

Proceedings of the Berkeley Carroll



INDEPENDENT RESEARCH CONFERENCE

SCIENCE RESEARCH PARTICIPANTS

Seniors

Isabel Baum Eve Blank Max Bonthrone Savannah Der Maddie Eve Noah Finkelstein Eugenie Haring Matthew Justh Ella Novogratz Graham Stodolski Zach Tegtmeier Alayna Thomas Diego Turturro Charlie Welch Brendan Zelikman Juniors

Zachary Ashen Kirill Bektashev Luca Broder Ayden Cherry Phoebe Degn Lucas Diamond Leah Khovov Donald Minerva Zoe Rabinowitz Isabella Solimine Sophia Sondey Hope Swetow Justin Walker

Sophomores

Alma-Lilas Anzalone Shreva Balaji Jonathan Barley Lila Blank Antonio Domonkos Caleb Dunn Luke Ehrenfreund Kai Ellis **Nichole Fernandez** Lucy Graham **Charlotte Hampton** Lucija Jurevics Harry Justice **Michael Khoury Christopher Kretschman** Desiree Naini Ukmar Gabi Prieto Elliott Radin Natasha Rasnick Sade Robinson Adeleen Rosner **Daniel Safian** Margaret Schaffzin Ava Sonies William Swartz Eden Weissmann Andreas Welch

Facilitators

Danny Gentile 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 Upper School Robert Vitalo Head of School

Welcome from the Editor

n the 1760's, scientists and explorers undertook synchronized voyages throughout the world to observe the transit of Venus and thereby determine the distance from Earth to the Sun. The undertaking was vast and exhaustive, and helped launch the careers of explorers like James Cook, Jeremiah Mason, and Charles Dixon.

This worldwide experiment – conducted right in the middle of the Age of Enlightenment – exemplifies the precept that scientific inquiry mirrors the population and the times in which it is conducted.

In other words, if you want to know what our community has been thinking about, look at the questions our researchers are asking.

So, what's been on our minds at Berkeley Carroll recently?

Some of our young scientists have been working on the big issues facing our country and the world at large. Max Bonthrone, for example, built and tested several wind turbines, Isabel Baum investigated cognitive gender-based bias, and Maddie Eve looked for the underlying cause for academic dishonesty.

Others were inspired by the research equipment in our institution. Savannah Der used a PCR machine to look for fast food fraud, and Diego Turturro mixed concrete to see what would happen when you heat it while varying pore space.

The majority of the experimental questions this year have been about communication.

Some of this communication has been non-human. Eve Blank, for example, studied the mycorrhizal networks through which stressed plants send signals to one another, Eugenie Haring carefully monitored how zebrafish behaviorally communicate, and Ella Novogratz spent the year looking for evidence that animals reduce stress in humans.

Matthew Justh looked inward to see if mindfulness (communication with the self, if you will) improved test results, while Zach Tegtmeier investigated if it is possible to communicate with your own future self.

Charlie Welch, on the other hand, looked outward to determine the consequences of deceptive business practices of websites telling their subscribers that they have figured out Daily Fantasy Sports and the stock market. Similarly, Noah Finkelstein evaluated the effectiveness of attack ads in Senatorial campaigns.

A bit closer to home, Brendan Zelikman and Alayna Thomas both looked at communication in the classroom. Brendan asked whether teachers unwittingly produce false memories during lessons and Alayna examined how race influences student-teacher relationships. Meanwhile, Graham Stodolski's homemade robot Elsa attempted with some success to teach her own Swedish lessons.

Clearly, we've been thinking about communication this year at Berkeley Carroll. That's as it should be; we live in a time when it's increasingly difficult to see eye to eye with one another. Perhaps we can take inspiration from the attempts made by the fish, trees, and robots that you'll read about in the pages of this journal.

Until next year, Scott Rubin Upper School Science Chair

Science Research and Design SENIOR RESEARCHERS



Isabel Baum '19

For the past three years, Isabel has been exploring the field of implicit cognitive bias; specifically, the bias of gender leadership. During her senior year, she further narrowed her area of

study to exploring the bias behind authoritarian behaviors in both male and female teachers, respectively in the field of science, technology, engineering, and math (STEM). In the future, Isabel hopes to continue her study of authoritarian behavior by expanding her experiment to the humanities field. Additionally, she hopes to continue studying cognitive bias in college and to participate in campus research.



Eve Blank '19

Eve Blank will be presenting on the impact of stress on the likelihood of the formation of mycorrhizal networks. She has been interested in interplant communication since 10th grade when

she began researching mycorrhizal networks and their ability to transfer resources between plants in times of stress. Due to the extensive body of literature on the ways these networks function, Eve decided to test the way stress affects the formation of networks, a question applicable to the subjects of forest regrowth and drought. In the future, she hopes to be able to directly test for the formation of networks to support her findings in this study.



Maxwell Bonthrone '19

Max Bonthrone's interest in the environment is founded upon creating a better, cleaner, and safer world. His study focuses on the orientation of small-scale-wind turbines in urban

environments. He spent his junior year constructing anemometers which measure wind speed in order to determine where to place his wind turbines on the Berkeley Carroll roof.



Savannah Der '19

Savannah has always been interested in biotechnology as well as food and how it is sold to consumers. With her study, she combined these interests by testing the claims of food products

using PCR and sequencing methods. She began by investigating food fraud in fast food chains and has conducted a follow up study on White Tuna. She hopes to continue using biotechnology as a means of exploring topics relevant to our society.



Maddie Eve '19

Maddie has been studying the motives behind lying in academic and social contexts, inspired by an article that looked into the permanent effects of frequent lying on the

chemical makeup of the brain. In her study, Maddie investigated specific motivations Berkeley Carroll high school students find acceptable when telling a lie or when committing an act of academic dishonesty. Maddie would like to continue investigating different situations in which lying is commonly accepted and how that acceptance changes with age or between different genders.



Noah Finkelstein '19

Noah Finkelstein will be presenting on the science of smear campaigns, specifically how attack ads influence polling data in U.S. congressional races. Noah is fascinated by the psy-

chological phenomenon of the halo effect, a cognitive bias in which people attribute a specific quality to an individual based on that individual's non-specific essence. As a junior, Noah investigated how the halo effect influences the perception of foreign countries by asking high school students to evaluate specific attributes of a nation based solely on photographs. As a senior, Noah is researching how effectively attack ads transferred a negative halo effect to candidates for the U.S. Senate in the 2016 midterm elections. In the future, Noah intends to expand his senior study to include more Senate races.



Eugenie Haring '19

Eugenie Haring will be presenting her research on the shoaling behavior of wildtype zebrafish. Eugenie's interest in marine biology and the effect of climate change on our oceans led her

to read several studies involving fish. She was drawn to one in particular on the behavior of physically different zebrafish of the same species. Inspired by this study, Eugenie set out to devise a strategy for training wildtype zebrafish to overcome imprinted social preferences in favor genetically modified green fluorescent protein zebrafish using a food reward. Eugenie has really enjoyed her time in the Science Research and Design program at Berkeley Carroll and she intends to pursue scientific research in college but will be taking a break from studies involving zebrafish.



Matthew Justh '19

Matthew will be the first to tell you he likes to find ways to work smarter in school. From early on in his time in the science research program, he has wondered about how certain actions

affect one's problem solving ability. After reading the book *Why We Sleep* by Matthew Walker, Matthew decided he wanted to focus his study on sleep and problem solving. However, after running into limitations such as his inability to control his subjects' sleep habits, Matthew recalled learning the benefits of spending 5 minutes mindfully meditating before taking tests in science class. His current study looks to see if there is an effect on problem solving directly after spending time mindfully meditating. He hopes to expand his study to take a broader look at the lasting effects of habitual mindfulness.



Ella Novogratz '19

For the entirety of her high school experience, Ella Novogratz has had a deep interest in both psychology and animals. In an attempt to combine

these interests, she has read many articles regarding the effects of animals on stress and mental health. Noticing that the majority of these papers were focused on dogs specifically, Ella decided to expand her own investigation to other domesticated animals. After many sessions working closely with students and animals, Ella will be presenting her research on the impact of domesticated animal presence on the stress and performance of high school students. Ella hopes to continue her exploration of the link between animals and psychology in the future.



Graham Stodolski '19

Despite originally being drawn to the impacts of music on the brain, Graham found an interest in educational robotics after attending a lecture at the Yale Young Global

Scholars program. He has spent the majority of the last two years in the Beta Lab constructing his robot, Elsa, who uses speech recognition and eSpeak textto-speech to teach Swedish. His research compares Elsa's teaching with his own and attempts to determine if there is a difference between the two. He hopes to continue working with robotics in the future as he studies engineering in college.



Zach Tegtmeier '19

Zach has been interested in understanding and measuring Future Self Continuity (FSC), the connection and empathy one feels with their future selves, ever since being introduced to

the idea in a PSAT two years ago. Past studies have shown that FSC can directly affect how much people are inclined to do things that will affect them in the future such as studying or saving money. Zach became interested in whether conscious FSC matches subconscious FSC. He spent the majority of his junior year researching the topic, programming an Implicit Association Test (IAT) to measure subconscious FSC with Mr. Byrne, and running a preliminary experiment on 10th grade SRD students using a questionnaire and his IAT. His senior year, he debugged his IAT, expanded his subject pool, and filtered his data. Zach hopes to continue his research to directly test behavior of students against FSC using the IAT.



Alayna Thomas '19

Alayna's interests in psychology and the inequalities of the education system led her to study teacher bias in the classroom. She started off looking at the intersection of race and gender,

but now she is investigating whether teachers treat students of the same racial identity differently than students of a different racial identity. Her study comprises a classroom observation and a survey for both teachers and students. In the future, Alayna wants to continue to study this subject and possibly look more at the origins of racial bias.



Diego Turturro '19

Since 10th grade, Diego has had an obsession with fire, and it has led him to an interest in heat. Today, Diego will present his findings on the heat resistance of concrete when the pore

space is varied. He is comparing blocks with no pores to blocks with man-made pores, and he has been working closely with mentor Dr. Gentile to heat the blocks in a controlled environment. He has spent most of the past 2 years perfecting how to make a perfect block. In the future, Diego will continue to explore the real life ramifications of his study.



Charlie Welch '19

Charlie Welch will be presenting his research on the emerging industry of Daily Fantasy Sports. As someone who has always been interested in sports statistics, Charlie worked

closely with Mr. Rubin during 11th grade to design an experiment that was tailored to this interest. He has compared the performance of third-party advice sites in this industry based on whether the site was paid or free, but has since expanded his investigation to other industries including the stock market, and he hopes to continue researching this comparison in different fields to help better inform customers.



Brendan Zelikman '19

Since the 10th grade, Brendan Zelikman has been keenly interested in the topic of false memories. During his time in the Science Research and Design program, he has been at-

tempting to determine how prominent false recall is in the classroom setting. As a junior, Brendan worked closely with mentor Dr. John Mariani to adapt the Deese-Roediger-McDermott paradigm to investigate whether word lists related to stressful classes led to increased false memories in students. As a senior, he worked closely with BC faculty to observe whether false memories can be created within classes themselves. He hopes to continueresearching and to determine the full extent of false memories within classes and schools as a whole.

Table of Contents

Cognitive Bias of Teacher Suitability Between Male and Female Teachers	7
by Isabel Baum '19	
Comparing the Efficiency of Vertical and Horizontal Axis Wind Turbines	15
by Max Bonthrone '19	
The Impact of Stress on the Likelihood and Composition of Mycorrhizal Network Formation	23
by Eve Blank '19	
Testing the Claims of Fast Food Chains Using PCR and Sequencing Methods	35
by Savannah Der '19	
Academic Dishonesty at Berkeley Carroll	51
The Effect of Attack Ads on Political Campaigns	62
Shoaling Preferences of Wildtype Zebrafish Trained to Associate with GFP Zebrafish	67
by Eugenie Haring '19	
The Effect of Mindfulness on Problem Solving Ability in High School Students	74
by Matthew Justh '19	
An Investigation of the Effects of Animal Presence on the Performance and Stress Levels of High School Students at Berkeley Carroll	
by Ella Novogratz '19	

Table of Contents (cont'd)

The Feasibility and Effectiveness of	
Educational Robotics: Comparing the Swedish Teaching	
Abilities of a Robot and Human	88
by Graham Stodolski '19	
Conscious and Subconscious Future Self Continuity	
in High School Students	97
by Zachary James Tegtmeier '19	
Race Matching Between Teachers	
and Students: Who Do Teachers Treat Differently?	107
by Alayna Thomas '19	
Effects of Pore Space on the Structural Integrity of Concrete	118
by Diego Turturro '19	
Daily Fantasy Sports and the Stock Market:	
An Investigation Into the Deception of Online Advice	129
by Charlie Welch '19	
Testing Alternative Presentations	
of Word Lists with the DRM Paradigm	141
by Brendan Zelikman '19	

Cognitive Bias of Teacher Suitability Between Male and Female Teachers

by Isabel Baum

Abstract

omen with the same credentials and work experience as men do not receive as many job opportunities or leadership positions as men. This is especially evident for women in the field of science, technology, engineering, and math, as there is a strong bias formed against them. Two math teachers, a male and a female, from the Berkeley Carroll School were recorded while teaching the same eight math problems: two Geometry problems, two Algebra II problems, two Precalculus / Trigonometry problems, and two Calculus problems. Twenty subjects watched a video of one of the two teachers teaching two math problems, based on the level of math they were placed in at their high school. Subjects then filled out a survey regarding teacher suitability. There was a statistically significant difference between how subjects perceived the two teachers (p=0.000457). The male math teacher was perceived as having greater suitability than the female math teacher.

Background

It has been well studied that women with the same credentials and work experience as men are not getting as many jobs in leadership positions as men are (Jackson-, Hillard, A. L., & Schneider, T. R., 2014). It is often found that when a company, business, or work force is presented with multiple equivalent resumes or vignettes from both men and women, the job is more likely to go to the male applicant (Jackson et al., 2014). The suggested reasoning behind this is cognitive bias (Jackson, S. M., Hillard, A. L., & Schneider, T. R., 2014). Cognitive bias can best be defined as "a mistake in reasoning, evaluating, remembering, or other cognitive process, often occurring as a result of holding onto one's preferences and beliefs regardless of contrary information" (Chegg, 2003-2019). In other words, a preconceived notion, your peers, or your surrounding environment are all factors that contribute to decisions you make. Therefore, you might not be making a true, unbiased decision in general.

It has been discovered that there is a greater cognitive bias around women in the Science, Technology, Engineering, and Math (STEM) fields (Jackson, S. M., Hillard, A. L., & Schneider, T. R., 2014). Our society as a whole already has a preset idea that men, rather than women, should be dominating the STEM fields. A general implicit negative attitude is expected towards women in the STEM fields, whereas a positive attitude is expected towards men in the same STEM fields (Jackson, S. M., Hillard, A. L., & Schneider, T. R., 2014). In the scientific paper *"Using implicit bias training to improve attitudes toward women in STEM"*, the authors note that through creating an artificial environment where subjects learn about bias before reviewing hypothetical applicants, the subjects are more likely to decrease their cognitive bias to a degree, further allowing the consideration of women over men in STEM fields (Jackson, S. M., Hillard, A. L., & Schneider, T. R., 2014).

Not only does work experience, ability, and gender contribute to the hiring of new applicants in an intended field, but behavior is a large component as well. A paper entitled *"What mediates gender bias in work behavior ratings"* explores the concept of how behavior relates to cognitive bias. Again, it is not only a question of if men are more suitable for a certain leadership position, title, or job, but if they also have the behavior to comply to the new role. (Jackson, S. M., Hillard, A. L., & Schneider, T. R., 2014). Leaders are expected to have an authoritarian behavior; to make decisions and take control (Jackson, S. M., Hillard, A. L., & Schneider, T. R., 2014). Some may perceive this behavior as outgoing, violent, dominant and / or aggressive. These are common behaviors that are typically associated with a male/masculine role rather than a female/feminine role (Martell, 1996). This is a cognitive bias: people believe men have a more authoritarian persona. This study has helped to further support the idea that men are viewed as having better leadership behaviors than women (Martell, 1996).

When considering an applicant for a job position, one needs to consider the context of the applicant's resume or vignette, where they are now, and where they will be in the future if given the position. In other words, when hiring a new member, it is important to take into account their future behaviors and if they are fit for the role for which they are applying (Hogue, 2016). The study *"Gender bias in communal leadership: Examining servant leadership"* focuses on the cognitive bias around future leadership behaviors. It has been shown through artificial and non-artificial environments that men are viewed as more capable and can be seen as taking on a more commanding leadership position (Hogue, 2016). After distributing a cover letter from a new potential boss to twenty subjects, the subjects had to fill out a survey to prove they had sufficiently read the letter. Afterwards, they filled out an additional survey regarding how they thought the potential boss would engage in future behaviors (1-5 scale survey) (Hogue, 2016). It was shown that there was a set notion that men are a more demanding force and more likely to be respected and listened to if placed in a high status job position. On the other hand, women were not associated with an authoritarian behavior pattern (Hogue, 2016). In other words, there is already a set idea that women are not viewed as powerful leaders, and men are viewed as powerful leaders (Hogue, 2016).

This current study tests the cognitive bias behind the decision making of women in leadership positions. It is useful to have an understanding of what the preconceived capabilities and behaviors of men are, as it is also useful to have an understanding of what the preconceived capabilities and behaviors of women are. It is interesting to compare and contrast the two within the greater scientific world because it gives us a better understanding of the bias placed around leadership positions within a job, business, or field. This experiment consisted of subjects watching either a male or female math teacher teach a few problems, and then filling out a survey regarding teacher suitability.

Procedure

Both math teachers (male and female) used in this experiment had an equivalent level of capability to take on leadership positions within their field of being a high school math teacher. This study consisted of several parts. Both teachers had similar demeanor and teaching style, to prevent additional bias. The second part was having subjects sit through one of the teacher's lessons of a math problem, then filling out a survey regarding teacher suitability and authoritarian behavior. Lastly, the third part was evaluating how the results of each survey compared to the gender of each applicant.

The goal of this experiment was to determine if there is an authoritarian cognitive bias of females versus males in leadership positions. The independent variable was the gender of the teacher giving the lesson (male, female). The dependent variable of this study was the teacher suitability (based on the results from a 1-5 scale survey). The control variable of this experiment was the teaching style and the problems the teachers taught the class through a projected video. Both teachers had the same eight problems written out on a whiteboard. They both taught by lecturing into the camera and demonstrating what they were saying by writing on a whiteboard. Subjects were recruited through an email that was sent out to the freshman, sophomore, junior, and senior grades at the Berkeley Carroll School asking for willing participants.

There were 20 subjects in total that participated in the experiment. Ten of the subjects individually watched the video of the male teacher teaching a set of two math problems, and the other ten subjects independently watched the video of the female teacher teaching a set of two math problems. Prior to conducting the experiment with subjects, one male teacher, from the Berkeley Carroll School, stood in front of a camera that recorded him teaching two Geometry problems, two Trigonometry / Algebra II problems, two Precalculus problems, and two Calculus problems. This procedure was repeated with the female teacher, also from the Berkeley Carroll School, using the same math problems. Subjects were distributed randomly to eliminate any possible influence on data. This distribution was done by alternating which subjects watched the male video and which subjects watched the female video, splitting the group in half. Subjects watched two math problems being taught based on the class level they were placed in at the Berkeley Carroll School. After watching the video, subjects filled out a survey (a short, 1-5 scale survey) which asked questions regarding teacher suitability. Data was collected from the survey, placed into a spreadsheet, and then analyzed to determine if there was a statistically significant difference between the suitability of the male teacher and the suitability of the female teacher.

Survey Questions

Question I: Do you think this teacher is a suitable math teacher? **Question II:** In the future, do you think this teacher will be successful in this position? **Question III:** How much would you like to be taught by this teacher for a full academic year? **Question IV:** Are you able to retain the lesson?

lmage 1

Survey Subjects Received after Watching the Math Videos

ole math teache	r? •
3 4 5	They are suitable
	Will be very successful
ght by this teac	her for a full academic year? *
3 4 5	I would like it a lot
3 4 5	I can solve another problem similar to the ones I learne
	3 4 5 3 4 5 3 4 5 3 4 5 3 4 5 3 4 5 3 4 5 3 4 5 3 4 5 3 4 5 3 4 5 3 4 5 3 4 5

lmage 2

Math Problems Taught in the Video Subjects Watched

	SRD Experiment- Math Problems
	Geometry .
0	1) The neurollale areas shown in the figure halow has a neurostar
	1) The parallelogram shown in the figure below has a perimeter
	of 44cm and an area of 64cm ² . Find angle I in degrees.
	3x + 2
	<u>λ</u> τ
	5x + 4
	2) Triangle DEF is similar to triangle PQR. Solve for $k.$
	E
	12
	24 26 R 26
	Algebra II
	1) Simplify using rational exponents.
\bigcirc	<u>(x^{1/2} y)²</u>
	X ''/2 Y 5/4
	2) Divide using synthetic division.
	$(2x^4 + 2x^3 - 10x - 9) \div (x - 2)$
	1) Find a quadratic function f(x) which has 5 and 1 as the
	x-intercepts of its graph and which has a minimum value of -12.
	2) Graph the following equation:
	$(x-2)^2 + (y+3)^2 = 16$
	Calculus
	1) $f(x) = 16 x^2$
	Find the exact rate of change at t=2 sec.
	2) How rapidly will the fluid inside a vertical cylindrical tank drop
	if we pump the fluid out at the rate of 3,000 liters/min.

 \odot

Data Analysis

Table 1

Average Survey Responses from Subjects (1-5 scale)

(1 = negative response, 5 = positive response) Total subjects: 20 Group I (male teacher) number of subjects: 10 Group II (female teacher) number of subjects: 10

	Group 1 Average	Group II Average
Question I	4.8	4.3
Question II	4.8	4.0
Question III	4.8	3.7
Question IV	4.6	4.4

Table 2

Adjectives Used to Describe Teachers

If a subject checked off that they have already been taught by this teacher, they were asked to list some adjective they would use to describe them. These adjectives were scored on a scale of -3 to 3 (-3 being the worst teacher, 0 being average, 3 being the best teacher), and all the numbers were then added up and compiled to give the overall adjective score.

*Sample adjectives are shown below

Gender	Male	Female
Adjectives	Instructional Comprehensive Engaging	Cheerful Average Fine
Total Score	34 → average: 1.4	$6 \rightarrow average: 0.857$

Table 3

Subject Gender Insight

SM: subject male SF: subject female TM: teacher male TF: teacher female

This is used to analyze the gender of the subject pool and compare which gendered subjects watched which gendered teachers in the video. This gives insight as to how subject gender is controlled.

SM TM	SM TF	SF TM	SF TF	Total	
4	6	6	4	20	

While compiling the data from the survey, it became evident that the male teacher received more positive feedback than the female teacher. Subjects found the male math teacher to have greater teacher suitability than the female teacher. Table 1 demonstrates the average rating the subjects gave each math teacher, based on if they watched the lesson taught by the male teacher or if they watched the lesson taught by the female teacher. The average response for each survey question is recorded in the table. By way of VassarStats, an independent, two sample, two-tailed, t-test was run to determine the statistical significance of the data in Table 1. There was a statistically significant difference between the two group averages, with subjects showing a bias towards the male math teacher (p=0.000457). This provides evidence that there is a statistically significant bias towards the male math teacher. At the end of the survey, there was a section that allowed subjects who had been previously taught by the math teacher they watched in the video to list adjectives they would use to describe the teacher based on prior experiences with them, as shown in Table 2. Sample adjectives are given in Table 2. Additionally, each adjective was ranked, using an internal system, on a scale from -3 to 3 (-3 being the worst teacher, 0 being average, 3 being the best teacher). Sample rankings are indicated next to each adjective in red. An independent, two sample, two-tailed, t-test was run, resulting in p=0.293445. Therefore, the results of the adjectives do not suggest a bias. *Table 3* clarifies the gender relationship between the subjects and the teachers. In total, there were 20 subjects. This table breaks down the gender of each subject and compares it to the gender of the teacher they watched in the math lesson video. This shows how the gender of the subject pool was split approximately evenly.

Discussion

Factors that were not able to be controlled while running the experiment were the mannerisms and physical appearances of the male and female teachers. Although both teachers taught the same math problems by a combination of lecture and demonstration on a whiteboard, there were distinguishable differences. Subjects might have perceived differences in voice tone, charisma, or pace. Additionally, the use of colored markers when writing on the whiteboard was a factor that was not controlled for. Both teachers used a combination of the black marker distributed to them and additional colored markers. Lastly, the ethnicity of the two teachers differed. These are all factors that could be a contribution to the bias of either teacher and the data results.

Going forward, I would like to expand my subject pool. I feel that it will be beneficial to my study to recruit as many subjects as possible, resulting in more accurate data and conclusions. I would also control for the race of the two teachers by using teachers that identify with the same ethnicity. In addition, I would like to run a related side experiment, which would solely include juniors at the Berkeley Carroll School. American Studies is a mandatory humanities class for all juniors at the high school to take. There are four American studies sections, each led by one female teacher and one male teacher. I would like to send out a survey regarding the authoritarian behavior of the two American Studies teachers in each of the four sections. I would then compile this data and determine if there is a bias and a statistical significance. Lastly, I will use this additional data and compare and contrast it to my current data.

References

"Cognitive Bias Lesson." Chegg, www.chegg.com/homework-help/definitions/cognitive-bias-13.

- Hogue, M. (2016). Gender bias in communal leadership: Examining servant leadership. Journal of Managerial Psychology, 31(4), 837-849.
- Jackson, S. M., Hillard, A. L., & Schneider, T. R. (2014). Using implicit bias training to improve attitudes toward women in STEM. Social Psychology of Education : An International Journal, 17(3), 419-438.

Martell, R. F. (1996). What mediates gender bias in work behavior ratings? Sex Roles, 35(3-4), 153-169.

Comparing the Efficiency of Vertical and Horizontal Axis Wind Turbines

by Max Bonthrone

Abstract

ertical wind turbines have the potential to generate power for small scale systems, making them extremely useful in urban environments due to the potential to connect directly (both physically and electrically) to a residence or apartment building. In this study, I determined which type of turbine (vertical or horizontal) performs better under controlled continuous conditions, and where on Berkeley Carroll's Lincoln Place campus roof I should put my two types of turbines. In the end, horizontal wind turbines turned out to be superior to vertical turbines under controlled conditions, and anemometer B (the southwest side of the roof of Berkeley Carroll) indicated a higher average wind speed than anemometer A.

Background

Our Earth is currently being polluted by greenhouse gases, most notably carbon dioxide (Green et al., 2014). These greenhouse gases come from the combustion of fossil fuels: natural gas, coal, and oil (Energy Sources: Fossil). 60% of said released carbon dioxide floats around in the atmosphere rather than being reabsorbed by Earth's plants through photosynthesis (Green et al., 2014). The heat that arrives here from the sun gets trapped in the Earth's atmosphere by the carbon dioxide and other greenhouse gases (Green et al., 2014). In the future, if we continue to overload Earth's atmosphere with carbon dioxide, we will form an inhospitable environment for humans and other living things on Earth (Green et al., 2014). Therefore, finding alternative sources of energy that don't harm the environment through the release of greenhouse gases is important to future generations and their survival on Earth.

Renewable energy sources (or "renewables" for short) are safe and clean alternatives that should be used more often to prevent the creation of an inhospitable Earth. These renewables make use of natural sources of energy which don't emit greenhouse gases upon combustion. In fact, most of them don't even require combustion, unlike their fossil fuel counterparts (Energy Sources: Clean Energy). The three main options of renewables that are most popular are solar power (obtained through solar panels), hydroelectric power (harvested from waterwheels), and wind power (harvested through the wind turbines) (Energy Sources: Clean Energy). According to the United States of America Energy Information Association (EIA), wind power is the third most generated renewable energy source right under hydroelectric power (Figure 1) (USA EIA, 2018).

U.S. Renewables Summary

The U.S. Renewables summary table gives an overview of the amount of renewable energy generated in the listed years. In addition, it also projects the amount of energy generated, measured in British Thermal Units, in 2018 and projected for 2019. A Btu is equivalent to 1055.6 Joules of energy (USA EIA, 2018).

U.S. Renewables & CO ₂ Emisions Summary								
2016 2017 2018 2019								
U.S. Renewables Consumption	(quadrill	ion Btu)						
Geothermal	0.210	0.211	0.220	0.229				
Hydropower ³	2.472	2.770	2.576	2.523				
Solar	0.569	0.774	0.964	1.094				
Waste Biomass	0.503	0.482	0.489	0.490				
Wind	2.094	2.345	2.525	2.671				
Wood Biomass	2.131	2.145	2.189	2.162				
Electricity Subtotal ^b	7.960	8.704	8.956	9.137				
Biomass-Based Diese	I 0.291	0.280	0.295	0.350				
Ethanol	1.182	1.192	1.190	1.197				
Biofuels Subtotal	1.473	1.473	1.487	1.548				
Other ^o	0.801	0.821	0.824	0.814				
Total	1 0.233	10.998	11.270	11.499				

According to the table, wind power is the second most generated renewable energy source in all of America, and the EIA has actually projected that wind power will overtake hydroelectric power by 2019 (Comstock, 2018b). It takes about 8,500 Btu's to power an oven for one hour (Jones C, 2018). Therefore, the wind power harvested within one year can power 246.35 billion ovens for one hour, to put things into perspective.

The EIA has also released data that shows the production amounts for all types of power from 2001 to 2017 (Figure 2), a graphic representation which shows how much wind power has grown over a decade. Net wind power generated has increased by 100% since 2009, and is responsible for the generation of 0.25 trillion kilowatt hours (85.304 trillion Btus) in the year of 2017 (Comstock, 2018a).

U.S. Net Electricity Generation (2001-2017)

This graph shows the net generation of all types of electricity sources, with fossil fuels and renewables included. It shows the amount of electricity generated per each type of energy per year (Comstock, 2018a).





In the past, windmills would spin until the force exerted onto the propeller of the windmill was enough to turn the axle to grind wheat into flour. This contraption was converted into a modern day application when scientists discovered that spinning an axle with a magnet attached in the center of a magnet wire core generated power (How does a wind turbine work?). When most people think of wind turbines, they think most wind energy comes from turbines that can be as tall as an office building and have three 200-foot-long blades (Wind power). When the propeller spins from the wind, it spins an axle, called the low speed shaft, spinning at about 30-60 revolutions per minute (rpms). The low speed shaft then goes into a gearbox (How does a wind turbine work?). This gearbox then connects to another axle which spins much faster, specifically within the range of 1,000-1,800 rpm, and is connected to a generator (How does a wind turbine work?). However, the type of wind turbine previously described can perform with varying results, depending on its location.

If you are in an urban area, micro wind turbines, scaled down versions of their larger 100 foot counterparts, are much more common. The two types of wind turbines are horizontal axis wind turbines (Fig. 3) and vertical axis wind turbines (Fig. 4). While the sizes and shapes of the propellers can be different within the two groups, the main distinction between vertical and horizontal wind turbines results from the orientation of the axle to which the propeller is attached. The orientation of the axle can make a difference in the generation of power based on the location and scale of the wind turbine (Makkawi, Celik, & Muneer, 2009).

VisionAIR series vertical axis micro wind turbine.

The picture below depicts a micro horizontal axis wind turbine, in which the axle which the propeller is attached to is horizontal.

Figure 4

A 1kW horizontal wind turbine for electrical generation.

The picture below depicts a micro vertical axis wind turbine, in which the axle which the propeller is attached to is vertical.



The current study focuses on the type of micro wind turbine (vertical axis or horizontal axis) that would produce more power in a medium density urban environment, specifically Park Slope, Brooklyn.

One drawback to micro horizontal axis wind turbines is the self orientation that the turbine needs to make whenever the wind blows in a new direction (Makkawi, Celik, & Muneer, 2009). This self orientation is something that the turbine does on its own. When the wind blows in any given direction other than the direction that the turbine is facing, the turbine will rotate in order to align its propeller to face into the wind in order to start generating power. This is a hindrance to the generation of power and is called an idling period. As stated in the paper, *The evaluation of micro-wind turbine aerodynamics, wind speed sampling interval and its spatial variation* (Makkawi, Celik, & Muneer, 2009), the type of micro wind turbine (horizontal or vertical) that should be deployed in locations where there isn't a reliable wind resource but rather where power is required, like urban areas, is in fact vertical wind turbines. One drawback of micro horizontal axis wind turbines illustrated in the paper and described above is the idling period. If there isn't a reliable wind resource, meaning that wind comes in quick 30 second gusts, the power generated by a micro horizontal axis wind turbine is essentially lost due to the idling period (Makkawi, Celik & Muneer, 2009).

A vertical axis wind turbine, in which the axle of the wind turbine is oriented vertically and has a differently shaped propeller, eliminates the idling period, resulting in a more efficient wind power generator.

While the objective of the paper, *Micro wind turbine performance under real weather conditions in urban environment* (Glass & Levermore, 2011), is to test the drawbacks of micro wind turbines and report them to manufacturing companies, the paper fails to take into account and test micro vertical axis wind turbines, and instead lists all of the problems with horizontal axis wind turbines. The researchers placed two different variations (three bladed and five bladed propellers) of five micro horizontal axis wind turbines on the roofs of a few suburban two story homes in order to measure efficiency and output from their wind resource under "real weather conditions." According to the researchers, lateral turbulence is a factor which decreases power output. Lateral turbulence and the aforementioned idling period are two factors which seem to limit the performance of micro horizontal axis wind turbines, creating the possibility for micro vertical axis wind turbines to prevail. I would like to test these two types of micro wind turbines in the context of Berkeley Carroll, a medium density, unobscured building located approximately 15 km from the coast (Glass & Levermore, 2011). Specifically, I am looking for which type of micro wind turbine generates more power at Berkeley Carroll.

Procedure

My first experiment revolves around a proof of concept for my larger application of wind turbines to Berkeley Carroll, specifically testing the turbines under controlled conditions. In this experiment, my independent variable is the type of turbine, and my dependent variable is average power generation. I measured power generation by taking the average voltage over 15 seconds. This experiment is designed to reduce the presence of all control variables associated with loss of power generation in wind turbines such as wind turbulence or inclement weather conditions. I purchased two scaled-down vertical and horizontal "nano-wind" turbines (the RoseSummer Div Kit Small DC Motor Vertical Micro Wind and Homvl DIY Horizontal Wind Generator Micro Wind Turbine; both can be purchased on Amazon for a combined total under \$30.00). I found a large wind machine brand fan around school to use as my artificial wind source, and purchased a Hold Peak brand anemometer (model number H-866B) to run the experiment. I created a Google spreadsheet and made a table in which there was a column dedicated to each type of turbine. I measured the wind speed of each fan speed setting by holding the anemometer 50 centimeters away from the fan to see which speed was most appropriate for measuring power generation. I found that 2.5 meters per second was the highest wind speed I could achieve. In order to prepare the wind turbines to record power generation data, I modified both nano wind turbines by installing voltmeters in place of the LED lights that were there. Before conducting the experiment and generating power, I made sure that each turbine was 50 centimeters away from the base of the fan. I measured and recorded all the power generated over fifteen seconds of constant wind blowing at the fan speed setting closest to 2.5 meters per second. After recording these results, I measured and recorded the overall power generated over 15 seconds at 50 centimeters at the fan speed setting closest to 2.5 meters per second with the horizontal axis nano turbine turned 90 degrees to the right to simulate an idling period. I averaged each column to see which type of turbine performed better under controlled conditions.

In order to find the optimal location to place the wind turbines, I wanted to measure the average wind speed over one month in two locations on the roof, specifically North and South. In order to measure the wind speed, I constructed anemometers, devices that have scoop-like elements attached to three arms on a vertical axle which measures wind speed when spun. I used a 3D printer with the required plastic filament to print the parts required. The parts list is available upon request. I constructed the anemometers and placed them on the roof to measure the wind speed over one month. The wind speeds were recorded via Raspberry Pi, a programmable computer that interfaced directly with the anemome-

ters and the school's wifi network. The Raspberry pi then recorded and compiled the data points into a spreadsheet which displayed the date and time during which both wind speeds were recorded. In the two columns to the right, I recorded the corresponding wind speeds. Then, at the bottom of the spreadsheet, I averaged each anemometer column to see which location is a more favorable spot for power generation.

Data Analysis

I ran an ANOVA (Analysis of Variance) test comparing three means: continuous wind blowing on both the vertical and horizontal turbines, and offset horizontal turbines, which involved me turning the horizontal turbine 90° so the propeller wasn't facing directly into the fan, to simulate an idling period. I only tested for the idling period in the horizontal turbine since there are no idling periods with vertical turbines. The p-value was approximately 1, leading me to accept the null hypothesis, that vertical wind turbines are not superior to horizontal wind turbines. As shown in the table below, the horizontal continuous turbine produced the most power, with 11 volts on average. The offset horizontal wind turbine generated 9 volts on average, while the vertical turbine produced 2.24 volts on average, meaning that under controlled conditions with a continuous wind source, the horizontal wind turbine generated more power.

Table 1

This table shows that horizontal axis wind turbines produced more power than the vertical turbine, even when offset by 90^e. All units are measured in volts. The averages are taken over 15 seconds.

Power Generated (Offset) HAWT 9.211 Volts (Average over 15 Seconds)	Power Generated (Continuous) HAWT 11.011 Volts (Average over 15 Seconds)	Power Generated (Continuous) VAWT 2.247 Volts (Average over 15 Seconds)
0	0.01	0.83
0	2.38	1.93
0	5.72	2.23
0.08	9.18	2.62
4.35	11.79	2.98
9.89	13.02	2.42
11.14	13.39	2.26
13.01	13.61	2.04
14	13.81	2.48
14.63	13.67	2.48
14.56	13.94	2.59
14.51	13.84	2.04
14.14	13.61	2.29
14.08	13.7	2.31
13.78	13.49	2.23
9.21	11.01	2.25

In my second experiment, regarding the wind history in Park Slope, my anemometers collected around 1080 data points each, leaving me with a p-value of less than 0.0001, comparing the averages of two different data sets from two different anemometers. Anemometer A, placed on the Southeast side of the building, measured an average wind speed of 8.96 meters per second, and anemometer B, placed on the Southwest side of the building, measured an average wind speed of 9.33 meters per second, meaning that the wind turbines should be put on the side of the roof where anemometer B was located, the southwest corner of the roof.

Conclusion

In summation, my first two experiments have served as supporting studies for the third and final experiment which is discussed below. The first experiment illustrated that horizontal turbines do in fact perform better than vertical turbines in controlled continuous conditions. However, it is important to consider that these conditions are where horizontal wind turbines perform best, and that when the wind resource is unreliable, vertical wind turbines outperform horizontal wind turbines. My second experiment has shown that the optimal location for wind turbines in the next installment of this study is the southwest corner of the roof.

Discussion

For my next steps, I would like to mount both my micro vertical axis wind turbine (meaning that the orientation of the axle attached to the propeller and gearbox is vertical) and my micro horizontal axis wind turbine (meaning that the orientation of the axle attached to the propeller and gearbox is horizontal) on the roof to truly find out which turbine will perform better in the environment of Park Slope, Brooklyn. This will require me to measure average power generation over one month in daily or hourly intervals. One flaw in my first experiment that I would like to acknowledge is the scaling of the wind turbines to the strength of wind source. The fan was much bigger than both types of wind turbines, providing the perfect wind conditions for the horizontal wind turbine to prevail over the vertical wind turbine. A flaw I would like to acknowledge in my second experiment is the fact that the anemometers were 3D printed, meaning that the dimensions of some of the parts were inconsistently sized, which could have skewed data or produced incorrect results.

References

- Comstock, O. (2018a). Electricity generation from fossil fuels declined in 2017 as renewable generation rose [Webpage]. Retrieved from: www.eia.gov/todayinenergy/detail.php?id=35412
- Comstock, O. (2018b). Wind expected to surpass hydro as largest renewable electricity generation source [Webpage]. Retrieved from: www.eia.gov/todayinenergy/detail.php?id=34652

Energy Sources: Clean Energy [Webpage]. Retrieved from: www.energy.gov/science-innovation/clean-energy

Energy sources: Fossil [Webpage]. Retrieved from: www.energy.gov/science-innovation/ energy-sources/fossil

- Glass, A., & Levermore, G. (2011). Micro wind turbine performance under real weather conditions in urban environment. *Building Services Engineering Research and Technology*, 32 (3), 245-262.
- Green, H., Gonzalez, E., De Pastino, B., Langner, H., Jenkins, N. & Aranda, M. CrashCourse. (2014). The global carbon cycle – crash course chemistry #46 [Video file]. Retrieved from: www.youtube.com/ watch?v=aLuSi_6018M
- Hanania, J., Heffernan, B. & Donev, J. (2018). A 1 kW horizontal wind turbine for electrical generation [Photograph]. Retrieved from: http://energyeducation.ca/encyclopedia/Types_of_wind_turbines_for_ micro Generation
- How does a wind turbine work? [Webpage]. Retrieved from: www.energy.gov/maps/how-does-windturbine-work
- Makkawi, A., Celik A. N., & Muneer, T. (2009). Evaluation of micro-wind turbine aerodynamics, wind speed sampling interval and its spatial variation. *Building Services Engineering Research and Technology*, 30 (1), 7-14.
- UGE (2014). VisionAIR series vertical axis micro wind turbine [Photograph]. Retrieved from: https://cleantechnica.com/2014/04/22/uge-raises-bar-vertical-axis-micro-win d-turbines/
- United States Energy Information Administration. (Oct 10, 2018). Renewables and Carbon Dioxide Emissions [Webpage]. Retrieved from: www.eia.gov/outlooks/steo/report/renew_co2.php?src=Environment -b1
- Wind Power [Webpage]. Retrieved from: www.nationalgeographic.com/environment/global-warming/ wind-power/

Jones, C. (2018). Average BTU of a home oven [Webpage]. Retrieved from: homeguides.sfgate.com/ average-btu-home-oven-87147.html

Special thanks to:

Mr. Byrne, Ms. Sefchovich, Mr. Rubin, Isabel Baum, Diego Turturro and Matthew Justh.

The Impact of Stress on the Likelihood and Composition of Mycorrhizal Network Formation

by Eve Blank

Abstract

hen connected by common mycorrhizal networks (CMN's), plants are able to support each other by transferring nutrients and defense signals from one to the other, increasing their chance of survival in times of stress (Simard et al., 1997). In my experiment, I investigated the impact of stress (drought) on the likelihood of the formation of mycorrhizal networks as well as on the composition of CMN's. To do so, I set up pots of 2 oak or 2 maple trees. Half of the pots containing each species were stressed. By determining which species of mycorrhizae form associations with the plants over a 4 month period, I was able to calculate how similar the species on one plant was to the species on the other (fungal overlap ratio) to find out if there is a difference between the likelihood of the formation of CMN's and if the composition of fungal communities changes when plants are grown in stress. However, because of the large number of species of mycorrhizae found overall and because only one species was able to be identified per sample, the fungal overlap ratio did not prove to be meaningful. Still, this study may provide a new alternative method for approximating the presence of mycorrhizal networks with limited resources.

Background

Plants communicate with and support each other in various ways. One form of communication is through airborne chemical signals. These signals are elicited when a plant is attacked by a pathogen or herbivore and preemptively elicits the production of toxic chemicals in neighboring plants (Kant, Bleeker, VanWilk, Schuurink, & Haring, 2009). However, these defense signals are dependent on atmospheric conditions and are therefore not a reliable form of plant communication (Song et al., 2010). There are also many forms of below ground communication. For example, in one study researchers found that stressed plants can receive limited resources from non-stressed plants through root grafts, points at which tissue from separate plants have grown together (Fraser, Lieffers, & Landhäusser, 2006). The most prominent form of interplant communication and support is through mycorrhizal networks – underground fungal networks that connect individual plants to each other (Song et al., 2010; Simard et al., 2012).

Plants' roots have a tendency to form a mutualistic relationship with mycorrhizal fungi in the soil (Song et al., 2010). This relationship is necessary for both the plants and fungi, since the plants cannot fixate nitrogen from the soil and fungi rely on the plants for other resources (Xin-Hua, Critchley, & Bled-

soe, 2003). However, these relationships are much more complex than previously thought. Mycorrhizae not only form relationships with plants, but actually form networks with each other. These networks are formed when part of one fungus called the hyphae, or mycelium, attaches to the hyphae of another fungus on a separate plant (Simard et al., 1997). Hyphal networks connect over 80% of all terrestrial plants (Song et al. 2010). They do not only occur between plants of the same species, but between different species compatible with the same type of mycorrhizae (Simard et al., 1997). We are just beginning to understand the extent to which common mycorrhizal networks (CMN's) impact plant systems as well as their surrounding environment. Mycorrhizal networks benefit plant populations by facilitating nutrient and defense signal transfer (Simard et al., 1997; Song et al. 2010), the larger ecosystem by encouraging seedling establishment and plant diversity (Booth & Hoeksema, 2010), and even have a global impact since they cause forest regeneration, which counteracts disturbance caused by climate change (Bingham & Simard, 2011). On every scale, CMN's offer countless benefits and are an important part of the environment.

Most importantly, though, CMN's support plant populations by transporting nutrients between plants (Simard et al., 1997; Song et al., 2010). In the laboratory, carbon, nitrogen, phosphorus, and water have been shown to be exchanged through networks (Simard et al., 1997: Song et al., 2010). Simard et al. (1997) showed that carbon is transferred between trees in the field as well. She set up pairs of 'donor' and 'receiver' trees half a meter apart. After two years, she found seven different species of mycorrhizae were on both species of trees and covered 90% of their roots. This signified a possible presence of networks connecting the plants. In addition, she used reciprocal carbon isotope labeling, a procedure that has since become standard in mycorrhizal network research, in which different isotopes of carbon are provided to connected plants in order to monitor transfer between them. On average 6% of the total carbon isotope was exchanged between plants. This result not only indicates the presence of CMN's, but illustrates that nutrients can be exchanged bidirectionally. Furthermore, a significantly greater net gain of carbon was found in the 'receiver' trees in deep shade than in partial shade or full ambient light, indicating a source-sink relationship in which the magnitude of nutrient transfer is dependent on the level of need of the receiving plant (Simard et al., 1997).

It is also important to note that CMN's can affect the surrounding ecosystem by facilitating seedling growth (Booth & Hoeksema, 2010). It can sometimes be hard for seeds to grow due to the shade of already established trees and lack of soil space (Booth & Hoeksema, 2010). However, researchers have observed hyphae from these trees colonizing, or attaching to and forming a symbiotic relationship with the roots of seedlings (Booth & Hoeksema, 2010). This raises the hypothesis that CMN's may counteract the negative effects of these larger trees (Booth & Hoeksema, 2010). Evidence for this claim was provided in a 2010 study by Booth and Hoeksema. By comparing seedlings connected to and seedlings not connected to older trees by CMN's, they found that seeds had a 56% greater chance of survival when connected to a mycorrhizal network. So, although overstory trees, or taller trees that make up most of the canopy, have negative effects on seedlings due to competition, they also help establish seedlings and increase their chance of survival through their connections to CMN's. Additionally, the researchers found no difference in facilitation of local or nonlocal genotypes. This observation is interesting since plant populations are specifically adapted to their local environment. The lack of difference in facilitation of local and nonlocal genotypes allows us to draw the conclusion that CMN's support plant diversity.

While it has been demonstrated that mycorrhizal networks have countless benefits in already established systems, less is known about the formation of networks themselves. In this study, I sought to determine how drought affects the formation and composition of mycorrhizal networks in both oak and maple trees. The results of this study could be pertinent to forest regrowth or primary establishment, as in these scenarios no networks exist prior to forest growth.

The carbon isotope labeling method used by Simard et al. in 1997 has since been adopted as the primary procedure to determine if mycorrhizal networks are present. However, this method requires the use of radioactive substances, which are not available in a high school setting. So, as it has been shown that plants within 0.5 meters will reliably form networks in 4 months (Simard et al., 1997), my study involves determining the probability of network formation by looking at the overlap in the species of mycorrhizae present on the roots, as well as the percent colonization for each plant.

Due to difficulty distinguishing species of mycorrhizae morphologically by identifying characteristics such as color, texture, and presence of clamp connections and comparing them with descriptions in a key, DNA analysis was used as the primary form of identification. The procedure for the DNA analysis of mycorrhizae is described by lotti and Zambonelli (2006) and consists of amplifying the internal transcribed spacer (ITS) region, a segment of DNA present in all fungi but different for each species, through polymerase chain reaction (PCR), sequencing the region, and comparing it with a database to classify the specific species of fungi.

Methods

Growing period

21 Northern red oak and 21 red maple trees between the heights of 8 and 16 inches were collected on May 11th, 2018 from Fordham's research station in Armonk, NY (Address: 31 Whippoorwill Rd, Armonk, NY 10504). Over the following two weeks, forest soil was brought in by various Berkeley Carroll faculty. The plants were then potted on June 5th. The 21 3-gallon pots were filled halfway with styrofoam packing peanuts to allow for proper drainage. The remaining volume was filled with a mixture of forest soil (from Boxborough, MA) and potting soil in a ratio of 1:3 by volume, optimizing the amount of forest soil used. 10 pairs of oak trees were transferred into 10 of the pots while another 10 contained pairs of maple trees. The final pot contained both an oak and a maple tree and acted as a test pot, allowing me to practice mycorrhizal identification procedures at the end of the growing period. Using both oaks and maples would allow me to draw conclusions about both ectomycorrhizal fungi, which forms associations with the maples. 5 of the oak pots and 5 of the maple pots were stressed while the other 10 pots as well as the test pot were not.

Over the next 4 months, non-stressed groups were watered with 1 liter of water and the stressed groups with 250 milliliters. Plants were watered as soon as the non-stressed pots became visibly dry. Time between watering varied significantly since the room's temperature changed throughout the summer. For the first three months of the experiment, pots were rotated counter clockwise each time they were watered to ensure that they received a more even amount of sunlight throughout the experiment. However, at the start of the school year, the pots had to be moved to avoid heat and air conditioning and were not rotated thereafter. Every twelve days the heights of the plants were recorded with the exception of on June 18th and August 17th. The growing period ended on October 3rd.

Data collection period

Over the following month, the plants were carefully uprooted and separated. Any attached styrofoam was taken off the roots. Then the root system was cut off, placed in a 6.5 in by 6 in plastic bag, and stored in the freezer. A soil sample from each pot was also obtained and stored in a 1.5 in by 2 in plastic bag in the freezer. The remaining soil and styrofoam was discarded.

Morphological Identification of Mycorrhizae

Safety Precautions: Gloves, a lab coat, and goggles were worn at all times during this experiment as 10% KOH solution (wt/v) is an extremely basic solution. Additionally, the KOH solution must be stored in a glass container.

In order to identify mycorrhizae visually, a staining procedure was adapted from the paper "Ink and Vinegar, a Simple Staining Technique for Arbuscular-Mycorrhizal Fungi" by Vierheilig et al. (1998). Roots from plants in the test pot (on which the procedure was designed), pot 1 (oak), and pot 11 (maple) were boiled in a 10% KOH solution (wt/v) for 8 minutes in order to clear the roots. After 8 minutes, the roots became more transparent and the KOH solution turned a deep red color. Then the roots were rinsed and boiled for 3 minutes in a 5% vinegar solution in order to neutralize any remaining KOH. Finally, the roots were rinsed and boiled in a 2.5% black Pelikan ink solution for 5 minutes. After rinsing again, the roots remained partially transparent and any fungi present would have been stained with the ink.

This procedure was not executed on the remainder of the plants for two main reasons. First, the staining procedure was not always consistent. Not all roots were fully cleared and many remained dyed after rinsing. Second, even when mycorrhizae were noticeably present, it proved incredibly hard to differentiate between species, even when comparing characteristics to a key.

Genetic Identification of Mycorrhizae

DNA Extraction

Note: Gloves were worn at all times during this experiment so as not to contaminate samples.

Due to the difficulty of the morphological identification, a DNA barcoding approach was used to ensure more accurate results. Using this method, plant and fungal DNA was extracted using the DNeasy Plant Pro Kit (QIAGEN) as was done in "A quick and precise technique for identifying ectomycorrhizas by PCR" by lotti and Zambonelli (2006). DNA was first extracted from pots 1, 15, 17, and 18. 2 samples of plants from pot 17 were assumed to be high in phenolic compounds. DNA was extracted from the remainder of the pots over the subsequent two weeks. The protocol from the kit reads as follows:

Notes before starting:

- Ensure that the Tissue Disruption Tubes rotate freely in the centrifuge without rubbing.
- If Buffer APP contains precipitates, heat at 60°C until precipitate dissolves.
- Perform all centrifugation steps at room temperature (15-25°C).
- 1. Add 5–100 mg of fresh or frozen plant tissue and 500 μL of Solution CD1 to a 2 mL Tissue Disruption Tube. Vortex briefly to mix.

Note: We recommend that the tissue be cut into small pieces before loading into the bead tube. **Note:** If the sample is high in phenolic compounds, add 450 μ L Solution CD1 and 50 μ L Solution PS. For very difficult samples, the volume of Solution PS can be increased up to 100 μ L and the volume of Solution CD1 decreased correspondingly.

Added note: As an alternative to vortexing samples for 10 min, liquid nitrogen may be used to freeze roots which can then be crushed. This method requires extra safety precautions.

- 2. Homogenize by vortexing samples for 10 min.
- 3. Centrifuge the Tissue Disruption Tubes at 12,000 × g for 2 min.
- Transfer supernatant to a clean 1.5 mL Microcentrifuge Tube (provided in kit). Note: Expect 350–450 μL. The supernatant may still contain some plant particles.

				<u> </u>
5.	Add 200 µL Solution CD	2 and vortex for 5 s.		\dashv
	Note: For problematic s	sample you can add up to 250 μL Solu	ition CD2 at this step.	-
6.	Centrifuge at 12,000 × g	g for 1 min at room temperature. Avo	ding the pellet, transfer the	
	supernatant to a clean	2.5 mL Microcentrifuge Tube (provide	d in kit).	_
	Note: Expect 400–500	μL.		
7.	Add 500 µL of Buffer Al	PP and vortex for 5 s.		
8.	Load 600 µL of the lysa	te onto an MB Spin Column and centi	ifuge at 12,000 × g for 1 min.	D
9.	Discard the flow-throug	h and repeat step 8 to ensure that al	l of the lysate has passed	
	through the MB Spin Co	olumn.		н
10	Carefully place the MB	Spin Column into a clean 2 mL Collec	tion Tube (provided in kit).	
	Avoid splashing any flo	w-through onto the MB Spin Column.		
11.	Add 650 µL of Buffer AV	N1 to the MB Spin Column. Centrifug	e at 12,000 × g for 1 min.	Z
	Discard the flow-throug	h and place the MB Spin Column bac	k into the same 2 mL Collection Tube.	D
12	Add 650 µL of Buffer AV	N2 to the Spin Column. Centrifuge at	12,000 × g for 1 min.	
	Discard the flow-throug	h and place the MB Spin Column bac	k into the same 2 mL Collection Tube.	
13	Centrifuge at up to 16,0	$100 \times g$ for 2 min. Carefully place the	MB Spin Column into a	
	new 1.5 mL Elution Tub	e (provided in kit).		Т
14	Add 50–100 μL of Buffe	er EB to the center of the white filter i	nembrane.	
15	Centrifuge at 12,000 × g	g for 1 min. Discard the MB Spin Colu	mn. The DNA is now ready	R
	for downstream applica	ition.		
PCR	and Electronhoresis			ŝ
N	. Claura	Il times during this summing the		
Note	: Gloves were worn at a	ii times during this experiment so as	not to contaminate samples.	A
Tabl	o 1			æ
	<u>, (D. 11</u>			\bigcirc
Sequ	iences of Primers Us	ed During PCR.		
Pri	mer	Sequence]	
		ocquence		\bigcirc
ITS	1	TCCTCCGCTTATTGATATGC	-	

A MasterMix solution was made containing the primers ITS1 and ITS4, distilled water, and the Master Mix (containing Taq DNA polymerase, dNTPs, MgCl2, and other buffers necessary for PCR) in a ratio of 1:1:79:80 respectively. 25 μ L of this solution was then combined with 10 μ L of extracted template DNA and 15 μ L distilled water. PCR was conducted on the samples in order to amplify the ITS region of the fungal DNA. Samples were first subjected to a 6 min denaturation period at 95°C, followed by 36 amplification cycles at 94°C for 30 s (melting), 55°C for 30 s (annealing), and 72°C for 1 minute (extension). Finally, the samples were held at 72°C for 10 minutes for a final extension step.

CTTGGTCATTTAGAGGAAGTAA

ITS4

Gel electrophoresis was then performed on the end product of plants in pots 1, 15, 17, and 18 to ensure that DNA had successfully been extracted and the correct segment had been amplified. Gel was made by preparing a 1% agarose solution in 1x TAE buffer (wt/v). Part of the agarose solution was boiled, poured into a 10 cm by 15 cm mold containing wells for the placement of samples, and allowed to set. The gel was placed in the center of the sub-cell and the DNA samples were combined with UV

Z F

F

 \mathbb{Z}

 \bigcirc

-

dye and inserted into the gel. 1x TAE running buffer was used. Electrophoresis was run for roughly 20 min at 120 V in order to separate DNA regions by size (smaller fragments move further, larger ones cover less distance).

Once it was confirmed that the intended region of DNA had been amplified, the remaining amplified DNA sample (PCR product) was sent to Sequetech for sequencing. In order to determine the species of mycorrhizae present on the roots of each plant, the sequences were run through the BOLD fungal identification database, a system that compares a given sequence to the ITS regions in the database to identify the species of fungi.

Data Analysis Period

Likelihood of the formation of mycorrhizal networks was to be determined by calculating the similarity in the composition of mycorrhizal colonization between plants in each pot. This similarity was represented by the overlap ratio:

Number of species in common
Total number of species present in the pot

Two tailed, independent, comparison
Two tailed, independent, comparison
Two tailed, independent, comparison
Total number of species present in the pot
Total number of species present in the pot
Two tailed, independent, comparison
Two tailed, independent, comparison
The second plants is a difference between stressed oaks and non-stressed plants. These tests would indicate whether or not there is a difference between the likelihood of the formation of mycorrhizal networks in stressed vs non-stressed oaks, maples, and both, as well as whether or not there is a difference in likelihood of the formation of arbuscular mycorrhizal networks (between maple trees) and ectomycorrhizal networks (between oak trees).

Data

Growth Data

Throughout the growing period, there was average height increase of 2.24 cm. Interestingly, this growth was not consistent among all plants. As can be seen in Figure 1, the difference in height between plants in pots increased throughout the duration of the experiment. Despite this increase, the average difference between differences in heights of plants in the same pot at the start and at the end of the experiment was not statistically significant (p = .25334).

In order to determine the magnitude of the impact of the drought treatment on stressed plants, I compared the average difference in heights of stressed plants in the same pot to the average difference in heights of non-stressed plants in the same pot using an independent, one-tailed, comparison of means t-test. This test yielded marginally significant results (p = 0.079028), meaning there was a slight difference in heights of stressed plants in the same pot compared to the difference in heights of non-stressed plants in the same pot compared to the difference in heights of non-stressed plants in the same pot compared to the difference in heights of non-stressed plants in the same pot. These results indicate that stressed plants were in fact stressed and is supported by the poor conditions of the plants at the end of the growing period. In the stressed group, 5 plants had died and 3 were close to death, while in the non-stressed group, no plants had died and 4 plants displayed an immense amount of outward growth which was not taken into account in the numerical data.

Difference in Heights between Plants in the Same Pot Over Time.

Using a correlated, one-tailed, comparison of means t test, I found that the final difference in heights (+/- SEM) was not significantly greater than the initial difference in heights (+/- SEM) at the start and at the end of the study (p = .25334).



Figure 2

Difference in Heights between Stressed and Non-Stressed Plants in the Same Pot.

Using an independent, one-tailed, comparison of means t-test, I found that the difference in heights (+/- SEM) between plants in stressed groups was not larger than that of non-stressed pots (p = .079028).



Morphological Identification

The results of the staining procedure were incredibly variable. When subjected to the same conditions, some roots were very clear while others remained pigmented (see Figures 3a-d). This difference made it difficult to determine whether fungi was present or not, much less to differentiate between species. Additionally, while it is possible to identify mycorrhizae morphologically using the guide "Identification of Ectomycorrhizas" by Ingleby, Mason, Last, & Fleming (1990), even with a key it is extraordinarily difficult to locate certain characteristics necessary for identification or even differentiation. Consequently, mycorrhizae were identified through DNA barcoding to ensure higher accuracy and reliability.

Figure 3a-d

Stained Roots from Plant 1 (right) and Plant 2 (left) in Pot 1.

(a) Fully cleared root from plant 1, pot 1 with possible mycorrhizae present
(b) fully cleared root from plant 2, pot 1 with possible mycorrhizae present
(c) partially cleared root from plant 1, pot 1
(d) partially cleared root from plant 2, pot 1



(a)







Gel Electrophoresis and DNA sequencing of the ITS region

Once DNA had been amplified using the PCR procedure, the electrophoresis was run on the PCR product to confirm that the ITS region had in fact been amplified. This was determined by the length of the fragment of DNA. As the ITS region is between 600 and 700 base pairs (bps), any fragment appearing in the gel in that approximate range was considered to be the ITS region. PCR was seemingly unsuccessful for many samples, especially soil samples, as can be seen by the lack of fluorescent marks in the gels. Despite this lack of success, all samples were sent for sequencing with the possibility that there was still a small amount of DNA that would suffice for sequencing.

Figure 4a-d

Gels for Plant and Soil Samples.

Blue fluorescent marks represent DNA fragments. (a) DNA from fungi in pots 1, 15, 17, and 18 (b) DNA from fungi on plant 1's in remaining pots, sequentially (c) DNA from fungi on plant 2's in remaining pots, sequentially (d) DNA from fungi in soil from remaining pots, sequentially. Many soil samples here demonstrate that PCR was unsuccessful.



(b)



00

Table 2

Species of Fungi and Percent Match for Plants and Soil.

Only species in pot 1 on both plants are the same.

Pot Number	Species	Stress?	Soil Type	Percent Match	Plant 1	Percent Match	Plant 2	Percent Match
1	М	No	Sordariales Chaetomiaceae	79.45	Agrocybe erebia	99.06	Agrocybe erebia	99.06
2	М	No	No match	N/A	Agrocybe erebia	83.75	Penicillium citrinum	99.28
3	М	Yes	Galactomyces candidum	99.72	Clavariopsis aquatica	88.58	Dictyochaeta fertilis	95.13
4	М	Yes	Conocybe macrospora	89.37	Plectosphaerella oligotrophica	99.28	Trichoderma asperellum	99.83
5	М	No	No match	N/A	Clavariopsis aquatica	76.88	Fusarium proliferatum	81.89
6	М	Yes	Scedosporium dehoogii	73.75	llyonectria robusta	97.78	Fusarium oxysporum	74.76
7	М	Yes	Rectipilus idahoensis	96.95	Fusarium oxysporum	84.25	Agrocybe erebia	96.39
8	М	Yes	Cladorrhinum bulbillosum	94.74	Agrocybe erebia	96.39	Mycena flavescens	79.24
9	М	No	Pholiota gummosa	99.07	Trichoderma hamatum	72.58	Cylindrocarpon pauciseptatum	93.28
10	М	No	No match	N/a	Fusarium oxysporum	95.72	Agrocybe erebia	88.46
11	0	Yes	Scedosporium dehoogii	88.19	Fusarium oxysporum	96.42	Agrocybe erebia	95.14
12	0	No	Penicillium sp. FF15	88.32	Agrocybe erebia	99.06	Prosthemium orientale	80.51
13	0	Yes	Chaetomium globosum	85.62	Dictyochaeta fertilis	96.4	Fusarium oxysporum	97.86
14	0	Yes	Leptographium sp. HS-ch	65.66	Clitopilus cf. passeckerianus	92.19	Cylindrocarpon pauciseptatum	79.37
15	0	No	Scedosporium dehoogii	65.01	Clitopilus cf. passeckerianus	84.57	llyonectria anthuriicola	84.97
16	0	No	Coprinellus callinus	76.1	Chaetomium globosum	94	llyonectria robusta	75.47
17	0	No	Clitopilus cf. passeckerianus	76.65	Agrocybe erebia	97.65	llyonectria anthuriicola	97.65
18	0	Yes	Zopfiella latipes	95.52	Cylindrocarpon pauciseptatum	99.63	Fusarium oxysporum	99.63
19	0	No	Phialemonium inflatum	88.5	Fusarium oxysporum	73.07	Cylindrocarpon pauciseptatum	81.91
20	0	Yes	Fomitiporia punctata	96.67	Clavariopsis aquatica	91.53	Agrocybe erebia	92.42

Once samples were sequenced, they were run through the BOLD database in order to identify the species of mycorrhizae that were present. Because sequencing only returned the most common sequence per sample, only one species could be identified from the results. This species can safely be assumed to be the most prominent species on or in each sample. No statistical tests were performed as sample sizes were not large enough for calculations. A total of 26 distinct species were identified (see Table 1). 12 of those species resided solely in soil samples, while another 11 were only found in plants with an overlap of 3 species residing both on plants and in the soil. The small size of the overlap between species found on soil and plants suggests that there were in fact associations forming between mycorrhizae and plants, as one would expect the two sets to be similar had specific species not formed relationships with plants. However, there were a total of 14 different species on the plants, of which 5 were found only on maple samples, 3 were found only on oak samples, and 6 were found on both maple and oak samples. This large size of this overlap is surprising since oak trees were expected to only form associations with ectomycorrhizae and maples with arbuscular mycorrhizae. If associations had formed, there should ideally be no fungal species in common between maple and oak samples. The overlap found here thus contradicts the previous evidence of associations forming, instead indicating that associations between mycorrhizal fungi and plant roots were not present.

Discussions and Conclusions

Because only one species was identified for each respective plant, I was not able to approximate the likelihood of network formation using overlap ratios. Due to the large number of species present overall, the likelihood that the species identified on each plant was the same was very low, making the overlap ratio an ineffective way to quantify likelihood of network formation. Consequently, I was not able to determine the impact of stress on likelihood and composition of network formation. Given more time, it would be possible to identify less prominent species of fungi by looking at bases with lower frequencies and extrapolating secondary sequences from those bases. However, this is a very time consuming process and would not yield enough species for a meaningful overlap ratio.

While the results of this study were not significant, it may nonetheless provide an alternate procedure for approximating the presence and composition of mycorrhizal networks. The primary method for CMN identification is carbon isotope labeling established by Simard et al. in 1997. However, radioactive substances are not available in all settings, the procedure laid out here can serve as a safe alternative. I was not able to determine whether fungi present on roots indicated associations with plants and consequently the formation of networks. With further investigation, though, it may be possible to find a new method for the identification of mycorrhizae that will lead to quantifiable results and therefore an approximation of the presence and composition of mycorrhizal network formation.

References

- Bingham, M. A., & Simard, S. W. (2011). Do mycorrhizal network benefits to survival and growth of interior Douglas-fir seedlings increase with soil moisture stress? *Ecological Evolution*, 1(3), 306 - 316.
- Booth, M. G., & Hoeksema, J. D. (2010). Ectomycorrhizal networks counteract competitive effects of canopy trees on seedling survival. *Ecology*, 91(8), 2294 - 2302.
- Fraser, E. C., Lieffers, V. J., & Landhäusser, S. M. (2006). Carbohydrate transfer through root grafts to support shaded trees. *Tree Physiology*, 26, 1019 - 1023.
- Ingleby, K., Mason, P. A., Last, F. T., & Fleming, L. V. (1990). Identification of ectomycorrhizas. Institute of Terrestrial Ecology, 5.
- Iotti, M., & Zambonelli, A. (2006). A quick and precise technique for identifying ectomycorrhizas by PCR. Mycological Research, 110(6), 60 - 65.
- Kant, M. R., Bleeker, P. M., VanWilk, M., Schuurink, R. C., & Haring, M. A. (2009). Plant Volatiles in Defense. Advances in Botanical Research, 51 (9), 613 - 666.
- Simard, S. W., Beiler, K. J., Bingham, M. A., Deslippe, J. R., Philip, L. J., & Teste, F. P. (2012). Mycorrhizal networks: Mechanisms, ecology and modelling. *Fungal Biology Reviews*, 26, 39 - 60.
- Simard, S. W., Perry, D. A., Jones, M. D., Myrold, D. D., Durall, D. M., & Molina, R. (1997). Net transfer of carbon between tree species with shared ectomycorrhizal fungi. *Nature, 388*, 579 - 582.
- Song, Y. Y., Zeng, R. S., Xu, J. F., Li, J., Shen, X., & Yihdego, W. G. (2010). Interplant communication of tomato plants through underground common mycorrhizal networks. *PLoS ONE*, 5(10), e13324.
- Xin-Hua, H., Critchley, C., & Bledsoe, C. (2003). Nitrogen transfer within and between plants through common mycorrhizal networks (CMNs). *Critical Reviews in Plant Sciences, 22*(6), 531-567.
Testing the Claims of Fast Food Chains Using PCR and Sequencing Methods

by Savannah Der

Abstract

n this study, the experimental question (EQ) was: Which fast food chain is selling what they claim? To determine that, I tested the fish sandwich of three different fast food chains (McDonald's, Wendy's, and Burger King's) who each claim their fish is 100% a certain species. I performed PCR on each of the fish samples which gave me a large amount of the fish's DNA so I could then analyze it by barcoding it which told me the species. Barcoding showed me what percent similar the fish sample was to the species of fish it claimed to be. My independent variable was the different restaurants and my dependent variable was the accuracy of the product claim, operationalized by the percent species indicated by the barcode. This method of analysis suggested that each fast food chain was selling what they claimed to be selling. I conducted a second study using the same procedure in which the EQ was: What species of fish is "White Tuna"? Is it the same between each restaurant? The type of restaurant was my independent variable and the species of fish it matches with, according to barcoding, was my dependent variable.

Background

Food is a basic necessity for all humans. Therefore, it makes sense that the food industry is the largest in the planet (Monique, 2016). However, like any other industry, it is run by money. This leads many people to question if the food we buy is healthy and actually what it is labeled to be. Food fraud, meaning detection of undeclared food products, happens all over the world and depending on its severity, has many negative effects (Ripple et al., 2016).

In West and Central Africa, where food is not as accessible, bushmeat (meat from wild animals that are hunted for food) has become extremely popular (Ripple et al., 2016). Orders with the most species threatened by hunting include primates, even-toed ungulates, diprotodont marsupials, rodents, and carnivores. The most commonly hunted of these animals are monkeys and bats. The hunting of these animals has had many detrimental effects including the spread of deadly viruses, the increased risk of extinction for many animals, and the deterioration of many ecosystems (Ripple et al., 2016). Eating, and especially preparing raw bushmeat, is an easy way to transfer deadly diseases from animals to humans. It is thought that this is to blame for the emergence of illnesses like ebola and HIV 1 and 2 in humans (Ripple et al., 2016). Additionally, if a particular species were to become extinct, the entire ecosystem that they're a part of could suffer since that species won't be able to fulfill its ecological roles (Ripple et al., 2016). Due to the high rates of international trade, this can quickly become a global health crisis.

Bushmeat hunting becomes a big problem when bushmeat is traded and sold in markets illegally. For example, in 2012, a BBC investigation found that butchers at a market in London were selling bush-

meat (Rathi, 2016). However, despite this finding, their license was not revoked. This fraud not only puts the consumer at risk for contracting deadly diseases, but keeps the bushmeat market running.

Food frauds also happen more locally, and in a smaller scale, in urban areas and cities in the U.S. (Kimberly, W. Ph.D., Timme, W., Lowell, B., Hirshfield, M. Ph.D., 2013). There is always some skepticism around some commercially sold food products, especially ones sold a very low cost. For example, "White Tuna", a commonly sold fish at sushi restaurants, isn't actually tuna--white tuna does not exist. A study done in 2013 found that 59% of all the samples labeled "White Tuna" were in fact a species of fish named escolar (Kimberly, W. Ph.D., Timme, W., Lowell, B., Hirshfield, M. Ph.D., 2013). Escolar is priced less than most tunas per pound (Fitts Seafood Prices, 2009). However, by using "tuna" in its name, sushi restaurants can sell escolar for the price most tunas are sold.

Unfortunately, it is almost impossible to detect undeclared food products (food fraud) with plain sight. Only through analyzing the product at the molecular level can the claims be accurately tested (Mafra et al., 2007). There have been a few studies conducted to test the claims of food labels using a technique called Polymerase Chain Reaction (PCR). PCR is a quick and accurate technique used to duplicate a specific part of a specimen's DNA in order to obtain large amounts of it (Saik, R. K., Gelfend, H. D., Susanne, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., 1987).

DNA is a very long molecule and it is similar between each living species. There are only certain genes within that larger DNA molecule that are different between even closely related species, or that are only in the DNA of a certain species (Fish DNA Barcoding Kit Curriculum Manual). The point of PCR is to make many copies of that particular gene so you can later analyze it. The result serves as a genetic fingerprint that you can compare to those of other species. For that reason, PCR can be used as a means of authenticating food and its contents and nearly all studies that test food fraud use it (Mafra, et al. 2007). For example, a literature review done in 2007 aimed to understand the large variety of food groups that commonly have fraud. Mafra et al. (2007) suggested that PCR was the best method for food authentication as well as for detecting genetically modified organisms (GMOs) and allergens. The researchers also reported the data of hundreds of studies, all of which used PCR as a means of determining whether or not there was fraud in their food sample. In each category of food being analyzed (meat. plant products, dairy, seafood), they found food fraud (Mafra, et al. 2007). For example, one study used PCR to amplify a specific goat's gene in ovine cheese samples to see if there was contamination/dilution of that cheese with cheaper goat's milk. They found as little as 1% contamination in their samples. Another study used PCR to amplify the gene of Venerupis pullastra (a species of clam) in R. decussatus (a more expensive species of clam) samples to see if there was substitution of a less expensive clam for a more expensive one (Mafra., et al. 2007).

Fraud is not only unjust, but also dangerous. For example, if someone with a food allergy were to consume a food product that contains an allergen that is not labelled, anaphylaxis could occur and that can be life threatening. However, it isn't just people with food allergies that can be seriously affected by food fraud. More recently, a study investigated how food fraud can affect people following religious practices that include food restrictions. Barakat et al. (2014) used a species-specific PCR procedure to test for a specific pork gene in halal sausages and found these to contain up to 10% of pork meat in them. Looking at the results of these studies, we can see two things: that food fraud is not uncommon and can occur in many different types of food products. We can also see that PCR is the technique of choice commonly used as a means of authenticating the food.

While in all of these studies PCR is used to test for a specific fraud, I wanted to explore the idea of looking for fraud in general by determining the specific species of origin. For this, I used PCR and DNA sequencing.

PCR amplifies a specific gene in the sample's DNA that can distinguish it from other species. In animal samples, the most commonly used gene is the Cytochrome c oxidase subunit I gene (COI) found in the mitochondria. Mitochondrial DNA is an ideal choice to serve as the barcode region for animals given that it is different even between closely related species. This is because mitochondrial DNA mutates at a faster rate than nuclear DNA (Fish DNA Barcoding Kit Curriculum Manual). To perform PCR, you need to extract the DNA by using the proper solutions to break down and isolate it. PCR provides enough of that DNA to sequence and compare between species. The first main step in PCR is denaturation of the DNA which separates the two strands (Fish DNA barcoding Kit quick guide). Then specific molecules called primers find a target gene, COI in this case, on the separated strands to determine which segment of DNA will be amplified. Then using polymerase and nucleotides, a complementary strand is built on the two separated ones making two copies of the original gene. The heating and cooling of the machine allows this process to repeat until you end up with millions of copies of the gene, called the amplicon. In this case, PCR produced a COI amplicon.

Sequencing a gene simply finds the exact order of its nucleotides (which are the building blocks that make up DNA: Adenine (A), Thymine (T), Cytosine (C), Guanine (G)) (Fish DNA Barcoding Kit Curriculum Manual). The order of nucleotides distinguishes one gene from another, but it can also distinguish one species from another or even one sample from another. The Sanger sequencing method is used to generate the specific sequence of nucleotide bases that make up a specific section of DNA (COI gene in most animal sample cases).

In the first step of the Sanger sequencing method, four tubes go through PCR as shown at the top left of Figure 1. When each of those tubes go through PCR, the COI genes are essentially going through DNA synthesis. In DNA synthesis, one strand of DNA is used as a template to make new strands. Deoxynucleotriphosphates (DNTPs), or nucleotides, which make up DNA, match up to the template and form a new, complementary strand (Sturm, 2015). The COI amplicon, made through PCR, is in each of those tubes, along with regular nucleotides (DNTPs), one of the four Dideoxynucleotriphosphates (DDNTPs), and other mixtures you need when performing PCR like polymerase, specific primers, and salts. A DNTP is a single nucleotide which, when bonded together, makes up genes, A DDNTP is different in its structure which allows it to have a different function. The most important difference is that a DDNTP does not have an —OH group on its 3 carbon. A free 3 OH is required for DNA synthesis, during which process the 5 phosphate of a free nucleotide is added to the 3 OH of the growing chain (this can be seen in Figure 2). A DDNTP can still match up to its complementary nucleotide on the template, but once a DDNTP is added to the sequence strand, the strand will terminate there. There is also a certain fluorescent color that is attached to DDNTPs and corresponds to each type of DDNTP (red for T, blue for A, yellow for G, and green for C). During PCR, the nucleotides (DNTPs) will attach to the corresponding nucleotide of the COI amplicon/template (labeled "DNA sequence" in Figure 1). For example, on the second to smallest/lowest part of Figure 1, there is a green C and a grey T. The grey T represents a normal nucleotide (DNTP) that is attaching with its corresponding nucleotide on the COI gene (which is A because A and T are complementary). Next to it, there is a C that fluoresces green because it is a DDCTP (the cytosine version of DDNTP). It attaches normally to its corresponding nucleotide on the COI amplicon (which is G) but the sequence stops there because, as mentioned earlier, once a DDNTP attaches to the sequence, nothing else can attach to it. In this instance, the sequence is very short because a DDNTP attached early. The DNA synthesis steps will repeat multiple times, thereby ensuring that at the end of each sequencing reaction, single-stranded DNA fragments of every possible length are generated. Eventually, you are left with many different lengths of the copies of the sample gene where the smallest chain corresponds to the first nucleotide in the sample gene, and the longest chain corresponds to the

Figure 1 (Karki, 2017)

Sanger Sequencing Method.

The Sanger Sequencing Method is done in three parts. The first, which includes PCR, is shown on the left. This first part will allow the COI amplicon to be duplicated multiple times to produce many strands of varying sizes that each end with a DDNTP. Because they each end with a DDNTP, the end product of each of the PCR tubes together will be the order of nucleotides of the COI amplicon. The second step, electrophoresis, is shown on the bottom right. This will put the strands of PCR in the correct order and excite the fluorescent colors on the DDNTP, giving us the nucleotide sequence of the COI gene. The final step, transferring the data onto an electronic device, is shown on the top right. This allows for further analysis of the sequence to be done (such as comparison to other COI gene sequences in a database).



Figure 2 (Lang, 2016)

Difference between Deoxynucleotriphosphate (DNTP) and Dideoxynucleotriphosphate (DDNTP).

On the left side, there is a normal DNTP and on the right side, there is aDDNTP. These differ in their 3' carbon. DNTP has a hydroxyl group (-OH) that is used to bind the phosphate of the next nucleotide in the strand. DDNTPs lack the hydroxyl group (-OH) which will terminate synthesis of the growing strand.



ddNTPs Terminate DNA Synthesis

last nucleotide in the sequence. Importantly, after each tube has gone through PCR, each fragment terminates with one of the four DDNTPs, which are labeled with a different fluorescent tag (Fish Barcoding Bioinformatics). Therefore, if you only read the fluorescent DDNTPs going on a diagonal in Figure 1, you end up with the corresponding nucleotide arrangement of the sample gene.

The next step is to separate the products by electrophoresis. According to Chasteen (2005) this is a method of separation and analysis of macromolecules based on size and charge. This specific type is called capillary electrophoresis. Capillary electrophoresis is the same as normal electrophoresis except there is a window with a laser going through it that excites the different color fluorescents on the DDNTPs. In electrophoresis, a current is running through a sheet of gel in which the products are placed. This creates a negative side towards the bottom and a positive side towards the top. Since DNA is slightly negative, the products with the most DNA will stay towards the top because they will be most repelled by the negative side and attracted to the positive side. The smaller products to be separated by size order where smallest goes first. Therefore, the smallest sequence corresponds to the first nucleotide in the sample gene, and we will get the correct order of nucleotides. Since the laser will only excite the DDNTPs because they are the only ones with fluorescence, those are the only part of the sequences that the laser will recognize. That information is then transferred through a machine that recognizes what color corresponds to what nucleotide and gives you the final order of nucleotides of the sample gene.

Once the order of nucleotides (the barcode) has been acquired , the quality of the results of the sequencing needs to be analyzed and determined to be usable or acceptable. This is determined by the equation for quality: $Q = -10 \log 10 P$ where P is the probability that there is an incorrect base call. If the quality is high (>50), the results can be plugged into a database like BOLD, a list of the sequences of many known species on file. Plugging your sequence into this type of database, will compare your sequence to every sequence on file. A percent similarity graph shows the similarity between the gene of interest and the gene of a specific species in the database. This process allows me to determine if there is a fraud by showing me what specific species my sample came from (Fish DNA Barcoding Kit Quick Guide).

I decided to use this technique to analyze the contents of very commonly questioned products – fast food. Fast food chains are frequently questioned because of their extremely low cost. However, since fast food chains serve about 50 million people in America per day, it is still expected that customers get what they paid for (Johnson, 2016). In my study, I tested food from three different fast food chains and compared their claims to the real contents of their food. Each chain claims their fish is 100% a certain species. My experimental question is: Which fast food chain is selling what they claim? In order to determine this, I compared the sequences of each fish sandwich to others in the database to see if they are actually made from with what they advertise. My independent variables are the different restaurants and my dependent variable is the accuracy of the product claim, operationalized by the percent species indicated by the barcode.

I also conducted a second study in which I used the same method to try to determine what species of fish "White Tuna" is made from. I collected two fish samples from three sushi restaurants, one piece of raw tuna and one piece of raw white tuna. The regular tuna piece served as a control. My experimental question was: what species of fish is "White Tuna"? Is it the same between each restaurant?

Procedure

For each sample, I performed the following procedure exactly as instructed by the Biotechnology DNA barcoding PCR kit (Fish DNA barcoding Kit quick guide). I used the cytochrome oxidase subunit I gene (COI). Once PCR was complete, I performed gel electrophoresis to confirm that I successfully amplified DNA and then sent this DNA fragment out to be sequenced.

Preparing Fish Samples

- 1. One capped 2 ml microcentrifuge tube was labeled for each of fish sample (that is, "1" for fish sample 1, "2" for fish sample 2, etc.)
- A piece of fish muscle up to 100 mg in mass, approximately the size of a pencil eraser head, was cut from each fish sample and placed in their own weigh boats. It was finely minced with a razor blade. That was then transferred into the appropriately labeled microcentrifuge tube (refer to Figure 3)
- 3. The razor blade and gloves were properly disposed of and changed between each sample.

Figure 3



Extracting DNA

- 1. 200 µl of Resuspension was added to each of the microcentrifuge tubes containing the minced fish and flicked several times to ensure full submersion of the fish sample in the resuspension solution.
- 2. 250 µl of Lysis was added to each tube and gently mixed by inverting tubes 10 times.
- 3. The samples were incubated at 55°C for 10 min.
- 250 µl of Neutralization was added to each microcentrifuge tube and mixed gently by inverting tubes 10 times to mix contents.
- 5. The tubes were centrifuged for 5 min at top speed (12,000–14,000 x g) in the microcentrifuge. A compact pellet formed along the side of the tube. The supernatant (liquid surrounding the pellet) contained the DNA. If there were a lot of particulates remaining in the supernatant after centrifugation, the tube was centrifuged for 5 additional min.
- 6. The bottoms of spin columns were snapped off, inserted into a capless 2 ml microcentrifuge tube, and labeled (1 for Fish 1 and a second spin column 2 for Fish 2, etc.)
- The entire supernatant (500–550 μl) of each fish sample from step 5 was transferred into the appropriately labeled spin column.

- 8. After being thoroughly mixed, 200 µl of resuspended Matrix was added to the each column containing fish extract and pipetted up and down to mix. 9. The columns were centrifuged for 30 sec at full speed. 10. The spin columns were removed from the 2 ml microcentrifuge tube, and the flowthrough at the bottom of the 2 ml tube was discarded, and replaced back in the same tube. 500 µl of Wash \mathbf{Z} was added to wash the matrix by centrifugation for 30 sec. U 11. Repeat step 10 to wash samples again. 12. The spin columns were removed from the 2 ml microcentrifuge tube, and the flowthrough at the bottom of the 2 ml tube was discarded, and replaced back in the same tube. The columns were Н centrifuged for a full 2 min to remove residual traces of ethanol and dry to out the matrix. -13. Clean 2 ml capless microcentrifuge tubes were labeled with each fish sample's name. \mathbf{Z} 14. When the 2 min spin was completed, the spin columns were removed and the 2 ml microcentrifuge wash tubes were discarded. 15. Each spin column for each sample was placed into a new capless 2 ml microcentrifuge tube from step 13. \mathbf{Z} 16. Using a fresh pipet tip for each sample, 100 µl of distilled water was added to each spin column. The DNA was eluted by centrifuging for 1 min at full speed. 17. Clean 2 ml microcentrifuge tubes (with caps) were labeled with Fish 1, Fish 2, etc. 18. The eluted DNA was transferred into the appropriately labeled 2 ml microcentrifuge tube with caps and stored at 4°C until it was ready to proceed. S Using PCR, I amplified the DNA that I extracted in the previous steps. This allowed me to have many copies of the COI gene that I could then analyze. PCR Amplification of DNA 1. PCR tubes were labeled with sample name (1 for fish sample 1, 2 for fish sample 2, etc.) and \bigcirc one tube labeled: (+) for the PCR positive control DNA, and one labeled: (-) for the PCR negative T control. The tubes were kept on ice for the remaining steps. 2. A fresh aerosol filter pipet tip was used each time to add 35 µl of CMM (COI master mix) reaction \cap mix to each PCR tube. Each tube was capped immediately after the addition of liquid. \bigcirc 3. A fresh aerosol filter pipet tip was used for each tube to add 5 µl of the appropriate DNA sample directly into the CMM liquid in each PCR tube as indicated by the labels on the tubes, and pipetted up and down to mix. -4. The PCR tubes were placed in the thermal cycler and ran with the program following cycling ____ conditions: 1 94°C – 2 min F 2.94°C-30 sec 3. 55°C – 2 min
 - 4.72°C 1 min
 - 5. Repeat steps 2–4 35x
 - 6. 72°C 10 min
 - 7. $4^{\circ}C$ hold

The tubes were stored at 4°C after the thermal cycling was complete.

-

00

Electrophoresis is a way to make sure the samples amplified correctly in PCR. A blue line where the sample was placed indicates that sample was amplified correctly and there was a successful product of PCR (refer to Figure 4). Once we determine that from electrophoresis, we can send the samples out to be sequenced where they will go through the Sanger sequencing method.

Preparing PCR Samples for Electrophoresis and Sequencing

- 1. 2 ml microcentrifuge tubes were labeled with both a number that represents what fish sample it is (also, one tube (+), and one tube (–)), and E. E stands for electrophoresis.
- 5 µl were removed from each PCR reaction and deposited into the 2 ml microcentrifuge tube corresponding to that sample.
- 3. 2 ml microcentrifuge tubes were labeled with both a number that represents what fish sample it is (also, one tube (+), and one tube (–)), and SEQ. SEQ stands for sequencing.
- 30 µl were removed from each PCR reaction and deposited into the 2 ml microcentrifuge tube corresponding to that sample.
- 5. All samples were stored at 4°C until electrophoresis and sequencing.

Gel Electrophoresis

- 1. To each 5 µl sample of PCR, 5 µl of sterile water was added.
- 2. 2 μ I of UViewTM 6x loading dye was added to each sample and pulse-spun.
- 3. Gel electrophoresis was set up.
- 4. The agarose gel was loaded in the following lane order and volumes.

Lane Sample:

- 1 Empty
- 2 Empty
- 3-20 µl PCR molecular weight ruler
- 4 12 µl (+) E
- 5 12 µl (–) E
- 6 12 µl 1 E
- 7 12 µl 2 E
- 8 EMPTY
- 5. The gel was run at 200 V for 20 min.
- 6. The gel was visualized on a UV transilluminator or imaging system.

On the left-most side (the first column) is the marker which is a reference for how long the DNA is. The top line of the marker column represents 1000 base pairs (base pairs are the compounds that make up DNA), and the second line on the marker column represents 700 base pairs (which is slightly shorter than 1000 base pairs). As you can see in Figure 4, the samples are around the 700 base pair row because the COI gene that we amplified in PCR is a little less than 700 base pairs. The third, fourth, and fifth columns are the three samples from the restaurants. Since we see a blue line is the correct row, according to the marker, the COI gene was successfully amplified during PCR. This sample is ready to be barcoded.

Figure 4

Gel Electrophoresis for Each Fish Sample.

The left-most side column is the marker. After that the samples were put in order of their number (first column after the marker is sample 1, second column after the marker is sample 2, and so on). Each sample was successfully amplified as shown by their fluorescence in the 700 base pair region according to the marker.



Sequencing

- 1. The SEQ samples were sent out to be sequenced.
- Samples were sent out via US post to Sequetech (935 Sierra Vista Ave. Ste C, Mountain View CA 94043) using the prepaid envelope that comes with the Biotechnology Fish Barcoding Kit.
- 3. Sequences were then accessible via the Sequetech login system and can be downloaded in various formats for further analysis.

Online databases like BOLD can compare the sequence you plug in to the known sequences of species it has on file.

Plugging into BOLD

- 1. The forward and backward sequence for each sample was copied and pasted into the BOLD identification section.
- 2. The "Species Level Barcode Records" was selected under Search Databases before "submit" was pressed
- 3. The species it was most similar to and what that similarity percent was was determined by looking at the charts and graphs provided.

For my second study with white tuna, I completed the exact same steps as listed above: preparing fish samples, extracting DNA, PCR amplification of DNA, preparing PCR samples for electrophoresis and sequencing, gel electrophoresis, sequencing, plugging into BOLD. The only difference in this second study comes in the data analysis; instead of comparing the results from BOLD for each sample, to the claims of the restaurant it came from, I compared the BOLD results (what species of fish white tuna is) between each restaurant to determine not only what species white tuna is, but if it is the same in all places.

Study 1 Data

The sequence of the COI gene of each independent variable was plugged into the BOLD system database to find a match to any of the known sequences of fish species on file. The BOLD database identified the species with COI sequences that are the most similar to the sample sequence submitted. This can be seen in "The TOP 20 Matches" in Figure 5. BOLD also provide a similarity percent between the sample imputed in and the matches generated.

Figure 5 (Fish DNA Barcoding Kit Quick Guide)

Example of the Data BOLD Provides when a COI Gene is Submitted.

Example of all the types of data BOLD provides; top 20 matches, and similarity graph of ranked matches to the sample.



Similarity Scores of the Top 99 Matches

TOP 20 MATCHES

Tree Based Identification

Display Option: Top 20 🔻

Phylum	Class	Order	Family	Genus	Species	Subspecies	Similarity (%)	Status
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	macrocephalus		96.73	Published 🗳
Chordata							96.72	Private
Chordata							96.72	Published 🛃
Chordata							96.72	Published 🛃
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	macrocephalus		96.72	Published 🚱
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	macrocephalus		96.71	Published 🗗
Chordata							96.67	Private
Chordata	Actinopterygii	Gadiformes	Gadidae	Boreogadus	saida		96.59	Early-Release
Chordata							96.53	Published 🚱
Chordata							96.53	Published 🗗
Chordata							96.52	Published 🛃
Chordata							96.5	Published 🚱
Chordata							96.38	Published 🚱
Chordata							96.32	Published 🚱
Chordata							96.28	Published 🚱
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	macrocephalus		96.17	Published 🗳
Chordata							96.09	Published 🗳
Chordata	Actinopterygii	Gadiformes	Gadidae	Gadus	macrocephalus		96.01	Published 🗳
Chordata							95.99	Private
Chordata							95.99	Private

Figure 6

Comparison of COI Sequence Data Generated by BOLD for the Three Fast Food Restaurant's Fish Sandwich and the Control.

In the middle column, is the taxonomy (classification) of the sample and the probability that it was placed in the correct group. The right-most column has the percent similarity graph of the sample and the top 100 species that it matched with as well as the species and percent similarity of the highest ranked match. A and B is the data for McDonald's which shows that their fish sandwich is 99.83% similar to Gadus Chalcogrammus (Alaskan Pollock), C and D is the data for Burger King which shows the same about their sandwich. E and F is the data for Wendy's which shows that their sandwich is 96.73% Gadus Macrocephalus (Pacific Cod). The control data, G and H, showed that its COI gene is 98.68% similar to Goodei (Palometa). E and G are blank because BOLD did not provide the taxonomy graph for that sample.



Study 1 Data Analysis/Conclusions

The results from BOLD database search inform the experimental question: Which fast food chain is selling what they claim? The results suggest that all three fast food chains sell what they claim. Figure 6 shows the data of fish sample 3, McDonald's, and sample 4, Burger King. Both McDonald's and Burger King claim on their website that their fish sandwich is made of 100% Alaskan Pollock. The COI gene identified from both sandwiches in the BOLD database showed to be 99.83% similar to the COI gene of *Gadus chalcogrammus* (Alaskan Pollock). This suggests that the McDonald's Filet O' Fish and the Burger King Big Fish Sandwich are in fact made of Alaskan Pollock and do not contain any food fraud.

Figure 6 also shows the data of fish sample 5, Wendy's. When their fish sandwich's COI gene was imputed into the BOLD database, the sequence showed to be 96.73% similar to the sequence of *Gadus Macrocephalus* (Pacific Cod). Although this similarity percent isn't as high as that of Burger King and McDonald's, the lower percent is within the acceptable error range determined by the positive control. From looking at the positive control, there are still some differences between the sequence of the control and the sequences it matched to on BOLD. This is because of error in the procedure that is unavoidable; this error can be from unclean DNA extraction or bad quality sequencing. For the negative control, water, instead of a fish DNA sample, was used with the same procedure as the rest of the samples. However, since it's water, no DNA was amplified during PCR. If it were to match with something on BOLD, we could assume there was contamination and our results would not be trustworthy.

Study 2 Data

This study's aim was to determine what species of fish white tuna is made from. I collected two fish samples from three sushi restaurants, one piece of raw tuna and one piece of raw "White Tuna". The regular tuna piece served as a control. My experimental question is: what species of fish is "White Tuna" made from? is it the same between each restaurant?

Figure 7

Comparison of COI Sequence Data Generated by BOLD for Both Samples of Each of the Three Sushi Restaurants, as well as the Positive Control.

In the middle column is the taxonomy (classification) of the sample and the probability that it was placed in the correct group. The right-most column has the percent similarity graph of the sample and the top 100 species that it matched with, as well as the species and percent similarity of the highest ranked match. A and B is the data for the tuna of restaurant 1 which shows that their sample is 100% similar to Thunnus Albacares (Yellowfin Tuna), C and D is the data for the white tuna of restaurant 1 which shows that their fish sample is 94.84% Pseudomonas Fluorescens (a bacteria). E and F is the data of the tuna sample from restaurant 2 which shows to be 99.85% similar to Thunnus Obesus (Bigeye tuna), G and H is the data of the white tuna sample from restaurant 3 which shows that their fish sample is 100% similar to Thunnus Albacares, K and L is the data of the white tuna sample from restaurant 3 which shows that their fish sample is 100% similar to Thunnus Albacares, K and L is the data of the white tuna sample from restaurant 3 which shows that their fish sample is 100% similar to Thunnus Albacares, K and L is the data of the white tuna sample from restaurant 3 which shows that that sample is 94.73% similar to Pseudomonas Fluorescens, and M and N is the positive control data which shows to be 100% Goodei. C, G, and K are blank because BOLD did not provide a taxonomy graph for those samples.



continued

Fish Sandwich Source	Тахопоту	Top Similarity Scores			
Restaurant 2 (White Tuna)	G. N/A	H. Similarity Scores of the Top 99 Matches 900 900 900 900 1 12 23 34 45 55 67 78 89 100 Rended Matches Pseudomonas Flourescens 94.66%			
Restaurant 3 (Tuna) Restaurant 3 (White Tuna)	I. Identification Summary Taxonomic Taxon Probability Level Assignment of Placement (%) Phylum Chordata 100 Class Actinoptergii 100 Order Scombridormes 100 Family Scombridae 100 Genus Thunnus 100 K. N/A	J. Similarity Scores of the Top 99 Matches 101.0 90.0 91			
Positive Control	M. Identification Summary Taxonomic Taxon Probability Level Assignment of Placement (%) Phylum Chordata 100 Class Actinoptergii 100 Order Scombridormes 100 Family Scombridae 100 Genus Thumus 100 Species Sebastas Goodei 100	N. Similarity Scores of the Top 99 Matches 985 980 975 970 1 12 23 34 45 56 67 78 89 100 Ranket Matches Goodei 100%			

Study 2 Data Analysis/Conclusion

The results don't help answer the experimental question: What species of fish is "White Tuna"? Or is it the same at each restaurant? Each white tuna sample collected matched to a bacteria that causes a deadly disease in fish called *Pseudomonas Fluorescens*. None of the fish matched to a fish species, so we can assume that PCR amplified the DNA of the bacteria and not of the fish. We can't assume that this is because of the procedure because the tuna from each restaurant amplified correctly and sequences were matched, with a very high percent, to a species of tuna (either Bigeye or Yellowfin Tuna); as seen in Figure 7. The positive control also matched as it was supposed to. Therefore, from these results, it is impossible to suggest the species of fish that makes up "White Tuna" in sushi restaurants and if they are the same at each restaurant.

Discussion

According to BOLD, each of the three fast food chains are honest about what they are selling, contrary to what many believe. The results of the second study, despite not answering the EQ, show some interesting results. Pseudomonas Fluorescens is a bacteria that causes a deadly disease in fish. The bacteria can be found in plants and in water and some predisposing factors for fish are: "Physical stresses such as excesses of pH, Injuries or damage to skin or scales, Reduced dissolved oxygen, Presence of toxic substance in the water, Overcrowding of the fishes in farm, Malnutrition, Heavy infestation with parasites" (Pseudomonas Septicaemia). This suggests that the white tuna samples came from fish that didn't live in healthy conditions. This brings up another question: "Does Pseudomonas Fluorescens cause sickness in humans?" According to Dan Washmuth (Pseudomonas Fluorescens Infection: Symptoms & Treatment), this bacteria usually doesn't have an effect on healthy adults. However, there have been some cases in the past of the bacteria causing an infection in patients with compromised immune systems. This includes children, pregnant women, and people who are already sick. Another interesting fact about this bacteria is that it has been found to be antibiotic resistant. This means it would likely take the use of multiple antibiotics to rid the body of *Pseudomonas Fluorescens* (Pseudomonas Fluorescens Infection: Symptoms & Treatment). I think it is important to note that these three restaurants, whose other fish were not infected by *Pseudomonas Fluorescens*, each had white tuna with the bacteria. This may suggest that whatever species of fish "White Tuna" is, it is common for it to live in unhealthy, or unclean conditions. Taking this information into consideration, I would have to agree with Dan Washmuth when he states, "The best form of treatment for Pseudomonas fluorescens is to prevent the bacterium from entering the body in the first place. People who are immunocompromised, such as cancer patients, should be extra cautious when eating and drinking" (Pseudomonas Fluorescens Infection: Symptoms & Treatment).

References

Barakat, H., El-Garhy, H., Moustafa, A. S., & Mahmoud, M. A.(2014). Detection of pork adulteration in processed meat by species-specific PCR-QIAxcel procedure based on D-loop and cytb genes. *Appl Microbiol Biotechnol*, (98), 9805–9816.

Biotechnology Explorer™ Fish DNA Barcoding Kit Bioinformatics Guide. 5115EDU, 76.

Biotechnology Explorer™ Fish DNA Barcoding Kit Curriculum Manual. 5100EDU, 68.

Biotechnology Explorer™ Fish DNA Barcoding Kit Quick Guide. 166-5100EDU, 12.

Dalmasso, A., Sacchi, P., Bottero, M. T. (2011). Development of a real-time PCR assay for the detection of cow and donkey milk. *Eur Food Res Technol*, (235), 47–52.

Fitts Seafood Prices (2009).

- Formosa, R., Ravi, H., Happe, H., Huffman, D., Novoradovskaya, N., Kincaid, R. (2010). DNA-based Fish Species Identification Protocol. J Vis Exp, (38).
- Johnson, Monique. More Shocking Fast Food Statistics You Should Know (2016). https://blog.partnersfor yourhealth.com/blog/more-shocking-fast-food-statistics-you-should-know
- Karki, Gaurab. Sanger's Method of Gene Sequencing (2017). www.onlinebiologynotes.com/sangers-methodgene-sequencing/
- Kimberly, W. Ph.D., Timme, W., Lowell, B., Hirshfield, M. Ph.D. (2013). Oceana Study Reveals Seafood Fraud Nationwide. Oceana.
- Lang, Steven. (November 29, 2016) In terms of biology, what is the main and most important difference between ddNTP and dNTP? www.quora.com/In-terms-of-biology-what-is-the-main-and-mostimportant-difference-between-ddNTP-and-dNTP
- Mafra, I., Ferreira, I. M. P. L. V. O. O., Beatriz, M. P. P. (2007). Food authentication by PCR-based methods. *Eur Food Res Technol* (227), 649–665. Molecular genetic analysis. *Segen's Medical Dictionary* (2011).

Pseudomonas Septicaemia. www.aun.edu.eg/developmentvet/fish%20diseases/2_9.htm

- Rathi, Akshat. The illegal trade in wild-animal meat could cause the next global pandemic (2016).
- Ripple, W. J., Abernethy, K., Betts, M. G., Chapron, G., Dirzo, R., Galetti, M., Levi, T... (2016). Bushmeat hunting and extinction risk to the world's mammals. *R Soc Open Sci*, 3(10).
- Saik, R. K., Gelfend, H. D., Susanne, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. (1987). Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase. *Science*, (239).

Sturm, Noel. DNA Synthesis (2015) www2.csudh.edu/nsturm/CHEMXL153/DNASynthesis.htm

T. G. Chasteen. Capillary Electrophoresis (2005). www.shsu.edu/~chm_tgc/primers/pdf/CE.pdf

Washmuth, Dan. Pseudomonas fluorescens infection: Symptoms & Treatment. https://study.com/academy/lesson/pseudomonas-fluorescens-infection-symptoms-treatment.html

Academic Dishonesty at Berkeley Carroll

by Madison Eve

Abstract

n this study, I was looking to investigate, when given different reasons to lie in a fictional scenario, whether there are specific motivations for lying that Berkeley Carroll high school students find more acceptable. I attempted to do this by creating a survey with three separate scenarios, each containing a situation in which an act of dishonesty is committed, and the acceptability of the dishonesty is ranked on a scale from 1-5 based on the person's motive for lying, as well as a free response question in which the participant said what they would do in that situation. I found, throughout all scenarios, participants found it significantly more acceptable to lie when the motive seemed reasonable in the context of the scenario (e.g. Sam lets Brad copy off of his homework because he knows Brad is very busy with extracurriculars and other commitments). The free response answers, specifically for the first and second scenarios, showed more complex methods of avoiding dishonesty that reflected an unwillingness to tell other-serving lies, as well as a concern for the well being of others in the long term.

Background/Introduction

Dishonest acts play a role in our everyday lives. From white lies to plagiarism, it is difficult to avoid dishonesty throughout our lifetime. While the primary focus is often to tell whether or not someone is lying in a moment, it is equally important to determine the motive for lying. Jensen, L. A., Jensen, J. A., Feldman, S. S., & Cauffman, E. study, *"The Right to Do Wrong: Lying to Parents Among Adolescents and Emerging Adults"*, investigates the motivation to lie to parents amongst high school and collegestudents (2004). The conclusions of this study were similar to that of another study by Aaron Bolin, titled *"Self-Control, Perceived Opportunity, and Attitudes as Predictors of Academic Dishonesty"*, which looks into what motivates college students to commit academic dishonesty (Bolin, 2004). Both of these studies conclude that motivation for lying among young people is highly dependent on how acceptable they believe the lie is.

In an effort to study how the brain functions as a result of dishonesty, scientists examined the responsiveness of the amygdala after continuous dishonesty. In their study titled *"The Brain Adapts to Dishonesty,"* the researchers indicate a significant decrease in emotional response from the amygdala when the participant continuously told self-serving lies (p = .027). The subjects in this study were asked to look at a picture of a jar filled with pennies and estimate the amount of money in the jar. They were told to act as "financial advisers" to a partner in another room and were given different incentives to lie (e.g. they were told they would receive a reward) if they purposely overestimated the amount of money in the jar. Functional Magnetic Resonance Imaging (fMRI) was used to track signal reduction from the

amygdala resulting from the subjects repeatedly telling lies over the course of the experiment. This effect was concluded to be permanent, showing the danger that could result from continuous dishonesty (Garrett, N., Lazzaro, S. C., Ariely, D., & Sharot, T., 2016). These results suggest that a person's moral attitude towards dishonesty greatly affects what they are willing to lie for (e.g. is it more acceptable to lie in order to avoid being punished, even at the expense of someone else, or lie to benefit someone else at your own expense), and this acceptability of lying can increase due to permanent changes in the brain that result from self-serving lies, such as a decrease in signals to your amygdala to illicit the emotional response (e.g. guilt) that people usually associate with lying.

Many other studies have expanded on Garret. et. al., looking deeper into motivated dishonesty. Throughout the course of research on this topic, many speculations as to why people lie have been discounted. The General Theory of Crime states that "lack of self-control, perceived opportunity, and the interaction between the two are the major causes of all deviant behavior," (Bolin, 2004). This theory attributes most criminal and dishonest acts simply to "lack of self-control," however, a study conducted by Aaron Bolin concluded that self-control and perceived opportunity played a minimal role in academic dishonesty, which is not only a dishonest act but potentially a crime as well. Due to these findings, it is apparent that the General Theory of Crime is incomplete. Jensen et. al. states that it is commonly believed that adolescents often lie to their parents in order to incite rebellion, which was concluded from the results of a survey sent out to high school and college students that assessed the student's acceptability of lying in different circumstances, such as lying to parents, self-restraint, deviance, and family cohesion (Jensen et al., 2004). This conclusion was dismissed, however, since the researchers found that teenagers were more likely to lie in order to avoid conflict, and inciting rebellion would only create conflict (p<0.05) (Jensen et al., 2004). The methods of this study directly influenced how I decided to gather data for my study. The main difference was that this study looked at a wider range of scenarios and they looked at 19 motives for each scenario.

At the center of these studies is the notion that the main motivation towards dishonesty is a person's acceptability of it. Jensen et al. states that "Positive, prosocial, and altruistic motives for lying are viewed the most favorably by adolescents and emerging adults" (Jensen et al., 2004). A similar idea is stated in Bolin's study; " that attitude towards academic dishonesty plays a critical role in the explanation of academic dishonesty" (Bolin, 2004). Using multiple ANOVA (Analysis of Variance) tests, Jensen et al. were able to conclude that motive does have a very significant influence over someone's acceptability of lying, some of the motives for the protagonist in the vignette being " 'knew her family was having trouble and didn't want to add additional stress,' and 'knew her boyfriend was having trouble and needed her support' " (p < 0.001) (Jensen, 2004). Bolin specifically looked into academic dishonesty on college campuses and gathered his data by creating a survey that assessed the subjects' perception of opportunity for cheating, their attitude towards cheating, and their capacity for self-control. Bolin conclude that attitude towards academic dishonesty had a more significant effect, compared to lack of self-control and perceived opportunity, on whether or not someone committed academic dishonesty (p < 0.01). Through the results of these studies, it can be seen that teens and young adults are more or less susceptible to dishonesty based on the moral attitude towards it and acceptability of it.

Garrett et al. (2016) concluded that after repeated lying the brain goes through a permanent change in which it becomes less receptive to experiencing emotions while telling lies. However, this brain adaptation only occurs when a person repeatedly tells self-serving lies. In other words, when given the chance to lie for their own benefit, regardless of the benefit or deficit of another, the experimenters saw that those who chose to lie had a significantly diminished response in their amygdala the more they lied compared to their amygdala response before the experiment (p = .027). As stated in the study; "Our

results also suggest that dishonesty escalation is contingent on the motivation for the dishonest act. Specifically, while the magnitude of dishonesty was driven by considerations of benefit both to the self and to the other, the escalation of dishonesty, as well as the amygdala's response to it over time, was best accounted for by whether dishonesty was self-serving." (Garrett et al, 2016). These findings are significant because they serve as a partial explanation for the elevation in dishonest acts, specifically crime. Essentially, since people stop feeling emotions linked to dishonesty over time from repeated lying, they are more likely to commit more serious dishonest acts over time.

In order to better understand dishonesty, we have to look to the root of the problem. Looking at how attitudes towards dishonesty are formed could be an effective way of preventing it altogether, especially if these attitudes have been shown to have a significant influence on whether people commit dishonest acts or not. Ideas of what is moral and immoral are not suddenly decided upon during adolescence; rather much of our education and social experiences gradually teach us what is right and what is wrong (Jensen et al., 2004). The studies by Bolin and Jensen et al. illuminate some of the factors that seem to contribute to dishonesty, with a specific focus on the acceptability of dishonesty and its significantly greater effect than self-control and perceived opportunity. Since the Garrett et. al. study demonstrates the effect of self-serving dishonesty on your brain, it is import to understand how a person's attitude towards dishonesty affects their acceptance of self-serving lies, as well as how the severity of self-serving lying causes changes in our brains.

Though experiences outside of school at times cannot be controlled, what is taught and enforced in school can be. Adjusting the way we approach dishonesty, specifically in schools, can prevent students from being desensitized to their dishonest acts at a young age. In this study, I aim to answer the question: when given different reasons to lie in a fictional scenario, are there specific motivations for lying that high school-aged students find morally acceptable? Initially, I tested this by collecting responses from a survey, which consisted of three fictional scenarios in ranked these motives on a 1-5 Likert Scale, 1 being "not acceptable" and 5 being "very acceptable." My independent variables, in this case, are the motives for lying in each scenario, and the dependent variable is the average acceptability of the lie based on the motive. From the results of the survey, I hypothesize that participants will find other-serving motives to lie more acceptable than self-serving motives. Going forward from the survey, I would like to design an in-person component to my study in order to supplement the results from the survey and gather more data pertaining to my experimental question.

Experimental Procedure

Through this procedure I was looking to investigate, when given different reasons to lie in a fictional scenario, are there specific motivations for lying that Berkeley Carroll high school students find more acceptable? Berkeley Carroll high school students were surveyed, ranging from 9th-12th grade. The participants (170 total) were asked to provide their grade and gender. Participants were asked to read through a survey containing three scenarios, each consisting of a situation in which a person tells a lie. In the first scenario, one character lies on behalf of a friend who has missed an important exam, telling the teacher that the friend was too sick to come to school. In the second scenario, one character lets his friend copy his Spanish homework. In the third scenario, one character covers for another character who lied about her whereabouts, telling her mother that they had had a fun sleepover, which was the story her friend had told her parents. Three motives for each character's lie were listed below each scenario, and the participants were asked to rate the acceptability of the lie based on the motive. The acceptability was measured on a 1 (not acceptable) – 5 (very acceptable) scale. Scores were tabulated and

analyzed using a one-way correlated samples ANOVA test with vassarstats.net. After each scenario, students were additionally given the opportunity to provide a written free response stating what they would do as the character telling the lie in each scenario. Since participants used different language to describe each scenario, responses communicating the same idea were assumed to be similar and grouped together, even when responses were phrased differently (e.g. "I wouldn't let them copy" vs. "I would not give my homework away"). These responses were assessed by finding the most common response (e.g., for the second scenario, an overwhelming majority of the participants said that they would not let their friend directly copy from their homework) as well as how this correlated or differed from how participants ranked the acceptability of the lie.

Survey Grade 9th 10th 11th 12th Gender

Male
Female
Non-binary

Scenario 1

Sally's friend Maggie doesn't show up to calculus on the day of a very important exam. When the teacher asks Sally why Maggie isn't there, Sally says Maggie is sick and won't be able to come to school, even though this isn't true. Rate how acceptable Sally's lie was based on the following motives:

#1: Sally knows Maggie is having family trouble and hasn't had time to prepare for the test. Not Acceptable 1 2 3 4 5 Very Acceptable **#2:** Sally knows Maggie is skipping the test because she didn't study the material, but Sally doesn't want her to get in trouble because she's her friend. 2 3 4 5 Very Acceptable Not Acceptable 1 #3: Sally doesn't know where Maggie is, but she's covering for her because she assumes Maggie would do the same for her. Not Acceptable 1 2 3 4 5 Very Acceptable What would you do in this situation? _____

Survey (continued)

Scenario 2

Brad's friend Sam came up to Brad in a panic, and asked him if he could use his answers on last night's Spanish homework. After contemplating whether or not to let Sam cheat, Brad lets Sam copy his homework. Rate how acceptable Brad's act was based on the following motives:

#1: He knows Sam has been really busy with extracurriculars, and understands he probably doesn't have enough time to do all of his homework.

Not Acceptable	1	2	3	4	5	Very Acceptable
#2: Brad has cheate returns the favor.	ed off of S	am's homew	ork in the	past, so he b	elieves it's	s only fair that he
Not Acceptable	1	2	3	4	5	Very Acceptable
#3: Brad knows Sar but since Sam is his	n barely e s friend, E	ever does his Irad feels obl	homeworl igated to l	k (simply bec nelp him.	ause he do	besn't want to),
Not Acceptable	1	2	3	4	5	Very Acceptable

What would you do in this situation? _____

Scenario 3

Quinn told her parents that she was sleeping over at Ali's house. However, Quinn actually went to sleepover at Riley's house, but she knows her parents don't like Riley, which is why she told them she was going to Ali's. The next day, Ali sees Quinn's mom at the grocery store, and she asks Ali how the sleepover went. Ali, knowing Quinn lied about sleeping over at Riley's, told Quinn's mother that the sleepover went well and that her and Quinn had a lot of fun. Rate how acceptable Ali's lie was based on the following motives:

#1 She believes Quinn's parent's dislike of Riley is irrational, and supported Quinn's lie. Not Acceptable 1 2 3 4 5 Very Acceptable **#2** Quinn's mom makes her nervous, and she didn't want to upset her by telling her Quinn lied. 2 3 Not Acceptable 1 4 5 Very Acceptable **#3** She worried that Quinn would be mad at her if she told her mom the truth. 4 Not Acceptable 1 2 3 5 Very Acceptable What would you do in this situation? _____

Data

Figure 1

Mean Acceptability of Motives for Scenario 1.

The blue bar represents the average acceptability rankings for motive 1, the red bar represents the average for motive 2, and the yellow bar represents the average for motive 3. Error bars indicate standard error of the mean for each motive.



Table 1

Differences in Acceptability between Motives Calculated with Post-hoc Tukey t-tests.

p-values between the motives in scenario 1.

m1 vs. m2	p < 0.01
m2 vs. m3	p < 0.01
m3 vs. m1	p < 0.01

Figure 2

Mean Acceptability of Motives for Scenario 2.

The blue bar represents the average acceptability rankings for motive 1, the red bar represents the average for motive 2, and the yellow bar represents the average for motive 3. Error bars indicate standard error of the mean for each motive.



Table 2

Differences in Acceptability between Motives Calculated with Post-hoc Tukey t-tests.

p-values between the motives in scenario 2.

m1 vs. m2	nonsignificant
m1 vs. m3	p<0.01
m2 vs. m3	p<0.01

Figure 3

Averages for Motives in Scenario 3.

The blue bar represents the average acceptability rankings for motive 1, the red bar represents the average for motive 2, and the yellow bar represents the average for motive 3. Error bars indicate standard error of the mean for each motive.



Table 2

Differences in Acceptability between Motives Calculated with Post-hoc Tukey t-tests.

p-values between the motives in scenario 3.

m1 vs m2	p<0.01
m1 vs m3	p<0.01
m2 vs m3	nonsignificant

Analysis

For scenario 1, I found that there was a significant difference in acceptability for each motive. Specifically, motive 2 (*Sally knows Maggie is skipping the test because she didn't study the material, but Sally doesn't want her to get in trouble because she's her friend*) was significantly less acceptable than the other two motives (#1: Sally knows Maggie is having family trouble and hasn't had time to prepare for *the test*, (difference=1.492, p<0.01) and #3: Sally doesn't know where Maggie is, but she's covering for *her because she assumes Maggie would do the same for her*, (difference=0.4884, p<0.01)). This data would suggest that people were more willing to lie for someone if they believe they have a reasonable excuse. When going over the write-in responses for this scenario, I found that an overwhelming majority of the responses said they would either not say anything to the teacher to begin with, or they would say that they did not know where Maggie was. According to the data, the participants found it significantly more acceptable for Sally to lie if she knew Maggie was having family trouble, or that she knew Maggie would cover for her in a similar situation, however, in the write-ins, there is a disparity in the selflessness displayed in the scenario versus how the participants responded. Most indicated that they would act in a way that would not implicate them in any sort of wrongdoing by not speaking to the teacher at all, or saying they did not know where Maggie was.

For scenario 2, there was no statistically significant difference in acceptability between the first two motives, however, the third motive (*Brad knows Sam barely ever does his homework (simply be-cause he doesn't want to), but since Sam is his friend, Brad feels obligated to help him*) was found to be significantly less acceptable than the first two (difference between motives 1 and 3=1.3016, p<0.01, difference between motives 2 and 3=1.1046, p<0.01). For the write-in portion, I found that there seemed to be a consistent middle ground throughout the responses. Though a few said they would let their friend cheat, and an almost equal number said they would flat out say no or say they had not done the homework, many responded with more thoughtful alternatives. Though the wording was varied throughout the responses, many said they would try to do the assignment with their friend instead of simply giving him the answers because they didn't believe in blatant cheating. Along with this, some said they would give their friend the answers but would try and explain the homework as well so their friend is not confused in the future.

For scenario 3, though there are significant differences between motives 1 and 2 (difference=0.3082, p<0.01), as well as 1 and 3 (difference=.5059, p<0.01), the numerical differences are relatively small, and all the average ratings are relatively high in comparison to the other scenarios (motive 1 avg.=3.7326, motive 2 avg.=3.4244, motive 3 avg.=3.2267), so it seems that all the motives in scenario three were generally acceptable. In the write-in portion, an overwhelming majority stated, confidently, that they would lie for Quinn. A few specifically said that since this wasn't an instance of academic dishonesty, they would be more willing to lie. Most said that they would definitely lie to Quinn's mom, but would be sure to check in with Quinn later to get the full story. Along with this, a handful of people said that they would try to avoid Quinn's mom altogether, and although that possibility didn't really fit within the parameters of the scenario, that sense of self-preservation from scenario one was seen through these responses. Finally, the people that said they would tell Quinn's mom the truth gave responses that brought Quinn's safety into question, and how lying to her mom would be wrong if there was a possibility that Quinn was not safe with Riley.

Conclusion

From the first scenario, I concluded from the quantitative results that Berkeley Carroll high school students found it most acceptable to lie for a friend if they knew they had extenuating circumstances, such as family trouble (avg.=3.8, p=<.01). However, in the written response, the majority of participants stated they would rather not implicate themselves in a lie, even if it was for a friend. This discrepancy highlights the difference between willingness to tell other-serving vs. self-serving lies, and this is apparent in the other scenarios as well. Essentially, the results from the first scenario indicate that Berkeley Carroll high school students are, more often than not, unwilling to tell other-serving lies, even when they've indicated that they believe the circumstance warrants a lie they deem acceptable.

Though the second scenario has similar motives to the first scenario, the write-in responses overwhelmingly convey a more selfless perspective. Though the majority of the participants ranked letting Brad cheat off of Sam's homework moderately acceptable if he is exceptionally busy, or if Brad has let Sam copy his homework in the past, the written responses indicated a strong aversion towards cheating. Most said they would help Brad work through the problem rather than letting him copy the answers without understanding the content. In the first scenario, lying about a person's excuse for missing an exam does not allow for the same opportunity to help the person in need as scenario two does. The first includes telling a lie directly to someone, while the second deals with copying answers, which are two different types of dishonesty. While the first scenario doesn't allow for many options in which both Maggie and Sally win, the second gives Sam a chance to stop an act of dishonesty before it happens, which majority of the participants indicated there are circumstances in which allowing a friend to cheat is acceptable, Berkeley Carroll high school students would rather avoid academic dishonesty by helping their friend comprehend the material instead.

Finally, for the third scenario, I concluded that, in a social situation, a significant number of Berkeley Carroll high school students found lying about a friend's whereabouts (to that friend's parent) acceptable. This was clearly reflected in the written responses, where most participants stated simply that they would lie for Quinn. The only notable difference amongst the written responses was from those who said they would not lie to Quinn's mom if they were worried about Quinn's safety. For teenagers, lying can be seen as a way of retaining autonomy from parents (Jensen et. al., 2004), so it makes sense that Berkeley Carroll high school students would find it acceptable for Quinn and Ali to lie, and the discrepancy with concern to safety also makes sense when looking at results from scenario two, where many students indicated concern for Brad's understanding (well-being) in the long run. These results indicate that, in social settings, lying is generally deemed acceptable when someone's safety is not a genuine concern.

Discussion

Due to the nature of this study, an area which there could be concern for bias is free-response interpretation. Since these responses were almost completely unique from one another in phrasing, word choice, etc., much of the interpretation is left up to my understanding of the responses, and what I believe the participant was attempting to communicate through their response. Along with this, at least a few responses for each scenario stated they would not tell a lie at all, no matter their motive or motive. Furthermore, it is possible that some participants could have overstated their honesty since students do not always trust the anonymity of Google forms. In a follow-up study, I would want to find a way to create a more open-ended experiment in which participants have the option to avoid commiting a dishonest act. That would allow me to see in what scenarios participants would prefer not to implicate themselves whatsoever, even if it would mean they are abandoning the opportunity to help another person. Finally, I would like to design a study that would look more into the acceptability of lying in different social situations, since I believe the parameters for those are somewhat more complex than academic dishonesty. One way I could do this is by possibly trying the experiment used by Garrett et. al. in "The brain adapts to dishonesty," in which they have participants act as "financial advisers" to a partner in another room. They then give their partner a rough estimate on how much money they believe is in a jar, and they are incentivized to overestimate or underestimate based on whether they or their partner will be given a prize or receive some type of punishment. One adjustment that could be made is changing the punishments or rewards to be more realistic for teenagers rather than adults (e.g. a chocolate muffin as a reward or "mandatory afternoon study hall" (MASH) as a punishment). In this way, I'd be able to look more specifically at Berkeley Carroll students and what they believe is valuable enough to lie for.

References

- Bolin, A. U. (2004). Self-control, perceived opportunity, and attitudes as predictors of academic dishonesty. *The Journal of Psychology*, 138 (2), 101-114.
- Garrett, N., Lazzaro, S. C., Ariely, D., & Sharot, T. (2016). The brain adapts to dishonesty. *Nature Neuroscience, 19* (12), 1727-1732.
- Jensen, L. A., Jensen, J. A., Feldman, S. S., & Cauffman, E. (2004). The right to do wrong: Lying to parents among adolescents and emerging adults. *Journal of Youth and Adolescence, 33* (2), 101-101.

The Effect of Attack Ads on Political Campaigns

by Noah Finkelstein

Abstract

tried to determine whether or not attack ads negatively influence U.S. political campaigns by examining how U.S. Senate candidates' poll numbers changed after attack ads aired. I averaged opinion polls for each attack ad issuer within one month before and after each ad to calculate the change in poll numbers. When a candidate issued an attack ad, it resulted in an average increase of 1.26 percentage points (+/- 0.84). I concluded these results were not statistically significant (p = 0.15) because the measured effect was not significantly different than the mean hypothetical value of zero.

Background

Since the dawn of American politics, the goal of a campaign has always been the same: secure more votes than the other side. However, there have been a wide variety of political strategies to accomplish this goal: voter registration drives, canvassing voters in a neighborhood, and campaign speeches to name a few (Lerche, 1948). In 1800, Adams and Jefferson both appealed to specific delegates because no popular vote existed at the time (Lerche, 1948). Though little in modern politics resembles that election from two centuries ago, there is a strategy that has endured: the smear campaign. Political strategists understand that it is not enough to talk about why one's own candidate is right for the job, they must also describe why the opposing candidate is wrong. Their campaign has a two-pronged goal: fire up their supporters while detracting their opponents. Within this political messaging, we can find a very powerful psychological phenomenon that influences voters' perceptions, known as the *halo effect* (Thorndike, 1920).

The *halo effect* is the idea that our global evaluations of a person affect how we judge specific aspects of their character (Nisbett & Wilson, 1977). Psychologists define a global evaluation as an identification of the nonspecific essence of a person or thing; a few examples of global evaluations are whether a person is seen as likable, attractive, or trustworthy (Nisbett & Wilson, 1977). Unlike specific evaluations, global evaluations are more qualitative than quantitative. In addition, global evaluations are shared by a majority of evaluators (Nisbett & Wilson, 1977).

Though Edward Thorndike is credited with identifying the *halo effect*, University of Michigan researchers Richard Nisbett and Timothy Wilson were the ones who popularized the concept (Thorndike, 1920). In their landmark study, two groups of college students watched two separate videos of the same professor teaching and interacting with his students. In one video, the professor answered all of the questions in a generous and informative way and in the other video, the professor acted rude and unhelpful towards the students. After watching the lesson, the subjects were then asked to rate the

professor on three specific attributes – his mustache, his accent, and his mannerisms – none of which related to his competency as an instructor. The group that saw the nice professor rated these attributes significantly higher than the group who witnessed the rude professor, showing how a global evaluation (the professor's attitude) influenced how subjects viewed specific attributes (his mustache, accent, and mannerisms) of the professor (Nisbett & Wilson, 1977).

Not confined to university campuses, the *halo effect* has also been studied in a political context. With his focus on character-based attacks aimed at political candidates, researcher Scott Clifford investigated the role of moral principles in elections. In the second of three separate experiments. Clifford examined how a politician's policy decisions influence how potential voters perceive the moral principles of that politician (Clifford, 2014). Clifford recruited subjects through Amazon's Mechanical Turk (MTurk); on this technological platform, individuals can outsource a task to a network of workers, which is an increasingly common and relatively diverse way to acquire subjects for political science research (Arceneaux, 2012; Berinsky, Huber, & Lenz, 2012). In Clifford's second experiment, the subjects were split into two groups based on whether they strongly supported or strongly disapproved of the death penalty. The subjects were given limited information about Sam Brownback, a Republican governor in Kansas who opposes the death penalty; they were then asked to rate how well certain character traits applied to him. When comparing the two groups, Clifford found that the subjects who were against the death penalty (as Brownback was) perceived that the governor possessed higher levels of two key traits: "authority" and "care." Subjects who were in favor of the death penalty perceived that Brownback possessed lower levels of "authority;" however, there was little change in this group's perception of his "care." (difference =, p < 0.01) (Clifford, 2014). Clifford's experiment confirms that for voters, policy stances bleed into moral principles. In addition to examining "care" and "authority," Clifford also demonstrates how other attributes, such as compassion and leadership, are affected by a particular policy stance. This is an extension of the halo effect; instead of a global evaluation affecting the perception of a specific attribute (e.g. that person is likeable, so he must be intelligent), it is an agreement or disagreement on policy that influences the perception (e.g. she favors enhanced interrogation, so she must be evil).

In his examination of the 1996 Taipei Mayoral election, French researcher Eric Sautedé observed a number of growing trends that made this election different from past ones in Taiwan, specifically how candidates amplified their own moral principles. In this election between Ma Ying-jeou and Chen Shuibian, both candidates spent unprecedented sums of money (\$310 million each) on marketing campaigns to both enhance their own reputation and attack their opponents (Sautedé, 1999). On all of his campaign posters, Ma Ying-jeou depicted himself running in a red, white, and blue silhouette to match the Taiwanese flag. This patriotic image not only appealed to conservatives, but it also showed off his "vitality and dynamism" (Sautedé, 1999). For a campaign slogan, Ma Ying-jeou chose "Taipei first, Taiwan first," which was designed to distract the voters from the fact that he was born in mainland China. Ma Ying-jeou was also aided by the current president, who branded Ma Ying-jeou's campaign "new Taiwanese" (Sautedé, 1999). Ma Ying-jeou applied a third slogan, calling himself "xiao Mage," which translates to "little big brother Ma" (Sautedé, 1999). All of these marketing tools created the image of Ma Ying-jeou's closeness to Taiwan, both as an active politician and as a family member. Creating this familial image is something that has been done in many countries around the world, serving to make voters feel personally tied to the candidates (Sautedé, 1999).

Even more revealing than the self-branding were the candidates' attacks on each other. During the race, a rumor emerged that the progressive candidate Chen Shui-bian had allegedly visited prostitutes in the Macau, China. Despite these allegations, the conservative candidate, Ma Ying-jeou, decided not to

attack the incident, but rather to question his opponent's character. Ma Ying-jeou's campaign leaflets changed Chen Shui-bian's first name to "suibian," which means "thoughtlessly" (Sautedé, 1999). In response, Chen Shui-bian nicknamed Ma Ying-jeou "mamahuhu" (meaning "so-so") in an attempt to characterize his opponent as mediocre and lukewarm. Instead of referring to specific events or damaging rumors, the candidates created catchy nicknames to smear their character. The *halo effect* is evident in both smear campaigns: the intention was to connect Chen Shui-bian's ignorance and Ma Ying-jeou's mediocrity to the specific trait of their respective fitness for the mayor's office. In the end, Ma Ying-jeou beat Chen Shui-bian by 5.22% of the vote (Sautedé, *1999*).

A common criticism of attack campaigns in politics is that these campaigns decrease voter turnout for elections and increase the hatred of politics in general (Jackson, Mondak, & Huckfeldt, 2009). In response to the latter argument, a 2009 study by researchers Jackson, Mondak and Huckfeldt attempted to determine whether attack ads would damage voter opinions of the U.S. political system during the 2002 midterm elections. The researchers called nearly 2,000 voters across the country and polled them on their approval of Congress, the performance of Congress, their opinions of House Democrats and Republicans, and the perceived efficacy inside and outside of Washington, D.C. They correlated the voters' opinions with their exposure to ad campaigns, considering the number of positive and negative ads that they would likely have seen. The researchers found no significant effects of increased exposure to attack ads in any of these dependent variables. This conclusion suggests that critics may be giving attack ads too much credit for souring voters on the election process (Jackson, Mondak, & Huckfeldt, 2009).

My experiment focused on U.S. Senate races in the 2018 midterm elections, and on the attack ads put forth by various campaigns. Attack ads are designed to attach a negative *halo effect* to an opponent, so I was curious how strongly an attack ad transferred this negativity. I tried to determine whether attack ads had an effect on public opinion in terms of poll numbers for U.S. Senate campaigns. The *halo effect* is a mental phenomenon; therefore, my experiment examined the opinions of voters rather than voter turnout. I attempted to determine the influence of attack ads on voter opinion by identifying how Senate candidates' polling numbers changed in the one-month aftermath of their respective attack ads.

Procedure

In order to gather content for the study, I found advertisements on the official YouTube channels of the various 2018 U.S. Senate campaigns. I only found three races that had both candidates issue a minimum of three attack ads: Florida (Bill Nelson v. Rick Scott), West Virginia (Joe Manchin v. Patrick Morrissey), and New Jersey (Bob Menendez v. Bob Hugin).

I defined what constitutes an attack ad with three criteria:

- 1. The opponent must be the subject of the ad;
- 2. The opponent must be referenced by both name and picture;
- 3. The ad cannot have been issued by a third party, such as a Super PAC (Political Action Committee). A Super PAC is a political group that can receive an unrestricted amount of capital from "individuals, corporations, labor unions and other political action committees" as long as they reside within the US (Registering as a Super PAC, 2019).

I deemed the third criterion particularly important because it ensures that voters associate the ads directly with a single campaign, not a third party organization. Under the Bipartisan Campaign Reform Act of 2002, official campaign ads must feature a candidate's photo and voice at the end of the message

approving of the advertisement; this helps viewers to distinguish between messages from multiple PACs and those from the actual candidate (Jackson, Mondak, & Huckfeldt, 2009).

Also, Super PACs typically advocate for very specific issues, and might not be an accurate representation of the campaign as a whole. They don't have to coordinate with the campaign they are helping, as their interest is served by showing the power of the Super PAC and their voters (Registering as a Super PAC, 2019).

To find the polls that correspond to each of the ads, I collected polling data from the website Real-ClearPolitics. I chose this aggregator site because it featured links for polls conducted during each Senate campaign, organized chronologically. RealClearPolitics was useful because the site is transparent about its polls' Democratic or Republican leanings; this ensured that the polls I used were reliable and not skewed by partisanship. The site also provides information about sample size, which I used to calculate the uncertainty involved.

Just as criteria was set for the attack ads, I also set limits for the polling data. In order to make sure that polling data was influenced by a specific attack ad, I included only polls conducted within a month of the ad's air date, which made it difficult to use attack ads in the slower, early stages of the election cycle.

For each attack ad issuer, I averaged the 3 nearest preceding opinion polls and subsequent opinion polls as long as they were within 1 month of the air date. I defined the difference between those averages as the change in poll numbers.

Data Analysis

The differences associated with each attack ad were put into a single sample two-tailed t-test with a hypothetical mean of zero. A statistically significant result would find that the candidate's approval percentage changed after the attack ad aired. However, the t-test showed attack ads to not have an effect on polling (p = 0.147).

Figure 1

Change in Poll Numbers after Attack Ads.



After the attack ad, the candidate was on average 1.26% points better off in the polls (SEM of +/- 0.84). Measured in a single sample two-tailed t-test against a hypothetical mean value of 0, the results were not significant (p = 0.15).

Interpretation/Conclusion

I found no significant evidence to support the conclusion that attack ads in U.S. Senate campaigns have an effect on poll numbers. This lack of evidence further disputes the assumption that attack ads are influential.

After finding the results from the main experimental question, I examined the effect of several other variables associated with attack ads on poll numbers. These included differences in the dates when the attack ads aired (before August vs. after October), whether the ad issuer was the incumbent or the challenger, and whether the ad issuer eventually won or lost. None of those criteria resulted in a statistically significant difference when subjected to the same single sample two-tailed t-test as the main experimental question. In the end; the main experimental question tested with the lowest p-value.

When analyzing the study as a whole for ways to improve upon it in the future, it became clear that a study would need to include more attack ads for a more definitive answer to the experimental question. Over the course of the three Senate campaigns I had chosen, I found a total of 27 attack ads issued by the candidates, which may be too small a sample size to suggest a significant difference even if one is present.

References

- Arceneaux, K. (2012). Cognitive biases and the strength of political arguments. American Journal of Political Science, 56 (2), 271–85.
- Berinsky, A. J., G. A. Huber, & G. S. Lenz. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis*, 20 (3), 351–68.
- Clifford, S. (2014). Linking issue stances and trait inferences: A theory of moral exemplification. *The Journal of Politics, 76* (3), 698-710. Government Publishing Office. (n.d.). Retrieved from www.govinfo.gov/content/pkg/PLAW-107publ155/html/PLAW-107publ155.htm
- Jackson, R. A., Mondak, J. J., & Huckfeldt, R. (2009). Examining the possible corrosive impact of negative advertising on citizens' attitudes toward politics. *Political Research Quarterly*, 62 (1), 55-69.
- Lerche, C. (1948). Jefferson and the election of 1800: A case study in the political smear. *The William and Mary Quarterly*, 5(4), 467-491
- Nisbett, R. E., & T. D. Wilson. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35* (4). 250–256. Registering as a Super PAC. (n.d.). Retrieved from www.fec.gov/help-candidates-and-committees/filing-pacreports/registering-super-pac/
- Sautedé, E. (1999). Political marketing in latest style. French Centre for Research on Contemporary China, 21 (1), 38-41.
- Talamas, S. N., Mavor, K. I., Perrett, D. I. (2016). Blinded by beauty: Attractiveness bias and accurate perceptions of academic performance. *Plos One*, 11 (2).
- Thorndike, E.L. (1920). A constant error in psychological ratings. Journal of Applied Psychology, 4(1), 25-29.

Shoaling Preferences of Wildtype Zebrafish Trained to Associate with GFP Zebrafish

by Eugenie Haring

Abstract

his study attempts to find whether shoal preference training affects the association time (sec) of conspecifically reared wildtype focal (WF) zebrafish with two separate stimuli shoals. WF were trained to prefer a shoal of green fluorescent protein modified (GFP) zebrafish stimuli over the wildtype (WT) zebrafish stimuli they were predisposed to shoal with. WF were reared with other WF to ensure imprinting, meaning they should instinctively shoal with WT which is supported by preliminary data. WF were trained using a T-maze tank, live stimuli, and a food reward, to associate the GFP with a positive stimuli (food). After training, WF were tested under the same conditions and association time with either shoal was measured. Color of shoal was the independent variable and association time (seconds) was the dependent variable. Thus far, WF tested after seven days of training showed no alteration in preference (WF spent more time with WT than GFP) but the time spent with WT decreased from pre-training data.

Background

When small fish, including the zebrafish (Danio rerio), choose to shoal (join a social group), two phenomena are at play. The first, the 'confusion effect', states that it is harder for predators to target a single fish in a large shoal of similar fish (Snekser, McRobert, Murphy, & Clotfelter, 2006). The second, the 'oddity effect,' states that fish who physically stick out in a shoal are much more likely to be targeted by a predator (Snekser et al., 2006). It can therefore be assumed that given the choice and self-awareness, fish would choose a shoal they perceive to be similar to them. Zebrafish have been shown in fact to make shoaling choices based on physical appearance alone (Rosenthal, & Ryan, 2005) including both color (Park et al., 2016) and pattern (Rosenthal & Ryan). Rosenthal et al. found that conspecifically reared zebrafish chose to shoal with the fish most physically similar to themselves based solely on pattern (striped vs. bare). But it has also been shown that shoal preferences are imprinted upon fish and are not innate, meaning that zebrafish will choose to shoal with the fish that they are most familiar with or have the most experience around and not the fish that they themselves are most similar to (Spence, & Smith, 2007). Spence and Smith cross-reared different color zebrafish strains and then performed a shoal preference test and showed that zebrafish choose to shoal with the fish they were reared with regardless of their phenotypic similarity showing that it is possible to imprint shoal preference via rearing. Though there is abundant research on zebrafish shoal preference, there is hardly any research on using

behavioral training to change fish shoal preference. I am interested in whether it is possible to train already-imprinted wildtype zebrafish (WF) to prefer a heterospecific green fluorescent protein (GFP) zebrafish over the physically identical wildtype zebrafish stimuli (WT) shoal using a T-maze tank and food rewards. Most studies involving T-maze tanks and zebrafish are researching the effects of drugs and/or chemicals, but I used the T-maze tank as a way to force the fish to make a shoaling decision. I hypothesize that the wildtype focal fish (WF) with an established preference for a wildtype stimuli fish (WT) shoal can be trained to shoal with a green fluorescent protein modified zebrafish (GFP) shoal over a WT shoal.

Methods

Animal Subjects

Wildtype and GFP focal and stimuli zebrafish were obtained from PetSmart (WT n = 4) (WF n = 9) (GFP n = 4). WF were housed apart from WT and GFP stimuli fish in a 10 gallon tank with a clear acrylic partition separating the tank into pre- and post-training session sections. WF were housed together to ensure imprinting. WT and GFP stimuli were housed in a second 10 gallon tank. The two tanks were physically separate from each other so that there could be no interaction between the WF and WT/GFP outside of testing/training. None of the zebrafish in this study were ever gendered. All fish were fed flake food on non training days; on training days, GFP stimuli were fed before the session as usual while WF were fed individually in T-maze tank. This study on live subjects was granted approval by the Berkeley Carroll Institutional Review Board (IRB) as well as Dr. Jennifer Snekser, a biologist well versed in scientific research involving zebrafish and their well-being.

Testing and Training Tanks

Three 10 gallon aquaria (51.3 x 26.7 x 33 cm) were adapted for this study. The T-maze tank for testing and training was constructed by securing four pieces of white acrylic (two 20cm x 22cm, two 10cm x 22cm) in the form of the letter T to the aquarium (see Figure 1 for details). Threshold lines were drawn on the outside bottom of the tank with permanent marker in order to define association with either shoal: if the WF passes the line into that end of the T it counts towards association time with that shoal. The tank was set on top of a white surface and white laminated sheets of paper were placed on exposed outer walls of tank so that the entire T-maze was one color. This was done so that the stimuli fish placed in either end of the T would stand out and thus their appearance could be assumed to prompt the WF's decision. During the testing and training sessions, stimuli fish were placed in cylindrical glass containers (r = 5.5in, h = 14in) and had no olfactory communication with the WF. WF were housed in a 10 gallon tank with a clear acrylic partition dividing the tank unevenly into $\frac{1}{3}$ and $\frac{2}{3}$. Fish spent the majority of their time in the larger section but after each WF was individually trained or tested it would be temporarily placed in the smaller section until all fish had undergone training or testing and could be moved back to the larger section. This was done so that the same fish would not be trained or tested twice. The stimuli fish were housed together in an unmodified 10 gallon tank and were moved into separate glass containers at the start of each training and testing session. All tanks were filtered and heated but for the T-maze tank, all equipment was removed during sessions.

Behavioural Observations and Training

WF were trained to associate food with GFP stimuli via daily food-training sessions. WF were individually relocated to the T-maze tank and fed usual flake food while the GFP stimuli swam in their cylindrical glass container on one end of the T-tank. The GFP container was switched from one side to the other between feeding-training sessions to ensure that WF were being trained to associate the positive reward with the presence of GFP fish and not the left or right side of the tank (Figure 1.2). Food-training sessions were 180 seconds long to ensure that WF had plenty of time to eat food and observe surroundings and

Figure 1

T-Maze Testing and Training Tank.

White acrylic pieces (represented by white thin rectangles in blue section) were glued into T shape using aquarium cement. The entire tank, including sectioned off areas, was filled with water to maintain equal pressure on acrylic pieces and prevent any sort of buckling. Dotted lines represent drawn-on thresholds and black-outlined thin rectangles on outer edges of tank represent laminated paper. GFP and WT stimuli are shown in cylindrical glass containers and WF is shown in "stem" of tank in starting position.



Figure 1.2

Training Configuration.



WF were food-trained once a day for 7 days. After training, WF underwent testing. The preliminary and final testing procedures were identical to one another. WF were placed into the stem of the T-maze tank and both WT and GFP stimuli fish were placed separately into cylindrical glass containers which were placed in either end of the T. Location of GFP container was opposite of last training session. The 120 seconds long sessions were video recorded and association time was measured. It has yet to be determined whether initial movement/decision will taken into consideration when analyzing shoal preference.

Statistical Analysis

For each data comparison, I conducted a one tailed correlated t-test. The test was correlated because I am measuring two association times for each fish in each test.

Results

Figure 2

The preliminary testing revealed that the WF did in fact have a predisposition to shoal with the WT as predicted by the confusion and oddity effects as well as the nature of imprinting (p = 0.0018) (Figure 2). The 7 days of training did not yield any difference in the WF's preference for WT over GFP. WF spent significantly more time on the WT side than the GFP side, as they had in preliminary training (p = 0.027) (Figure 3). However WT pre- and post-training times were compared and the WF spent significantly less time on the WT side after training (p = 0.037) (Figure 4). The same was done for GFP but no significant difference was observed (p > 0.05) (Figure 5).

Preliminary Testing p = 0.001767100 75 50 25 0 Wildtype Side GFP Side
Figure 3

Post-Training Testing

p = 0.026445



Figure 4

Association Time with WT Shoal: Pre- vs. Post-Training.

p = 0.037254



Figure 5

Association Time with GFP Shoal: Pre- vs. Post-Training.

p = 0.254856



Discussion

This study will hopefully be the first iteration of an attempt to create an effective method for training zebrafish to make social choices based off visual cues alone. However, after the first 7 days of training, I accepted my null hypothesis which states that there is no difference in WF's shoal preference from preliminary to final testing. There were several areas of uncertainty in this study. One of the aspects of ichthyology I am most unfamiliar with is the olfactory communications that fish use. While I controlled for communication between the stimuli and the WF, I don't know what kind of communication/scent the WFs were leaving behind in the water since I did not filter water between sessions; for example, would the next fish to be trained know if the previous WF spent more time on the WT time? Additionally, I did not account for acclimation time, as many studies I read had talked about. In each session, the WT was transferred to the T-maze tank and training or testing started immediately from there. From a timing and logistical standpoint this was convenient because it was fast and I didn't have to worry about having the WF in the T-maze and then putting the glass stimuli containers in. However, this could account for the darting (back and forth between shoals) behaviour that many fish displayed.

While the statistical results of the study show that the WF have not been significantly influenced by my training, the results of the pre- vs. post-training data is promising. In the future, I plan to experiment with several different methods of training, possibly using a different food delivery system or creating a lock system to trap the fish on the GFP side after it has chosen to shoal with them or just extending the training period. As of now, I am extending the study so that I can train the WF for 28 consecutive days. I also would like to take a closer look at the individual fish behavior. For instance, when I was doing preliminary testing, it became clear that when I introduced the net into the tank to retrieve the fish after time was up, several WF would swim over to the WT side of the tank. The net was always introduced to the middle of the tank—it was never pushing the fish to one side or another when it was first introduced. I continued to keep track of this reaction as I trained the fish but there were no

noticeable changes. I also kept track of the WF's initial shoal choice for both the pre and post training. In both testing sessions 6 out of the 9 WF chose the GFP side to shoal with first. This initial instinct as well as the idea of first choice is something I would like to delve into deeper in my coming iterations. As for this study, the lack of a significant alteration in WF preference pushes me to look for other ways to induce an affinity for the GFP shoal without imprinting the WT. However, it may actually be impossible to train the WF against their imprinted preference in which case no adjustment of experimental design could sway them otherwise.

References

- Engeszer, R. E., Ryan, M. J., & Parichy, D. M. (2004). Learned social preference in zebrafish. *Current Biology*, *14* (10), 881-884.
- Gentile, D. (2018, January). [Personal interview by the author].
- Park, J. S., Ryu, J. H., Choi, T. I., Bae, Y. K., Lee, S., Kang, H. J., & Kim, C. H. (2016). Innate color preference of zebrafish and its use in behavioral analyses. *Molecules and Cells*, 39 (10), 750.
- Rosenthal, G. G., & Ryan, M. J. (2005). Assortative preferences for stripes in danios. *Animal Behaviour, 70* (5), 1063-1066.
- Snekser, J. L., McRobert, S. P., Murphy, C. E., & Clotfelter, E. D. (2006). Aggregation behavior in wildtype and transgenic zebrafish. *Ethology*, 112 (2), 181-187.
- Spence, R., & Smith, C. (2007). The role of early learning in determining shoaling preferences based on visual cues in the zebrafish, Danio rerio. *Ethology*, *113* (1), 62-67.

The Effect of Mindfulness on Problem Solving Ability in High School Students

by Matthew Justh

Abstract

his study looks to find if a 5 minute session of mindfulness before an assessment has an affect on problem solving ability in high school students. To do so, the researcher created two problem solving tests of similar difficulty with three sections and tested them on high school students ranging in age from 14 to 18. The results supported the researcher's initial hypothesis that there would be a significant difference in problem solving ability between when students practiced mindfulness and when they did not (p value = .0024).

Background

Meditation is an ancient practice that predates written history. It originates in prehistoric religions that involved rhythmic chanting, known as mantras. It was first cataloged in the Vedas, a set of Sacred Hindu texts around 1700 BCE (Bushak, 2016). Meditation itself is not a specific action, but rather a group of actions that aim to make one more conscious and able to see reality (Bushak, 2016). Among the group of practices in meditation is mindfulness, which is the focus on the present, specifically on an object or action (Mindful, 2018). Buddhists believe that through this mindful meditation one can achieve a better grasp of reality and get one step closer on the road to true clarity, or *nirvana* (Bushak, 2016).

Despite its religious roots, mindful meditation has become a large trend in secular Western society over the last twenty years (Dam, et al. 2017). More and more, people are regularly practicing mindful meditation separated entirely from any religious practice. This has been attributed to recent evidence pointing to multiple health benefits related to mindfulness, including stress reduction, increased relationship satisfaction, and improved focus (Davis & Hayes, 2011). Studies have also found that habitual mindfulness can reduce blood pressure in subjects with prehypertension (higher blood pressure than desirable) (Hughes et al., 2013). These benefits, paired with blog posts and social media updates spreading awareness, have led to the phenomenon we see today.

Though the health benefits of mindfulness have been well documented, it is less well known if mindfulness is beneficial for problem solving. However, a few studies have done research on this.

A study in the journal "Psychonomic Society" attempted to look at the effect of mindfulness on problem solving (Dillbeck, 1982) by establishing two groups of subjects: one group that completed regular mindful meditation, and one that did not. Each subject completed two visual perception tests. One of these tests involved letter identification, which had subjects report how to spell nonsense words they had just seen. Another looked at playing card identification, where subjects had to report on playing cards they had just seen. For problem solving, an anagram test was used. Subjects had to find an English word out of a five letter anagram. From the data, the researcher found that there was a significant

difference in visual perception average test score between the meditation group and the control group, with the former performing better. The results were also significant on the card test; once again the meditation group performed better. There was no significant difference, however, on the anagram test between the control and the mediation group. This study showed that there are benefits to visual perception from mindfulness but was inconclusive on problem solving (Dillbeck, 1982).

A more recent article entitled "Meditation Promotes Insightful Problem-solving by Keeping People in a Mindful and Alert Conscious State" from Chinese Life Sciences, also dealt with this question (Ren et al. 2011). The researchers hypothesized that a 20 minute session of mindful meditation would increase problem solving ability. They tested this by having three groups of 16 subjects attempt to solve 10 insight problems, problems which force one to change perspective in order to find a solution. After this test was complete, two of the groups were instructed to mindfully meditate for twenty minutes. One group was told to count to 10 breaths in their head and then restart, while another was told to count to 100 breaths in their head and then restart. The control group waited outside the test room. After the twenty minutes, all the subjects were then shown the problems they originally could not find the solution to and were asked to try and solve them again. There was no significant difference in the number of problems the three groups of subjects completed in the first test (before the meditation/breathing exercises). However, the meditation groups performed significantly better in their problem solving rate and number of questions finished (P<0.05). From this the researchers concluded that meditation adds better insight and can improve problem solving (Ren, et al. 2011).

Some teachers are now practicing mindfulness in the classroom, with many different expectations for the effect it may have (Garey, (2107). Beyond potentially improving problem solving, teachers are also looking for students to be more ready to engage with class materials and to be less distracted in general (Garey, (2107).

As mindfulness has become more mainstream, digital applications have sprung up offering guided mindful meditation, as opposed to a more traditional unguided style (Types of Meditation, 2018). The "guide" is often a pre-recorded voice, instructing the practitioner to focus on something specific and on breathing. Beyond personal preference, there is not much evidence on which of these methods is more effective.

The current study looks to grapple with the question "What is the effect of focused facilitated mindful meditation on problem solving ability?" This study differs from those previously mentioned in that it is within subject, examines results directly after mindfulness, and testing is done on high school students specifically.

Procedure

This study looks to understand the direct effects of mindfulness on problem solving ability in high school students. Subjects were acquired through an email that reached the whole of the upper school. This email explained to students that, if they chose to participate, they would come to one five-minute and one ten-minute session. Details on the actual study and problem solving tests were kept to a minimum to help ensure the subject pool was not biased. Once the subject pool was acquired, subjects were divided into two groups, split evenly by gender and age. Two groups are needed due to the fact that two tests of similar difficulty are required for the study to be effective. This study looks to measure if there is a change in problem solving ability when subjects practice mindfulness before an assessment. The control session and the mindfulness session must have tests with similar results without mindfulness or the results will not be conclusive. The tests themselves included three parts. The first section was

pattern recognition; these are often used to measure intelligence quotient (IQ Test Labs), but have also been found as an effective way to measure changes in problem solving ability (Dillbeck, 1982). The second section was verbal problem solving, which involved subjects attempting to make words out of anagram nonsense words. This is another type of problem solving that requires the subject to find the familiar in something unfamiliar (Dillbeck, 1982). The last section was two insight problems, defined above. They have been used in studies assessing problem solving in the past (Ren. et al. 2011) and are unique in the creativity they require to get the right answer. These three types of problems were selected because they all assess different type of problem solving: visual for the pattern recognition, verbal for the anagrams, and logical for the insight problems. In the first session, both groups took the problem solving assessment in a guiet environment. This was ensured by having subjects take the test in an empty classroom. Subjects had 5 minutes to complete the test to the best of their ability. After the preliminary data from the first session was collected, it was ensured that there was no significant difference between scores on the two tests (p=0.41). To attempt tro keep the results as accurate as possible, subjects completed the second session at the same time of day as the first. This specific time for both sessions ranged between subjects, which is fine in the case of this study as change in score, not score achieved, was being measured. For the second session, group one and group two took the problem solving tests again, this time switching which group took which test. In the second session, though, both groups practiced mindfulness for five minutes before the test. The following script was read to subjects before they began mindfulness: "Before you take this assessment, you will spend five minutes mindfully meditating. Jon Kabot-Zin describes mindfulness as "A means of paying attention in a particular way: on purpose, in the present moment, and non-judgmentally." While this may seem like a complicated or abstract idea, it is as simple as focusing your mind on your breath, closing your eyes and keeping your feet flat on the floor. If you have any questions please ask now." After answering any questions, the insight app timer was used to have subjects spend five minutes mindfully meditating. After the groups finished their five minutes of mindfulness, they completed the problem solving test, once again with five minutes to complete it.

Data Analysis

Table 1

Average Scores Achieved on Each Section by Each Group

Control Pattern Score	Mindful Pattern Score	Control VR	Mindful VR	Control Insight	Mindful Insight	Control Total	Mindful Total
2.06	2.41	6.29	6.94	.47	.88	8.82	10.24

Graph 1

Average Pattern Recognition Score

This graph shows the difference in pattern recognition ability (p = 0.054).



Graph 2

Average Verbal Reasoning

This graph shows the difference in verbal reasoning ability (p value = 0.085).



Graph 1, shown above, details all data collected so far in this study (n = 17). Data was analyzed using multiple two-tailed correlated t-tests. As noted in the procedure, the problem solving tests subjects took were broken into three sections: pattern recognition, verbal reasoning and insight problems. Of these individual sections, all average scores increased post mindfulness, but the only significant difference found was in insight problems (Graph 3, p=.049). The least significant difference was found in verbal reasoning (Graph 2, p=.085) and pattern recognition was marginally significant with a p value of .054 (Graph 1). The tests as a whole, though, showed an average increase in score of nearly 2 and a p value of .0024 (Graph 4). This showed that there was a significant increase in score received post-mindfulness

00

Graph 3

Average Insight Problem Score

This graph shows the difference in insight ability (p value = 0.049).



Graph 4

Total Test Score Average

This graph shows the difference in total test score (p value = 0.0024).



when compared to pre-mindfulness. The data was significant due to the tests being within subject, meaning they measured the change of each single subject's test scores, and not the averages of the group as a whole. Those averages were not significantly different (hence the wide error bars; which indicate +/- standard error of the mean).

The data shows a significant increase in problem solving scores after 5 minutes of mindfulness, and thus, the alternative hypothesis is supported. Despite this, two out of the three individual question types did not show a significant difference. Also, despite the overall trend, 3 subjects got the exact same score on their assessments with and without the mindfulness and 2 subjects performed worse after mindfulness. Even so, it can be seen that the mindfulness had an overall benefit on problem solving ability. However, this study is not fully complete and requires a larger subject pool to come to any definite conclusions. The data, so far, does lend to our growing understanding of the possible benefits to mindfulness. If the alternative hypothesis continues to be supported as more data is collected, then it can be concluded that this study adds evidence to support a more widespread use of mindfulness in the classroom. Once this study has been fully complete and more data is received, it can expand to look at habitual mindfulness and if it has an affect on problem solving ability.

References

Bonk, C. Insight Problems Retrieved from http://curtbonk.com/bobweb/Handout/d4.ips.htm

- Bushak, L. (2016, March 10). Mindfulness Vs Meditation: The Difference Between These Two Pathways To Well-Being And Peace Of Mind. Retrieved from www.medicaldaily.com/mindfulnessmeditation-differences-377346
- Dam, N. T. V., et al. (YEAR) "Mind the hype: A critical evaluation and prescriptive agenda for research on mindfulness and meditation." *Perspectives on Psychological Science* 0(0)
- Davis, D. M., & Hayes, J. A. (2011). What are the benefits of mindfulness? A practice review of psychotherapy-related research. *Psychotherapy*, 48 (2), 198-208. doi:10.1037/a0022062
- Dillbeck, M.C. Memory & Cognition (1982) 10: 207. https://doi.org/10.3758/BF03197631)
- Garey, J., (2107). New York Times, Los Angeles Times, American Library Association, & NPR Best. (n.d.). Mindfulness in the Classroom. Retrieved from https://childmind.org/article/mindfulness-in-theclassroom/
- A Guide to 7 Different Types of Meditation I Everyday Health. (2018, June 22). Retrieved from www.everydayhealth.com/meditation/types/
- Hughes, J. W, Fresco, D. M, Myerscough, R, Van Dulmen, M. H. M, Carlson, L. E, Josephson, R. (2013). Randomized Controlled Trial of Mindfulness-Based Stress Reduction for Prehypertension. *Psychosomatic Medicine*, *75* (8): 721 DOI: 10.1097/%u200BPSY.0b013e3182a3e4e5
- IQ Test Labs. (n.d.). Retrieved from www.intelligencetest.com/questions/precognition.html
- Ren, J., Huang, Z., Luo, J., Wei, G., Ying, X., Ding, Z., Wu, Y., Luo, F. (2011). Meditation promotes insightful problem-solving by keeping people in a mindful and alert conscious state Sci. China Life Science. 54: 961. https://doi.org/10.1007/s11427-011-4233-3)

What is Mindfulness? (2018, June 18). Retrieved from www.mindful.org/what-is-mindfulness/

Special Thanks:

Ms. Sefchovich, Diego, Max and Isabel without whom this paper would've been completed months earlier.

00

An Investigation of the Effects of Animal Presence on the Performance and Stress Levels of High School Students at Berkeley Carroll

by Ella Novogratz

Abstract

his study investigates whether the presence of common domesticated animals affects the stress and performance of students during a stressful situation. Students were given either a dog, hamster, fish, or turtle and made to rest and then count backwards from large numbers by smaller numbers while their heart rate and skin conductance were being monitored. Results suggested that there was no significant difference in heart rate between animal groups, but that students with hamsters had a significantly lower skin conductance response than students with a fish, turtle, or no animal.

Introduction

Pets are a common part of life in America. Many people consider their pets, domesticated animals such as dogs and cats, members of their family. There have been multiple studies done observing the effects of pets on their owners' health, including some that have measured their effects on stress. Allen and Blascovich (1991) used systolic and diastolic blood pressure, heart rate, and skin conductivity response as measures of physiological stress. They found that female subjects, with their pet dogs present, showed little or no physiological reactivity during performance of a stressful arithmetic task, whereas female subjects with their close friends present showed substantially greater physiological reactivity. Based on their findings the researchers hypothesized that subjects with their pets present were less psychologically threatened and that the presence of pet dogs during the stressful task provided social-emotional support that was critical to lowering physiological stress (Allen & Blascovich, 1991).

This idea, that dogs can lower stress, was further supported by another study done by Barker, Knisely, Barker, Cobb, and Schubert in 2012. They focused on the effects of dogs in the workplace. They took surveys and measured salivary cortisol, a steroid hormone produced by the adrenal cortex in response to stress (Crump & Derting, 2015). Their subjects were employees who were allowed to bring their dogs to work, employees who left their dogs at home, and employees who had no dogs. The researchers found that salivary cortisol was slightly increased in the group who had no dogs, and that participants who brought their dogs to work had significantly lower stress scores than other participants on the Stress Visual Analog Scale, a self-reported stress survey. The average score for participants who brought their dogs to work was 22.305, the average scores of participants who left their dogs at home was 36.025, and the average scores of participants with no dogs was 31.085. Their data supported the claim that the presence of a dog decreased stress throughout the day (Barker et al., 2012). Polheber & Matchock (2014) also used salivary cortisol and heart rate to measure stress levels in another study measuring the effects of dogs and human friends on stress. They met the participants in an unfamiliar lab environment, and exposed them to an unfamiliar therapy dog. To induce stress, they used the Trier Social Stress Test (TSST), which consists of a fake interview experience (designed to induce stress), and an arithmetic task. Participants in the Dog condition had an average cortisol level of 0.180 which was significantly lower than the average cortisol levels of participants in the control condition (0.296) and the friend condition (0.251). Furthermore, they found that the average heart rate of the participants with dogs was significantly lower than in the other groups during the experimental procedure. Based on these results, they suggested that exposure to an unfamiliar dog in an unfamiliar setting may be beneficial. (Polheber & Matchock 2014).

Research also suggests that dogs may even be more effective at stress reduction than other forms of relaxation. A study by Crump and Derting (2015) used the Stress-Arousal Checklist which measures the current psychological stress and arousal of participants, and the Perceived Stress Scale, a stress perception test intended to test longer-term psychological stress and increased psychological arousal, although interestingly they found no significant effects of animal-assisted activities on diastolic blood pressure, heart rate, or salivary cortisol level. They also observed that interacting with a dog reduced more stress than other forms of relaxation such as drawing and other sedentary activities (Crump & Derting, 2015).

Dogs are not the only animal to be used in stress related studies. Another study was done observing the effect of horses on stress in adolescents. In one such study working with 12-14 year olds with behavioral or learning difficulties, the experimenters told the human subjects to pet either a real horse or a model horse over the course of four 3-hour sessions. Interaction frequency, self-report anxiety, and self-esteem were measured with the the Rosenberg Self-Esteem Scale as well as the Spence Children's Anxiety Scale, a survey which answered questions on a four point scale to calculate anxiety levels. The researchers found that while self-esteem levels were not significantly altered, a significant reduction in anxiety was reported. They suspected that the increase in contact with horses may have affected their decrease in anxiety. (Holmes, Goodwin, Redhead, & Goymour, 2012).

While there are a number of studies detailing the effects of the presence of dogs (and a few on the effects of horses) on stress, there seem to be few studies considering the effects of other species of animals. Since both dogs and horses have been shown to reduce stress, it is reasonable to suspect that other animals might as well. This leads to the experimental question, "Does the presence of common domesticated animals affect the stress and performance of students during a stressful situation?"

Using different species of common pets, this study looked at whether other animals are effective stress-relievers in humans by comparing heart rate and skin conductance in participants who have been exposed to a dog, a hamster, a turtle, a fish, or no pet. These animals were selected for different reasons. I decided to use a dog due to the number of previous studies that support dogs being good stress-relievers. I decided to use a hamster because they are also mammals but a different species with very different structure. I decided to use a turtle because I was interested in the effects of reptiles and other animal classes on human stress. I decided to use a fish because I was interested in whether physical contact was necessary for causing lower stress levels in humans. For the purpose of this study, stress was measured by heart rate and skin conductivity. These measures of stress were chosen because it seemed that physical evidence may be easier to control than self-reported evidence in the case of this study. Heart rate and skin conductivity were also previously used in the Allen and Blascovich study (1991), which used a similar stress-causing method to this study. A USB polygraph device was used to measure heart rate consistently throughout the experiment.

Methods

The goal of this study was to determine if the presence of common domesticated animals affects the stress and performance of upper school students. Variables that were considered during this study were location, familiarity, and level of comfort with the animals involved, as well as responses to different animals, and improvement in the task. Before the experiment, I emailed the Science Research and Design teachers at Berkeley Carroll and asked them to share an online survey with their students in order to gather participants for my experiment. This survey collected information on current grade level, animal allergies, and any animal fears. Any student who indicated on that initial survey that they were afraid of any of the animals used in this experiment were removed from the subject pool, as this study aimed to study the responses of subjects who were already somewhat comfortable with the animals used.

Students were offered candy in order to convince them to participate in the experiment. Students with allergies to the relevant animals were excluded for their health and well-being. Because none of the subjects had previously met the animals involved in this study, familiarity and comfort with the particular animals were constant. Furthermore, the same individual animals were used throughout this experiment, controlling for any differences in response to different individuals of the same species.

This study took place over five days, and the order of the days was assigned randomly for each participant. Each day was different in that each day I tested the participants with a different animal or control. I chose to randomize the order of the days of the study so that it would not significantly affect the data if subjects became more proficient at the arithmetic task during the course of the experiment. The relevant materials used in this investigation were the USB Polygraph kit sold by Swami-ware L.L.C, a computer with the USB polygraph program downloaded, a second computer capable of acting as a stopwatch, a mobile camera phone with a functioning camera, printed number sheets, a mixed bag of nut-free candy, and four live animals. Number sheets were used so that I could easily keep track of the correct subtraction values while watching the heart rate and skin conductivity. The animals used in this experiment were a dog, a betta fish, a russian dwarf hamster, and a red-eared slider turtle.

On the control day of the experiment, the USB polygraph was plugged into the designated computer through the USB port and the downloaded polygraph program was opened. I attached the heart rate reader to the index finger of each subject's non-dominant hand, and attached the skin conductivity readers to the middle and third fingers. I began a timer for two minutes, and told the participant to sit still for that time. When the two minutes were up, I looked at the downloaded polygraph program and took a picture to record the heart rate and skin conductivity readings. The participants then began to count down from a four digit number in increments of a one-digit number for two minutes, and were told to start over when they got a wrong answer. I recorded whenever they got an answer correct or incorrect by marking with a pen next to the numbers on the number sheets. I recorded stress measurements (heart rate and skin conductivity) every twenty seconds while this was occurring by taking a picture of the polygraph program's readings. Once the two minutes were over, I removed the polygraph. Later, the data was recorded from the photographs and added to a spread-sheet for more convenient viewing.

The other four days of the experiment ran the same way as the first day with some modifications. On each day of the experiment, a different animal was given to the subject at the start of the initial two-minute resting period, and the subject was instructed to count down from different fourdigit numbers by different one-digit numbers. I made each number set more difficult (using larger intervals), in order to control for potential improvement. The first time a subject took the stress test they counted down from 4156 by sevens, the second time they counted down from 6023 by nines, and the third time they counted down from 5235 by eights. Location was kept constant through the experiment, so was unlikely to affect the outcome of the experiment.

Data

Test Efficiency

To test the efficiency of the mathematical stress task, I used a within subject Analysis of Variance (ANOVA) to compare the heart rates and skin conductance readings from before the stress task, the average reading during the test, and the highest reading during the test. For heart rate, the average initial reading was 72.2 bpm (beats per minute), the average reading during the test was 90.3 bpm, and the average highest reading was 111.3 bpm. The ANOVA test revealed a p-value of less than .0001, which suggested a significant difference, and follow up Tukey t-tests revealed a significant difference between all groups. For skin conductance response, the the average highest reading was 41.7%, the average reading during the test was 54.4%, and the average highest reading was 69.3%. The ANOVA test revealed a p-value of less than .0001, which suggested a significant difference between the initial and the during test average and between the initial and during test highest reading, but not between the during test average and the highest readings. This suggests that the stress test was successful in raising the stress levels of the experimental subjects.

Heart Rate

Figure 1

Average Heart Rates of Participants in Different Animal Groups During the Stress Task.

(Average ± Standard Deviation)



Animal Group

Figure 2

Rise in Heart Rates of Participants in Different Animal Groups during the Stress Task.

(Average ± Standard Deviation)



To test if the different animal groups significantly affected participants' heart rates, I used two between subject Analysis of Variance (ANOVA) tests. The first ANOVA test compared the average heart rates of participants in different animal groups during the stress task (Figure 1). The average heart rates were 90.8 bpm for the hamster group, 82.6 bpm for the dog group, 93.7 bpm for the fish group, 95.6 bpm for the turtle group, and 88.6 bpm for the control group. The ANOVA test revealed a p-value of 0.680244, which suggested no significant difference between the heart rates of any of the animal groups.

The next ANOVA test compared the rise in heart rates of participants in different animal groups during the stress task (Figure 2). The average heart rate increase was 49 bpm for the hamster group, 28.0 bpm for the dog group, 46 bpm for the fish group, 32.6 bpm for the turtle group, and 38.7 bpm for the control group. The ANOVA test revealed a p-value of 0.307998, which also suggested no significant difference between the rise in heart rates of any of the animal groups.

Skin Conductance Response

To test if the different animal groups significantly affected participants' skin conductance response, I used two between subject ANOVA tests. The first ANOVA test compared the average skin conductance of participants in different animal groups during the stress task (Figure 3). The average skin conductance was 27.1% for the hamster group, 44.3% for the dog group, 64.6% for the fish group , 66.0% for the turtle group, and 75.0% for the control group. The ANOVA test revealed a p-value of 0.003355, which suggested a significant difference, and follow up Tukey t-tests revealed a significant difference between the hamster group and the control, fish, and turtle groups but not between the hamster and dog groups or between any of the other groups.

The next ANOVA test compared the rise in skin conductance of participants in different animal groups during the stress task (Figure 4). The average rise in skin conductance was 13.4% for the hamster

Figure 3

Average Skin Conductance of Participants in Different Animal Groups during the Stress Task.

(Average ± Standard Deviation)



Figure 4

Rise in Skin Conductance of Participants in Different Animal Groups during the Stress Task.

(Average ± Standard Deviation)



group, 25.2% for the dog group, 25.0% for the fish group, 9.4% for the turtle group, and 23.0% for the control group. The ANOVA test revealed a p-value of 0.454109, which also suggested no significant difference between the rise in skin conductance of any of the animal groups

00

Performance

To test if the different animal groups significantly affected participants' performance on the stress task, I used another between-subject Analysis of Variance (ANOVA) test. The ANOVA test compared the number participants in different animal groups were able to reach during the stress task (Figure 5). The average score was 22.2 for the hamster group, 24.8 for the dog group, 24.8 for the fish group, 22.2 for the turtle group, and 26.7 for the control group. The ANOVA test revealed a p-value of 0.946321, which suggested no significant difference in the performance of any of the animal groups.

Figure 5

Number Participants in Different Animal Groups were able to Reach during the Stress Task.

(Average ± Standard Deviation)



Conclusions

The data suggests that the mathematical stress task succeeded in causing stress, as both the heart rate and skin conductance of the participants became significantly higher during the experiment. The data showed no significant difference in heart rate or rise in heart rate between any of the animal groups. There was no significant difference in rise of skin conductance or performance between the different animal groups, but the average skin conductance during the test was significantly lower in the hamster group than the fish, turtle, or control group. This means that the people who were holding hamsters had significantly lower skin conductance levels than people who were holding a turtle, watching a fish, or not interacting with an animal. Because the heart rate and skin conductance conclusions are inconsistent with each other, it is difficult to draw conclusions about the effect of interacting with these animals on the overall stress of the participants.

Discussion

While there was some significant data, there are a few sources of uncertainty in the study. While dogs were never shown to lower stress in my study, numerous other studies claim the opposite. Because the current data was largely inconsistent with the findings of other papers the results must be looked at through a critical lens. Because this experiment did not have the same levels of funding and technology as some of the previous experiments, it is possible that those other experiments had more accurate data. The sample size was also a factor that may have limited the viability of the results. Due to time restrictions, the sample size was relatively small and further research could benefit from collecting more data and participants. This also affected the statistical tests I was able to run. Not all participants were able to meet with all five animals, so instead of running the data as within-subject, I had to run it as between-subject for the stress measurements. Another possible limitation is technology. The USB polygraph reader was chosen because unlike other options available for this experiment, it was able to read heart rate and skin conductance at the same time, which was necessary for this experimental design. While convenient, it is possible that better equipment might provide more accurate data or alter the results. It is also likely that the age of the experimental subjects affected the results of this experiment. Many of the past studies that I looked at involved older participants, who may have reacted differently to the presence of stress or animals.

Every experiment has limitations of some sort, but with each new experiment and study, we get closer to a breakthrough. In the future, a similar study could benefit from having more participants and better technology. Future studies could continue this research and expand it to different animals or different psychological reactions.

References

- Allen, K. M., Blascovich, J., Tomaka, J., & Kelsey, R.M. (1991). Presence of human friends and pet dogs as moderators of autonomic responses to stress in women. *Journal of Personality and Social Psychology*, *61* (4), 582.
- Barker, R.T., Knisely, J.S., Barker, S.B, Cobb, R.K., & Schubert, C.M. (2012) Preliminary investigation of employee's dog presence on stress and organizational perceptions, *International Journal of Workplace Health Management*, 5(1), 15 - 30
- Crump, C., & Derting, T. L. (2015). Effects of pet therapy on the psychological and physiological stress levels of first-year female undergraduates. *North American Journal of Psychology*, 17(3), 575-590.
- Holmes, C. M., P., Goodwin, D., Redhead, E. S., & Goymour, K. L. (2012). The benefits of equine-assisted activities: An exploratory study. *Child & Adolescent Social Work Journal*, 29 (2), 111-122.
- Polheber, J. P., & Matchock, R. L. (2014). The presence of a dog attenuates cortisol and heart rate in the trier social stress test compared to human friends. *Journal of Behavioral Medicine*, 37 (5), 860-7.

The Feasibility and Effectiveness of Educational Robotics: Comparing the Swedish Teaching Abilities of a Robot and Human

by Graham Stodolski

Abstract

ttempting to compare the teaching abilities of a human and an interactive, educational robot, this experiment addresses the question "Is there a difference in language test scores of Berkeley Carroll high school students when taught five basic Swedish words and phrases by a human or by a robot?" A robot was created that was able to speak in English and Swedish and engage in an interactive question-and-answer session by asking students to tell it how to say a certain word or phrase in Swedish. Subjects either learned from the robot or a human teacher, which entailed first being told all five words and phrases and what they meant, then the interactive Q&A, and finally the entire lesson repeated. Their learning was assessed through a five question Swedish-to-English test online. The averages of the two groups were identical (=74.28571429), and a between-subject two-tailed t-test concluded that there was no difference between scores when taught by the robot or human (p=1.0; n=14).

Background

More than 72 million children around the world do not have access to an education. One reason for this is that some countries do not have the funding to build schools and recruit or pay teachers (Humanium, 2018). Another reason is that there is a shortage of teachers due to an increasing number of students worldwide and an increase in the number of retiring teachers. While simply putting more effort into teacher recruitment appears to be an easy solution, studies show that 40% of novice teachers leave education after only 5 years (Ingersoll & Smith, 2003). In order to combat this, a time and cost-efficient solution is needed. In these cases where many teachers leave after only a few years, if a school is unable to afford to pay or find long-term teachers quickly, one possible solution would be to use social robots in the classroom (Chang, Lee, Chao, Wang, & Chen, 2010; Hong, Huang, Hsu, & Shen, 2016; McLurkin, Rykowski, John, Kaseman, & Lynch, 2013). Although they may not be able to solve the lack of funding for schools worldwide, they could potentially replace teachers in locations where they're lacking and support the education needs of that community.

Robots have been successful in education for many years, helping teach a wide range of subjects, from STEM to foreign languages (Cooper, Keating, Harwin, & Dautenhahn, 1999). Their success can be seen in and attributed to three main reasons: 1. they are engaging and encourage students to be interactive, 2. they help yield higher test scores and a better understanding of material, and 3. they can be both effective and inexpensive (Chang et al., 2010; Hong et al., 2016; McLurkin et al., 2013).

Robots are effective in education because they are engaging, interactive, and help increase students' motivation to learn (Chang et al., 2010). One study programmed a robot to help a teacher teach

a foreign language in five different educational scenarios including telling a story, serving as a "cheerleader," and both giving commands to students and receiving basic commands from them. The researchers placed the robot in three different fifth-grade language classes for 5 weeks. The robot taught a different scenario each week. Although they did not have any quantitative data, the researchers concluded that while learning with the robot, "students responded loudly, raised their hands and spoke more often, asked questions, laughed and shouted, stood up, were thoughtful during the lesson, listened quietly, and discussed the robot with excitement." Their findings suggest that a key part of robots' success is their ability to engage students (Chang et al., 2010).

Because robots are such engaging educational tools, they help students attain higher test scores and a better understanding of course material. A study conducted by Hong et al. looked at the effectiveness of teaching English to Taiwanese fifth- and sixth-grade students using Robot-Assisted Language Learning (RALL; Hong et al., 2016). Researchers split classes in half and had one half of the class taught by a human teacher only (control group), while the other half was taught alongside a low-cost anthropomorphized, or human-like, robot assistant (experimental group). Similar to the previous study, the robot assisted in activities such as storytelling, reading aloud, and guestioning and answering. After administering an English comprehension test with listening, speaking, reading, and writing portions, the researchers saw that there was a significant difference between the experimental and control groups' scores, with the experimental group scoring 13 points higher on average (p = .003). In addition, Hong et al. (2016) administered a version of the Instructional Materials Motivation Survey, a questionnaire designed to assess learners' motivation towards educational material, that attempted to see if learning motivation was higher for students who were exposed to the RALL framework. They saw that because the experimental group's average survey score was 28 points greater, motivation to learn was higher for those in the experimental group (p = .0001). Thus, they concluded that the RALL framework helped engage students to be more motivated and receptive to learning English, thus resulting in higher test scores (Hong et al., 2016).

Educational robots are also successful because they can be inexpensive while remaining effective. In a study conducted at Rice University, researchers introduced a small, inexpensive, yet powerful robot into undergraduate and graduate STEM classes to explore the potential of using multi-robot systems in teaching engineering (McLurkin et al., 2013). The researchers were looking to see if their robot could replace textbooks, increase engineering literacy, and demonstrate common engineering topics to students in an introduction to engineering systems class for freshmen. After sending out surveys to students in the class, researchers found that 67% of students agreed that the class helped them figure out what engineering really is, 67% agreed strongly that the class improved their familiarity with several areas of engineering (i.e. engineering literacy), and 64% agreed strongly that the class increased their desire to select an engineering major. Furthermore and most importantly, McLurkin et al. stressed that the robots were affordable enough for every student in the class to have their own, showing that educational robots do not need to be expensive in order to work or teach well (McLurkin et al., 2013).

While these studies provide insights into the benefits of educational robots and how they are able to help in STEM or language classrooms, there is one important omission: none of these studies look into robots that are teachers as opposed to teaching supplements. It is generally supported by the aforementioned papers that robots are exceptional teaching supplements, but it is unknown whether they are effective enough to teach alone. Also, besides Hong et al.'s study, the majority of the studies' data focuses on the perception of the robots or the behavior they induce and not how much the students learned. The current study primarily focuses on whether there is a difference in the number of words and phrases learned by subjects when taught by either a robot or a human. Subjects' learning is assessed by testing their knowledge and not asking for their opinions on their teacher (either robot or human). In order to achieve this quantification, the current study uses a similar approach to Hong et al.'s experiment, which used test scores to operationalize learning, rather than other popular approaches that focus on perception. This investigation is a first step in determining and weighing the teaching abilities of humans and robots with the hope that, eventually, the experiment can diversify the already widely-explored field of robots in education.

In the fall of 2017, a preliminary study tested the experimental methods before building the robot. The experiment explored whether a video of a human or a cartoon could teach a made-up language more effectively. Data was gathered through a survey to Berkeley Carroll Science Research and Design (SRD) students that included a test on a 10-word invented language. Half of the surveys included a video in which students were taught by a human, while the other half received a cartoon video with the same script. Data showed a significant difference in test scores between the two videos (p < 0.01), with the cartoon average being significantly higher. The findings of the study helped determine that a test via a Google survey was an appropriate way to see how well subjects learned a language. Also, due to low scores overall (average score for human video was 27%, average score for cartoon was 46%), the preliminary study helped show that learning is more than just being lectured at or watching a video, leading to the inclusion of an interactive portion in the next iteration of the study.

In the spring of 2018, the first round of the official study was run. The experiment addressed the question "Is there a difference in a language test score between Berkeley Carroll high school students when taught five basic Swedish words and phrases by a human or by a robot for similar amounts of time?" Using either an educational robot prototype named Elsa 1.0 or a human teacher, subjects learned five Swedish words and phrases through an interactive question-and-answer lesson. Their learning was assessed through a five question Swedish-to-English test online. Swedish was chosen due to its difference from common Romance languages yet similarities to English. It was concluded that there was not a significant difference between scores when taught by the robot and human (average score of 60 for human, 100 for robot; n=6). The large difference in averages, though, suggested that a follow-up study should be run for further investigation.

The experimental question this current study addresses is the same as the first run in 2018. It was hypothesised that there would be a difference in language test scores between when students were

Figure 1



Elsa 1.0 as of April 27, 2018.

taught by the robot and by the human. The independent variable in the study is the teacher, which is either the robot or the human. The dependent variable is the test score, which is out of 100 and collected and calculated through a Google survey.

Procedure

In order to recruit subjects and determine their eligibility, a survey was sent that asked questions that helped control variables and ascertain subjects' interest in participating. To control for prior knowledge, subjects were asked if they had any experience with Scandinavian languages. Students who answered yes were considered ineligible. To control for the current number of language classes, only participants enrolled in one language class were considered. To control for grade level, the survey was only sent to high school students.

The preliminary survey was created using a Google form. This included three sections; the first section included four questions (the lettered bullet points below each question represent possible answers):

1.	How many languages do you speak fluently other than English?						
	a. O						
	b. 1						
	c. 2						
	d. 3						
	e. 4						
	f. 5						
	How many languages do you take at Berkeley Carroll or online?						
	a. 0						
	b. 1						
	c. 2						
	d. 3						
	e. 4						
	f. 5						
8.	Have you had exposure to Swedish, Danish, Norwegian, or other Scandinavian languages?						
	a. Yes						
	b. No						
ŀ.	Would you be willing to participate in my study?						
	a. Yes						
	b. No						

If the subject indicated "yes," they were moved onto section two of the survey. Section two included one question:

- 1. Please write your email below.
 - a. Short answer text box

If the subject indicated "no" to question four in section one, they were moved directly onto section three. Section three included one question:

If you selected no, please explain why.
 a. Short answer text box

Next, the survey was distributed via a link to grade deans, or faculty leaders of each grade, and SRD teachers, who then sent it to their students. This allowed for a potentially larger portion of the student body to be reached quickly.

The interested and eligible students were entered into a spreadsheet and assigned a number. The total number of subjects (n) was divided by two and entered into an online random number generator to spawn random numbers ranging from 1 to n. The subjects with the corresponding number were found and assigned to be taught by the robot. Exactly half of the subjects (n/2) were taught by the robot and the other half were taught by the human, all randomly assigned.

A robot was created that was able to teach Swedish. SolidWorks, an application used for 3D design, was used to draw its head, jaw, body, and arms. The head measured 11 cm x 9 cm x 10 cm, the jaw measured 5 cm x 4 cm x 1 cm, the body measured 13 cm x 6 cm x 13 cm, and the arms measured 10.5 cm x 8 cm x 1 cm. Next, each part was 3D-printed using a MakerBot Regulator 3D printer. This process took around 12 hours. The head, jaw, body, and arms were then assembled using hot glue and the head, jaw, and arms were connected to their servos, small motors that allow them to rotate. The servos were wired to a breadboard, a tool used to build circuits, and then to a Raspberry Pi, the small computer that executed the teaching program.

Figures 2 and 3

Solidworks rendering of Elsa 2.0; Elsa 2.0 as of January 10, 2019.



The Python coding language and Google Speech Recognition software were used to code the lesson taught by the robot. This started with an introduction of the robot and a description of the ensuing lesson and was quickly followed by the recitation of the five Swedish words and phrases. After this, the subject participated in a question and answer (Q&A) session where they were asked how to say something by the robot then answer in Swedish. Then, the entire lesson, excluding the introduction, was repeated. The human teacher followed the same script. Below is an excerpt from the script:

"Hello, and welcome to Swedish 101. My name is Elsa and I will be your teacher. Today, I will be teaching you 5 useful words and phrases in Swedish. First, I will say the word or phrase and what it means. Then, I will repeat once. Let's get started! The first phrase is 'God morgon' and it means 'Good morning.' 'God morgon,' 'Good morning.' The second phrase is 'God kväll' and it means 'Good evening.' 'God kväll,' 'Good evening.' ...How do you say 'I speak Swedish'? (If answer is correct, nod head and say 'Correct!'; if answer is incorrect, shake head and say 'The correct answer is 'Jag talar svenska.'). How do you say 'Good evening? (If answer is correct, nod head and say 'Correct!; if answer is incorrect, shake head and say 'Correct!'; if answer is incorrect, shake head and say 'Correct!'; if answer is 'God kväll.'...."

The lesson was coded by writing out each sentence on its own line and having the robot speak that line. Simple servo commands were used to rotate the motors when the robot spoke and when the subject answered a question correctly or incorrectly. It nodded if correct, using one motor, and shook its head if incorrect, using a separate motor.

Finally, a second survey was created to be taken after the study that determined subjects' test scores. This survey included seven questions, five of which included a YouTube video link where a native Swedish speaker recited each of the five words and phrases. After watching each video, subjects were required to indicate what that word or phrase meant in English. The seven survey questions were as follows:

- 1. Email address?
 - a. Short answer text box
- 2. Teacher?
 - a. Human
 - b. Robot
- 3. What does this word/phrase mean in English? (link to video 1 below)
 - a. Short answer text box
- 4. What does this word/phrase mean in English? (link to video 2 below)
 - a. Short answer text box
- What does this word/phrase mean in English? (link to video 3 below)
 a. Short answer text box
- 6. What does this word/phrase mean in English? (link to video 4 below)
 - a. Short answer text box
- 7. What does this word/phrase mean in English? (link to video 5 below)
 - a. Short answer text box

Table 1

List of Five Swedish Words/Phrases Used and their English Counter	rparts.
---	---------

English Word/Phrase	Swedish Word/Phrase		
Good morning	God morgon		
Good evening	God kväll		
Please	Snälla		
You're welcome	Varsågod		
l speak Swedish	Jag talar Svenska		

Once a subject had completed the initial survey to gauge their eligibility and interest, their email was taken from question one in section two of the first survey and a meeting was scheduled to run the study. The location was any empty classroom. The experiment was begun by beginning the lesson; for the human teacher, this simply involved reciting the same script the robot used off a laptop. For the robot, beginning the lesson required plugging in the Raspberry Pi, opening the script, and entering the password. After the script was finished and the lesson was complete, the second survey was administered by sending the subject an email with the link. Once the response was submitted, their answers were graded in the spreadsheet linked to the survey. The number of correct answers was divided by five and multiplied by 100, resulting in a score out of 100. The averages of the robot and human groups were compared using a t-test on vassarstats.net.

Data, Analysis, and Conclusions

The experiment was run 14 times: seven times with the human teacher, and seven times with the robot teacher. After their lesson, subjects took the Google survey quiz to determine their test score. The results were as follows:

Table 2

Comparison of Student Test Scores of Human and Robot Instructors Showing the Number of Subjects for Each Group, the Average Test Score, the Standard Deviation, and the Standard Error of the Mean (SEM).

	Human	Robot
Number of Subjects	7	7
Average (out of 100)	74.29	74.29
Standard Deviation	25.07	15.12
Standard Error of Mean	9.48	5.71

Figure 3

Comparison of Student Test Scores of Human and Robot Instructors. No Significant Difference in Test Score (\pm SEM) Between Human and Robot Teacher (p=1.0).



The experimental question this study attempted to answer is "Is there a difference in a language test scores of Berkeley Carroll high school students when taught five basic Swedish words/phrases by a human or by a robot?" After a between-subject two-tailed t-test, data shows that there is no significant difference between test scores when taught by the human or the robot teacher (p=1.0). The averages of the two groups are identical. The null hypothesis, that there is no difference in test scores, is accepted. This lack of a difference implies that the human and robot teachers were equally effective in teaching high school students the five Swedish words and phrases.

Discussion and Next Steps

Many studies and experiments have attempted to determine the feasibility of educational robotics. However, few have tested their feasibility as independent, primary teachers. This study intended to do just that, and has successfully done so by concluding that there is no difference between the Swedish teaching of a human and a robot. While it can be concluded from the data that it is plausible that educational robots could replace humans when necessary and teach Swedish just as effectively, a strong conclusion cannot be drawn due to the small sample size of this study.

It is important to understand how teaching is defined, recognized, and quantified in this particular study. First, it is assessed through subjects' score on a test. For people who are not good test takers, this is not a good measure of learning. It's also done immediately after the lesson, meaning that it's measuring short-term learning. Second, the lesson itself is very rigid and controlled. For an experiment, this makes sense, but for practical learning, this would likely not be effective. In my opinion, the ability to answer students' spontaneous questions and tailor material to what the student is struggling with is what's truly beneficial in teaching. With the use and power of artificial intelligence, though, robots could likely become advanced enough to have these functionalities, therefore being more effective teachers (Lemaignan, Warnier, Sisbot, Clodic, & Alami, 2016).

00

It is also important to realize that this experiment's results only pertain to Swedish. In order for more general conclusions to be drawn, more languages and disciplines must be tested. This is a possible next step to more thoroughly explore the teaching abilities of robots. Other next steps could include developing a more in-depth, interactive lesson through artificial intelligence or determining which teacher elicits more long-term learning. Actual next steps to be taken in the next few months will primarily include collecting more data, since the number of participants was very low (n=14).

Regardless, data from a preliminary study in 2017, a first trial in 2018, and a second, final trial in 2019 have all pointed towards the effectiveness of a robotic teacher when compared with a human teacher. These findings are a step in the right direction for practically implementing interactive robots to communities where human teachers are unavailable.

References

- Chang, C., Lee, J., Chao, P., Wang, C., & Chen, G. (2010). Exploring the possibility of using humanoid robots as instructional tools for teaching a second language in primary school. *Educational Technology* & Society, 13 (2), 13-24.
- Humanium (2018). Right to education: Situation around the world. *Humanium*. Retrieved from www.humanium.org/en/right-to-education/
- Hong, Z. W., Huang, Y. M., Hsu, M., & Shen, W. W. (2016). Authoring robot-assisted instructional materials for improving learning performance and motivation in EFL classrooms. *Educational Technology & Society*, 19 (1), 337–349.
- Ingersoll, R. M., & Smith, T. M. (2003). The wrong solution to the teacher shortage. Educational leadership, 60 (8), 30-33.
- Lemaignan, S., Warnier, M., Sisbot, E. A., Clodic, A., & Alami, R. (2016). Artificial cognition for social human–robot interaction: An implementation. *Artificial Intelligence*, 247, 45-69.
- Martyn, C., Keating, D., Harwin, W., & Dautenhahn, K. (1999). Robots in the classroom-tools for accessible education. Assistive technology on the threshold of the new millennium, 6, 448.
- McLurkin, J., Rykowski, J., John, M., Kaseman, Q., & Lynch, A. (2012). Using multi-robot systems for engineering education. *IEEE Transactions of Education*, 56 (1), 24-33. Retrieved from https://link.springer.com/article/10.1007/s10803-012-1645-2

Special thanks:

Ryan Byrne, Scott Rubin, Essy Levy Sefchovich, Alexandra Durham, Lycke Skau, and Elsa.

Conscious and Subconscious Future Self Continuity in High School Students

by Zachary James Tegtmeier

Abstract

revious studies have shown that the extent of the connection, and by extension empathy, that you feel with your future self (Future Self Continuity) can predict how likely you are to act in ways that benefit you in the future (Ersner-Hershfield, Wimmer, & Knutson, 2008). This study tests if there is a significant correlation between subconscious Future Self Continuity (IV) and conscious Future Self Continuity (FSC) in high school students (DV). The experimenter uses a survey that evaluates conscious FSC by asking students to rate how much they think about a topic on a 7-point scale. The experimenter measures subconscious FSC using the Implicit Association Test (IAT), a task that measures automatic association between two different sets of words at a time. The data did not suggest a correlation between conscious and subconscious FSC (p = 0.109716, p > .05, r = 0.2631).

Background

At first glance you are the same person now as you will be in ten years; you will probably still have the same name, face, and at least some of the same memories. In the eyes of others, you are the same person. However, your future self can feel like a completely different person. If you saw a picture of yourself ten years from now you might see that person as a complete stranger. It can be difficult to imagine what new memories this "stranger" has: new friendships, career path, lifestyle, and goals. When you know so little about your future self, it can be easy to dismiss that person as a stranger, and therefore not worry about their well-being. When this is true, there is no reason not to consume your available resources now rather than saving them for the far future (Ersner-Hershfield, Garton, Ballard, Samanez-Larkin, & Knutson, 2009). Conversely, if you think about your future self frequently and have a vivid idea of what they will look like and how they connect with you now, you are more likely to see the benefits of making decisions for the betterment of your future self (Ersner-Hershfield et al., 2009). The connection, and by extension the empathy an individual feels with their future self, is called Future Self Continuity (FSC), a term coined by Hershfield and his fellow authors (Ersner-Hershfield, Wimmer, & Knutson, 2008.; Ersner-Hershfield et al., 2009).

Hershfield's 2008 study looked to answer if an individual's FSC could predict how inclined they are to save money (Ersner-Hershfield, Wimmer, & Knutson, 2008). Most subjects reasonably want to have higher FSC then they might actually have (Ersner-Hershfield, Wimmer, & Knutson, 2008). Therefore, to remove subject bias from the data, this study used neurological measurements of the prefrontal cortex to measure FSC, choosing parts of the prefrontal cortex known to show greater activity when thinking about something related to one's self (Ersner-Hershfield, Wimmer, & Knutson, 2008.). Subjects were hooked up to a functional magnetic resonance imaging (fMRI) machine, and four different groups would appear on the screen, one at a time; current self, future self, current other, and future other (Ersner-Hershfield, Wimmer, & Knutson, 2008). When a group name appeared on the screen, an adjective appeared below, and the subject was prompted to rate how much the adjective describes the group on a scale of 1-4 (Ersner-Hershfield, Wimmer, & Knutson, 2008.). The two "other" groups acted as controls, and the current self group showed how much brain activity was normal for each person when thinking about their current self; this was compared to brain activity when being questioned about their future self (Ersner-Hershfield, Wimmer, & Knutson, 2008). For example, the top of the screen would say current self, and the middle would display an adjective such as "honorable". Along the bottom, buttons would appear that were labeled on a scale of 1-4 ranging from "Very Unlike" to "Very Like." The participant would choose one of these buttons and then be prompted with another group and adjective until that part of the experiment was completed. Subsequently, subjects answered survey questions asking them to choose between receiving an amount of money now or more money later with varying amounts of money and time (Ersner-Hershfield, Wimmer, & Knutson, 2008). The study found that subjects who had higher brain activity when thinking about themselves in the future (and therefore had higher FSC) were more likely to choose to save money for their future self (Ersner-Hershfield, Wimmer, & Knutson, 2008). This meant a positive and significant correlation between the likelihood of saving and brain activity (higher brain activity meant higher FSC) (r = .47, p < .05) (Ersner-Hershfield, Wimmer, & Knutson, 2008.). While the study did a great job of controlling for subject bias, using an fMRI greatly limited the subject pool; the study only had 18 subjects (Ersner-Hershfield, Wimmer, & Knutson, 2008.).

Hershfield's 2009 study attempted to rectify this limitation (Ersner-Hershfield, Garton, Ballard, Samanez-Larkin, Knutson, 2009). With the same experimental question, the study measured FSC in a two-step process. First, subjects rated what they believed their FSC was by choosing two circles that overlap each other enough to visually represent their FSC (Ersner-Hershfield, Garton, Ballard, Samanez-Larkin, Knutson, 2009). Second, subjects rated how much they believe adjectives describe both their current self and their future self; the more similarly a subject's current and future self were described, the more FSC the subject should have. A score was then created from these two data sets (Ersner-Hershfield, Garton, Ballard, Samanez-Larkin, Knutson, 2009). Hershfield then used the same methodology as in the 2008 study to measure how inclined a subject was to save money (Ersner-Hershfield, Garton, Ballard, Samanez-Larkin, Knutson, 2009). Hershfield again found that FSC was positively correlated with how likely subjects were to save their money (r = .34, p < .001) (Ersner-Hershfield, Garton, Ballard, Samanez-Larkin, Knutson, 2009). This study gathered more subjects (n = 260), however Hershfield did not avoid his original problem of subject bias; subjects would often be inclined to give themselves high future self continuity (Ersner-Hershfield, Garton, Ballard, Samanez-Larkin, Knutson, 2009).

A tool that allows a large subject pool and removes subject bias is the implicit association test (IAT) (ProjectImplicit). This test can be completed by anyone with a laptop and takes about 5-7 minutes, making it easy to administer (ProjectImplicit). Additionally the test is designed specifically to remove subject bias; the IAT relies on automatic reactions rather than self reporting, so participants are unable

to skew data based on what results they want or believe they should have (Greenwald et al., 2003.). One of the most famous uses of the IAT is to measure implicit preference among self proclaimed unbiased white people between lighter and darker skinned individuals (Greenwald et al., 2003.). The IAT in this study measures implicit preference between two groups (light and dark skin) by measuring automatic association between positive adjectives and each group (Greenwald, McGhee, & Schwartz, 2003). On a computer, subjects are given words (or sometimes pictures) that belong to each group and must sort them into either the left or right side of the screen by hitting a key (Greenwald et al., 2003). By measuring reaction time when positive adjectives and each group are on the same side, and then comparing those two reaction times, you can identify the group for which an individual seems to have an implicit preference (Greenwald et al., 2003). In Greenwald's study, subjects all stated they were unbiased regarding race. The four groups in the IAT were positive attributes, negative attributes, pictures of light skin faces, and pictures of dark skin faces. The study found an average 179 millisecond difference between positive attributes and light skinned people being associated together and positive attributes and dark skinned people being explicit and provide the additional set the study found an average 179 millisecond difference between positive attributes and light skinned people being associated together and positive attributes and dark skinned people being associated together and positive attributes and dark skinned people being associated together and positive attributes and dark skinned people being associated together and positive attributes and dark skinned people being associated together and positive attributes and dark skinned people being associated together and positive attributes and dark skinned people being associated together and positive attributes and dark skinned people being associated together and p

The IAT has also been used to measure implicit associations for topics similar to FSC. In 1998, Greenwald used the IAT to measure self esteem by measuring implicit preference between the self and other people. Utilizing the same method as in the previous studies he measured the difference in reaction time when each grouping was associated with positive adjectives (Greenwald & Farnham, 1998). For example, if the average reaction time for grouping positive words and self-related words was faster than when negative words and self-related words were recorded on the same side, subjects felt more positively about themselves (Greenwald & Farnham, 1998). Implicit measures of self esteem were compared with an explicit measure that consistently asked subjects to rate how well they felt about themselves in regard to different characteristics such as appearance, intellect, and personality (Greenwald & Farnham, 1998). The study found that there was a positive, albeit weak, correlation between implicit and explicit self esteem (r = .17, p < .05); (Greenwald & Farnham, 1998.).

FSC can help us learn many different things about people, ranging from how likely they are to save money for their future (Ersner-Hershfield, Wimmer, Knutson, 2008.; Ersner-Hershfield, Garton, Ballard, Samanez-Larkin, Knutson, 2009), to how likely they are to act ethically (Ersner-Hershfield, Cohen, & Thompson, 2007). Because of this it is important to see if FSC that is self-reported reflects a participant's actual FSC. The IAT is a great test to measure implicit FSC, removing subject bias while still allowing a large subject pool. Therefore, the current study attempts to answer the following question: Is there a correlation between conscious FSC (score on Likert scale) and subconscious FSC (measured reaction time difference on IAT) for high school students?

A similar model of the IATs used in previous studies is followed. The four groups are current, future, self, and other. This study measures the difference in reaction time between when current and self are on the same side of the screen and when future and self are on the same side. The pair with the faster reaction time has a closer association. It is expected that current and self always have a closer association (and therefore a faster reaction time), however the extent of that time difference reveals the specific implicit FSC of each subject. The difference in reaction time between the two pairs is measured as the independent variable. Explicit FSC is measured using a Likert scale, in which subjects rate on a scale of one to seven how much they care about certain aspects of their life now, compared to parallels of those aspects ten years from now; the difference in Likert scale score between current and future events is measured as the dependent variable. The hypothesis is that there indeed would be a correlation between conscious and subconscious FSC. 00

Experimental Procedure

Google Forms is used as a tool for gathering and recording data with eight questions answered by each subject to measure their conscious FSC. Four of the questions prompt the user to rate how often they think about a topic that is affecting them at the current time, and the other four ask the subject the same questions about similar topics ten years later. All questions have the subject rank how much they think about the topic of the question on a scale of one (never) to seven (all the time). The order of the questions is randomized using settings on Google Forms. This acts as a control to ensure that the order of questions does not affect how subjects perceive or think about each question. Results from each question are sent to a Google spreadsheet linked automatically to the Google Form when it is created. The questions used are listed below:

Survey Questions

- 1. How often do you think about your future career 10 years from now?
- 2. How often do you think about where you will live 10 years from now?
- 3. How often do you think about your social life 10 years from now?
- 4. How often do you think about your physical and/or psychological health 10 years from now?
- 5. How often do you think about your GPA?
- 6. How often do you think about your social life and friends?
- 7. How often do you think about your physical and/or psychological health?
- How often do you think about extracurriculars you are involved in (sports, music, debate, etc.)?

The subject answers each of these questions and is then sent to a second page on the Google Form to start the IAT. The IAT has been modified for this study to measure if words related to self are more closely associated with words related to the present or words related to the future by having users place words from all of these categories to its assigned side of the screen. The IAT changes which side of the screen categories should be placed in between rounds to gather more data.

The next page provides the subject with a link to the IAT on Google Sites, instructing the user to follow the directions on the website and paste their results from the IAT onto a space provided on the form page. The website with the IAT used in this study is available upon request. Google Sites is a free application that allows the experimenter to simply paste in the code needed for the IAT. Additionally, just like Google Forms, Google Sites can be accessed from any computer, allowing subjects to participate in the study in any location with wifi and a computer. In my study the four groups that were tested were self, other, current, and future. The items chosen for each group had to be commonly known single words. These requirements ensure that reaction times do not vary due to confusion about a word's meaning or time reading the word. In my study the two pairs of groups associated together and recorded to measure subconscious FSC are self & current, and self & future. The two other pairs recorded in the IAT, other & current and other & future, act as controls. If the reaction times of these two pairs are

significantly shorter or longer than that of the preceding two pairs, that would mean the words are affecting the reaction time rather than the current-self or future-self pairing, and the word choice should be reevaluated. A list of the words and their groups are presented at the beginning of the test in order to avoid confusion for the subject. The words used in my IAT can be found below:

Groups	Words in Each Group (Stimuli)					
Self	Me	Personal	Му	1	Own	Mine
Other	Them	Him	Her	Friend	Stranger	Celebrity
Current	Now	Present	Today	Instant	Here	Immediate
Future	Later	Distant	After	Delayed	Faraway	Eventually

The IAT in this study has eight stages. Instructions are given between nearly every stage to prepare subjects for the next round. The first two stages introduce opposite stimuli (*e.g.* self vs other) in the same round to introduce to the subject to how the test works by getting them familiar with organizing words on a specific side. The third and fourth rounds introduce all the categories at once, having two groups together on each side. These two rounds appear to be the same round to the subject since they have the same rules. However, unbeknownst to the user, the IAT does not start recording data until the fourth round. All rounds that are not recording data are meant to help the subject habituate to the test so that differences in reaction time are not due to confusion about the test or any of the words. The same method is repeated for rounds five through eight except that future and current words are switched onto different sides so a different category of words can be tested on the same key as self words. The "D" and "K" keys are used to place words onto each side. The exact keys used are not important; however subjects must use a different hand to hit each key. The exact design of the IAT is shown below:

Implicit Association Test Stages

- 1. 20 trials with self words on the D key and other words on the K key.
- 2. 20 trials with current words on the D key and future words in the K key.
- 3. 20 trials with with self and current words in the D key and other and future words in the K key.
- 4. 40 trials with with self and current words in the D key and other and future words in the K key. *(Reaction time measurements recorded).*
- 5. 20 trials with self words on the D key and other words on the K key.
- 6. 20 trials with future words on the D key and current words on the K key.
- 7. 20 trials with with self and future words in the D key and other and current words in the K key.
- 8. 40 trials with with self and future words in the D key and other and current words in the K key. *(Reaction time measurements recorded).*

*No instructions or pause are given after rounds 3 and 7 since the rules for rounds 4 and 8 are the same as their preceding rounds.

The IAT was programmed using three languages common in web design: Hyper Text Markup Language (HTML), Cascading Style Sheets (CSS), and Javascript. HTML was primarily used to place text on the page. This includes all text starting with the instructions in the beginning, the words in the center and on the side that change throughout the test, all the way through to the results at the end of the test. CSS is used to change where all the text is placed and how it looks, including font, size, opaqueness, etc. The program most heavily relied on was Javascript, which actually manipulates the text on the screen to change when a key is hit, decides if the subject assigned the text to the wrong side, and randomly chooses the middle word (the word organized by the subject) based on the groups available in that round. The program uses Jquery, a Javascript library, to set up the process of randomly choosing a word that belongs in that round, putting it on the screen, recording the reaction time of the user for that word, and changing the categories if appropriate. The program has a main function that activates every time a key is hit. Each time a key is hit, the function finds the ID number of that trial and uses "if else" statements (statements that only run specific lines of code if certain conditions are met) to make sure the next randomly chosen word is in one of the categories in that round. If the ID number determines a new trial should begin, the screen will instead display text using HTML that introduces the rules for that round. The program additionally uses an "if else" statement to check if the word for that trial was associated with the correct side. For rounds four and eight, if the conditional statement determines the user answered correctly, the program will record the ID number of the word; the word itself; the key the subject hit; and the reaction time of the subject. The program puts this data into a list that will be shown to the subject at the conclusion of the IAT. If the "if else" statement determines the word was associated incorrectly, the program activates a sleep function which runs an empty "while" loop (a function that runs a line of code repeatedly while certain conditions are true) that repeats until 2 seconds have passed. This discourages subjects from getting future answers wrong. Additionally, the number of incorrect associations from a participant is recorded; if the subject made more than 20% of their associations incorrectly, it would be clear the subject was not fully focused on the IAT, invalidating the results of the test. In this case, their data would be removed from the data set. Luckily, this was never the case with any subject. A word generator function is needed to randomly choose a word that is in either category on which the subject is currently being tested. At the end of the test, the program displays the list that places all the recorded data together, along with text instructing the subject to copy and paste the data back onto the survey. Having the data organized in a list in this fashion is necessary for converting the data into a csv (comma separated values) file. Source code for the IAT is available upon request.

Once the user has inserted the results from the IAT back onto the second page of the survey, the survey takes them to a third page asking if the subject found any of the words and their categories confusing. If the user says no, the survey ends and they are thanked for their help. Otherwise they are taken to a fourth page to specify which words they found confusing. Subjects must actively decide that they found a word confusing before they are prompted to type what the word is. This ensures that subjects are not trying to think of a word simply to answer the question. If subjects consistently find specific words confusing, this indicates the word should likely be changed.

All data for each participant is put into a single Google Survey to measure the difference in conscious and subconscious FSC for each subject. Participants are kept anonymous to avoid any potential emotional harm to them. Once all data is collected for a single participant, an average is taken for all current issue questions, followed by an average for all future issue questions. Next, the average for current issue questions is subtracted by the average for future issue questions. This value is used for conscious FSC when analyzing data with all subjects. To measure subconscious FSC, the data from the list on the Google Form is copied and pasted into textEdit, saved as a csv file, and imported into a Google Spreadsheet with the survey question results. This allows all data to be kept together in an

organized manner. All reaction time results from the IAT when the word is associated with the self group are averaged separately for rounds four and eight. The average for round eight is subtracted by the average for round four. This value is used for subconscious FSC when analyzing data with all subjects. These two values are reported for all subjects on a scatterplot using a Linear Correlation and Regression stat test.

Data Analysis

Conscious FSC was measured using the questionnaire scores at the beginning of the experiment. For each subject all scores related to questions about their current self and their future self were averaged separately, giving each subject (n=37) a single score for current self and future self questions. A within subject two-tailed t-test was run on this data set, showing that the score for present questions was significantly higher than the score for future questions, as shown in Figure 1 (p = 0.002286, p < .05). In other words, in the questionnaire, participants consciously stated that they think more about what happens to them in the present than what happens to them in the future.

Figure 1

Comparison of Present and Future Self Scores.

The average present score is 5.07 + -0.19, while the average future score is 4.14 + -.23. All questions answered by subjects in this part of the study ranged from 1 (I never think about this) to 7 (I think about this all the time). Error bars +/- Standard error of the mean (SEM)



Subconscious FSC was measured using reaction time measurements from the IAT. Recorded reaction times that were greater than three seconds were removed from the average, since a reaction that takes longer than three seconds is not based on an automatic association from the participant. Between 0 and 2 reaction time data points were removed per participant (31 times total). A within subject two-tailed t-test was run on this data set, showing that the reaction time when present and self words were associated was significantly faster than when future and self words were associated, as shown in Figure 2 (p = <.0001). This shows that the subjects have a greater automatic association for their current self than their future self. It is important that data from Figure 1 and 2 shows that subjects have a clear

00

Figure 2

Comparison of Present-Self and Future-Self Implicit Association Test Reaction Times.

The average self-current reaction time is 941.14 +/- 42.21 milliseconds, and the average self-future reaction time is 1344.82 +/- 53.33 milliseconds. Error bars +- Standard Error of the Mean (SEM).



Figure 3

Comparison of Reaction Time vs Questionnaire Score Difference for Each Participant.

The average reaction time difference and score difference respectively for all participants was 408.33 +/- 41.85 milliseconds and .93 +/- .24.



preference for their current self over their future selves, because otherwise the FSC hypothesis among these subjects would not be confirmed, invalidating the experimental question.

For each subject the difference between their two average questionnaire scores was taken. Because each person cares about specific issues differently (regardless of when they are taking place), the difference between both averages provides a value that can be compared to other participants. Since everyone has different reaction times, the difference between the two averages for each subject was taken. This gives each subject two new values that show the difference between each of their scores from the questionnaire and the IAT. A linear correlation and regression test was run on this data set, showing no correlation between reaction time difference and score difference, as shown in Figure 3 (p = 0.109716, p > .05, r = 0.2631).

Conclusion

The goal of this study was to answer the experimental question "Is there a correlation between conscious FSC and subconscious FSC for high school students?" Figure 1 shows subjects reporting through their surveys that they care more about current issues than future issues, meaning participants consciously care more about their current self than their future self. Similarly, Figure 2 shows that subjects associate more with their current self than their future self, allowing the conclusion that subjects subconsciously care more about their current selves. From Figure 3, it can be seen that there is no significant correlation between reaction time difference and score difference. The reaction time difference shows the subconscious FSC of participants while the score difference shows the conscious FSC of participants. Therefore together these two points demonstrate that while there is a slight positive trend for the line of best fit, there is no significant correlation between conscious and subconscious FSC.

Discussion

This preliminary study allows us to accept the null hypothesis, that there is no correlation between conscious and subconscious FSC. Perhaps with further research there could be a significant correlation, however as of now this research shows us that the FSC high school students say and believe is not an accurate indicator of FSC they actually subconsciously have. This suggests that studies attempting to correlate FSC with behavior need to measure FSC subconsciously, otherwise they will not get an accurate measurement of FSC from participants. Hershfield's 2008 study already attempts to do this; however, future studies should consider using the IAT as an apparatus to measure FSC without severely reducing subject size with tools similar to an fMRI (Ersner-Hershfield, Wimmer, & Knutson, 2008.).

The biggest limitations when conducting my study were diversity of my subject pool and having subjects complete the experiment properly. First, all subjects were high school students who attend the same school. It would be valuable to conduct this study on people of different ages (particularly adults) and people from different schools, especially because it is difficult to predict how these two factors will affect the results. Because of the subject pool, these results only apply to students at the Berkeley Carroll School. Additionally, while the questionnaire and the IAT allowed participation in the experiment without the experimenter being present, 32 of the 69 participants who took the study could not be used for data analysis because they did not finish the IAT. This suggests that to gather more subjects and have a higher yield rate for usable data, it would be helpful to have participants conduct future iterations of this study while being proctored. This additionally helps to ensure not only that subjects will finish the entire experiment, but that subjects will not be distracted and have their reaction time slowed down by external variables while taking the IAT.

References

Ersner-Hershfield, H., Wimmer, G. E., & Knutson, B. (2008). Saving for the future self: Neural measures of future self-continuity predict temporal discounting. *Social cognitive and affective neuroscience*, 4 (1), 85-92.

Ersner-Hershfield, H., Cohen, T. R., & Thompson, L. (2007, November 7). Short horizons and tempting situations: Lack of continuity to our future selves leads to unethical decision making and behavior. Retrieved from http://halhershfield.com/wpcontent/uploads/2016/06/Hershfield_Cohen_Thompson_2012_0BHDP.pdf

Ersner-Hershfield, H., Garton, M. T., Ballard, K., Samanez-Larkin, G. R., & Knutson, B. (2009). Don't stop thinking about tomorrow: Individual differences in future self-continuity account for saving. *Judgment and Decision Making*, 4 (4), 280.

- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: An improved scoring algorithm. *Journal of personality and social psychology*, 85 (2), 197.
- Greenwald, A. G., & Farnham, S. D. (2000). Using the implicit association test to measure self-esteem and self-concept. *Journal of personality and social psychology*, 79 (6), 1022.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of personality and social psychology*, 74 (6), 1464.
- Lee, E. (2005). The relationship of motivation and flow experience to academic procrastination in university students. *The Journal of Genetic Psychology*, *166* (1), 5-14. Retrieved from https://search.proquest.com/docview/228500284?accountid=35114

ProjectImplicit. (n.d.). [Website] Retrieved from https://implicit.harvard.edu/implicit/takeatest.html

Special thanks to:

Dr. Banerjee for helping me develop my procedure and giving me the idea to use the Implicit Association Test.

Mr. Byrne for meeting with me very often to help program my Implicit Association Test.

Mr. Stevens, Mr. Rubin, and Ms. Sefchovich for teaching me and guiding me through the scientific process.

All the Science Research and Design teachers and Grade Deans for helping collect subjects.
Race Matching Between Teachers and Students: Who Do Teachers Treat Differently?

by Alayna Thomas

Abstract

his study asked if the race of a teacher affects how they interact with students of the same or different race. Subjects were recruited from a faculty meeting and through email. The experiment was made up of two parts: a classroom observation, and a survey for teachers and students. There was no significant difference in how students of color and White students are treated. Students of color and White students were called on with and without their hands raised approximately equally, and raised their hands at similar rates. This study can open doors to make more teachers aware of their own biases and can lead to other helpful experiments that can look into the intersection of race and gender, or who gets reprimanded the most in class and why.

Background

New York City is one of the most racially and ethnically diverse cities in the United States and the world. However, this diversity isn't properly represented in our schools. Around 84% of public school students are Black, Latino, or Asian, compared to only 39% of teachers (Disare, 2018). Specifically in New York City: 16% of public school students are Asian, 6% of public school teachers are Asian; 27% of students are Black, 18% of teachers are Black; 41% of students are Latino, 15% of teachers are Latino, 15% of students are White, 58% of teachers are White (Disare, 2018). This trend can also be seen in different parts of the United States. Around 90% of Black students in Washington D.C. attend segregated schools. In this case, a segregated school means that the school is less than 10% White (Yin, 2017).

There is an increasing amount of evidence that having a teaching force that is as diverse as its students can be beneficial for students. (Cherng & Halpin, 2016). Cheng and Halpin used data from the Measure of Effective Teaching Study (METS) to assess how students view their teachers. Self-reporting survey data was collected from public schools nationwide and included over 200 teachers in the study. The results of the study suggest that both students and teachers have more positive views of each other when both their gender and race are the same, or when teachers and students are race matched. While Cherng and Halpin found that race matching could help raise a student's grades, they also found that kids of all races generally prefer teachers of color. For instance, students said that their Black and Latino teachers are better at clarifying information than their White teachers (Cherng & Halpin, 2016). In addition, minority students that have a minority teacher benefit regardless of race matching. For example, Cherng and Halpin found that Asian students have especially high perceptions of African-American or Black teachers. Asian students. The researchers suggest that this perception may be due to the fact

that teachers of color are likely to be inherently culturally aware and understand cultures that are different from their own, as they have to move through a society that treats them as "others" (Cherng & Halpin, 2016).

This all contributes to a phenomenon that Thomas S. Dee calls the "Role Model Effect". The "Role Model Effect" is when a minority student sees someone demographically similar to them and subconsciously has better self-esteem, as they are seeing someone who is also disenfranchised in a higher ranking position (2005). This, in turn, makes a student more likely to attend class and complete homework (Dee, 2005). The role model effect also works outside of the classroom. Young girls who see women politicians are more likely to pay attention to politics (Campbell &Worbecht, 2006). We can still apply this to other prominent figures, such as Oprah Winfrey and Barack Obama. Even popular movies like *Black Panther* and *Crazy Rich Asians* could have an influence on young people to fight for better media representation.

Teachers have a more direct impact on teenagers than politicians or movie characters, as this is a period where many students look to adults other than their parents for moral support and guidance. In her study *"We make it controversial: elementary preservice teachers' beliefs about race, "*Lisa Brown Buchanan, a teacher, wanted to examine what effect having predominantly White female teachers had on her and on her predominantly White female colleagues, who are also all teachers (Buchanan, 2015). She learned that they all grew up in predominantly White neighborhoods and attended predominantly White schools, so they did not like to bring up topics pertaining to race in the classroom because they didn't talk about it during their childhood. During the interviews, many of the teachers acknowledged that if they had more diverse surroundings as children of a different race. They concluded that having more diverse teachers benefits everyone, not just kids of color (Buchanan, 2015).

Race matching isn't good for just the self-esteem of students; it can potentially help raise their grades. A study tracked public school students in Florida using the Florida Comprehensive Assessment Test (FCAT) to see if student-teacher race matching would improve student test scores in reading and math (Egalite, Kisida, & Winters, 2015). When students have teachers of the same race or of a similar background, they score a bit higher on standardized testing. Black, Asian, and White students score higher on reading and math tests when their teacher is of the same race (Egalite et. al., 2015).

The improved test scores may be linked to not having a stereotype threat when a student has a teacher of the same race. Stereotype threat is an awareness that puts a student at risk of conforming to a stereotype of their social/racial group (Dee, 2005). For example, the pressure could be tougher on Black and Latino students to be better at math and science because stereotypes of their race and ethnicity say otherwise. On the flip side, Asian students might have different pressure to be good at math and science because stereotypes of their race say that they should be good at those subjects.

Thomas S. Dee's 2005 study analyzed the National Longitudinal Education Study of 1988 (NELS:88), in which he took teachers' ratings of students and concluded that teachers of the same race as their students rate them as less disruptive, more attentive, and likely to complete homework (Dee, 2005). Students can sense when a teacher doesn't like them, which in turn, can lead to a stereotype threat and disengagement from class. Taking inspiration from Dee, another study used NELS:02 to gauge teacher-student expectations, specifically if teachers of the same race as their students are more likely to expect their students to go on to college and get a degree (Gershenson, Holt, & Papageorge, 2015). Non-Black teachers are significantly less likely to expect Black students to attend and graduate college with a four-year degree (Gershenson et al., 2015). White males especially are 10-20 percentage points

more likely to have low expectations for Black female students than other teachers of other backgrounds, and in general, non-Black teachers are 12 percentage points less likely to believe that a Black student will complete a 4 year college degree than are Black teachers (Gershenson et al., 2015). It is likely that this biased judging can severely affect how a teacher behaves in the classroom.

Berkeley Carroll is a predominantly White institution (PWI). However, the school has a strong focus on diversity and equality for their students. This research study is investigating whether students truly are getting treated equally in classrooms. Inequality doesn't have to take on blatant and obvious discrimination, such as the use of racial slurs. What many of these studies looked at was unconscious or implicit bias, or the biases that people have that they aren't completely aware of (blind spots). These stereotypes are automatic and subconscious (Dee, 2005). The current study looks at simple actions teachers do (calling on students who raise their hands or cold calling students, as well as who is raising their hands more in class) that can inadvertently alienate some students in the classroom. This will be operationalized by the race of the students and teachers, hand raising and getting called on by the teacher, and a survey that will ask students and teachers about their classroom dynamics.

Procedure

The experimental problem posed is: How does the race of a teacher affect the way they interact with the students of the same or different race? This experiment is observational and correlated. The subjects were recruited by going to a faculty meeting and asking for volunteers; teachers were also emailed to ask for their participation. A total of four teachers volunteered: two White and two teachers of color. Teachers were observed while teaching a class for a total of 50 minutes and during that period, the number of cold calls, how many students raised hands, and the number of kids called on while their hands were raised, according to their race were recorded on a table. Teachers were also told that the study is about "effective teaching methods in the classroom" to combat the problem of hyper-awareness of how they may be treating students, which may cause them to not conduct a class how they normally would. Every class was observed for only 50 minutes because during the day there are three 75 minute periods and one 50 minute period. This ensures that every observation takes the same amount of time. A cold call is defined as calling on a student to participate in class- for example, answering a question, while their hand is *not* raised. This technique is often employed when a teacher suspects that a student is disengaged, as it may force them to pay attention. The reason why observational points such as "cold called" was used is because other variables like language and tone of voice can be interpreted differently by different people. What one person may deem as rude may not be rude to another.

After the classroom observation, a survey was sent out to both teachers and students, which asked them about their race and the class dynamic. The questions about the class dynamic (e.g. "Do you reach out to your students?" for the teachers) were rated on a scale of 1 (no/not at all/not connected) to 5 (very/very connected/yes). There were also short free response questions in the survey. For example, students and teachers are asked to identify methods that could be used to improve class relations. All four teachers and all 39 students answered the surveys. The survey was sent out after the classroom observation because the purpose of the study becomes very apparent, and it would bias the experimental finding if teachers and students were hyper-aware of race and who is being treated differently.

Templates for Recording Data

Raised Hand and Called On:

White	POC

Raised Hand:

White	POC

Cold-Called:

White	POC

Surveys

Teacher

- ____ How interested do you think your students are in the class? (1-5)
- ____ Do you feel any connection with your students? (1-5)
- ____ Do you try to reach out/relate to your students? (1-5)
 - How do you try to reach out/relate to your students? (short answer)
- ____ Do you think that your students reach out to you? (1-5)
- ____ How do you think student/teacher relations in your class can be improved? (short answer)
- Please state the race/ethnicity that you identify as:
- □ White or European American
- Black or African American
- Asian American/Pacific Islander
- Hispanic or Latin American
- Native American

Student

- How interested are you in the class? (1-5)
- ____ Do you try to reach out/relate to your teacher? (1-5)
- ____ How do you try to reach out/relate to your teacher? (short answer)
- ____ Is there anything your teacher does that makes you feel alienated?
 - Do they make you feel included? (short answer)

How do you think student/teacher relations in your class can be improved? (short answer)

Please state the race/ethnicity that you identify as:

- □ White or European American
- Black or African American
- Asian American/Pacific Islander
- Hispanic or Latin American
- Native American

Data

3 separate t-tests were used for each criteria: raised hand and called on, raised hand, and cold called. Because the number of White kids and kids of color in these classes were not approximately equal, the data had to be normalized with a 0-1 scale. Table 1 shows the number of times a student of color or White student raised their hand, raised their hand and got called on, or got cold called.

Table 1

Comparison of Teachers for Number of Students who Raised their Hand and Got Called On, Raised their Hand, and Got Cold Called On in their Class.

These tables are just the comparisons of what happened in each classroom during the 50 minute periods. Teachers 1 and 4 are teachers of color, whereas teachers 2 and 3 are White. This is raw data- no statistical test is done.

1.			2.		
	White	POC		White	POC
Raised Hand and Called on	15	3	Raised Hand and Called on	8	6
Raised Hand	20	5	Raised Hand	18	8
Cold Called	5	3	Cold Called	1	0
			_		
3.			4.		

э.		
	White	POC
Raised Hand and Called on	13	3
Raised Hand	14	5
Cold Called	0	0

4.					
	White	POC			
Raised Hand and Called on	23	7			
Raised Hand	28	9			
Cold Called	0	0			

Table 2

Comparisons of Teachers of Color (1) and White Teachers (2) for Students Who Raised their Hand and Got Called On, Raised their Hand, and Got Cold Called on in their Class on the Basis of Teacher Race.

The teachers had their numbers compiled-- this is similar to table 1, but provides an overall look of teachers of color vs White teachers instead of looking at each teacher separately. This is also raw data with no statistical test done.

1.		2.			
	White	POC		White	POC
Raised Hand and Called on	38	10	Raised Hand and Called on	21	9
Raised Hand	48	14	Raised Hand	32	13
Cold Called	5	3	Cold Called	1	0

Figure 1

Normalized Number of Times White Kids and Kids of Color Raised their Hands and Got Called On.

This measures the number of times White kids and kids of color raised their hand and got called on by the teacher (p value:0.427787). This is the average of the classes. There is no significant data. Students are not more likely to be called on after raising their hand due to either their race or their teacher's race.



Figure 2

Normalized Number of Times White Kids and Kids of Color Raised their Hands.

This measures the number of times White kids and kids of color raised their hands (p-value: 0.48852). There is no significant data. Students are not more likely to raise their hand due to either their race or their teacher's race.



Figure 3

Normalized Number of Times White Kids and Kids of Color Got Cold Called.

This measured the number of times White kids and kids of color got cold called (p-value: 0.4447895). There is no significant data. Students are not more likely to be cold called on due to either their race or their teacher's race.



From the surveys, teachers reported that they try to use humor to connect with their students. One teacher stated that they are "trying to go to an event that each of them [the students] is involved with this school year--athletic games, concerts, plays, etc. I also relate to them in the sense that my own children are their age so I can let them know that I know the challenges they face with balancing a busy schedule since we live it too". When suggesting how they improve class relations, some teachers responded that they try to enforce class participation by "suggesting" that certain students should speak- although they are not outright cold calling them. One teacher expressed their frustration about students not emailing when they need support. Emailing seems to be a common thread among students, as many responded that one way to become closer to teachers is to email them for help.

Many students outright said that they have never tried to connect with their teachers, but that they feel included regardless. The students who did try to establish a connection with their teacher most often used meetings outside of class to talk with their teachers more. Some students also said that a simple greeting in the hallway makes them feel closer to their teacher because they are acknowledged outside of the classroom.

There are more varied answers on how relations can be improved. Some students say that they themselves need to put in more work, while others think that the relationship is already good. However, it is noteworthy that a few students have an emphasis on knowing little personal stories about their teacher and talking outside of classroom about things not school related, with one student saying: "I think that I don't know anything about my teacher, no personal anecdotes or anything, and it would be nice if I knew something about him other than the class he teaches." This aspect of knowing personal stories about ones' teacher is something that hasn't been mentioned much in previous papers. This may have to do with the smaller setting in Berkeley Carroll and how much easier it is, and to an extent that it is expected for teachers and students to form closer relationships.

Some students gave a few extremely specific responses. A very small number of students said that in science class they don't necessarily feel included, as that class is not very discussion based. Perhaps it is easier to make connections in classes where you have to be personal and share stories, such as a humanities class. Some students included tidbits such as a teacher nodding and saying "mhm" too much, or that a specific teacher was sarcastic.

This study had no significant results, but the anecdotal evidence gives some insight into how students and teachers feel about their classes, what can be improved in terms of student and teacher relationships, and how those relationships can be improved. Ultimately, the scientific question posed was not answered. The data does not support the hypothesis that if a student is race matched with their teacher, those students will want to raise their hand more, will get called on more after their hands are raised, or will get cold called on more. However, this reflects the makeup of the classes. For example, a class of 16 kids may have only 4 kids of color. If those 4 kids raised their hands 5 times compared to the 15 times to the rest of the class, it cannot be said that those 4 kids raised their hands less- there is simply just fewer students of color in Berkeley Carroll. That problem is why the data had to be normalized. The four classes that were observed did not have an equal number of kids of color compared to White kids. Essentially, these results do not answer how a teacher's race affects their interactions with students of the same or different race.

Discussion

The templates that were used in this study did not capture all behaviors in the classroom. An interesting follow-up or continuation of this study would be to add more categories to the templates, based on what was observed from this study. One aspect that came up during observations is that certain kids do call out, and on a completely surface level analysis, it was mostly White kids, regardless of gender. Could it be due to cultural difference? Are there fewer reprimands for White students when they talk out of turn? That could be the subject of a future study to look at discipline methods for students of different races, especially in older children and teenagers.

There was also a problem with the survey. There was no option to choose more than one race, which speaks to a growing issue that multiracial and biracial people face. Someone who is biracial may be treated differently than someone who belongs to one race because of simple things such as skin tone. This also poses the question of visibility vs identity. For example, someone who may be visibly Black to their classmates and teachers may be biracial or multiracial. When observing in the classroom, there may be no way of knowing that person's racial identity because the survey is sent out after the observation.

Additionally, more controls could have been introduced but would be hard to manage in the setting of Berkeley Carroll. For example, it would be better if the classes all had the same number of students, or if the experiment could have been conducted for one subject or with an even mix of subjects. For example, would this dynamic of race in the classroom be different when it is a science class or math class? Would it change when the subject is history or English?

Lastly, taking into account intersectionality would be the most beneficial way to see who is being treated different and how- a Black male and a Black female may be treated differently from each other, likewise, a Black female and an Asian female may be treated differently, even though the Black female shares some traits with both people. Looking into how race and gender affect how one is treated in the classroom can bring us one step closer to overcoming bias.

References

Buchanan, L.B.(2015). "We make it controversial": Elementary preservice teachers' beliefs about race. <i>Teacher Education Quarterly 42</i> (1), 3-26.
Campbell, D.E. & Wolbrecht, Christina. (2006). See Jane run: Women politicians as role models for adolescents. <i>The Journal of Politics, 68</i> (2), 233-247.
Cherng, H.S. & Halpin, P.F. (2016). The importance of minority teachers: Student perceptions of minority versus white teachers. <i>Educational Researcher</i> , 45 (7), 407-420.
Dee, T. S.(2005). A teacher like me: Does race, ethnicity, or gender matter? American Economic Association, 95 (2), 158-165.
Disare, M. (2018, January 08). How diverse is the teaching force in your district? A new analysis highlights the gap between students and teachers of color. Retrieved from www.chalkbeat.org/posts/ny/2018/01/08/how-diverse-is-the-teaching-force-in-your-district-a-new- analysis-highlights-the-gap-between-students-and-teachers-of-color/
Egalite, A.J, Kisida, B., & Winters, M.A. (2015). Representation in the classroom: the effect of own-race teachers on student achievement. <i>Economics of Education Review, 45</i> (2015), 44-42.
Gershenson, S; Holt, Stephen B.; Papageorge, Nicholas W. (2015). Who believes in me? The effect of student-teacher demographic match on teacher expectations, <i>IZA Discussion Papers</i> , No. 9202
Yin, A. (2017, September 08). Education by the Numbers. Retrieved from www.nytimes.com/2017/09/08/magazine/education-by-the-numbers.html

Effects of Pore Space on the Structural Integrity of Concrete

by Diego Turturro

Abstract

y study is aimed at the question: Does adding pore space increase a 68.6 cm³ concrete block's resistance to heat? Molds were made of a 7.62 cm side length cube using Quikrete[®] fast setting concrete. The block set overnight and then was heated the next day on a hot plate. The change in temperature (Δ T) over one hour was defined as the heating rate. After doing two tests per pore size, I found that there was no correlation (p=0.98) between pore space and heating rate. A one-tailed, between subject, analysis of variance (ANOVA), f-test was used.

Background

Concrete is one of the most common materials that buildings are made out of, and each brand is vetted for structural durability before it is used (Han, Fu, Wang, Xie, 2018). In many desert cities such as Dubai, hot temperatures are abundant throughout the year, and the buildings have to maintain a certain level of structural integrity for decades at a time (Han, Fu, Wang, Xie, 2018). Thus, it is important to make sure that the materials inside the building can withstand and endure long stretches of heat. However, not all cities are that hot. The primary concern about heat when building with any material is fire safety (Tsuil and Chow, 2004). Although concrete is generally considered to be an extremely resistant material, a recent study suggested that there is only a small amount of data on concrete's properties at high temperatures (Han, Fu, Wang, Xie, 2018). These properties include strength, thermal conductivity (heat resistance), creep (cracking networks), and porosity. Furthermore, methods have only recently been standardized to measure thermal and mechanical properties of concrete. Newer types of concrete are focused more on economic efficiency, and they have performed poorly at higher temperatures (Han, Fu, Wang, Xie, 2018). With the uncertainties that come with the newer types of concrete, tests must be run to indicate their thermal conductivity and cracking networks under heat pressure.

Concrete is a mixture of three major components: cement, water, and aggregate. Aggregate consists of sand, small rocks, and gravel (PCA). Different concrete mixes have different amounts of aggregate and cement, so the amount of water added changes drastically depending on the brand. There are small spaces in the aggregate, so after the concrete sets, it is not uniformly dense (Lammi and Zhou, 2016). These spaces are formed because the ingredients don't fit together perfectly, creating separation between small rocks and pieces of sand. These small spaces are called pores, and they contain water and gas. When concrete is heated, the water evaporates as the vapor molecules speed up, expanding

00

the pore space. As that happens throughout the concrete, its structure starts to weaken, and if the pores become large enough the concrete will begin to crack (Lammi and Zhou, 2016). A study conducted by students at Georgia Tech investigated the formation of cracking networks in concrete by creating computer models of the interior of their HPC (high performance concrete) blocks, measuring stress and pore pressure that was caused by the evaporation of water (Lammi and Zhou, 2016). Aggregate size and volume were also varied. Each block was run through a series of algorithms to determine their reaction to rapid heating. The authors concluded that smaller aggregate sizes and less dense volumes caused the blocks to be weaker when heated, as the blocks' makeup increased the temperature gradient, which in turn damaged the block. Afterwards, the authors suggested that in a concrete block with no pores, the gas would have nowhere to move, thus making cracking, and possibly combustion (explosion), more likely to occur (Lammi & Zhou, 2016). However, a concrete block with excessive pore space would clearly be too weak from the start, so there must be a certain amount of space that is optimum for maximum resilience, or heat resistance.

A different study focused primarily on thickness, and how it affects heat resistance (Harada, Uesugi, & Masuda, 2006). To determine the resilience of wood boards, the researchers tested different thicknesses, noting the time it took each board to reach 200° Celsius. They measured the temperature at the furthest point from the side which was closest to the heat. The authors concluded that the thicker the board, the more heat resistant it is (Harada, Uesugi, & Masuda, 2006).

Instead of just using one measurement on a board to figure out heat resistance, it is possible to use multiple measurements if a gradient is used. A temperature gradient describes how temperatures change across a surface. Measuring it requires taking multiple temperatures at different points on the block, and seeing how each part of the concrete heats independently from other parts. In a study conducted to see how accurate a method is at predicting temperature, the researchers used gradients as one of their variables (Bobko, Zadeh, & Seracino, 2015). They conclude that most of the method was better than had been stated, but that there were still a few shortcomings with it. The authors mention how cracking is the result of a temperature gradient throughout the concrete, created as the hot gas leaves the pore space, cooling it down and creating different temperatures inside the block. When concrete is heated, it releases heat from its edges, but when some of the edges are cooler than others less heat is released throughout the concrete, causing a buildup of gas and energy. That, paired with increasing stiffness along the surfaces of the block, creates stresses in the concrete which leads to cracking. The authors also talk about how to control a temperature gradient. Gradients can be controlled by changing the proportions of the concrete mix, insulation sheets, and cooling methods before the concrete is heated (Bobko, Zadeh, & Seracino, 2015). Han, Fu, Wang, and Xie talk about gradients in their article as well. They state how as the temperature gradient increases, the stress on the concrete increases, which leads to cracking networks. Cracks are created when the stress on the concrete is greater than the tensile strength of the concrete. As stress increases, structural integrity starts to decrease, giving cracks the chance to expand. Once the stress is greater than the strength of the block, cracks are able to break open (Han et al., 2018).

The previously mentioned studies use ideas that I combined in my study. Many variables have been studied on concrete using mathematical models to predict concrete's strength, heat resistance, or something similar. I am looking at pore space in concrete, and how variability can increase or decrease the heat resistance of the blocks.

In the Spring of 2018, I ran a study to answer the experimental question: "Does adding holes in a 6 by 6 by 6 inch concrete block strengthen its resistance to heat?" That study attempted to find the optimum amount of pore space for maximum resilience. For the study, I used Quikrete[®]'s fast-setting concrete, because I wanted to make sure the concrete was able to set completely before being heated. The difference between fast-setting concrete and normal concrete is that fast-setting concrete sets within an hour, where normal concrete can take days to completely set. Using fast-setting concrete eliminates a possible variable, how much the concrete has set, that would have appeared had I used regular concrete. The concrete reaches its maximum strength a day after it has been mixed, reducing uncertainty in whether or not it is ready for heating. After gathering data on 8 blocks, no correlation was found between pore space and heat resistance (p=0.41). While constants such as volume, mass, and density were thought to be accounted for, it was clear that each of those variables were not always constant, as the blocks did not resemble what real concrete looks like as it is poured, and thus the data from that experiment can only be thought of as preliminary. The pore space was made by balloons, whose volume could only be measured approximately, adding to the uncertainty.

The current study addresses the experimental question "Does adding pore space inside a 442.5cm³ concrete block strengthen its resistance to heat?" Ping pong balls were used instead of balloons to reduce uncertainty in pore space. Mass and volume remained constant, as each pouring was checked for rocks that would be too large to mix. The independent variable was still pore space (number of ping pong balls), and the dependent variable was heat resistance (ΔT in each section of the gradient).

Procedure

Figure A

Two of the Larger Pieces Connected to the Base, Which is the 7.62 x 7.62 x 2.54 cm Piece.



Preparation of Concrete Blocks

Concrete blocks were constructed using plywood box molds with the following dimensions: 7.62 x 7.62 x 7.62 cm. The halfway point up the plywood box (3.81cm) was measured and marked with a thick pen or marker.

The concrete was sieved to attain even consistency by holding an aquarium net blazer above the container that the concrete mix was being poured into (Figure B). Figure C shows the concrete after it has been sieved. For each block 1062 grams of dry concrete was sieved. The concrete was poured from the container into a large bucket, which was used for mixing. A large tarp or garbage bag was placed under the bucket to ensure none of the concrete got on the floor. Separately, 255 mL of tap water was measured using a 1000 mL graduated cylinder. The water was poured into the bucket and mixed with the concrete using a plastic rod or a substitute mixing tool for 90 seconds.



Figure C Concrete Post-Sieving



Preparation of Pore Setting in Concrete Blocks

When the concrete was ready to be poured, the plywood box was stabilized on level ground before carefully pouring the concrete into it. If the concrete block was not going to contain ping pong balls, then the concrete was poured until it reached the top of the box, and then it was set aside to set. If the concrete block was going to contain ping pong balls, then the concrete was poured up to the labeled halfway point. If the box had one ping pong ball, a ruler was used to place the ball in the center of the square, so that when it set, the ball would be in the center of the cube. The ball should have been 1.75 cm above the bottom of the box and 1.75 cm to the left or right of each side. Figure 1 shows approximately what it should look like without the concrete. If the box has two ping pong balls, they were placed in opposite corners of the square to maximize their distance from one another. The balls should be 1.5 cm from the closest corner of the box, and have a slight separation (<1 cm) from one another. Figure 2 shows approximately what it looked like, without any of the concrete being poured yet. The additional concrete was poured until the box was full and the ping pong balls were completely covered. Each box was numbered on its side to identify the number of ping pong balls that it contained.

After the box was fully poured, it was placed on a shelf, and it set overnight. After the concrete was completely set a crowbar was used to take apart the sides of the plywood box mold, carefully as to not damage the concrete block. The plywood box mold would not be reused so the wood could be recycled.

Figure 1

Placement of Ball in a 1-Ball Box.

Figure 2

Placement of Balls in a 2-Ball Box.



Heating the Concrete

Two hot plates (Flinn Scientific) were placed inside a fume hood. One concrete block was put in the center of each hot plate. Each of the blocks had the same number of ping pong balls in them so as to have a duplicate of each data set. Points along the edges of each concrete block were marked at the 0 cm, 2.54cm, 5.08cm, and 7.62cm measurements. These points would allow the temperature gradient to be measured. A blast shield was positioned in front of the hot plates to protect from possible debris. An infrared thermometer (Etekcity Lasergrip) was used just outside the hood to measure and record the initial temperatures at the marked points for each block. Measurements were taken around the blast shield by pointing the thermometer at the marked points once every 10 minutes.

The hood's ventilation was turned on, then each hot plate was turned to its maximum setting (marked by the number 10 on the dial). The temperature of each marked point every 10 minutes was recorded. This was done for 1 hour, as after 1 hour the bottom most point had virtually reached its maximum temperature. Time, location and temperature for each observation were also noted. A template for the way the data was recorded is shown below.

Table 1

Heating of Concrete Block with Two Pores.

Block 1 2 Ping Pong Balls	0 cm	2.54 cm	5.08 cm	7.62 cm
Initial Temperature (C)				
10 min				
20				
30				
40				
50				
60				
ΔΤ				

After collecting the data, the hot plates were turned off and the hood stayed on for 20 minutes. Any observations at the completion of the heating process were recorded. After the hour, the blocks were kept on the hot plates for at least a 30 min to allow them to cool down. Then, using protective gloves, the blocks were removed from the hotplate and disposed of. These steps were repeated for each block. Concrete block samples contained 0,1 or 2 ping pong balls. Each was done in duplicates.

Data

The temperature of each point on Block 1, with no ping pong balls. Temperatures went up at a constant rate during the final 20 min, but the heating rate was not constant during the first 30 min (as seen by some points cooling down briefly).

The temperature of each point on Block 2, with no ping pong balls. The temperature increase was more constant in the highest three points on the block, while the bottom two points showed some variation.

Block 1



The temperature of each point on Block 3, with 1 ping pong ball, is shown below. Other than one distinctly low measurement at the 20 min mark, all points on the block displayed a temperature increase at a rate close to that of blocks 1 and 2. In minute 59 out of 60, the box exploded from the heat build up (shown in Figure D). The graph above on the right represents a second degree polynomial that predicts the temperature of the 7.62 cm point at 60 minutes. It is the trendline of the first five points in a second degree polynomial. The predicted value for temperature at the 7.62 cm point after 60 minutes was 117.3^e, meaning that the Δ T for that point would have been 95.8^e.



Block 3





The temperature of each point on Block 4, with 1 ping pong ball. The block exploded after 35 minutes, and at that point the temperatures were not very differentiable from one another. Thus, the data from this block cannot be used in a graph of ΔT after 60 min.

Figure D

Block 3 after Exploding



Block 4



The change from initial temperature to final temperature (ΔT) was compared for each condition using an ANOVA f-test to determine if the number of ping pong balls affects the heat resistance of the concrete. The ΔT for the 7.62 cm point was compared between the two conditions, with p equal to 0.98. There is no significant difference in heat resistance as a function of pore space.



Conclusions

The p-value of 0.98 suggests that the null hypothesis, which states that the amount of pore space in concrete blocks doesn't affect its resistance to heat, should be accepted. However, since both blocks with 1 ping pong ball exploded, logic suggests that pore space greater than or equal to one ping pong ball structurally weakens the concrete. No conclusions can be drawn about pore space that is less than one ping pong ball, as no conditions like it were tested. The limited number of samples (4) is a possible cause of the high p-value. With a more samples, a more accurate idea of how much the ping pong ball affects the structural integrity could be drawn.

Discussion

The data for the boxes with 2 ping pong balls has not yet been collected. After the block with one ping pong ball exploded, the procedure was altered so that only one block would be heated at a time. The p-value is suspect because of the limited number of samples that were collected, thus more samples for each block type will be collected. In the future, more graphs of predicted values will be created, as equations will be used to model the maximum temperature, and the amount of time needed, for each block before it explodes based on the trendline of previous points. As more data is collected, a proportion will be found between the pore space and time a block takes to explode on the hot plate's max setting. Using pore space as a volume in this proportion, and modelling it with a best fit line of a polynomial, an amount of pore space that maximizes the blocks' resistance to heat will be predicted.

The largest source of uncertainty in the study was the density of the concrete blocks. In previous iterations, the Quikrete wasn't sieved, and the sieve was larger than the aquarium net blazer. All the data in this data has been sieved. Even with the new sieve, the concrete lost the majority of one of its major components: small rocks. Almost all of them were sieved out, causing the block not to have any of

them during the mixing period, and causing the concrete not to be precisely what Quikrete talks about on their site. Also, after doing dimensional analysis to figure out how much water was needed to mix with the concrete, the thin mixture of sieved concrete resulted in the amount of water to be tripled, from 85 mL to 255 mL. The previous mixture did not let the concrete reach the correct consistency for mixing, and the new amount of water gave the concrete a better consistency. This is not the proportion that Quikrete suggests, and thus the concrete created was a custom build, and not affiliated with any one brand. Since the mix was custom, conclusions cannot be drawn about real world concrete until a variety of brands are tested under the custom boundaries. Another source of uncertainty was the hot plates. While the same two plates were used for the duration of the study, when the study is replicated the hot plates used could have a different setting than the plates that were used, leading to different temperatures and different predicted values. Lastly, the placement of the ping pong balls could differ slightly among the blocks. This is not expected to cause a major discrepancy, as the goal of the balls was to create pores, and they accomplished that wherever they were placed in the block.

References

- Bobko, C. P., Zadeh, V. Z., & Seracino, R. (2015). Improved schmidt method for predicting temperature development in mass concrete. ACI Materials Journal, 112 (4), 579-586.
- Han, Y., Fu, S., Wang, S., & Xie, Z. (2018). Study on adiabatic temperature rise reflecting hydration degree of concrete. Advances in Materials Science and Engineering, 2018, 12.
- Harada T, Uesugi S, Masuda H (2006) Fire resistance of thick wood-based boards. Journal of Wood Science, 52, 544–551
- Kodur, V. K. R., Dwaikat, M. M. S., & Dwaikat, M. B. (2008). High-temperature properties of concrete for fire resistance modeling of structures. ACI Materials Journal, 105 (5), 517-527.
- Lammi, C., & Zhou, M. (2016). Multi-physics modeling of fire-induced damage in high-performance concrete. *The International Journal of Multiphysics*, 8(1).
- Östman, B., & Boström, L. (2015). Fire protection ability of wood coverings. *Fire Technology, 51,* 1475-1493.
- Portland Cement Association (n.d.). Retrieved from www.cement.org/cement-concrete-applications/ concrete-materials/aggregates
- Quikrete (n.d.). Retrieved from www.quikrete.com/productlines/fastsettingconcretemix.asp
- Tsuil, S. C., & Chow, W. K. (2004). Legislation aspects of fire safety management in hong kong. Facilities, 22 (5), 149-164

Special thanks to:

Dr. Gaines, Dr. Gentile, Mr. Rubin, Mr. Byrne, Ms. Sefchovich, Max B. Matty J. and Isabel B.

Daily Fantasy Sports and the Stock Market: An Investigation Into the Deception of Online Advice

by Charlie Welch

Abstract

aily Fantasy Sports (DFS) is an emerging industry that involves creating lineups of athletes on a nightly basis to win money based on their real-life performance. I investigated whether there was a significant difference between the average lineup scores of paid advice, free advice, and my algorithm on the Yahoo DFS platform. I inputted lineups from the three groups into the Yahoo DFS Website first for basketball and then re-ran the experiment for American football and baseball. A one-way independent Analysis of Variance (ANOVA) test revealed there was no significant difference between the average lineup scores of the three groups for basketball (p = 0.305, n = 35), the four groups for baseball (p = 0.18, n = 35), or the three groups for American football (p=0.68, n=25). The study was then expanded to compare third-party advice sources in the stock market, which resulted in no significant difference between the performance of five portfolios (p=0.99, n=30).

Background

In any industry where customers are taking risks with their money, there are going to be those who provide advice. Two industries, Daily Fantasy Sports (DFS) and the stock market, fall under this category. In both industries, there are plentiful ways customers choose to participate. These options range from not using any outside sources (making your own decisions) to using free or paid advice (advice you pay a subscription to access). In both industries, it can be unclear what is the best way to play. Is it worth your money to pay for advice? Customers are faced with this gamble as they put their own money on the line in an attempt to make more money. To understand how people make these decisions, one must first understand the context for third-party advice sites in both DFS and the stock market.

DFS is a rapidly growing industry which is relatively new in the field of fantasy sports (Leishman, 2016). DFS involves creating a lineup (a set of players) from the pool of real life athletes playing on a given night, and winning money if this lineup performs well compared to other lineups. In contrast with traditional fantasy sports leagues which "require a fan to commit many hours to drafts, weekly analysis, and consumption of game broadcasts over the course of a simulated season" (Leishman, 2016), DFS offers a nightly experience with less commitment. The industry has experienced massive growth in users, with Fanduel, a DFS platform, experiencing an increase from 40,000 users to over one million in two

years (Leishman, 2016). Draftkings, the rival site of Fanduel, experienced similar growth over this time (Leishman, 2016).

In addition to this dramatic growth, the new daily format has resulted in customers spending more money than in season long formats (Weiner & Dwyer, 2017). The amount of money in this new, exciting version of fantasy sports has led its users to seek an upper hand. As Burke describes, "The growth of the industry has given way to additional services that players can use to gain an edge" (Burke, Meek, & Norwood, 2016). According to Fantasylabs, a DFS site, these additional services include DFS sites that provide lineups to anyone who takes the time to access them online (Cabezas, 2016). However, some sites have a subscription fee that appears to set them above other sites. The abundance of these online services has created an environment where these resources may no longer give players an edge (Cabezas, 2016). However, this claim is an unsupported opinion. Currently there are no known studies that support this claim by statistically analyzing the performance of a paid site in comparison to a free site. I set out to investigate if this claim by Fantasylabs has merit by comparing the performance of lineups created by a paid site (Rotowire.com) to the performance of lineups created by a free site (Rotogrinders.com).

Rotowire.com portrays itself very highly to its customers and potential customers. In its extensive section titled "Rotowire Industry Awards", it lists its most recent award as "Best News and Analysis Site: Multiple Sports 2017" (Rotowire, 2017), which was given out by the FSTA (Fantasy Sports Trade Association). Rotowire.com being called the "Best Analysis Site", with analysis referring to their ability to predict player performance, implies that its lineups are amongst the best in the industry, a very attractive accolade to customers. If they were actually the "Best Analysis Site", you could reasonably expect their lineups to outperform the lineups of other lineup tools, such as a free site (Rotogrinders.com) and/or a simple algorithm lineup.

Daily Fantasy Sports is not the only industry where customers turn to third- party platforms for advice. The stock market also lends itself to a host of resources customers can access as they make decisions on how to invest their money. One of these options is investing money in a mutual fund, which is an investment program that is professionally managed (Marekwica, & Steininger, 2014). Mutual funds allow people to invest their money with professionals who, for the most part, have experience in investing. By 2010, there were almost 25 trillion dollars under management by mutual funds, which was a 39% increase from 2005 (Marekwica, & Steininger, 2014). This growth of money in mutual funds has coincided with the staggering growth of the number of different mutual funds, and by 2010, investors had almost 70,000 funds to choose from (Marekwica, & Steininger, 2014). The advantages of mutual funds are that they provide the structure for people to invest in well-diversified portfolios (which minimizes risk) and they allow for the restructuring of an investment portfolio at a cheaper price than a private investor (Marekwica, & Steininger, 2014). However, the glaring downside of mutual funds is the management service fees (usually as a percent of the money invested) that are imposed onto customers, which can significantly cut back on a customer's portfolio return. Customers are tasked with deciding if the fee is worth their money when considering the annual stock market return of a given mutual fund.

An alternative way people choose to invest their money involves doing external research about stocks and making decisions based off of this research. One of the most popular investment advice services is The Motley Fool, an internet stock chat site which gives stock market advice and is part of the America Online (AOL) website (Hirschey, Richardson, & Scholz, 2000). The Motley Fool (TMF), which originated as an offline newsletter, became incredibly popular as a stock advice provider and chat room online service. TMF accumulated over 17 million members by the year 2000 (Hirschey et al., 2000). As a service, TMF is as comprehensive as it is popular, with features including a daily portfolio update that keeps track of day-to-day stock performance and an immensely important nightly performance recap

(Hirschey et al., 2000). Additionally, what separates TMF from mutual funds is that it only costs a yearly subscription of fewer than two hundred dollars, compared to mutual funds which tend to take between 0.25% and 1.5% annually of the money you invest with them (Hirschey et al., 2000).

The core difference between these two services is the level of involvement on the part of the investor. Mutual funds allow you to hand your money over to trained investment professionals who place this money in established portfolios, hence the costly management service fees. The Motley Fool, on the other hand, allows its customers to act as the investors and gives advice on which stocks to pick instead of actively managing one's assets. However, with this freedom comes potential deception on the part of TMF. When TMF was founded in 1993, they claimed that the purpose of their service was to "educate, enlighten, and entertain investors" (Hirschey et al., 2000). They later continued by explaining that "the whole point of Foolishness (TMF) is to make your own decisions... You'll never actually learn anything... if you're copying others" (Hirschey et al., 2000).

The irony in this statement is that, for many customers, TMF is at its essence about copying others, or at least closely following advice from others. Whether this is through a chat room or an article, TMF allows customers to take most of the decision making out of the investing process and put their faith in the experts. While TMF can be used as a learning tool (there are chat rooms and radio talks held by them), it can also be used as a service to directly copy the stock picks they make when crafting your own portfolio (Hirschey et al., 2000). In my opinion, TMF's message seems to contradict the reality of how it is used and presents a potentially deceiving situation for customers, as they copy the stock picks of an online source without understanding the reasoning behind the picks.

Between the 70,000 mutual funds, TMF, other stock advice sites, and the option of investing without outside influence, it is clear there are many ways to approach investing money. How does a potential investor choose between these options, and how can they know which one to trust? Is TMF truly giving its customers insights that will elevate them above the rest of the market?

In his 2003 paper Internet Investment Advice: Investing with a Rock of Salt, Michael Dewally explored the quality of investment advice distributed by major newsgroups between 1999 and 2001 (Dewally, 2003). After collecting the stock picks from the news sources, he categorized them into three groups: buy, sell, and neutral. He found that the stock recommendations had an overwhelmingly positive buy to sell ratio of 7 to 1 (Dewally, 2003). For the broader conclusions of the study, he found that over both the short and long term scenarios, the advice did not perform significantly above the stock market's corresponding performance, and therefore offers little informational value to an investor (Dewally, 2003).

This study hopes to build off of Dewally's finding and update the concept of the study to be applied to the immensely complex modern investment scene. Dewally's study looked at the performance of advice from different news sources when compared to the stock market as a whole, but the current study will expand this by comparing the performance between five different groups: TMF (a paid advice service), a free advice service, a mutual fund, a randomly generated portfolio, and the stock market as a whole. With the multitude of investment vehicles available to investors, perhaps this could provide perspective on how these different methods compare to each other, and help customers make better informed decisions.

While the Daily Fantasy Sports (DFS) Industry and the stock market are seemingly very different, it is the many underlying similarities between the two that allows them to be used together. For both industries, this study explores the quality of advice given by third-party sources, and seeing how this performance compares to other resources available to customers. For DFS, this comparison is between a paid site, a free site, and an algorithm. For stocks, the comparison is between the five groups listed above, a similar range of options. The algorithm lineup for DFS and the randomly generated portfolio for

the stock market serve a similar purpose, which is to act as a baseline group to compare the other groups. For example, if TMF advice can't outperform a randomly selected portfolio, then clearly TMF is not worth your money.

Another similarity between the two industries is how both can be viewed as a form of gambling. For DFS, customers make decisions on which players to pick, and the performance of this player determines if the customer wins or loses money. Similarly, investors choose stocks to buy and sell, and these decisions can lead to the investor gaining or losing money based off the performance of these stocks. It is this monetary risk involved in both industries that funnels people towards expert advice sources, and allows customers to eradicate their own hesitancy by putting the decisions in the hands of professionals.

For DFS, this study investigates whether the performance of lineups created by a paid site (Rotowire.com) is significantly different from the the performance of lineups created by a free site (Rotogrinders.com), an expert lineup (only for baseball, created using free advice of Mac McClure, CBS Sports), and a simple algorithm lineup. This comparison hopes to provide insight in an industry where it is unclear who is the best. For the stock market portion of the study, a similar question is investigated, with the stock performance of five groups being compared: The Motley Fool (paid advice), The Street (free advice), the Fidelity Contrafund (a randomly selected mutual fund), a portfolio of ten randomly selected stocks from the S&P 500, and the S&P 500 index as a whole. This study investigates third-party advice sources in both industries and helps better inform customers in each. While seemingly unrelated, the stock market and the DFS industry logically follow together in a study that hopes to uncover similar problems in both industries.

Procedure

The study was split into two parts. The first part dealt with the industry of Daily Fantasy Sports (DFS), and the second part with the stock market. While treated separately during the experimental process, the overarching themes between them (third-party advice sources) allow connections to be drawn between industries. My main experimental question is whether there was a significant difference between the average lineup scores of paid advice, free advice, and my algorithm on the Yahoo DFS platform.

PART ONE

Overview

Part one investigates the performance of different third-party advice sources in DFS, and specifically compares the quality of advice between paid (subscription based) and free advice sources. Three DFS sports were investigated separately: basketball, baseball, and American football. For all three sports, the independent variable was the type of lineup (type of lineup dictated lineup creation), and the dependent variable was the average points scored by that lineup (performance). The method is only described for the basketball portion of the study, since the baseball and football portions are the same except for a few indicated differences. Control measures included entering all lineups in a given trial into the same time slot. This helped keep the player pool consistent across lineup types. Another control measure was inputting all lineups of a given trial at the same time. This controls for the fact that DFS sites have more information about players (injuries for example) closer to the start time of the games; therefore, it would be inconsistent to create lineups at different times throughout the day.

Methods

For the basketball portion of my study, three separate lineups were compared: a free lineup (rotogrinders.com), paid lineup (rotowire.com), and an algorithm lineup (lineup I created accounting for variables of home/away, nights of rest, and strength of opposing defense). A lineup is defined as the set of ten players selected to represent your entry on a given night. The website used to compare lineups was the Yahoo DFS platform (yahoo.com). One trial was defined as comparing the scores of the three lineups for one night's games (this is how DFS works, one "round" per night). Lineups acquire a score (points) based on the real-world performance of the athletes in a lineup on that given night. At the same time on a given day, the lineups of the different websites and my algorithm were entered on a spread-sheet. The following morning, the performance (points scored) of these lineups was recorded, with each trial containing a single number (points scored) for each lineup type. This process was repeated for as many times as possible (thirty five trials for basketball and baseball, twenty five trials for American football).

The process was mostly the same for the subsequent baseball and American football portions of the study. For baseball, trials were once again conducted on a nightly basis, with the only difference being a fourth lineup type was added (Expert Advice lineup from Mike McClure of CBS sports). For American football, three lineups were compared once again (free: Rotogrinders, paid: Rotowire, and algorithm). Trials were conducted each Sunday (the day with the most games), rather than on a nightly basis like the other two sports. Except these differences, all three sports were investigated with the same process.

PART TWO

Overview

The second part of the study investigated the quality of advice sources in the stock market. Five different stock portfolios were compared (detailed below). The independent variable was the type of portfolio and the dependent variable was the portfolio performance (the average daily percentage increase or decrease of the portfolio). A control measure was creating all portfolios at the same time. This controlled for the varying performance of the markets during different parts of the year, and therefore all portfolios had exposure to the same days of the market.

Methods

In this portion of the study, five portfolios were compared. These different groups were paid advice (The Motley Fool, fool.com), free advice (The Street, thestreet.com), a randomly generated portfolio, the Fidelity Contrafund (a large mutual fund), and the performance of the market as a whole (market refers to the S&P 500 index, as this index is comprehensive and widely used). The first three groups' portfolios were created through Investopedia's stock market simulator (investopedia.com). This simulator allows users to make stock trades using virtual currency and the stocks' performance depends on the performance of the real stock market. This platform was used because it simulates the market without the investment of real money, and it is a flexible platform which allows the creation and comparison of several different portfolios.

For these three portfolios, each stock made up an equal portion of the portfolio on Investopedia. Additionally, each portfolio on Investopedia remained constant for the duration of the experiment after the initial portfolio creation. For the Motley Fool, the portfolio consisted of ten randomly selected "buy" recommendations by the site. For "The Street" portfolio, nine stocks that are listed under an "A" rating were randomly selected and make up the portfolio. The portfolio consisted of ten randomly selected stocks from the S&P 500. For each random selection, all items from the appropriate site are entered on a spreadsheet, and then a random number generator selected ten numbers which corresponded to a row number. The stocks in these rows comprised the portfolio. Because the performance of the Fidelity Contrafund and the S&P 500 can be tracked online daily without entering it through Investopedia, they were not tracked on the Investopedia platform.

All portfolios were created and tracked at the same time and the daily performance of each portfolio was recorded in a spreadsheet at the end of each day the market was open (all days excluding weekends and some holidays). Each trial was defined as the performance of the five groups on a given day. Portfolio performance is defined as the percent decrease (or increase) of the portfolio on a given day. For example, if at the end of the first day, the Motley Fool portfolio had increased by one percent, that is entered as a data point (+1).

The setup of both parts of the experiment, the DFS sports section and the stock market portion, allows for an independent one-way ANOVA test between the respective groups (run on vassarstats.net). For the DFS portion, the average points scored for each lineup was compared, and for the stock market portion, the daily average percent change was compared between the portfolios. If the paid group performed significantly better than the other options, the ANOVA test allows this difference to become clear.

Stock Market Portfolio Composition

Stock Index: https://binged.it/2J7lppG Mutual Fund: https://binged.it/2XVBV0z

For the three portfolios created on the Investopedia stock market simulator, each stock comprised an equal portion of the portfolio.

Random Portfolio (10 stocks):

JPMORGAN CHASE & CO ACCENTURE PLC CLASS A CELGENE CORPORATION LOWE'S COMPANIES, INC. CISCO SYSTEMS, INC. GILEAD SCIENCES, INC. UNITEDHEALTH GROUP INCORPORATED NVIDIA CORPORATION BRISTOL-MYERS SQUIBB COMPANY DANAHER CORPORATION Free Portfolio: The Street (9 stocks)

TEXTRON INC. ALPHABET INC. CLASS A CONSTELLATION BRANDS, INC. CLASS A GARMIN LTD. CUTERA, INC. ULTA BEAUTY INC FIREEYE, INC. DELUXE CORPORATION AT&T INC.

Paid Portfolio: The Motley Fool (10 stocks)

STITCH FIX, INC. CLASS A APPLE INC. BOOKING HOLDINGS INC. ARISTA NETWORKS, INC. SHOPIFY, INC. CLASS A GRAND CANYON EDUCATION, INC. WIX.COM LTD. UNION PACIFIC CORPORATION AMGEN INC ZSCALER, INC.

Data Analysis

DAILY FANTASY SPORTS

Basketball Data Analysis

To answer the main experimental question, an independent one-way ANOVA test was run between the three basketball groups. A one-way ANOVA test was used because it compares the averages of more than two groups, and it is independent because the design is between-subject. The ANOVA test yielded no significant difference between the averages of the paid, free and algorithm lineups (p = 0.305, n = 35). Therefore, the null hypothesis, which is that there is no significant difference between the groups, is accepted. Figure 1 displays the results of the first ANOVA test.

Figure 1

Lineup Type vs. Average Points Scored.

Lineup performance is being compared by lineup type (p=0.305, n=35). Error bars ±SEM



Along the lines of customer deception, I was interested in whether there was a significant difference between the predicted points for a lineup and its actual score. If the predicted point averages were higher than the actual score averages, this would indicate that DFS sites are overselling their lineup creation abilities, and deceiving their customers by doing so. A higher actual score average, while unlikely, would mean that the sites are underselling their ability.

For basketball, three separate independent two-tailed t-tests were run, with each t-test comparing the average predicted points to the average actual points of a given lineup creation technique. The t-tests are independent because the experiment is between-subject, and they are two-tailed because I was testing for a difference in either direction. The tests yielded no significance between the predicted score averages and actual score averages for the free site (p = 0.07, n = 35) and the algorithm (p = 0.76, n = 35), but did yield a significant difference for the paid site (p = 0.027, n = 35). Therefore, the alternative hypothesis is accepted for the paid site t-test, but the null hypothesis is accepted for the free site and algorithm t-tests. Figure 2 shows the results of the three t-tests. The result of the paid site t-test is particularly interesting because it reveals that Rotowire.com is telling its customers that its lineups will produce a high amount of points, but the reality is that its lineups produce significantly less.

Figure 2

Lineup Type vs. Average Points Scored.

Predicted and actual lineup performance is being compared for each lineup type (paid: p=0.27, n=35, free: p=0.07, n=35, algorithm: p=0.76, n=35). Error bars ±SEM



Baseball Data Analysis

To answer the main experimental question for baseball, an independent one-way ANOVA test was run between the four baseball groups. A one-way ANOVA test was used because it compares the averages of more than two groups, and it is independent because the design is between-subject. The ANOVA test yielded no significant difference between the averages of the free, paid, algorithm, and expert lineups (p = 0.18, n = 35). Therefore, the null hypothesis is accepted which is that there is no significant difference between the results of the first ANOVA test.

Figure 3

Lineup Type vs. Average Points Scored.

Lineup performance is being compared by lineup type (p=0.18, n=35). Error bars ±SEM



In addition to comparing the lineup averages of the four groups, the predicted scores of the lineup (provided by the site) were compared to the actual scores of the lineup to see if the sites were promising a performance that they could not match. Four independent two-tailed t-tests were run, one test for each lineup type. All four t-tests yielded no significant difference between the predicted average points and the actual average points (Free: p = 0.12, Paid: p=0.09, Algorithm: p = 0.08, Expert Advice: p = 0.26, n = 35 for all). Figures 4 shows the results of the four t-tests. There is no significant difference between the average predicted points and the average actual points for all groups, so the null hypothesis is accepted in all four cases.

Figure 4

Lineup Type vs. Average Points Scored.

Predicted and actual lineup performance is being compared for each lineup type (free: p=0.12, n=35, paid: p=0.09, n=35, algorithm: p=0.08, n=35, expert advice: p=0.26, n=35). Error bars ±SEM



Type of Lineup

American Football Data Analysis

Like for baseball and basketball, an independent one-way ANOVA test was run between the three football groups. A one-way ANOVA test was used because it compares the averages of more than two groups, and it is independent because the design is between-subject. An independent one-way between-subject ANOVA test yielded no significant difference between the averages of the algorithm, paid, free lineups (p = 0.68, n = 25). Therefore, the null hypothesis is accepted which is that there is no significant difference between the groups. Figure 5 shows the results of the ANOVA test. This result falls in line with the findings of baseball and basketball, and suggests that the paid lineup is not offering superior advice to the free and algorithm lineups. Football did not have the feature of predicted performance so, unlike baseball and basketball, this could not be analyzed.

Figure 5

Lineup Type Average Points Scored.

Lineup performance is being compared by lineup type (p=0.305, n=35). Error bars ±SEM



Stock Market Data Analysis

An independent one-way ANOVA test was run between the five portfolios. A one-way ANOVA test was used because it compares the averages of more than two groups, and it is independent because the design is between-subject. The ANOVA test and yielded no significant difference between the averages of the S&P, Random, Paid, Free, and the Fidelity Contrafund (p = 0.99, n = 30). Therefore, the null hypothesis is accepted which is that there is no significant difference between the groups. Figure 6 shows the result of the ANOVA test. The reason the error bars seem large in relation to the averages is that the averages include negative and positive but the error bars are an absolute value. The results from the stock market portion fall in line with the results of the DFS portion. Paid advice sources are not providing advice that is performing significantly better than the other groups, and this is consistent across all three sports and the stock market.

Figure 6

Type of Portfolio vs. Average Daily Percent Change.

Portfolio performance is being compared by portfolio type (p=0.99, n=30). Error bars ±SEM



Discussion

In viewing the results from each portion together, it appears that across the board, the paid advice service is not performing significantly better than the other advice options. Whether it be in DFS or the stock market, these findings indicate that perhaps customers should be careful about paying for advice that may, in fact, not be effective. However, there are several limitations in the scope of this finding. First off, since the study was conducted on the Yahoo DFS platform, it would be unfair to generalize any finding to other DFS platforms. The performance of the paid site (Rotowire.com) may vary by platform. Another limitation of scope is with the specificity of the sites used. For DFS, Rotowire.com was the paid site and Rotogrinders.com was the free site, but these are just two of the many different sources people use when playing DFS. Therefore, it would be premature to claim that paid site and free site perform the same across the board for DFS. Rather, in the specific context of my study, Rotowire did not perform significantly different from Rotogrinders.

The same limitation applies to the stock market portion of the study, in that The Motley Fool and The Street were the paid and free source, respectively. Even more than DFS, there is an immense variety of ways to get advice on stock picks, and findings should not be generalized. An additional limitation for the stock market portion was that the data was gathered during a specific set of months in which the market did not perform well, as the return of all portfolios was negative, and the stock market historically has increased over time (Hirschey et al., 2000). Therefore, the data is limited to a context when the stock market performed poorly, and a better time in the market could produce different results.

As an extension of these limitations, some potential follow-ups to my study include running the DFS portion on different platforms (such as DraftKings or FanDuel) and/or using different sources for the paid and free lineups. For the stock market portion, running it again during a different time of the year

(and possibly with different sites) could produce different results and provide an interesting contrast. All these possible follow-ups would help broaden the scope of the study and generalize its findings.

Despite the limitations, it is important to remember the connecting link between the DFS and stock market portions. Both sections examine the quality of third-party advice sources, and this is a concept which extends beyond both industries and to many more, from elections to real estate. It is exciting to see how analyzing the quality of these sources can provide insights to customers who have little data to make their decision.

References

- Burke, D. D., Meek, B. P., & Norwood, J. M. (2016). Exploring the legality of the lucrative world of fantasy sports. *Journal of Legal, Ethical and Regulatory Issues*, 19(1), 38-55.
- Cabezas, J. (2016, March 03). Polling the FantasyLabs team: What is the edge in DFS? Retrieved June 03, 2018, from www.fantasylabs.com/articles/polling-the-fantasylabs-team-what-is -the-edge-in-dfs/
- Dewally, M. (2003). Internet investment advice: Investing with a rock of salt. Financial Analysts Journal, 59 (4), 65-77. Retrieved from https://search.proquest.com/docview/ 219184242?accountid=35114
- Hirschey, M., Richardson, V. J., & Scholz, S. (2000). How "foolish" are internet investors? *Financial Analysts Journal*, 56 (1), 62-69. Retrieved from https://search.proquest.com/docview/ 219223730?accountid=35114
- Leishman, N. C. (2016). Daily fantasy sports as game of chance: Distinction without a meaningful difference? *Brigham Young University Law Review, 2016* (3), 1043-1080.
- Marekwica, M., & Steininger, B. I. (2014). The tradeoff between mutual fund and direct stock investments: A theoretical analysis involving different types of investors. *Review of Managerial Science*, 8 (2), 197-224. doi:http://dx.doi.org/10.1007/s11846-013-0101-9
- Roto Sports Inc. (n.d.). Fantasy baseball, football, basketball, hockey and more. Retrieved March 1, 2018, from www.rotowire.com/
- Weiner, J., & Dwyer, B. (2017). A new player in the game: Examining differences in motives and consumption between traditional, hybrid, and daily fantasy sport users. *Sport Marketing Quarterly, 26* (3), 140-152.

Testing Alternative Presentations of Word Lists with the DRM Paradigm

by Brendan Zelikman

Abstract

his study attempted to determine whether or not high school students are more likely to recall false memories through the Deese-Roediger-McDermott (DRM) paradigm when presented with word lists in a classroom setting as opposed to just hearing them read aloud. The procedure involved two distinct methods: part one was the DRM paradigm, involving the recall of word lists and a recognition test (the list section), and parts two and three involved the teaching of a class structured around a word list and a recognition test as well (the class sections). There was no evidence indicating that there was a significant difference between false recall rates in list- and class-sections (p=0.3336). However, the high recall rates of lures within classes implies that they can indeed foster the creation of false memories to some extent.

Background

False memories have long been a point of interest in the field of psychology. Professionally, they are defined as "a mental experience that is mistakenly taken to be a veridical representation of an event from one's personal past" (Johnson, 2001). Researchers such as Sigmund Freud and Pierre Janet investigated false memories in the early 20th century, and they continue to be intensely studied today (Gleaves et al., 2004). Experimental procedures have ranged from a controlled recall of word lists to the suggestion and provision of misinformation to manipulate memories (Deese). False memories have important implications, providing a possibility for false or distorted testimonies in court or false memories remembered in therapy. Therefore, it is crucial to understand how false memories are created and function so they can be controlled for during memory-reliant experiences or situations.

Though memories as a whole are too vast to cover entirely here, the relevant kind of memory operating in this context is "associative memory," the ability for humans to mentally connect items like ideas or feelings (Kohonen, 2012). Items can be connected either spatially if they occur simultaneously or temporally if they occur in close succession (Kohonen, 2012). For example, one might remember their location when recalling an accident (spatially) or be able to piece together the plot of a movie chronologically (temporally). Additionally, items can be connected semantically, if they are either similar or contrary to one another (Kohonen, 2012). A word such as "hot" would naturally invoke thought of related words like "heat" or opposing words like "cold." Information is thus accessed from memory based off of "some measure of similarity relating to the key pattern" (Kohonen, 2012). Though simple, this is the foundation of associative memory.

There is no one part of the brain solely dedicated to memory; rather, memory employs and leaves traces over several areas (Kohonen, 2012). However, the hippocampus is well known as a crucial region for associative memory, serving as a "hub" to bind information from distinct sources together. (Wang et al., 2014). In a study called *"Targeted Enhancement of Cortical-Hippocampal Brain Networks and Associative Memory,"* researchers found that the hippocampus successfully interacts with multiple regions of the brain to create associative memories (Wang et al., 2014). For each subject, they targeted a region of the parietal lobe that showed maximum connectivity with a given part of the hippocampus. The researchers applied electro-magnetic stimulation to the parietal lobe for several days and found persistent improvements in memory and increased functional connectivity in subjects (Wang et al., 2014). Their findings, supported by functional magnetic resonance imaging (fMRI), novelly demonstrated the necessity of the cortal-hippocampal network for associative memory (Wang et al., 2014).

The hippocampus is equally important for the creation of false memories. A region heavily involved in fragmenting and piecing together recollections of memories would show activity regardless of the validity of the memory (Okado & Stark, 2005). And indeed, there is greater brain activity during the study of subsequently remembered items than for forgotten items, otherwise known as the "difference due to memory (Dm) effect" (Okado & Stark, 2005). For example, if a subject is presented with the words "car" and "fish" but can only recall "car," the subject would have had greater brain activity during the study of "car" than during "fish."

In a study entitled "Neural activity during encoding predicts false memories created by misinformation," subjects were instructed to study eight vignettes (composed of fifty still images) of an original event and, shortly afterwards, eight vignettes of the event with altered critical details in what was dubbed a "Misinformation Phase" (Okado & Stark, 2005). The subjects were then called in two days later to assess their memory recall. Subjects took a two-part test in which they recalled events from the vignettes and then attributed them to their proper sources. It was found that subjects displayed increased activity in the left hippocampus during the encoding of original and false memories that they later recalled (Okado & Stark, 2005). Encoding is the initial process in forming memories in which one "receives and registers information" into the brain; it is necessary in creating "memory representations" of experienced events and is dependent on attention paid during an event ("Memory", University of California San Francisco [2019]).

Since subjects had to both recall the events during the recognition tests and accurately attribute them to their proper phases, the researchers affirmed that the left hippocampus is significantly involved with the encoding of both items and their sources (Okado & Stark, 2005). It could either strongly encode a source but not its item, leading to a hazy memory, or strongly encode an item but not its source, leading to a false memory. Additionally, the latter could arise if an item is remembered and then falsely put into an association, given that the hippocampus is also involved with grouping items into associative memories. For instance, one could potentially see a car accident in real life and then falsely think it had been part of an action movie they had seen that day, since the movie would have likely created associations of shocking or intense events; the hippocampus, being involved with the encoding of the accident and the movie, could have inadvertently associated them together and created a false memory.

Furthermore, the creation of false memories by association can be easily tested through the Deese Roediger-McDermott (DRM) paradigm, a major breakthrough in procedure for the study of false memories. Refining James Deese's earlier findings, Roediger and McDermott (1995) found that having subjects listen to and study semantically related words causes them to recall a related but non-presented word or "lure." A lure, in this context, would be a specific word that is not present in the list but that the list would hint towards. For example, a word list consisting of "tired, bed, rest, drowsy" would lead towards
a lure of "sleep". Upon questioning, half of the subjects were positive that they remembered hearing the lure, implying that they could mentally relive the experience (Roediger & McDermott, 1995). Thus, a false memory had been successfully created with a word that wasn't read aloud at any point.

The scope of the study is limited, since the false memories had originated from deliberately crafted word lists being read aloud instead of the actual recollection of experiences or events, a more realistic and practical scenario. However, the significance of the study cannot be understated, since subjects indeed remembered an experience, even a brief auditory one, that they hadn't undergone. Associative memory played an important role in the experiment, since subjects were consistently exposed to semantically related words that inevitably create such a mental connection. In his original paper from 1959, Deese (as cited in Roediger & McDermott, 1995) showed that subjects both recalled the list lures in free association tests (when they freely wrote down words) and had a greater chance of falsely recalling them later if they did. Since Roediger and McDermott (1995) used longer, fifteen-word lists, the mental associative processes were only amplified with even more words to semantically hint towards the lure, making a false recall all the more likely.

It is of interest to determine whether association can be strengthened even further while still using word lists. We can attempt to have subjects evoke images of objects, which are widely regarded as the "mental representative of meaning," or what we "see" in our minds (Paivio, 1969). As such, images are likely connected with associative experiences involving concrete objects, since they help symbolically connect them together (Paivio, 1969). They can also vary in strength of association, as more concrete words are more likely to evoke images to aid in associative learning and memory (Paivio, 1969). For example, a word such as "house" has a clear image associated with it, whereas a word such as "truth" has no definitive depiction. Paivio et. al (1966, as cited in Paivio, 1969) even found that learning scores were highest for noun pairs when a mental image could be evoked.

What if we were to move beyond word lists and consider a more realistic setting? Continuing the trend of maintaining a learning environment, this study investigates the practicality of adapting word lists into specialized lessons to be taught to subjects. Closely resembling the DRM paradigm, an instructor teaches using semantically related words hinting towards a lure that is never actually said. This experimental format makes use of the natural objective of a class to combine broad concepts with concrete definitions and visualizations in order to increase association amongst terms. More specifically, this study asks: "Can presenting word lists in a classroom lesson environment lead to a difference in false recall as opposed to only reading them aloud?" The previous literature of associative learning suggests that a similar pattern will arise in this study: the increased association and facility of mental imagery that comes with a lesson will lead to an increase in false recall.

Procedure

This study was conducted in three parts. Part one involved running the DRM paradigm and parts two and three involved a live lesson taught by a teacher. Both the test and the lessons were structured around word lists created from a teacher's lesson plan in math (lure-heavy lesson) and physics (test and lesson), as well as three additional neutral lists for the test.

PART 1: List-section

This part of the procedure was conducted using the DRM paradigm. Subjects in two 10th grade Science Research and Design (SRD) classes (n=17) were told they would hear lists of words and then be tested after every list by writing down the words. They were instructed to write the last few items first, a standard test procedure, and then to recall the rest in any order. Additionally, they were told to write down as many words as they could remember, but not to guess. An instructor then read out four word lists one by one: three neutral control lists of "sleep," "cold," and "window" from Roediger–McDermott (1995) (Lists 1, 2, and 3), and the physics-related list (Physics List) (Table 1). After every individual list, the instructor would say "recall," and the subjects then had about two minutes to recall each word on the list.

After the final list, subjects had a two minute break, then took a recognition test (Table 2). During the break, the instructor explained the difference between two pairs of concepts to understand for the test. The first pair was "old" and "new," defined as "the word appeared on the list," and "the word did not appear on the list," respectively. The second was "remember" and "know," defined as "you can mentally relive the experience of hearing the word," and "you cannot mentally relive the experience of hearing the word," respectively.

Table 1

Physics List	List 1	List 2	List 3
Rotational	Bed	Hot	Door
Radius	Rest	Snow	Glass
Angular	Awake	Warm	Pane
Degrees	Tired	Winter	Shade
Radians	Dream	lce	Ledge
Rotation	Wake	Wet	Sill
Period	Snooze	Frigid	House
Arc Length	Blanket	Chilly	Open
Angle	Doze	Heat	Curtain
Frequency	Slumber	Weather	Frame
Lap	Snore	Freeze	View
Motion	Nap	Air	Breeze

The Word Lists that the Instructor Read Out to Subjects During Part One of the Procedure.

The recognition test consisted of four sections, one for every list read aloud by the instructor. Each section contained seven words: two words that appeared on the list, two related words that did not appear, two unrelated words that did not appear, and the semantic lure. For every word, subjects had to indicate whether a word was "old" or "new;" if they indicated the former, they had to additionally make the distinction of whether they "remember" or "know" that they heard it. In total, there were three options to choose from for every word: "NEW," "OLD REMEMBER", and "OLD KNOW."

Table 2

The Words Used for the Recognition Test.

Word order for each section was randomized, but with the lure always at the end, following Roediger-McDermott (1995). For every word, subjects could answer "NEW," "OLD REMEMBER," or "OLD KNOW." Sections 1, 2, and 3 use words from Lists 1, 2, and 3, respectively. The recognition test was given as a Google Form.

Word Type	Section 1	Section 2	Section 3	Physics
Old 1	Bed	Hot	Door	Rotational
Old 2	Tired	Freeze	Closed	Radians
Related 1	Room	Drink	Sill	Orbit
Related 2	Mattress	Wind	Air	Cycle
Unrelated 1	Position	Coat	Triangle	Satellite
Unrelated 2	Mattress	Fight	Pie	Body
Lure	Sleep	Cold	Window	Revolution

PARTS 2 & 3: Class Sections – the Standard Lesson and Lure-Heavy Lesson.

For part two of the study, a physics teacher conducted a lesson to a group of students (n=14) as normal – the standard lesson. The lesson was structured loosely around a word-list lesson plan created beforehand, but there was an active attempt to avoid artificiality and any irregularities in pedagogy. Adhering to the list, the teacher made sure to mention every word aloud at some point during the lesson and avoid the semantic lure. A confederate was additionally present in the class to ensure the lure wasn't uttered.

After the lesson was completed, subjects then took a recognition test similar to the test from part one (Table 3). Here, there was only one section on the related subject list, but with eight words that were said aloud, eleven new words, and the semantic lure, so as to have twenty words total. Subjects were told to make the same distinctions between "old" and "new" if the teacher had uttered the word at any point, and between "remember" and "know." 00

Words on the Standard Lesson Recognition Test.

Word order was randomized, but with the lure at the end, following Roediger-McDermott (1995). For every word, subjects could answer "NEW," "OLD REMEMBER," or "OLD KNOW." The recognition test was given as a Google Form.

Word Type	Word	Word Type	Word	Word Type	Word
Old 1	Rotational	New 1	Flow	Lure	Revolution
Old 2	Radius	New 2	Paper		
Old 3	Angular	New 3	Exponent		
Old 4	Radians	New 4	Surface Area		
Old 5	Degrees	New 5	Frequency		
Old 6	Rotation	New 6	Heat		
Old 7	Arc Length	New 7	Period		
Old 8	Angle	New 8	Relativistic		
		New 9	Plant		
		New 10	Newtonian		
		New 11	Wave		

Part three of the study – the lure-heavy lesson – was instead a math lesson taught by a math teacher (n=11), but operated virtually the same. However, the word list was slightly modified; there were six old words, five lure words, and nine new words. Everything else was conducted exactly like the standard lesson.

The false memory recall rates, meaning the rate of answers marked either "OLD, REMEMBER," and "OLD, KNOW" were then compared between lesson and test groups. The answers from the recognition test were obtained and compiled into a Google Sheet. "P-lure" and "M-lure" variables were created and used for analysis, taken from the recall rates of the physics lure and math lures, respectively. A significance of the difference between the two proportions test was then run on OLD rates for all words on Vassarstats.net.

Words on the Lure-Heavy Lesson Recognition Test.

Word order was randomized, but with the lure at the end, following Roediger-McDermott (1995). For every word, subjects could answer "NEW," "OLD REMEMBER," or "OLD KNOW." The recognition test was given as a Google Form.

Word Type	Word	Word Type	Word	Word Type	Word
Old 1	Solution	New 1	Flow	Lure 1	Formula
Old 2	Equation	New 2	Water	Lure 2	Substitute
Old 3	Sine	New 3	Cycle	Lure 3	Functions
Old 4	Cosine	New 4	Surface Area	Lure 4	Isolate
Old 5	Trig	New 5	Heat	Lure 5	Factor
Old 6	Identity	New 6	Relativistic		
		New 7	Plant		
		New 8	Rhombus		
		New 9	Wave		

Data

For the list-section, a significance of the difference between two proportions was conducted on lure vs. old word recall rates and on lure vs. new word recall rates (data on Table 5). There was a significant difference between the lure and new word recall rates (p<0.0001), but no significant difference between old and lure word recall rates. This indicates that the procedure was conducted satisfactorily as the recall rates replicate those of previous literature, like Roediger-McDermott (1995).

Additionally, there was a significant difference between both P-lure vs. old word recall rates and P-lure vs. new word recall rates (p<0.0001 and p=0.0313, respectively). The former indicates that the lure was designed convincingly enough not to be completely confused with a new word, meaning that the paradigm could be conducted properly without doubt in the lure. However, the latter indicates that the lure was not as effective in inducing false memories as the words that were on the list. This makes sense, though, as the lure was created by the teacher few days before a specific class, instead of originating from carefully revised/peer reviewed lists that intend to optimize recall rates, like in Deese (1959). Overall, P-lure recall fell roughly in between new and old word recall, which seems to be a lower rate than Roediger-McDermott (1995); the original study yielded lure word recall rates that mirrored old word recall rates.

List-Section Responses.

Old, new, lure, and P-lure words divided by Old, Remember; Old, Know; and New. Denominators for the proportions are obtained from the number of students that responded and the number of questions for the word type.

	Response		
Old	Remember	92/136	0.68
	Know	30/136	0.22
	New	14/136	0.10
		·	<u>.</u>
New	Remember	13/272	0.05
	Know	34/272	0.13
	New	225/272	0.83
Lures	Remember	27/51	0.53
	Know	17/51	0.33
	New	7/51	0.14
P-Lure	Remember	0/17	0
	Know	6/17	0.35
	New	11/17	0.65

Lure word recall rates were compared with new and old word recall rates from the standard (physics) lesson, as seen in Table 6. Conducting a significance of the difference between two proportions, it was found that there was a significant difference for both P-lure vs. new word recall rates and P-lure vs. old word recall rates (p < 0.0001 for both). This matches the trend from the list-section, falling roughly in between old and new word recall rates.

Standard Class Responses.

Class-section old, new, and lure words divided by Old, Remember; Old, Know; and New. Denominators for the proportions are obtained from the number of students that responded and the number of questions for the word type.

	Response		
Old	Remember	98/111	0.88
	Know	11/111	0.1
	New	1/111	0.01
New	Remember	5/139	0.04
	Know	7/139	0.05
	New	127/139	0.91
P-Lure	Remember	2/14	0.14
	Know	4/14	0.29
	New	8/14	0.57

Comparing the P-lure recall rate between the list- and standard class-sections (Table 7), I found there was no significant difference between the two (p=0.3336). This data suggests that there is no difference in false recall rate when presenting word lists in a class as opposed to just reading them aloud.

Table 7

Standard Class-Section Responses.

Old recall rate proportions for all words.

Environ.	Word Type	Old R+K	Proportion
Class	Old	0.99	111/112
	New	0.09	12/140
	P-Lure	0.43	6/14
List	Old	0.92	94/102
	New	0.16	32/204
	Lures	0.86	44/51
	P-Lure	0.35	6/17

Finally, I compared old word recall rates between old, new, and M-lure words for the lure-heavy class-section (Tables 8 and 9), using the same test as before; there was a significant difference between old words and M-lure recall rates, and M-lure and new word recall rates (p<0.0001 for both). This once again matches the trend from before, where the class-related lure rate falls in between new and old word recall rates. However, M-lure word recall rates were still well above half for the given sample, revealing an interesting trend where half of students consistently recalled material that wasn't mentioned at any point during the class.

Table 8

Lure-Heavy Class-Section Responses.

Class-section old, new, and lure words divided by Old, Remember; Old, Know; and New. Denominators for the proportions are obtained from the number of students that responded and the number of questions for the word type.

	Response		
Old	Remember	47/66	0.71
	Know	15/66	0.23
	New	4/66	0.06
M-Lures	Remember	15/55	0.27
	Know	21/55	0.38
	New	19/55	0.35
P-Lure	Remember	4/99	0.04
	Know	6/99	0.06
	New	89/99	0.9

Table 9

Lure-Heavy Class Responses.

Old recall rates for all words.

Word Type	Old R+K	Proportion
Old	0.94	62/66
New	0.65	36/55
P-Lure	0.1	10/99

Conclusions

Because the evidence supports the null hypothesis (that there is no difference in recall rate between class- and list-sections), it has profound implications. The P-lure was recalled statistically equally as often throughout both sections, indicating that the sections' different procedures yielded comparable results. A Berkeley Carroll class was virtually just as effective in creating false memories as the DRM paradigm, a procedure literally designed to induce false memories. This phenomenon was especially visible with the lure-heavy class-section, where a considerable portion of the class (over half) had misremembered several words that were never mentioned. The claim is limited, however, due to the class' small size of 11 students, but the trend was present with the class-section as well. This implies that, though not more effective than the DRM procedure, classes can become spaces that easily foster the creation of false memories. As stated earlier, the classes were taught as regularly as possible; there was an active effort to avoid any irregularities in pedagogy. Therefore, students being taught normally in a class had, by the end of it, misremembered hearing a concept or term from not more than an hour before.

Discussion

Some of the "false recall" can potentially be explained by stress. Both the standard and lure-heavy class-section were conducted on advanced math and science classes taken by 11th and 12th graders, often sources of difficulty and stress to high school students. When it came to M-lure/P-lure recall rates, students generally tended to respond more that they "knew" the words had been said, but not so much that they "remembered" them. Students might have doubt in themselves paying attention and accurately remembering what was taught. Thus, if presented with a word reminiscent of what was taught in class, as the lures are designed to be, students may become more willing to accept the idea that the word had indeed been said at some point. In this case, it would not be as much of a false memory as it would be an effect of self-doubt, but both factors may certainly be cooperating here.

A large roadblock that came up with this study was the coordination of classes. It was difficult to reach out to teachers and find the ideal class to conduct the study with; not every class is taught the same. Teachers often have classes for quiz review, classes based upon worksheets, classes based upon group work, etc. Not every class is a lecture, the optimal format for the study (since it explicitly consists of the presentation of concepts). Of the classes that are lectures, not all of them present topics that can have sharp, effective lures. Therefore, my attempts to find the perfect class slowed down progress significantly and prevented me from actually carrying out the study for weeks.

Looking forward, I simply wish to follow up on this study with more classes. The data and data trends are promising, and increasing the sample size of the study can only contribute to our understanding of these classes and the false memories that accompany them. Additionally, the two classes tested here were strictly math and science; perhaps a humanities class might present something else and offer a new insight. Overall, I believe the procedure was sound, and all that remains is to conduct the study on more subjects.

References

- Gleaves, D. H., Smith, S. M., Butler, L. D. and Spiegel, D. (2004). False and Recovered Memories in the Laboratory and Clinic: A Review of Experimental and Clinical Evidence. *Clinical Psychology: Science and Practice*, 11: 3-28.
- Johnson, M. K. (2001). False memories, Psychology of. *International Encyclopedia of the Social & Behavioral Sciences*, 5254
- Kohonen, T. (2012). Self-organization and associative memory (Vol. 8). Springer Science & Business Media.
- "Memory." *Memory and Aging Center*, University of California San Francisco, 2019, https://memory.ucsf.edu/memory
- Okado, Y., & Stark, C. E. (2005). Neural activity during encoding predicts false memories created by misinformation. *Learning & Memory*, 12(1), 3-11.
- Paivio, A. (1969). Mental imagery in associative learning and memory. Psychological Review, 76 (3), 241.
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21* (4), 803.
- Wang, J. X., Rogers, L. M., Gross, E. Z., Ryals, A. J., Dokucu, M. E., Brandstatt, K. L., ... & Voss, J. L. (2014). Targeted enhancement of cortical-hippocampal brain networks and associative memory. *Science*, 345 (6200), 1054-1057.

Special thanks to:

Scott Rubin, Amanda Finigan, Dr. Mariani, and The Parallel Lions

About Science Research and Design (SRD)

In this highly 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 wide range of papers and articles as they learn how to dissect and understand scientific writing. By the end of their first year, they are be able to read, understand, and explain journal level articles on one or two fields of study.

Second Year: Advanced Science Research and Design – Students learn the nuts and bolts of scientific research and become well positioned to embark on their independent research projects. They read seminal papers in different fields of science; conduct experiments, analyzing the results with an appropriate statistical method; perform scaffolded studies of their own; and visit external labs. By the end of their second year, students have either established a strong working relationship with external mentors or or are well on their way towards conducting an in-house scientific study.

Third Year: Science Research and Design Symposium – Students finalize their individual research; write a journal article reflecting their results; and mentor and help facilitate the research of younger SRD students. They publish their paper in the *Proceedings of the Berkeley Carroll Independent Research Conference*, a peer-reviewed journal, and present their results at the Berkeley Carroll Science Research Conference held in the spring.



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