Nichols, Loftus, 2011) defines the concept very clearly and explains its application to the current study. In a typical misinformation effect experimental design, "research subjects are shown materials (e.g. photographs) and are then exposed to deliberately misleading information about what they saw."

In a previous study (Shaw & Garven, 1997) researchers tested whether co-witness testimony affected the accuracy of eyewitness memory reports. Subjects would witness an event and then hear a co-witness recount the details of that event. One group of subjects heard a co-witness testimony with accurate information, while another group heard a testimony with subtle changes in the details of the event, intended to provoke retrieval induced forgetting. The results of this study showed that when co-witnesses gave inaccurate testimonies, eyewitness memory reports were similarly inaccurate. Another study (Kasprzyk, Montano, & Loftus, 1975) showed that memory can be manipulated through subtle

changes in language using an experimental procedure modeling a court scenario. One of the most notable ideas discussed in this paper is that "human memory involves an active constructive process, not a passive recording one." This statement explains the influence of biased leading questions on the memory of the subjects. The use of inaccurate questions manipulates the brain into actively constructing a false memory, pieced together using the false information hinted at through biased questioning.

Researchers are still unclear as to what underlying mechanism causes the misinformation effect phenomenon, but one possible explanation is the relationship between "retrieval-induced forgetting and the reporting of misinformation" (MacLeod & Saunders, 2008). Retrieval-induced forgetting describes the basic phenomenon of forgetting previously known information when asked to recall the information after retention. The reporting of misinformation refers to a subject recalling a false memory and believing it is correct, and therefore reporting that misinformation related to the formation of a false memory. It is proposed that the underlying cause for the report of misinformation is, in fact, the remembrance of other information. Researchers state that it is possible that the retrieval of a memory outside the target memory may hinder the retrieval of the memory people are seeking in a certain situation (MacLeod & Saunders, 2008) and cause retrieval-induced forgetting.

In their psychological study Chan and Thomas (2009) tested the effects of having subjects recount an event before the introduction of misinformation, and then again after misinformation was introduced. This phenomenon was named the "reversed testing effect". This experimental design was intended to reflect real life circumstances, such as making a phone call to 911 after witnessing a crime, and then recounting the situation again in court. The researchers note that "the reversed testing effect we observed was based on two mechanisms: First, immediate cued recall enhanced learning of the misinformation; second, the initially recalled details became particularly susceptible to interference from later misinformation..." These conclusions are particularly notable because they have a clear real-life application. The information drawn from this study may explain why humans are prone to forgetting information and replacing it with new false memories, even after previously recalling the information correctly.

My current study is intended to find the effects of inaccurate co-witness testimonies in conjunction with varied language in follow-up questions on the accuracy of memory recall, a design which has not been researched previously. Although both inaccurate testimonies and misleading questions have both been used independently in previous research to test the formation of false memories, to my knowledge they have never both been used simultaneously in an experimental setting. In my study, subjects are shown a short video which has been created for use in similar misinformation effect studies, and are then asked to answer a series of nine comprehension questions. Participants are then exposed to supplemental information, either misleading or accurate, in the form of a written co-witness testimony.

Subjects

Participants in the preliminary study included 143 9th and 10th grade Berkeley Carroll students, ages 14 through 16. 9th and 10th graders were used as subjects because of availability. Four distinct trial groups were created. The study was conducted in advising groups (small groups of students all from one grade level led by one or two teachers) containing around 12 students per group. Advising groups were randomly sorted into trial groups. The trial groups were each exposed to a different combination of testimony and questions. The trial groups included accurate testimony/accurate questions, inaccurate

testimony/inaccurate questions, inaccurate testimony/accurate questions and accurate testimony/inaccurate questions. Each trial group consisted of two or three advising groups in both the 9th and 10th grade. Informed consent was obtained from the subjects through the administration of a detailed informed consent sheet by the advisors. The specific video used in this study was a short (1 minute) clip from the television show *Friends.* This clip was chosen because the plot is brief and easy to follow, yet it contains a substantial number of details that can be manipulated in the subsequent testimonies and questions.

Subjects in the follow-up study included 34 9th grade Berkeley Carroll students. Subjects were once again chosen based on availability. The study was again conducted in advising groups, following the same procedure as the preliminary study. The participants in the follow-up study either received a combination of inaccurate testimony/accurate questions or inaccurate testimony/inaccurate questions. A new video was used in the follow-up study. The video used is brief (1 minute) and is intended to mimic security camera footage of a robbery. This video was chosen because it had been successfully used in previous misinformation effect studies and was created for this purpose. The video contained a basic plotline and many details that could be manipulated in the subsequent materials.

Procedure

9th and 10th grade advisors received a folder containing comprehension questions and follow-up questions as well as informed consent sheets. Comprehension questions and follow-up questions were printed on different color paper to prevent confusion. An email was sent to advisors containing instructions on the experimental procedure and a script to follow while conducting the study. The script and instructions served as a control for how the advisors conducted the study and the time allotted for each step in the procedure. The study began with the administration of informed consent sheets to the subjects, allowing for subjects to refuse to participate in the experiment. After this, subjects were asked to remain silent during the experiment and refrain from talking to one another. The video was then projected and the subjects were instructed to pay attention to detail during the duration of the video. After watching the video, the set of nine comprehension questions was administered. Subjects were given four minutes to answer these questions. Comprehension questions were collected after four minutes. After this, the designated testimony (inaccurate or accurate based on trial group), which was sent to advisors in the form of an email link, was projected. Subjects were given four minutes to read the testimony. After four minutes, the projector was turned off and the set of nine follow-up questions (inaccurate or accurate based on trial group) was administered. Misleading questions were created by incorporating inaccurate details in the questions. For example, if a character was wearing a red shirt a misleading question would state "Was the character's shirt green or blue?" Accurate follow-up questions were identical to the set of comprehension questions previously administered to all subjects. Subjects were given five minutes to answer these questions. Follow-up questions were collected after five minutes and all physical materials were placed in a folder marked with the name of the advisors and returned to the experimenter. Answers to follow-up questions were recorded and each subject's set of followup questions were given a score out of eight. Average number of correct answers for each trial group was calculated and recorded. Answers to comprehension questions were disregarded.

Data/Analysis

Figure 1

Results of Preliminary Study run on 143 9th and 10th Grade Students.

One-way independent ANOVA (p<.0001) and a follow-up Tukey HSD Test (p<.01 for all 4 samples) were run.

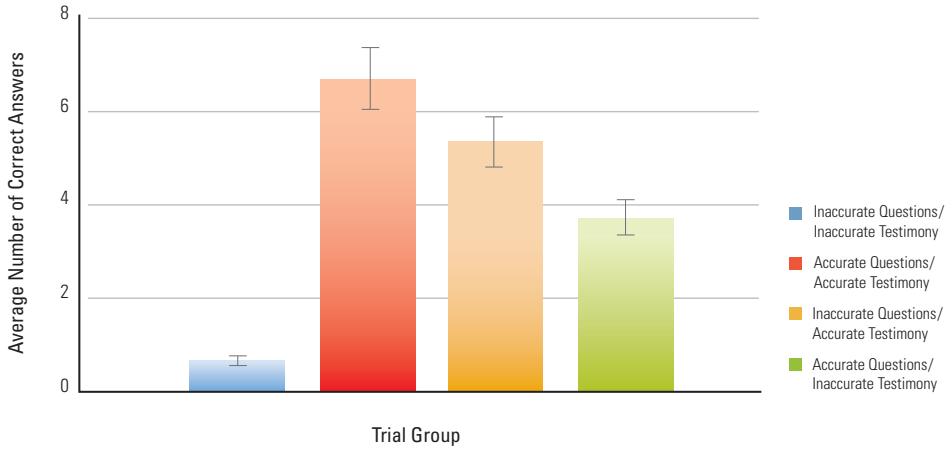
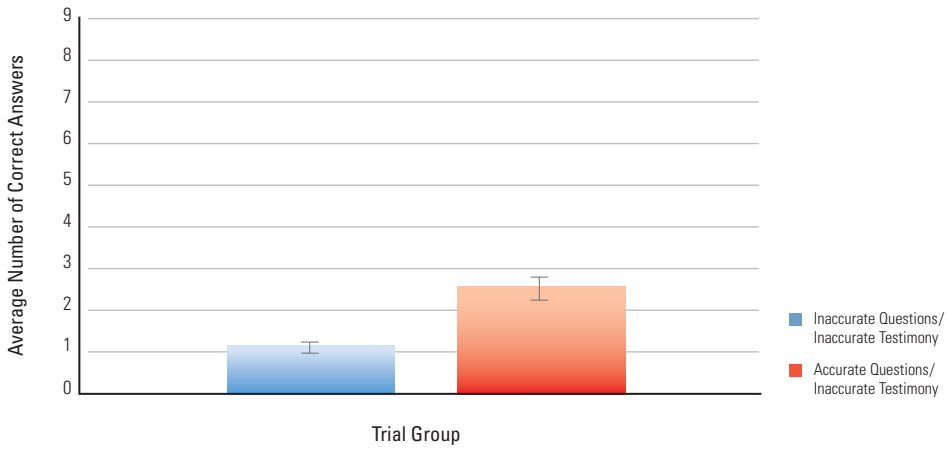


Figure 2

Results of Follow-up Study using 34 9th Grade Subjects.

A t-test for independent samples was run on the data and found no statistical significance (p=0.136731).



The results of the preliminary study conducted on 143 9th and 10th graders *(Figure 1)* are conclusive in many ways. The number of correct answers on the follow-up question set from each trial group were used to run an independent one-way ANOVA. The ANOVA summary treatment indicated that the data was significant ($p < 0.0001$). The follow-up Tukey HSD test indicated that the data was significant in comparing the means of all 4 different trial groups. Thus, the average number of correct answers of the accurate testimony/accurate questions group was significantly higher than all the other trial groups and the average of the accurate testimony/inaccurate questions group and the average of the inaccurate testimony/accurate questions group were both significantly higher than the average of the inaccurate testimony/inaccurate questions group. This data also shows that the average number of correct answers for the inaccurate questions/accurate testimony group was significantly higher than that of the accurate questions/inaccurate testimony group. This demonstrates that inaccuracies in testimonies are more effective in the formation of false memories than inaccuracies in questions. The data and statistical analysis from this preliminary study suggests that the alternative hypothesis should be accepted. Thus, this data suggests that there is a difference in accuracy of memory recall during eyewitness testimonies both when language in follow-up questions and when accuracy of co-witness testimonies are varied.

The results of the follow-up study testing the impact of varied accuracy of questions paired with inaccurate testimony *(Figure 2)* did not show statistically significant results. Because of this, there are no significant conclusions that can be drawn from this data.

Discussion

There are a few notable limitations to both the preliminary and follow-up studies previously discussed. The video used in the preliminary study had not been used in previous misinformation studies and was very simple. Some subjects may have realized the intentions of the study after reading false information in the supplemental materials, and this may have impacted the data. The study was also conducted by advisors, and although they were provided with clear instructions, this creates opportunities for mistakes in the execution of the study. Some advisors may have been more strict than others regarding the rules of the study (no talking during the study, no sharing answers, etc.) which may have also led to biases in the data. For example, subjects may have asked each other for answers, leading to multiple subjects reporting the same misinformation. This also applies to the follow-up study using 34 9th grade subjects, which was also conducted in advising groups. This study used a much more appropriate video, yet because of scheduling issues only 34 subjects were able to participate in this iteration of the study. Because of this, the data collected was only from inaccurate questions/inaccurate testimony and accurate questions/inaccurate testimony trial groups. Both of these groups showed very low scores for average number of correct answers, meaning most subjects recalled false details (from the inaccurate testimony/questions) instead of accurate details from the video. Further studies will need to be done in the future using this same apparatus from the follow-up iteration to see conclusive results. In the future, I will conduct an experiment using the same procedure used in both the preliminary and follow-up studies. I will use the video used in the follow-up study to ensure more accurate data. Subjects will be divided into the same four trial groups used in the preliminary study.

References

Bregman, N., & McAllister, H. (1982). Eyewitness testimony: The role of commitment in increasing reliability. *Social Psychology Quarterly, 45*(3), 181-184. Retrieved from www.jstor.org/stable/3033652

Chan, J., Thomas, A., & Bulevich, J. (2009). Recalling a Witnessed Event Increases Eyewitness Suggestibility: The Reversed Testing Effect. *Psychological Science, 20*(1), 66-73. Retrieved from www.jstor.org/stable/40065031

Frenda, S., Nichols, R., & Loftus, E. (2011). Current Issues and Advances in Misinformation Research. *Current Directions in Psychological Science, 20*(1), 20-23. Retrieved from www.jstor.org/stable/23045709

Godfrey-Smith, P. (1989). Misinformation. *Canadian Journal of Philosophy, 19*(4), 533-550. Retrieved from www.jstor.org/stable/40231668

Kasprzyk, D., Montano, D., & Loftus, E. (1975). Effect of leading questions on juror's verdicts. *Jurimetrics Journal, 16*(1), 48-51. Retrieved from www.jstor.org/stable/29761515

MacLeod, M., & Saunders, J. (2008). Retrieval Inhibition and Memory Distortion: Negative Consequences of an Adaptive Process. *Current Directions in Psychological Science, 17*(1), 26-30. Retrieved from www.jstor.org/stable/20183243

Shaw, J., Garven, S., & Wood, J. (1997). Co-Witness Information Can Have Immediate Effects on Eyewitness Memory Reports. *Law and Human Behavior, 21*(5), 503-523. Retrieved from www.jstor.org/stable/1394330

Emotional Response to Fragrance Smell and Marketing—Using Heart Rate and Self-Reported Questions to Measure Gender Differences in Physiological and Psychological Impact

by Sophia Sondey

Abstract

The presented experiment assessed whether there are differences in the way males and females emotionally respond to the scent and marketing information (name, bottle, video advertisement) of perfume and cologne. 23 students at the Berkeley Carroll School in grades 10 and 11 were recruited as subjects. All experiments consisted of two rest periods and two fragrance exposure periods. The order of perfume and cologne periods was alternated between subjects. In order to quantify emotional response, the Root Mean Square of Successive Differences (RMSSD) between heartbeats and average heart rate were measured and analyzed, in addition to self-reported responses from an emotion questionnaire. Three experimental trials assessed the effectiveness of the heart rate monitor and questionnaire as measurement tools; data from all three trials were pooled together for final analysis. Results showed no gender-based differences in physiological response to the fragrances, but also showed gender-based difference in the self-reported responses. This data suggests that the gender-divide in the fragrance industry is primarily a marketing strategy based on gender stereotypes, rather than a strategy based on scientifically proven differences in the way men and women respond to fragrance.

Background

Emotion: Sense of Smell, Marketing Strategies, and Gender

Emotions are not something we can physically see, and as a result, it is difficult to grasp exactly why we have them or where in our bodies they come from. Emotions are reactions associated with activity in the Autonomic Nervous System (ANS)—the branch of the nervous system that controls bodily functions that happen subconsciously—particularly with heart rate, finger temperature, and muscle tension (Appelhans, 2006; Ekman, 1983). According to the Polyvagal Theory, emotional reactions additionally relate to the branches of the vagal nerve that control the heart, lungs, and digestive tract (Slonim, 2014; Porges, 2007). Thus, emotions are not simply abstractions of the mind, but physiological responses.

The connection between smell and emotion is a topic that has been studied by many since it is rooted in aspects of many people's lives, such as memory stimulation (Fox, 2009), aromatherapy (Cho, 2013), consumer behavior (Chebat, 2003), and romantic partner pheromone attraction (Hofer, 2018). People identify smells by activating the limbic system, a section of the brain known to be the "seat of

emotion," and relating that smell to a memory or emotion (Fox, 2009). Only after this emotional connection has been made are we able to name a particular scent, suggesting that our perception of smell is closely linked to, and even dependent upon, our emotions.

For centuries, people have been manipulating their personal scent either to change the way others perceive them (Fox, 2009; Roberts, 2009) or for their own liking (Milinski, 2013; Sell, 2006). A correlation has been found between the Major Histocompatibility Complex (MHC) genotype—a collection of peptides that determine the strength of a person's immune system—and the fragrances that they like (Milinski, 2001). This further demonstrates how the sense of smell is neurologically and emotionally complex. These connections demonstrate that smells and fragrances can manipulate our emotions and, correspondingly, alter the rhythm of activity within the ANS.

The marketing information that is used to sell a fragrance also plays a role in manipulating consumers' emotions. Studies on general marketing information have shown, for example, that the vowels within a brand name can draw consumers toward names because people subconsciously favor certain vowels over others (Klink, 2009). Additionally, the images on the packaging of a product (Underwood, 2001) and a person's ability to touch an item (Krishna, 2010) have been shown to increase shoppers' attention to a product, and likeliness of purchasing a product. However, little research has been done to show any emotional responses due specifically to the marketing strategies of perfume and cologne.

Furthermore, researchers have not explored the question of whether perfume and cologne evoke different emotional responses from men and women. Despite the gender-based divide in the fragrance industry, from the separation of the ingredients used in perfumes and colognes to the different colors used for their packaging, minimal research has been done to assess how men and women react emotionally to the smells of perfume and cologne and how they are influenced by the marketing strategies displayed by fragrance brands. In popular culture, it is accepted that sweet, fruity, and floral smells are among the more feminine scents while scents like musk, spice, and smoke are marketed as more masculine (Fox, 2009; Sell, 2006). Additionally, perfume is typically sold in bottles and boxes that are pink, purple, or other colors that are considered to be feminine while cologne is sold using colors that are considered to be masculine (Sell, 2006). There are countless other ways perfume and cologne adhere to gender stereotypes. Knowing if men and women truly respond differently to these scent formulations and marketing strategies would suggest that there is a difference in the biological makeup of men and women that causes them to be drawn to different scents and images. Additionally, it would imply that the reason for the gender divide in the fragrance industry isn't solely a marketing strategy or an attempt at enforcing gender-stereotypes. Advertisements and other aspects of fragrance marketing are intentional (Klink 2009), and have clear target genders which can be inferred based on the color or the packaging to the words used in the titles. Yet, similar to the effects of the smell of fragrance on consumers, little research has been done to assess how these marketing strategies influence consumers' decisions and specifically if men and women respond to them in the same ways.

Building off of what is known about the relationship between the sense of smell and emotional response, the physiological response associated with emotions, and the effects of marketing information on consumers' emotional response to a product, the current study attempts to measure how people's emotions are affected by fragrance scent and marketing, and more specifically, whether men and women respond differently to fragrance scent and marketing.

Measuring Emotion

Heart rate is commonly used to evaluate emotional state since it is an easily measurable function of the ANS (Choi, 2017). However, inducing a measurable physiological reaction requires a stimulus that will adequately provoke emotions within the subject (Geethanjali, 2017). One effective way of doing this is to present stimulation that combines the senses, particularly visual stimulation with stimulation of another sense (Geethanjali, 2017). Thus, the physical smell, pictures of the packaging, the descriptions, and the video advertisements for the perfume and cologne were presented to subjects in this study to elicit a stronger stimulation of the ANS and subsequent change in heart rate. Heart Rate Variation (HRV) is a quantity derived from heartbeat that demonstrates how much the rate of a person's heartbeat changes between successive beats; this can be measured in terms of time (the time, in milliseconds (ms), that elapses between each heartbeat) or frequency (the number of heartbeats with the same select valence that would occur within one second; measured in Hertz) (Appelhans, 2006; Choi, 2017; Moore, 2017). Since emotions generally correlate to changes within the ANS, HRV provides a more accurate assessment of emotional state than a measurement of average heart rate would (Appelhans, 2006; Choi, 2017; Farnsworth, 2019; Moore, 2017). However, measuring the change in a person's average heart rate can still indicate signs of emotional stimulation (Ekman, 1983; Haag, 2004; Koelstra, 2011). Therefore, both HRV and average heart rate were used as the primary variables of analysis for this study.

HRV can be calculated in many different ways, all of which evaluate different dimensions of heart rate changes that translate into different dimensions of emotional response (Appelhans, 2006; Farnsworth, 2019; Kanjo, 2015). The most commonly used calculation is known as the Root Mean Square of Successive Differences (RMSSD), a time-based analysis that calculates the overall variance in the time between each heartbeat (Farnsworth, 2019; Moore, 2017). The overall progression of a heartbeat is referred to as the "QRS Complex," with Q representing the beginning of a beat, R representing the peak of a beat, and S representing the end of a beat (Farnsworth, 2019). The distance (in milliseconds) between the peak of each heartbeat, the "R," is referred to as the "RR Interval" (Farnsworth, 2019). The equation to calculate RMSSD utilizes the RR Interval as follows:

$$\text{RMSSD} = \sqrt{\frac{(RR\ Interval\,1 - RR\ Interval\,2)^2 + (RR\ Interval\,2 - RR\ Interval\,3)^2 + (\ldots)}{\#\text{ of RR Intervals}}}$$ (Farnsworth, 2019)

RMSSD was selected to be the measurement of HRV used for heart rate analysis in this study.

Methods

Subjects

All participants in Trial 1a (n=8), 1b (n=6), and 1c (n=15) were Upper School students at The Berkeley Carroll School in grades 10 and 11. All participants in Trial 1a and 1b were students enrolled in the Science Research and Design (SRD) program, and all participants in Trial 1c were students from a health class that met when the experiments were being conducted. The recruitment processes varied between trials. For Trial 1a, the experimental objective and procedure were verbally explained to a 10th grade SRD class and students volunteered to participate in person. For Trials 1b and 1c, class instructors were contacted regarding their students' participation; information and sign-up sheets were sent to the instructors who then gave them to the students. Students voluntarily signed up for available time slots, the location of experimentation was specified, and experiments were carried out at assigned times. Additional subjects for trial 1c were obtained from a 10th grade health class; the procedure for finding volunteers was the same as the procedure used for trial 1a. The subject pool was small due to limitations in time for the experiment, and access to volunteers. The experiment needed to be performed

on all subjects in person and individually, and the average time for each experiment was around fifteen minutes, which unfortunately only made it possible to obtain a small group of subjects.

Subjects in all trials were exposed to the same conditions to ensure that no external stimulation would influence data. All experiments were either held in the Berkeley Carroll Science Prep Room or in room 405. During the experiments, there weren't any additional people present, and there was minimal sound produced inside and outside of the room. No potential harm was associated with any aspect of this experiment, and permission to conduct this study was granted by the Berkeley Carroll Institutional Review Board (IRB). Each subject was aware of their role as a participant prior to the experimental trial. Anonymity was maintained since subjects' names were never recorded; there was no way to connect data back to the individual subject.

Apparatus and Materials

Two fragrances, one perfume (Coco Mademoiselle, by Channel) and one cologne (Boss, by Hugo Boss)[1], were used in this experiment; both fragrances were used to demonstrate differences in smell, marketing information, and overall gender specificity. It was hypothesized that this information would evoke different emotional responses in male and female subjects. The fragrance was presented to the subjects in a clear bottle without a label, accompanied by selected marketing information: an image of the bottle that the fragrance would be sold in, the name of the fragrance, the description of it from the company website, and the video advertisement that was used to market it. The type of fragrance (perfume or cologne) was not specified to subjects so that they wouldn't begin the experiment with an idea of how they thought they should respond based on their gender.

The Mindfield eSense Pulse HRV Biofeedback Sensor *(Figure 1)* was used to measure subjects' heart rate in response to exposure to the fragrances and marketing information. This monitor was selected since it consisted of only one sensor, it was easy to use, it calculated RMSSD and average heart rate, and the data could be easily viewed and downloaded through the free Mindfield eSense Pulse App which was made specifically for the monitor. This app was downloaded from the app store onto an iPad provided by Berkeley Carroll. The eSense Pulse also came with electrode contact spray, which needed to be applied to the two electrodes on the band before use to increase registration of subjects' heartbeat *(Figure 2)*.

Since RMSSD and average heart rate have not yet been applied to studies with fragrance, a self-reported emotion questionnaire was also used to evaluate emotional response. These questions were meant to provide additional insight into how subjects were responding emotionally to the fragrances by giving them a chance to self-report their feelings and opinions. The first part of these questions was adapted from the Self-Assessment Manikin Test (SAM), a questionnaire that has been used in previous studies as an additional measure of emotion along with a physiological measure (Geethanjali, 2017; Morris, 1995; Stevens, 2016). The SAM Test is useful since it assesses three aspects of an emotional state: valence (type of emotion), arousal (level of emotional excitement), and dominance (strength of emotion). Under each of the categories, the SAM presents a number scale—typically 1-5 or

[1] The highest-rated perfume and cologne on the Sephora website (Sephora.com) during the time this experiment was being planned. One worry when planning this experiment was that the subjects simply would not like the smells of the fragrances presented to them, which would cause them to base their answers solely on that rather than the marketing information and the smell. Therefore, the highest-rated fragrances were chosen with the idea that if many other people liked them, the subjects in this experiment would hopefully not be repulsed by the smell, and would be able to base their emotional response on all information provided.

1-9[2]—accompanied by figures that visually depict the feeling that each number on the three scales is meant to measure. In this study, an additional set of questions was added to the SAM Test *(Figure 3).* These questions included two short-response questions regarding first impressions of the fragrances, one question that asked subjects to circle the emotion that was the closest to what they were experiencing, and three scales (referred to as "likeliness scales") meant to evaluate consumer tendencies of men and women and their awareness of how their gender is being targeted by the marketing strategies of the fragrances. Together, the measurements of HRV and the emotion

Figure 1

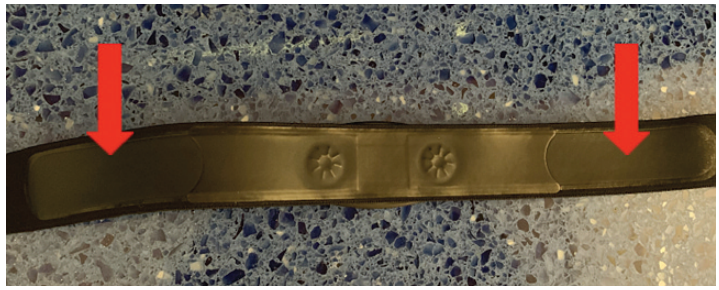
Picture of Mindfield eSense Pulse Apparatus.

Image includes the sensor band, and conductivity spray.

Figure 2

Application Points for Electrode Spray.

Diagram of where electrode spray is to be applied to electrodes on eSense Pulse band; shows the sensor band flipped over.



[2] Where 1 is the lowest, and 5 or 9 is the highest. For example, in this study, 1 represents "sad", and 9 represents "happy" under the category of valence. Labels change with three categories.

Figure 3

Emotion Questionnaire that was used in this Study.

"Question Set 1" includes the SAM Test, and "Question Set 2" includes the additional short response questions and the three likeliness scales.

Question Set 1
General Emotions – Self-Assessment Manikin Test

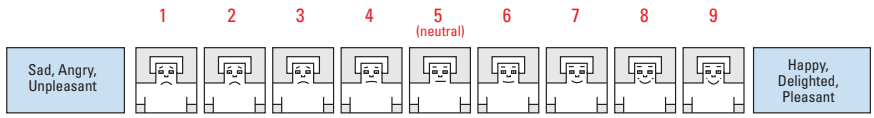
Overview

- This assessment allows you to report how you are feeling emotionally *in response to the fragrance and marketing information you were just exposed to*
- Asks about three dimensions of your emotion: valence, arousal, and dominance
- The images of the manikin are supposed to help you gauge how you are feeling

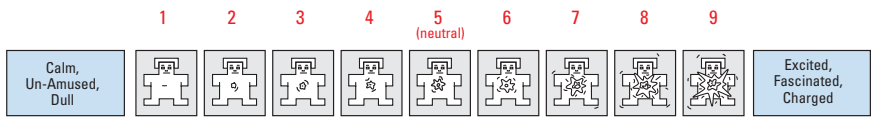
Directions

Circle the number that represents your emotional state (do not circle in between numbers, please).
You are assessing *how the fragrance smell and advertising made you feel.*

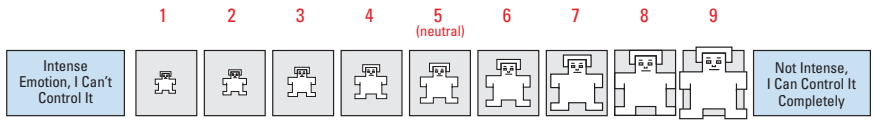
Valence: What type of emotion are you experiencing?



Arousal: How emotionally excited are you feeling?



Dominance: How intense is the emotion you are feeling/How much are you able to control it?



Question Set 2
Specific Emotional State Questions

Directions

1. For questions 1 and 2, give a brief written response. Please try to be as specific as possible.
2. For questions 3-5, circle the value that best describes your feeling. Please only circle one value, and don't circle any spaces in between numbers.

Questions

1. Briefly explain: What are your initial thoughts after smelling this fragrance and observing the marketing information? Focus on opinions, emotions, likes/dislikes, etc.

2. Circle two emotions that best describe how you are feeling in response to smelling the fragrance, and observing the marketing information:

   Happiness, Excited, Romantic, Soothing, Disgust, Irritation, Angry, Sad

3. Briefly explain: If you saw this fragrance in a store/online, would you purchase it for yourself? Why or why not?

4. How likely are you to buy a fragrance that does not specify gender in the marketing of the scent (circle one)?

   1 2 3 4 5 6 7 8 9
   Not Likely At All — Neutral — Very Likely

5. How likely are you to buy a fragrance that is marketed to your gender?

   1 2 3 4 5 6 7 8 9
   Not Likely At All — Neutral — Very Likely

6. How likely are you to buy a fragrance that is NOT marketed to your gender?

   1 2 3 4 5 6 7 8 9
   Not Likely At All — Neutral — Very Likely

questionnaire provided a complete assessment of both subjective and biometric emotional response to fragrance and data to assess any differences in the way males and females respond to perfume and cologne.

Procedure

Before the experimental process began, subjects put the heart rate monitor on. Subjects did this in privacy because the monitor needed to make contact with their skin and therefore be applied under their shirt. They were given the monitor, the electrode contact spray, and an instruction sheet with visuals before they put the monitor on. When subjects returned, it was ensured that the monitor was connected to the eSense App on the iPad. After the monitor was registered, subjects were given a brief overview of the experimental procedure and the opportunity to ask questions. There were four periods of heart rate measurement: rest 1, fragrance 1 exposure, rest 2, and fragrance 2 exposure. To control for olfactory fatigue[3], perfume and cologne were alternated between the fragrance periods (1 and 2) for each subject. The eSense monitor was started at the beginning of each period, and stopped at the end; there was no overlap in measurement between the four periods so that they could be analyzed separately. Each measurement period was labeled uniquely for distinction. During the first rest period, subjects followed a 60-second breathing exercise[4] (rest 1). The rest periods were meant to normalize subjects' heart rate and provide a control measurement for analysis. A breathing exercise was selected for both rest periods to control for any changes in heart rate that might have been attributed to deep inhalation of the fragrances. Following this period, subjects smelled the first fragrance and were shown the corresponding marketing information (fragrance 1 exposure). When this period ended, subjects were given coffee beans to smell to reset their olfactory receptors. For trial 1a, subjects then immediately began the second rest period (rest 2). For trials 1b and 1c, subjects were given the question sets after the first fragrance period; heart rate was not measured while subjects were answering questions. Subjects in trials 1b and 1c did the second rest period after finishing the questions (rest 2). Following the second rest period, subjects from all trials smelled the second fragrance, and were shown the selected marketing information (fragrance 2 exposure). When the second fragrance period was over, trial 1a ended, however, subjects in trials 1b and 1c were given the same set of questions as they were after the first fragrance period. A short feedback questionnaire was given at the end of the experiment for all trials.

Data

Since no studies have been conducted to assess emotional response to fragrance, it was relatively unknown which variables were the best to analyze for this experiment. Therefore, several variables were measured to assess which ones were the most effective in measuring emotional response to fragrance and would help answer the experimental questions. These variables included four different types of calculations derived from heart rate measured through the Mindfield eSense Pulse monitor and all of the questions provided on the emotion questionnaire *(Figure 3)*. Once all data was collected and statistical analyses were conducted, the most useful variables were selected and are presented below.

---

[3] When olfactory receptors in your nose can no longer detect smells sufficiently because they have been exposed to other smells previously (Sell, 2006).

[4] Exercise was conducted through a guided breathing app on an iPhone called, *The Breathing App*. Exercise consisted of a sphere expanding and contracting; subjects inhaled while sphere was expanding, and exhaled while sphere was contracting.

The effectiveness of these variables was determined by a few factors. RMSSD and average heart rate were the two calculations from the heart monitor that were selected because other scientific literature supports their accuracy (Appelhans, 2006; Choi, 2017; Ekman, 1983; Farnsworth, 2019; Haag, 2004; Koelstra, 2011; Moore, 2017), while the other variables are not widely used. The other variables were initially measured to see if they would provide insight different from that provided by the RMSSD and average heart rate data points, however, none of these measurements displayed any distinct results. All results from the SAM Test were excluded because many of the subjects expressed that they experienced difficulty filling it out. They felt that determining their own emotion was difficult to begin with and that the format of the test was also confusing. Because of this, the results from the SAM Test are most likely inaccurate. The two short response questions and the question that asked subjects to circle the emotion that best described how they were feeling were not meant to be statistically analyzed—they were initially meant to provide additional qualitative understanding of the subjects' response to the fragrances. For these reasons, these variables were excluded from the overall data analysis.

For each of the following analyses, data were obtained from 10th and 11th-grade girls and boys. The number of subjects varies between analyses—sample size is indicated before each data set. All subjects were exposed to the same fragrance scent and marketing information under the same experimental conditions. Error bars represent the estimated standard error of the mean (average ± SEM). A two-tailed t-test with independent samples was used to assess the significance of differences in data.

Heart Rate Data

For figures 4-7, data from 13 female subjects and 10 male subjects were included in analyses.

Figure 4

Male vs. Female: Change in RMSSD from Rest 1 to Perfume Exposure Period.

Bars represent the average amount that RMSSD changed from first rest period (rest 1) to the perfume exposure period between all male and female subjects (average ± SEM). No significant difference is present between the results (p= 0.23).

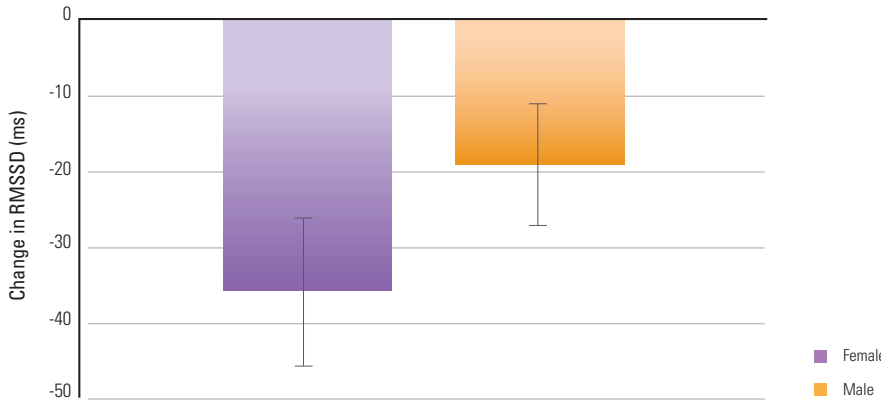


Figure 5

Male vs. Female: Change in RMSSD from Rest 2 to Cologne Exposure Period.

Bars represent the average amount that RMSSD changed from second rest period (rest 2) to the cologne exposure period between all male and female subjects (average ± SEM). No significant difference is present between the results (p= 0.50).

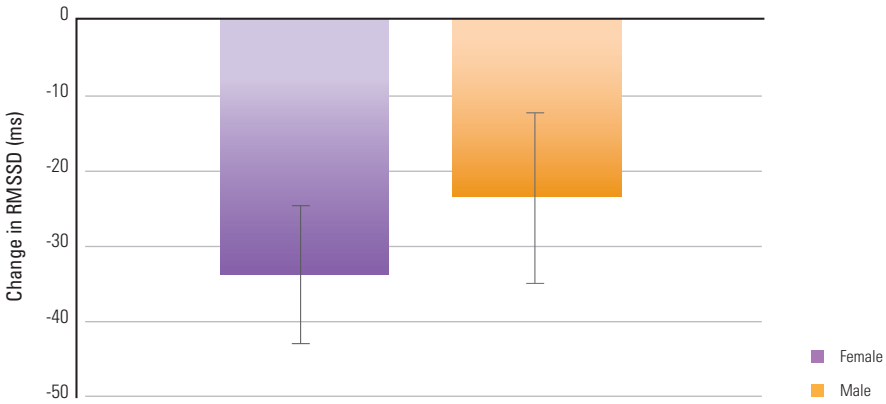


Figure 6

Male vs. Female: Change in Average Heart Rate from Rest 1 to Perfume Exposure Period.

Bars represent the average amount that the average heart rate changed from first rest period (rest 1) to the perfume exposure period between all male and female subjects (average ± SEM). No significant difference is present between the results (p= 0.36).

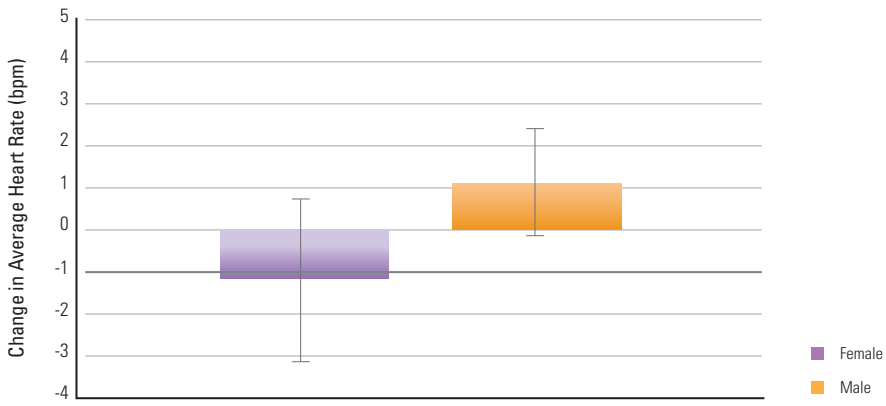
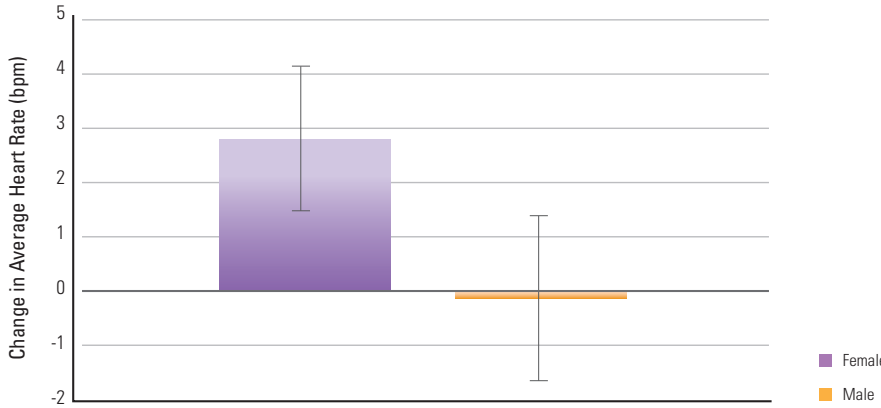


Figure 7

Male vs. Female: Change in Average Heart Rate from Rest 2 to Cologne Exposure Period.

Bars represent the average amount that RMSSD changed from second rest period (rest 2) to the cologne exposure period between all male and female subjects (average ± SEM). No significant difference is present between the results (p= 0.17).



Self-Reported Data

For figures 8-10, data from 8 female subjects and 6 male subjects were included in analyses. Additionally, data from perfume and cologne were combined.

Figure 8

Average Likeliness to Purchase a Fragrance that does not Specify Gender: Female vs. Male.

Bars represent the average likeliness ratings female and male subjects gave (average ± SEM). A significant difference is present between the results (p= 0.02).

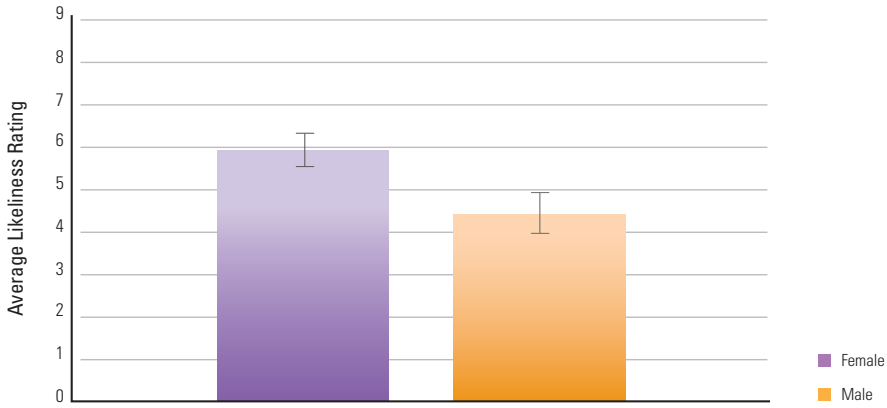


Figure 9

Average Likeliness to Purchase a Fragrance that is Marketed toward their Gender: Female vs. Male.

Bars represent the average likeliness ratings female and male subjects gave (average ± SEM). A significant difference is present between the results (p= 0.017).

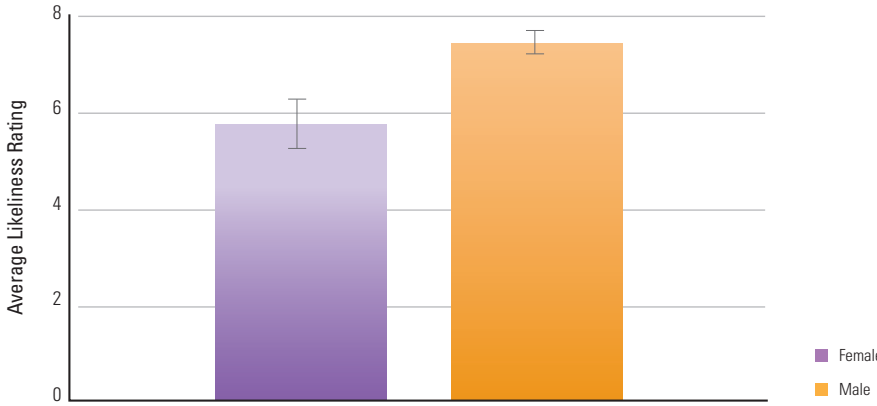
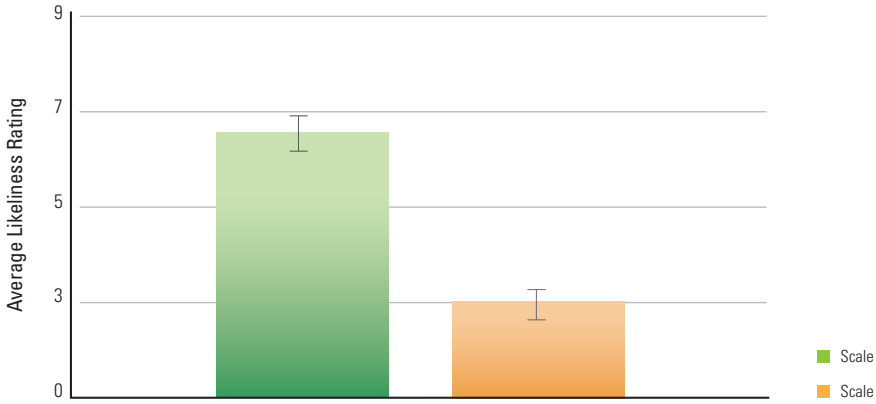


Figure 10

Average Likeliness Rating on Scale 2 vs. Scale 3 with Male and Female Responses Combined.

Bars represent the average likeliness ratings on scale 2 ("how likely are you to purchase a fragrance that is marketed toward your gender?") and scale 3 ("how likely are you to purchase a fragrance that is marketed toward the opposite gender?"); average ± SEM. A significant difference is present between the results (p= <0.0001).



Data Analysis and Discussion

The primary set of questions this experiment strived to answer was: Why is there a gender divide in the fragrance industry? Is there something in the biological makeup of men and women that causes them to be drawn to different scents and images, or is the divide primarily a marketing strategy based on gender stereotypes?

Two preliminary studies were conducted to answer this set of questions prior to this iteration of the experiment. In the first study, ingredients that are commonly used in perfume and cologne were selected, and presented to male and female subjects; the subjects were asked to select the emotions they felt when smelling each scent from a list of options. Each emotion on the list was categorized as either a positive or negative emotion. The proportion of number of positive to number of negative emotions was calculated for the male and female subjects separately, and then compared statistically; no results were significant. In an attempt to improve the experimental design, rule out factors that would have affected the data and answer the questions above more directly, the second experiment focused on how people decide to purchase a fragrance based on the marketing information they see. Subjects were split into two groups: one smelled perfume and cologne without any marketing information present, and the other smelled the perfume and cologne with all of the marketing information (name of fragrance, bottle, description) present. Male and female responses under the two conditions were compared, and none of the results were significant. This experiment was the third attempt to improve the experimental design to see if eliminating more outside variables would produce a different result; a result that suggested the reason for the gender divide in the fragrance industry goes beyond a social system. However, once again, the data collected did not produce any result that indicates men and women emotionally respond differently to perfume and cologne. The data from these three experiments suggested that gender divide in the fragrance industry is a marketing strategy based on gender stereotypes rather than physiological gender based differences in emotional response.

This iteration of the experiment was meant to further assess these questions from the perspective of the emotional response, both physiological and self-reported, that men and women have to fragrance. The data from this experiment supports the same conclusions that were drawn from the two preliminary studies. None of the presented analyses of male vs. female change in RMSSD and average heart rate data produced statistically significant results, indicating that there is no gender-based difference in physiological emotional response to perfume and cologne.

However, when the change in RMSSD data points were pooled together from all conditions—male and female, perfume and cologne—and compared against a change of zero (no change in RMSSD), the overall change in RMSSD was significantly greater than zero (single sample t-test, assumed mean of 0, $p=<0.0001$). In all conditions, RMSSD had decreased from the rest period to the fragrance exposure period, which means that both male and female subjects became more emotionally stimulated when they were exposed to fragrance than they were during a normal time of rest. The conclusion that the fragrance periods altered the emotional state of male and female subjects not only demonstrates that the experimental conditions were successful in evoking a physiological response, which was one of the main concerns of this experiment, but also shows that the results from the original gender-based analyses are accurate—fragrance clearly affects emotion, however, it does not affect males and female emotional physiology differently.

This trend was not seen in the average heart rate measurements—average heart rate of male and female subjects did not change in the same way under both fragrance conditions (single sample t-test,

assumed mean of 0, p= 0.403369). Overall, the average heart rate data did not provide any useful insights into the ways that the male and female subjects responded to the perfume and cologne, which indicates that RMSSD is a more effective measurement tool for assessing emotional response to fragrance than average heart rate.

The only results from this study that showed a statistically significant gender-difference in fragrance preference were seen in self-reported responses. The last portion of the questionnaire used in this study was the set of likeliness scales which evaluated how strongly the male and female subjects cared about the gender a fragrance was marketed toward (figure 3, "Question Set 2", questions 3-5). When subjects were asked how likely they were to purchase a fragrance that did not specify gender, the female subjects had a significantly greater average rating than the male subjects (figure 8, p=0.02). When they were asked how likely they were to purchase a fragrance that was marketed specifically toward their gender, the male subjects had a significantly greater average rating (figure 9, p=0.017). There was no gender-based difference present when subjects were asked how likely they were to purchase a fragrance that was marketed toward their gender, however, when male and female responses were combined, both genders were significantly more likely to purchase a fragrance marketed toward their gender than a fragrance that was marketed toward the opposite gender (two-tailed t-test, p= <0.0001). These results indicate that subjects use gender-targeted fragrance marketing to inform their purchasing decisions. More importantly, these results demonstrate that gendered marketing successfully draws consumers to purchase fragrances that match their gender. These conscious habits represent how the fragrance industry continues to maintain its gender-divide through the gender-stereotypes present in the marketing information of the products.

References

Appelhans, B., Luecken, L. (2006). Heart rate variability as an index of regulated emotional responding. *Review of General Psychology, 10*(3), 229-240.

Chebat, J., Michon, R., (2003) Impact of ambient odors on mall shoppers' emotions, cognition, and spending: A test of competitive causal theories. *Elsevier Inc., 56*(7), 529-539.

Choi, K., Kim, J., Kwon, S., et al, (2017). Is heart rate variability (HRV) an adequate tool for evaluating human emotions? A focus on the use of the International Affective Picture System (IAPS). *Psychiatry Research, 1*(251), 192-196

Cho, M. Y., Min, E. S., Hur, M. H., & Lee, M. S. (2013). Effects of aromatherapy on the anxiety, vital signs, and sleep quality of percutaneous coronary intervention patients in intensive care units. *Evidence-based complementary and alternative medicine: eCAM*. 2013;2013:381381

Ekman, P., Levenson, RW., Friesen, WV. (1983). Autonomic nervous system activity distinguishes among emotions. *Science, 221*(4616), 1208-1210

Farnsworth, B., (2019). Heart rate variability-how to analyze ECG data. Retrieved from: https://imotions.com/blog/heart-rate-variability/

Fox, K. (2009) The smell report. Social Issues Research Centre. Retrieved from: www. sirc.org/publik/smell.pdf

Fragrance. (n.d.). Retrieved from www.sephora.com/shop/fragrance

Geethanjali, B., Adalrasu, K., Hemaprabha, A., et al. (2017). Emotion analysis using SAM (Self-Assessment Manikin) scale. *Biomedical Research 2017: Complex World of Neuroscience,* Special Issue: S18-S24

Haag A., Goronzy S., Schaich P., Williams J. (2004). Emotion recognition using bio-sensors: first steps towards an automatic system. *Affective Dialogue Systems. Springer, Berlin, Heidelberg,* 3068, 36-48.

Hofer, M. K., Collins, H. K., Whillans, A. V., & Chen, F. S. (2018). Olfactory cues from romantic partners and strangers influence women's responses to stress. *Journal of Personality and Social Psychology, 114*(1), 1-9.

Kanjo, E., Al-Husain, L., Chamberlain, A., (2015). Emotions in context: examining pervasive affective sensing systems, applications, and analyses. *Personal and Ubiquitous Computing, 19*(7), 1197-1212

Klink, R. R. (2009). Gender differences in new brand name response. *Marketing Letters, 20*(3), 313-326

Koelstra, S., Muhl, C., Soleymanni, M., (2011). DEAP: A database for emotion analysis using physiological signals. *IEEE Transactions on Effective Computing, 3*(1), 18-31.

Krishna, A., Elder, R., Caldara, C. (2010) Feminine to smell but masculine to touch ? Multisensory congruence and its effect on the aesthetic experience. *Journal of Consumer Psychology, vol. 20*(4), 391-520.

Milinski, M., Wedekind, C. (2001). Evidence for MGC-correlated perfume preferences in humans. *Behavioral Ecology, 12*(2), 140-149.

Milinski, M., Croy, I., Hummel, T., Boehm, T. (2013). Major histocompatibility complex peptide ligands as olfactory cues in human body odour assessment. *Proceedings of The Royal Society B. 280*(1755).

Moore, J. (2017). Heart rate variability vs. heart rate. Retrieved from: https://hrvcourse.com/heart-rate-variability-vs-heart-rate/

Morris, J., (1995). SAM: The self-assessment manikin; An efficient cross-cultural measurement of emotional response. *Journal for Advertising Research, 35*(6), 63-68.

Porges, S. (2007). The polyvagal perspective. *Biological Psychology. 74*(2), 116-143.

Rockensüß, Niko. (2019) *Mindfield ESense Pulse Handbook. Mindfield ESense Pulse Handbook* Mindfield Bodysystems.

Roberts, C., Little AC, Lyndon A., Roberts J., Havlicek, J., Wright, RL. (2009). Manipulation of body odour alters men's self-confidence and judgments of their visual attractiveness by women. *International Journal of Cosmetic Science. 31*(1), 47-54.

Sell, C. (2006). The Chemistry of Fragrances. Cambridge, UK: *The Royal Society of Chemistry.*

Slonim, T., (2014) The Polyvagal Theory: Neuropsychological foundations of emotions, attachment, communication, & self-regulation. *International Journal of Group Psychotherapy, 64*(4), 593-600.

Stevens, F., Murphy, D., Smith, S. (2016). The self-assessment manikin and heart rate: Responses to auralised soundscapes. *Interactive Audio Systems Symposium, York* (2016).

Underwood, R., Klein, N., Burke, R., (2001). Packaging communication: Attentional effects of product imagery, *Journal of Product & Brand Management, 10*(7), 403-422.

The Significance of Dreaming: A Case Study Examining the Dreams of 13 High School Students

by Hope Swetow

Abstract

The current study explores the significance of dreaming in 13 high school students. For 14 nights, participants filled out a dream diary explaining their dreams and their perceived meaning. The experimental interpretation of the dreams was determined using the pattern matching technique; each dream was matched to 1 of 5 existing dream theories. In addition, a modified version of the word search method used in Bulkeley (2009) was employed to highlight the most frequent words and themes in each dream. The continuity hypothesis was found to best explain the significance of the adolescents' dreaming. The dreamers were often able to connect their dreams to specific events and emotions in their waking lives. The word search technique found that the most recurrent themes (eg. friends, family, and school), are themes that logically play major roles in the lives of teenagers. It was found that some themes in dreams were universal for the male and female participants, while other themes (eg. celebrities, transportation, drugs/smoking/alcohol, parties, love/sexual relationships) were more often linked to the dreams of females.

Background

The study of human dreams is based around two central questions: how do they occur (in regard to physiology) and what is their purpose? In the past, it was thought that the brain was in a state of dormancy and inactivity during sleep (Blake & Gerard, 1937). Although oneirology remains an undeveloped area of study, research has increased greatly since Aserinsky and Kleitman published their discovery that dreaming occurs in the rapid eye movement (REM) stage of sleep (Aserinsky & Kleitman, 1953). REM is characterized by "rapid, jerky, and binocularly symmetrical eye movements" (Aserinsky & Kleitman, 1953). The brain is most active during REM, which is why it is linked to dreaming (Aserinsky & Kleitman, 1953). Normally, REM first occurs 90 minutes after falling asleep and lasts for 10 minutes (Siegal, 2001). People usually experience 5-6 cycles of REM sleep in one night. More recent studies have found that while dreaming is most common during the REM stage of sleep, it can also occur during non-REM sleep (Solms, 2003).

Scientists currently have a greater understanding of the physiology of dreaming than of the explanation for dreaming. Why humans dream is a highly debated topic. Several scientists and philosophers have proposed widely accepted hypotheses explaining the significance of dreaming. The current study examines 5 of these hypotheses.

The activation synthesis hypothesis proposes that dreams don't serve a physiological purpose. The activation refers to the "random firing of neurons from the brain stem" and the synthesis refers to the "interpretation of these random signals into spontaneous thoughts by the cerebral cortex" (Hobson, 1999). Scientist Allan Hobson concluded that dreams are just "noise in the machine" (Hobson, 1999). Hobson proposed the activation synthesis hypothesis after examining different parts of the brain during dreaming. Hobson deduced that because only 14% of people remember dreams every night, it is unlikely that the content has value.

The continuity hypothesis suggests that waking states, concerns, and emotions are the primary images reflected in dreams (Bulkeley, 2009). Dreams accurately reflect emotional concerns but not always actual events. This theory states that one is able to evaluate someone's psychological well-being, personality, and priorities based on their dreams. According to this hypothesis, dreams do not serve an evolutionary function, but can still be used as a tool to understand the dreamer. The continuity hypothesis has prompted therapists and psychologists to discuss dreams with patients, since it suggests that great insight into someone's well-being can be accessed from dream content. In a longitudinal study testing the continuity hypothesis, scientists investigated the dreams of 28 female participants over the course of a 6-10 year period (Pesant & Zadra, 2005). They also took into consideration the participants' self-reported psychological well-being. The study concluded that lower self-reported levels of psychological well-being from participants corresponded with dreams containing more aggressive (as opposed to friendly) interactions and negative (as opposed to positive) emotions. They also concluded that the types of concerns most likely to be reflected in people's dreams are "predominantly interpersonal and emotional in nature" (Pesant & Zadra, 2005).

The emotional regulation theory suggests that dreams serve to process and confront emotions while dreaming (Wagner, 2001). This theory is very similar to the continuity hypothesis, except it postulates that dreams have evolutionary benefit. For example, if someone's emotional equilibrium in their waking life is disturbed, dreams dilute those feelings. Furthermore, dreams don't have to be reflections of daily life (a dream about zombies, for example). In the study conducted by Wagner, scientists compared memory formation for neutral and emotionally arousing materials during periods of early and late sleep of three groups of adult men (Wagner, 2001). All three groups had to read an emotional text and a neutral text. The first group went to sleep but subjects were woken before they could get to REM sleep. The second group went to sleep and were not woken until they had a few stages of REM sleep. The control group did not sleep. The two groups who went to sleep remembered more of the stories in general than the control group. However, during late-sleep, memory of emotional texts was significantly enhanced compared to the control and the early sleep group, suggesting that during REM (the stage in which most dreaming occurs), emotions were being processed more than other aspects of waking life. In another study, adult participants were asked about their depression before and after sleep. They found that those depressed before sleep reported significantly more positive dreams which led the scientists to believe that dreams are emotional stabilizers (Ruby, 2011).

The threat simulation hypothesis suggests that a biological function of dreaming is to simulate events that might have threatened the reproductive success of our ancestors in order to improve the probability that similar events in their waking lives would be handled successfully (Revonsuo, 2000).

According to Revonsuo, dreams let us "rehearse threat-avoidance skills in the simulated environment of dreams just like mental training and implicit learning have been shown to lead to improved performance on a wide variety of tasks". Because people today rarely face the same threats as our ancestors, the threat simulating function of dreams is likely not fully activated often. However, it is probably manifested in dreams that include recent traumatic events in a person's life that trigger negative long-term memories.

Freud's dream theory was not based on scientific evidence but rather philosophy. He believed that dreams reflect the desires of the subconscious that a person is unwilling, or unable, to confront in their waking life (Freud 1900). He believed that dreams must have meaning and claimed that even if the images in a dream seemed foreign, they were just distorted pieces of one's waking life. He states that dreams have two parts: the manifest content and latent content (Freud, 1900). What we remember of our dreams when we wake up is the manifest content, and the dreams that we have and don't remember (part of the subconscious) are the latent content. The id is an important aspect in Freud's model of the psyche. The id is defined as "the set of uncoordinated instinctual desires" of a person (Freud, 1900). Freud claimed that the latent part of dreaming is composed of aspects of the person's real life masking the id.

In this case study, I explore the significance of dreaming in 13 high school students. In most previous studies investigating the purpose of dreaming, scientists collected data quantitatively or by studying the brain directly. This case study format allowed for a large amount of in-depth qualitative data as opposed to a short form, survey, or brain scan as used in prior research, which was necessary to justify the conclusions made.

Several analysis techniques were used. The rigorous and accelerated data reduction (RADaR) technique was the framework of the analysis because it is a data organization tool commonly used for case studies to make sorting and arranging large amounts of data easier (Watkins, 2017). Two additional techniques were used for the analysis of the dreams. The pattern matching technique is used in different types of studies to connect data to previously proposed theories (Yin, 1984). In the current study, pattern matching allowed for linking the meaning of dreams with 5 already existing dream hypotheses. This was done in order to see which results from previous studies mirror the content of the dreams reported in this study. While other studies use data from dreams to develop theories, my study classifies dreams based on existing theories. A modified version of the word search method was also used to calculate the frequency of certain words/themes within each dream to highlight the most important concepts across the data set. To my knowledge, the word search method has never been used in such a small subject pool as it was designed to calculate the word frequency of thousands of dreams on DreamBank.net. The use of all three analysis techniques in this study maximizes the chance of reliable and justifiable results.

Methods

In order to get participants for this study, an email was sent out to 9th and 10th grade students at Berkeley Carroll asking if they would be willing to participate. In order to encourage participation, an incentive was implemented that gave students permission to come late to school on days when they had a first period free. However, the email only attracted 3 participants. The other 10 subjects were obtained in a few different ways. 8 subjects were recruited because their friends had participated in the study and they were interested and 2 participants were recruited through their Science Research and Design class. The participants' ages ranged from 15-17-years-old. Participants were not required to provide their name.

A dream diary was distributed to each of the 13 participants. The diaries for each participant were distributed on different days depending on when the person decided to join; this allowed for the acquisition of more data as subjects could be included when they became available. Participants were instructed to fill out their dream diary for 14 nights, after which it was returned. The instructions for dream entries were printed in the dream diary. Each night, the participants responded to two questions about their dreams:

Question #1: Please tell me about your dream(s) from last night in as much detail as possible. Include anything about people, places, conversations, actions, emotions, etc. that occured in the dream.

Question #2: Please tell me what you think the meaning of your dream(s) from last night is? Was it too confusing to understand or maybe it was reflective of a real event in your life/something that happened the previous day?

If the participant did not have a dream then they did not respond to the questions. The dreamless days did count towards the 14 required entries. A check-in email was sent 7 days into each participant's time to make sure participants were doing the study and did not have any questions. At the end of the study, participants were asked to answer 2 follow up questions before returning the diary.

Follow up Question #1:
Explain your experience filling out the dream journal.

Follow up Question #2:
Did you learn anything about yourself or your dreams?

Dream transcripts were analyzed using the rigorous and accelerated data reduction (RADar) technique. This method follows a 5 step process. The first step ensures that all the data transcripts are formatted similarly. This entails creating and transcribing all of the data from a participant into an organized transcript. The second step places the formatted data transcripts into an all-inclusive, Phase 1 data table. This means copying and pasting all of the text from the transcripts into a table with multiple rows and columns. This is the first "data reduction table" (ie. A table or spreadsheet that contains all the information from the transcripts (Watkins, 2012). The third step reduces data from all-inclusive data tables to produce a shorter, more concise Phase 2 data table. Information from the Phase 1 data table is removed/reduced so that what remains is only the text that is relevant to the overarching research question. The fourth step includes pattern matching and the word search technique. The pattern-matching technique identifies and compares the patterns evident in the data against one or more existing hypothesized patterns (Yin, 1984). In the current study, this involves aligning each dream to one of the 5 dream theories. The word search technique is a technique used to analyze large quantities of dream data on DreamBank.net (Bulkeley, 2009). The word search finds all the dreams recorded on the website containing certain words chosen by the analysts. The word search method has been modified for the purposes of this study, but the idea remains the same. The most important words were those that stood out in the dreams such as the emotions, people, and objects. These words were singled out and their frequencies were quantified *(Table 3).* The last step of the RADaR technique is drafting the final project deliverables (data tables and quotations), which are the most salient pieces of evidence from which the conclusions are drawn.

Results and Discussion

Table 1 includes the descriptive statistics calculated based on the dream diaries of the 13 participants over the course of 14 days. In total, 182 nights were recorded and of those 182 nights, 90 dreams (49%) were recorded. The participants did not recall their dreams for the other 92 days. The time of day the participants filled out the dream diary could have had an impact on the dream recall frequency. Sometimes dreams are forgotten throughout the day and sometimes pieces of the dream are recalled throughout the day. Dream recall between male and female subjects differed, although it was not statistically significant. One female subject did not remember any dreams at all, while the other female subjects tended to have at least five dreams during the 14 nights. It is important to note, however, that this is a trend found in previous research. In my previous study done at Berkeley Carroll regarding gender and dream recall, it was found that female high school students had a significantly higher dream recall than male high school students (p=0.0008). Table 1 also includes the average number of words per dream response of male and female participants. An independent two-tailed t-test determined that the female subjects recorded significantly more words per dream than male subjects (p=0.0082).

Table 1

Descriptive Statistics.

Total # of Nights Recorded	182
Number of Dreams Recalled	90
% of Nights with Recorded Dreams	49
Average # of Dreams per Participant	6.9
Average # of Dreams per Male Participant	5.8
Average # of Dreams per Female Participant	7.6
Average # of Words per Dream Response (male)	32.7
Average # of Words per Dream Response (female)	55.6

Table 2 shows the results of matching one of the already existing dream theories to the dreams recorded in this study. Of the 90 dreams recorded, 41 of the hypotheses aligned with the continuity hypothesis, meaning that they seemed to reflect the events and emotions of the dreamer in their waking life. I was able to come to this conclusion largely based on the dreamer's response to the second question which asked them if they believed there was a significance in their dreams. Many subjects knew exactly what their dreams meant and could connect them to events and emotions they were feeling in their waking lives. It was challenging to link a dream with the continuity hypothesis if the dreamer did not have a suspicion of their dream's meaning, because not enough about the dreamer's personal life was known. Therefore, it is possible that some dreams which aligned with the continuity hypothesis were not identified if the dreamer was not as in touch with what they believed their dreams meant. An example of a dream interpreted as continuity is: *I think I was either playing a video game with my brother or in a video game with him. We were trying to beat a level on the highest difficulty.* When asked what the dreamer perceived to be the meaning behind this dream, the dreamer stated: *I played this video game with my brother last week on the highest difficulty.* The dreamer was able to pinpoint that his dream was a response to an actual event that happened recently. In other cases, the dreamer was able to make more abstract connections. For example, one person described their dream: *Bret threw a party but his*

house looked just like mine. I went into my room to go to the bathroom and Carl followed me. Then he started screaming at me and called me creepy. In response to this dream, the person wrote: *The meaning of this dream meant that I was stressed out by my relationship with Carl.* This dream could potentially fall under the emotional regulation hypothesis as well. In this case, the dreamer did not actually live the scene in their dream, but was still able to identify the dream's emotional significance. Both of these examples align with the main principles of the continuity hypothesis. The first dream was reported by a male subject and the second by a female subject. Female respondents tended to report the emotional and big picture meaning of their dreams, while most male responses said that they did not know the meaning of their dream unless they could match it with a very specific moment.

The 12 dreams *(Table 2)* that fell under the emotional regulation theory could have also possibly been applied to the continuity theory because of high similarities between the two theories. However, for the purpose of this study, a dream was classified as emotional regulation when the dreamer reported strong emotional content and could not connect the dream to an event in their waking life. This was done to distinguish emotional regulation dreams from other dreams (specifically those in the continuity section that did not contain emotions) because there was no way to calculate if the dream actually benefited or impacted the dreamer at all. One person wrote: *It's the first day back from spring break and the teacher gives us a test that I am unaware of. We are in a room in the basement and when I look at the test I can't think. My mind is blank and I start panicking but I manage to fill out a few answers on the first page but it takes forever. The second and third page makes no sense to me. I can't figure out what the questions are asking or the format so I just stare at it freaking out for a while. I clearly am struggling and a senior girl taps me with her pencil and gives me this annoying look like you're doing it wrong and I feel pathetic. I look at her paper slightly and think I understand so I fill out a box and she looks back at me and shakes her head. Then I just have a panic attack until I woke up.* The dreamer then says: *I have had this same dream many times. I always have a dream like this before Chemistry tests.* It is because I get nervous for tests. This dream is not a memory of a specific time, but rather a manifestation of this person's anxiety surrounding tests. The person experiences a version of this dream before every test. The recurrence of correlation between exam stress and the dream supports the emotional regulation hypothesis, which states that emotions are processed during sleep. However, there is no way to know the impact this dream had on the person in their waking life because they did not share that information. Additionally, it may even be unlikely that they would be aware of whether it benefited them (calmed them down) or increased their nerves.

The activation synthesis hypothesis, which claims that dreams do not serve a purpose at all, was supported 21 times *(Table 2).* This theory was applied when the dreamer used phrases in their interpretations such as *the dream was completely random or the dream was way too confusing too understand,* because the dreamer could not identify the purpose of the dream. Many dreams were put in this category due to brevity. These shorter dreams also tended to be less detailed than the dreams that fell under the other hypotheses. Logically, if someone can barely remember a dream and can't find meaning in it, it seems like that dream would likely have no vital meaning. However, it is possible that some dreams were accidentally labelled as activation synthesis. It is possible that some participants did not feel comfortable explaining what they thought their dream really meant. For example, if the dream contained sexual or illegal activities, the participant might have written *I don't know* in response to why they had that dream, instead of confessing private parts about their waking life. It is also likely that some participants did not provide themselves enough time to fill out the dream diary, and instead of thinking through what their dreams meant, just wrote that they didn't know. While the participants were not required to write their names in the journal, strict anonymity when receiving journals was not

enforced by the principal investigator. In future studies, it would be advisable to have a double blind study so the participants feel comfortable being transparent. An example of one of the dreams that fell under activation synthesis was this one: *I think I might have been at a big outdoor concert.* The explanation was: *I don't know because I've never been to a concert.* Another example is: *I had a dream about going to a couch store. I remember something about a grey couch.* The explanation was: *There isn't really a connection to my life. I do have a grey couch.* In both of these dreams, not only can the dreamer not give enough information about the dream for it to have a purpose but the dreamer also cannot think of a valid explanation.

Freud's theory aligned with 8 of the dreams *(Table 2).* The dreams were placed under this category when the dream seemed to have a hidden message and the dreamer did not have an exact explanation for the dream, but they had a suspicion. Freud's theory postulates that dreams uncover emotions and thoughts that are not surface level. For example, one person wrote: *I had a dream that everyone needed to learn some secret code in order to get into people's houses when my friend group hung out. If you did not know the secret code, you could not get in.* Their analysis of the dream was: *I could have had this dream because sometimes my friend group can be kind of exclusive to people who we are not super close with.* The dream was not explicitly a reflection of an event but is likely uncovering guilt of the dreamer for having excluded people in the past. The dream could be the subconscious resurfacing previous events that the person has tried to suppress/forget about. The person does not let us know how this dream made them feel. It could have made them want to reform their ways or maybe they did not think about it at all after writing it down. Another study would benefit from anonymously gathering follow-up information regarding holes in the data so that the conclusions could be more concrete.

The threat simulation theory accounts for 5 of the dreams *(Table 2).* These dreams were the ones that included traumatic and violent events. This theory originated to be about our ancestors' survival, but now dreams under this category are mostly warnings or nightmares. One person wrote: *I also had two other dreams that were so, so scary. One of my close friends was sexually assaulting me and WOULD NOT stop no matter what I did.* The person then said: *I have no idea why I had that dream. I've had issues with this person in the past, but never anything about sexual assault, so I don't know why I dreamt that.* I think it is likely that the dream is not about the person who they experienced the sexual assault from in the dream, but more about a fear of being assualted haunting the person. Another example of a dream that fell under this category said: *I was with my family on a beach in Florida. It was rainy but I went for a swim anyway. Two fish started chasing me. One of them was a barracuda and I don't remember the other. I swam to shore but they started walking on land. After I got away from them, I went back to my family. There was a tidal wave approaching so we went inside. The dream wasn't*

Table 2

Dream Theory Pattern Matching.

Dream Theory Pattern Matching	Number of Dreams Matching the Theory
Continuity Hypothesis	41
Activation Synthesis	21
Emotional Regulation	15
Freud's Theory	8
Threat Simulation	5

scary at all. Interestingly, this person says this dream was not scary to them because in the analysis they say: *usually when I have chasing dreams or tidal wave dreams it's when I'm stressed out but I wasn't stressed at all last night*. The act of a barracuda chasing this dreamer seems to align with threat simulation although the dreamer denies that it was scary, because the dreamer explains that these chasing dreams recur and are normally very stressful.

Table 3 displays the results of the word search across all the dreams. Each of the words or themes commonly mentioned in the dreams were grouped together in five categories located in column 1 of the table. The main categories (characters, culture, social interactions, emotions, and nature) are all categories from the word search in the Bulkeley (2009) study. Under the main categories are the subcategories. The subcategories are exact words or themes mentioned in the dreams. For example, cars, trains, planes, and buses were all categorized under the theme of transportation. If a single dream mentioned transportation more than once, it was still counted as only one reference. This was meant to see how many dreams these themes appeared in, not how many times the individual word appeared in a dream. Most of the words found in the current study were also found in the Bulkeley word search paper.

The most common theme in the characters category was friends, which had 35 references in total. Friends were seen in all dream types alongside several other themes. In the 10 dreams that included drugs, smoking, and alcohol, 7 of those dreams included friends, signaling that the participants may associate drugs, smoking, and alcohol with their friends as opposed to other people. The continuity hypothesis supports the findings that teenagares dream a lot about their friends *(Table 2)*. In my opinion, teenagers are likely surrounded by their friends more than they are surrounded by their family due to the prominence of school and school-related activities. As a result, their dreams reflect images of their friends. Because social situations and relationships are important to adolescents, they are present in their waking minds, and therefore their dreams. Family was the category with the second highest number of references. 21 dreams included family members, whether it was someone's whole family or an individual member. Like friends, family is usually a big part of the lives of teenagers. The other characters (other acquaintances, celebrities, animals, strangers, and zombies) were referenced in dreams less than friends and family. According to the continuity hypothesis, this is logical. For example, there was only one mention of a fantastical being (a zombie). Fantastical beings are not characters that teenagers generally encounter on a daily basis which is why they were reflected in many dreams less frequently than friends.

The culture category includes words and themes that are likely common in the waking lives of Americans in the 21st century. Transportation had the most mentions. Most of the dreams that included transportation referenced cars, specifically Ubers. Trains were also frequently mentioned. These results align with the continuity hypothesis as cars, specifically taxis, are common in the lives of teenagers living in New York City. The train is a form of transportation utilized by many people in New York, which is likely why it was part of participants' dreams. School was mentioned 15 times and 9 of these dreams contained emotions such as stress, anxiety, anger, and loneliness. The continuity theory states that dreams contain the biggest concerns and emotions of the dreamer in their waking lives. The connection between school and negative emotions is rational, as school can induce academic stress and social pressure for many teenagers. Vacation and travel were mentioned by the participants 13 times. Unlike the previous themes, vacation is not always a major part of a person's daily life. The study was done while the participants were in school, not on vacation, though some participants did complete the dream diary the two weeks prior to a break, and some participants completed the dream diary directly following a break. This could explain why travel would be on the waking minds of participants. Travel was most often recorded in context with friends, family, or other acquaintances (classmates). Games, sports, drugs,

alcohol, smoking, parties, and tests were other frequently referenced themes. All of these are common in the lives of many teenagers. College, holidays, and death were referenced less frequently. The participants who mentioned college were seniors in high school. At the time of their participation in the study, they were going through the college process. The other participants were in lower grades, which likely explains why college was not on their minds.

Social interactions included violence, feeling alone or excluded, and love/sexual relationships. Violence manifested in dreams in different ways. There were dreams about sexual assault, guns and other weapons, physical fighting, and violence through the expression of anger. In these dreams, violence can be linked to fear and anger. The dreams referenced by female participants regarding violence often included fear, while those of male participants often included anger. I believe that violence is experienced and seen by different people in different ways, which could explain why there were many different emotional responses and manifestations of it in the dreams. In 6 dreams, the participants reported either being completely alone in a space or being excluded. The dreams in which the dreamer was completely alone were located in school. The dreams about exclusion were in context with friends. These dreams appear to exhibit the social and school-related anxieties of the dreamer. Love and sexual relationships were also mentioned 6 times. Relationships, love, and sexuality are topics that teenagers often explore and think about which may account for these dreams.

Many emotions were present in the dreams. Nervousness and stress were the most common emotions, with 7 mentions. 5 of the dreams about stress were linked to school, and the others were linked to relationships with parents/friends. I believe it is common for teenagers to be stressed about those things. Happiness was mentioned 6 times in different contexts. One person explained the happiness they felt in a dream regarding being with friends from camp, one person shared a dream of just her and her friend laughing, and one person shared a dream about feeling content with the high school they are attending. Dreams about being late did not specifically mention that they were stressed. However, these dreams could also potentially fall under the stress subcategory because in the dreams, the dreamers appear stressed. The other subcategories that fall under emotion (sadness, fear, failure, performing, yelling, confusion, and being chased) only had 1 or 2 references. These could be feelings less common in the lives of the participants.

There were not many references to nature. Summer and weather (rain and tornadoes) were each mentioned 3 times. Fire was mentioned once. Falling dreams, categorized as nature in the Bulkeley paper because "they provide literal and possibly metaphorical references to gravitational forces in nature" were only mentioned once, but are a phenomenon commonly studied in dreams (Bulkeley, 2009). Nature is certainly present in the lives of participants, but may not occupy much thought in the same way friends, family, and school might.

Column 3 of table 3 shows the ratio of male participants who mentioned a certain word/theme. Column 4 shows the ratio of female participants who mentioned a certain word/theme. There are some themes that do not seem to be associated with a specific gender. Friends, family, school, games/sports, being alone/excluded, happiness, tests, and stress, were themes that both male and female subjects reported. However, celebrities, transportation, drugs/smoking/alcohol, parties, love/sexual relationships, were more commonly seen in the female dreams. Drugs/smoking/alcohol, parties, and love/sexual relationships, were often seen in the same dreams of the female participants. The highest ratios for males were friends, family, school, sports/games, and being alone/excluded. The continuity theory may be able to explain the differences in the ratios based on gender. It is possible that the themes that didn't differ based on gender are more universal, while the categories that have the biggest gender difference represent different concerns/priorities of male and female teenagers in their waking lives. The

Table 3

Results of the Word/Theme Search and the Gender Breakdown.

Categories and Sub-categories	Frequency (# of Mentions)	Ratio of Male (n=5) Participants who Mentioned the Word/Theme	Ratio of Female (n=8) Participants who Mentioned the Word/Theme
Characters			
Friends	35	3/5	7/8
Family	21	3/5	7/8
Celebrities	5	0/5	3/8
Animals	5	1/5	3/8
Other Acquaintances	6	0/5	5/8
Stranger	5	2/5	3/8
Fantastic Beings	1	0/5	1/8
Culture			
Transportation	18	1/5	6/8
School	15	3/5	7/8
Vacation/Travel	13	2/5	6/8
Games/Sports	11	4/5	4/8
Party	10	0/5	3/8
Drugs/Smoking/Alcohol	10	0/5	4/8
Tests	6	2/5	4/8
College	3	1/5	2/8
Holidays	2	2/5	0/8
Death	1	1/5	0/8
Social Interactions			
Violence	8	2/5	4/8
Being Alone/Excluded	6	3/5	3/8
Love/Sexual Relationships	6	0/5	4/8
Emotion			
Nerves/Stress	7	2/5	4/8
Happiness	6	3/5	2/8
Being Late	3	0/5	2/8
Sadness	2	1/5	1/8
Fear	2	0/5	1/8
Failure	2	1/5	1/8
Performing	2	1/5	1/8
Yelling	2	1/5	1/8
Confusion	1	0/5	1/8
Being Chased	1	1/5	0/8
Nature			
Summer	3	1/5	2/8
Weather	3	0/5	2/8
Fire	1	0/5	1/8
Falling	1	1/5	0/8

statistically significant difference in average number of words per dream between male (32.7 words) and female (55.6 words) participants must also be taken into consideration (p-value=0.0082). The differences in gender could be due to the fact that the females were more in depth in their responses. Men could have experienced love/sexual relationships in their dreams, but they didn't include it. The sample size is another limitation of this study, because there were only 5 males and 8 females. It is possible that with more participants, these ratios would be different.

Despite limitations, the study concludes that the continuity hypothesis best explains the significance of dreaming of the 13 high school participants. This was concluded as the dreamers were often able to connect their dreams to specific events and emotions in their waking lives and the word search technique found that the most recurrent themes (friends, family, school, etc) in the dreams, are themes that are highly likely to play major roles in the lives of teenagers.

References

Aserinsky, E., & Kleitman, N. (1953). Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. *Science, 118*(3062), 273-274. doi: 10.1126/science.118.3062.273

Blake, H. & Gerard, W. (1937). Brain potentials during sleep. *American Physiological Society, 199*(4), 692-703.

Bulkeley, K. (2009). Seeking patterns in dream content: A systematic approach to word searches. *Consciousness and Cognition, 18*(4), 905-916.

Gerring, J. (2004). What is a case study and what is it good for? *American Political Science Review, 98*(2), 341-354.

Hobson, A. & R. McCarley. (1977). The brain as a dream state generator: An activation-synthesis hypothesis. *American Journal of Psychiatry, 134,* 1335-1348.

McNamara, P., Harris , E., & Kookoolis, A. (2007). Costly signaling theory of dream recall and dream sharing. *The New Science of Dreaming.*

Merced, M. (2012). Dreaming: Physiological sources, biological functions, psychological implications. *The Journal of Mind and Behavior, 33*(3/4), 173-193.

Pesant, N., & Zadra, A. (2005). Dream content and psychological well-being: A longitudinal study of the continuity hypothesis. *Journal of Clinical Psychology, 62*(1), 111-121.

Solms, M. Dreaming and REM sleep are controlled by different brain mechanisms." *Sleep and Dreaming,* pp. 51-58.

Strachey, J. (1953). The standard edition of the complete psychological works of Sigmund Freud, volume IV (1900): The interpretation of dreams (first part). i-xiii. *The Hogarth Press and the Institute of Psycho-analysis, London.*

Stickgold, R., Hobson, J., Fosse, R., & Fosse, M. (2001). Sleep, learning, and dreams: Off-line memory reprocessing. *Science, 294*(5544), 1052-1057.

Wagner, U., Gais, S., & Born , J. (2001). Emotional memory formation is enhanced across sleep intervals with high amounts of rapid eye movement sleep. *Cold Spring Harbor Laboratory Press, 8,* 112–119. doi: 10.1101/lm.36801

Watkins , D. C. (2017). Rapid and rigorous qualitative data analysis: The "RADaR" technique for applied research. *International Journal of Qualitative Methods, 16,* 1-9.

Yin, R.K. (1984), Case study research. Design and Methods, Sage, Thousand Oaks, CA.

About Science Research and Design (SRD)

In this selective three-year program, 10th-12th grade students conduct original scientific research and become experts in a field of study, writing their own essential questions and gaining a deep understanding of the dynamic, evolving nature of science. The goal of the program is for students to experience scientific research as scientists do. Students take this course in addition to their other science classes.

First Year: Introduction to Science Research and Design – Students read a variety of papers and articles as they learn how to dissect and understand scientific writing and expose themselves to many different fields of study. They then take part in a scaffolded research study on a prescribed topic as they learn the nuts and bolts of scientific research. This research study includes a Spring Intensive course in which students master scientific research methods and learn how to design detailed iterative procedures and record experimental results. By the end of their first year, they are able to read, understand, and explain journal level articles on one or two fields of study, and they should have a sense for the requirements of a complete scientific study.

Second Year: Advanced Science Research and Design – Students further hone their research skills by first verifying a published research study before designing and conducting their own independent project. Throughout the year, students continue to familiarize themselves with background literature, adjust their procedure as necessitated by their experimental observations, and record and analyze their results. By the end of their second year, they have completed several rounds of their own long-term study and they propose how that study could be modified and redone. They defend their findings in an end of year oral exam

Third Year: Symposium – Students finalize their individual research and learn how to statistically analyze their findings. They also refine their presentation skills before presenting their results in a variety of settings including an end of year Symposium.

the BerkeleyCarroll School
181 Lincoln Place
Brooklyn, NY 11217
www.berkeleycarroll.org

"We look with favor on all forms of learning, but with particular grace we encourage philosophical studies, especially those which by actual experiments attempt either to shape out a new philosophy or to perfect the old."

KING CHARLES

from the 1661
Charter for the formation
of the Royal Academy of Science;
the proceedings of which
are the oldest journal in existence