JOURNAL OF
EMERGING INVESTIGATORS

VOLUME 3, ISSUE 4 | APRIL 2020
emerginginvestigators.org

Root for the home team
How biases affect sports betting

Teen technology use
How social media usage affects teen's social lives

The right light for plants
How varying light conditions affect plant growth

Glass-transition temperatures
Machine learning to predict polymer properties

JOURNAL OF EMERGING INVESTIGATORS

The Journal of Emerging Investigators is an open-access journal that publishes original research in the biological and physical sciences that is written by middle and high school students. JEI provides students, under the guidance of a teacher or advisor, the opportunity to submit and gain feedback on original research and to publish their findings in a peer-reviewed scientific journal. Because grade-school students often lack access to formal research institutions, we expect that the work submitted by students may come from classroom-based projects, science fair projects, or other forms of mentor-supervised research.

JEI is a non-profit group run and operated by graduate students, postdoctoral fellows, and professors across the United States.

Contents

VOLUME 3, ISSUE 4 | APRIL 2020

JOURNAL OF
EMERGING INVESTIGATORS

Article

People's Preference to Bet on Home Teams Even When Losing is Likely

Matthew Weng[1], Russel Golman[2]
[1] Loomis Chaffee School, Windsor, Connecticut
[2] Carnegie Mellon University, Pittsburgh, Pennsylvania

SUMMARY

One intriguing phenomenon is when people make bets that seem to go against their better judgement. This can be seen in sports betting. In this paper, we report a survey-driven study that investigates if people bet more on their home teams, both in scenarios where the team is leading and scenarios where the team is likely to lose. We asked participants to imagine betting with $10,000 on different scenarios. We compared how much they bet on their home teams versus how much they bet on neutral teams in the same circumstance. On average, participants bet slightly more on their home teams than a neutral team when their home team was leading. Participants, however, bet significantly more on their home teams than the neutral teams when their team was facing a large deficit. This study can help explain some more impulsive betting behaviors that might be due to information avoidance.

INTRODUCTION

Sports betting has been well-studied in recent years. However, the impact of fan bias toward their favorite teams on betting decisions remains understudied. Consider, for example, that your favorite football team is down 14 with 5 minutes left. You are given the following bet: a 2-1 gamble if the team wins. Well it is against your better judgement, you might still choose to bet on your favorite team, and that desire is well documented in real-life scenarios. Most studies that outline these behaviors are most often big games because often no one will look at the betting numbers of a regular season comeback. During the Super Bowl LI, a significant number of gamblers bet for the Patriots when they were down 28-3 with the ball on their own 20-yard line and less than 20 minutes to play in the game (1). What makes this bet so shocking is that the odds of the Patriots winning were virtually zero.

There have already been several studies looking at trends in sports betting involving home teams receiving more bets. The home team receives a rather significant bias in betting if those betting are supporters of the home team (2). This phenomenon is mainly due to optimism bias as the ideal scenario the bettors imagine impacts the way they bet and causes a bias for their home team when the odds are even (3). Researchers have not yet studied the situation in which the gambler's team is facing a large deficit, such as the opponent having a 90% or higher chance of winning according to ESPN's chance of winning (an individual data for each game that changes as the competition goes on). This winning percentage chance is a data under every ESPN covered competition. It basically changes in favor of one of the teams as the game goes on and recalculates. (4) Can a gambler's team loyalty overshadow their logic and lead them to bet for the improbable comeback instead of the safe bet for their opponent?

Many studies have examined the reason behind the phenomenon of bettors favoring their home teams in betting behavior. One study hypothesized that risk aversion would drive fans to bet more on a situation that is positive to them when the odds are even than on a negative situation (5). Golman, Lowenstein and Gurney, in this 2017 study, in a between-subjects design, asked sports fans to bet on which of the two top hitters of a local baseball team would have more hits and to bet on which of the same two players would have more strikeouts (5). In the study, the mean bet in the hit condition was $2.30, while the mean bet in the strikeout condition was $1.16. This study found that participants were more willing to bet the batters would have more hits and less willing to bet that the batters would have more strikeouts.

Another study hypothesized that people avoid negative information (6). This study explained that people, when faced with negative information such as the diagnosis of a disease, would choose to avoid such information and try to turn it positive. This could potentially lead to fans betting more on their home team bet against and avoid predicting a negative outcome for the home team such as a loss.

In our study, we presented subjects with 20 scenarios of sporting events in which all the games featured one of the teams down a huge amount and a rather improbable comeback. Of the 20 scenarios, 5 included the specific home team of the city involved in the research and 15 featured teams that were neither liked or disliked to fans of the home team of the city. In each scenario, the survey-taker was asked to bet $10,000 on whichever team they wished. After collecting the data of how much each participant bet on each team in each scenario, we compared how much they favor their home teams versus other neutral teams in each circumstance. The study aimed to discover whether bettors will bet more on the home team when they are ahead than on a neutral team in the same situation. Additionally, we aimed to determine whether bettors tend to bet more on the home team when they are

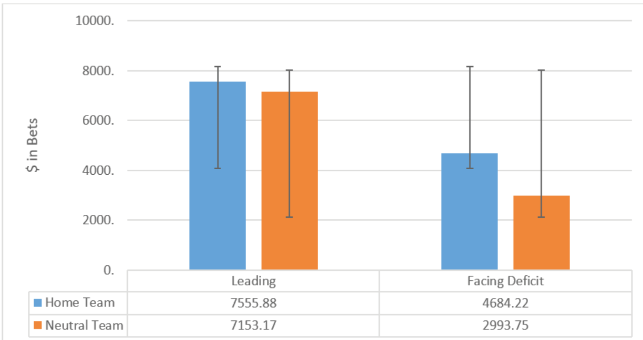


Figure 1: Bets from All participants. Participants (n=119 from 5 cities) were surveyed and average dollar amount that participants were willing to bet on home and neutral teams under leading and facing deficit circumstances was determined. Error bars represent the standard deviation which is the average difference between a value of the bet and the mean bet.

faced with drastic deficits than on other neutral teams in the same situation. In this study, we assessed whether fandom for a team can overshadow fans' judgment and lead those to bet significantly higher, i.e. take a risk that they otherwise would not, for their team.

RESULTS

We distributed a survey that contains several betting scenarios via Amazon Mechanical Turk to participants in cities whose teams were mentioned in the survey (New York, Boston, Chicago, Philadelphia, Los Angeles), and data were collected from those surveys.

Overall 140 results were collected (**Figure 1**) including 48 results from New York (**Figure 2**), 21 from Boston (**Figure 3**), 13 from Chicago (**Figure 4**), 16 from Philadelphia (**Figure 5**), 21 from Los Angeles (**Figure 6**), and 18 from other states. Though we did not record the gender, the median age group was 25-35, and participants claimed in the survey to have a median tendency of risk-taking of 3 on a scale of 1-5. The median annual income group was $50,000-100,000, and a majority (76%) of the participants watched sports at least once a week. The 18 responses that recorded participants that were not fans of the five cities' sports teams were disregarded as the data collected would not be appropriate.

To begin, the participants tended to bet more on their favorite teams than neutral teams in both scenarios (**Figure 1**). The average bet for a leading home team was $7,556 compared to $7153 for a leading neutral team (p=0.0035), a 5.6% advantage in favor of the home team. Participants were also more likely to vote for a trailing home team compared to a trailing neutral team. The participants' average bet was $4684 when betting for their favorite teams when they were down and $2855 when betting for a losing neutral team (p<0.0001). The bet for the home team increased by 64% compared to the neutral team. This reflected a strong trend of bettors electing to bet significantly more on their home teams when facing a serious deficit comparing to a neutral team.

Through a two-way ANOVA test, we determined that the factor of the score (whether the team was leading or behind) significantly impacted the bets (p<0.001) and the factor of the team (whether it was a home team or neutral team) significantly impacted the bets (p<0.001). In addition, the interaction of the two factors also significantly impacted the bets (p<0.001).

To prevent the numbers only reflecting the cities with larger fan-bases, such as New York which represented nearly half of the data collected, the data was separated by cities to reflect potential outliers or inaccuracies. This was also done to prevent a fan-base of a historically significantly more dominant franchise to vote much more in favor of their home teams than other teams.

According to our data, we ranked the amount of money bet on the participants' home teams when they were leading from the highest to lowest by cities in this order: Chicago, Boston, Los Angeles, New York, and Philadelphia. Furthermore, we ranked the amount of money bet on a neutral team when they were leading from the highest to lowest by cities in this order: Chicago, Boston, Los Angeles, New York, and Philadelphia.

The bet on the home team, when they were facing a significant deficit, as ranked from the highest to lowest in this order: Boston, New York, Chicago, Los Angeles, and Philadelphia. As for the amount bet on neutral teams when

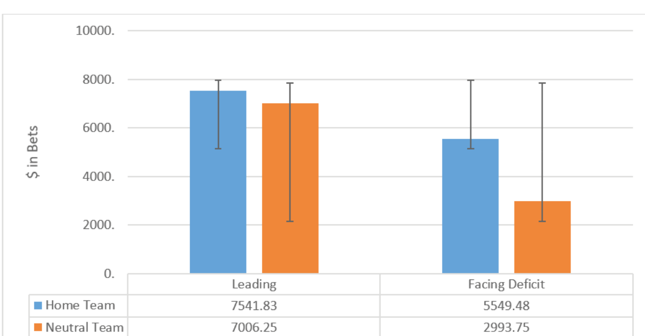


Figure 2: Bets from New York-based participants. Participants (n=48 from New York) were surveyed and average dollar amount that participants were willing to bet on home and neutral teams under leading and facing deficit circumstances was determined. Error bars represent the standard deviation which is the average difference between a value of the bet and the mean bet.

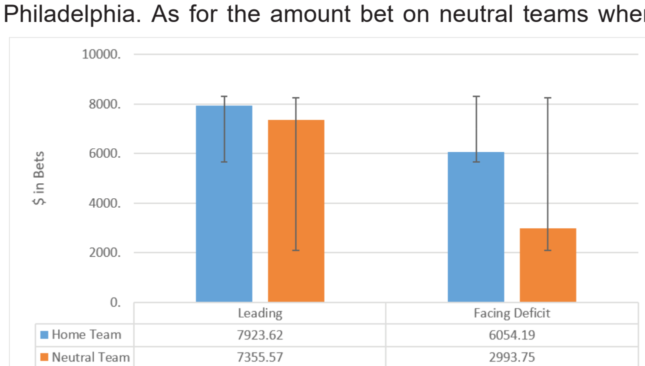


Figure 3: Bets from Boston-based participants. Participants (n=21 from Boston) were surveyed and average dollar amount that participants were willing to bet on home and neutral teams under leading and facing deficit circumstances was determined. Error bars represent the standard deviation which is the average difference between a value of the bet and the mean bet.

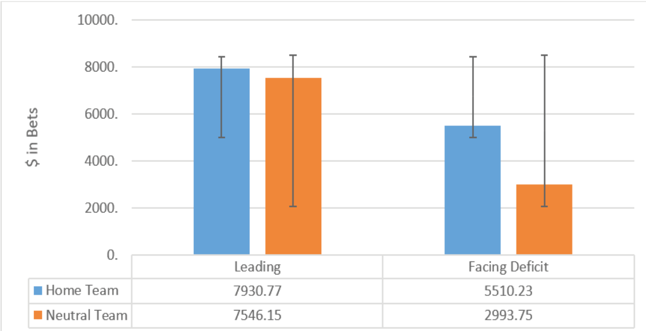


Figure 4: Bets from Chicago-based participants. Participants (n=13 from Chicago) were surveyed and average dollar amount that participants were willing to bet on home and neutral teams under leading and facing deficit circumstances was determined. Error bars represent the standard deviation which is the average difference between a value of the bet and the mean bet.

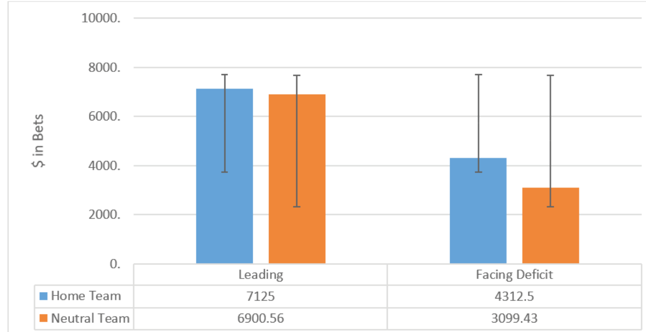


Figure 5: Bets from Philadelphia-based participants.Participants (n=16 from Philadelphia) were surveyed and average dollar amount that participants were willing to bet on home and neutral teams under leading and facing deficit circumstances was determined. Error bars represent the standard deviation which is the average difference between a value of the bet and the mean bet.

they were facing a deficit, the cities were ranked from the highest amount to lowest as follows, Philadelphia, New York, Los Angeles, Boston, and Chicago.

In the five cities, fans of Los Angeles sports teams had the highest difference between their bet on their home teams and neutral teams when leading (**Figure 6**), and fans of Philadelphia sports teams had the lowest difference between their bet on their home teams and neutral teams when leading (**Figure 5**). Fans of Boston bet the highest difference between the home teams and neutral teams when facing a deficit (**Figure 3**). Philadelphia fans bet the least difference between their home teams and neutral teams when they were facing a deficit (**Figure 5**). The standard deviation of all the bets was highest in New York (**Figure 2**) and lowest in Philadelphia (**Figure 5**).

The data demonstrate an overall trend that showed the bettors of all cities betting more on their home teams, both when facing a significant deficit or leading. Even though there was a difference in how much more each city bet on their home teams than neutral teams in each circumstance, they did overall tend to favor their home teams when facing a nearly impossible deficit. This demonstrated the existence of a bias favoring their home team that dictates the bettors' judgment, even when it was illogical to bet that way.

DISCUSSION

Here we present the results of a study that demonstrate the trend of fans betting more for the home team as opposed to a neutral team when the team was facing a significant deficit. This research specifically focused on how the bias fans have for their favorite teams might overshadow logical judgment when placing bets, sometimes betting for their favorite teams even during extremely unfavorable situations. In this study, participants bet in scenarios when a team was facing a significant deficit. We compared how they bet in those scenarios on their home teams compared to neutral teams.

Even though the results from the different cities varied slightly, we could see a general trend. In all the cities, the fans bet more on their favorite teams no matter if they were down or leading compared to neutral teams in the same circumstances. The fans bet especially higher on their favorite teams when they were down compared to when other neutral teams are down.

The participants in every city bet at least 50% higher on their home teams than the neutral teams when they were down, demonstrating a difference in betting preference. The large gap between the two values means that especially during high-risk circumstances when a team was facing a gigantic deficit, the fans' bias that favors their favorite team would lead them to disregard their better judgment to bet for the significantly fewer probable winners in the scenarios in the survey. Even though betting on the team that was down has a better potential return if they win, their victory was so improbable that even with these better returns, it was still the more reasonable choice to bet on the team that was leading. Yet, on average, the fans still bet nearly half of the $10,000 on their home team and even more in some circumstances (such as the fans of Boston who bet more on their home team when they were down than the team that is leading on average). These results support that the bias favoring their favorite team does overshadow logical judgment in extreme betting

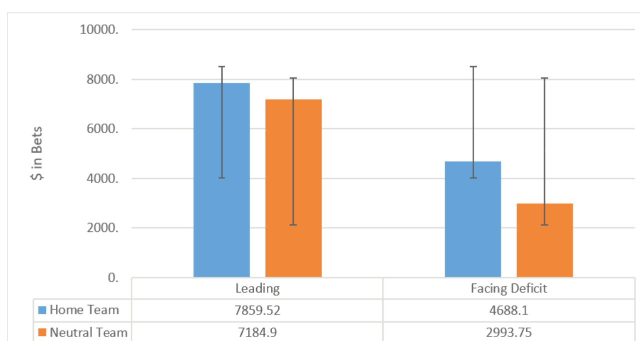


Figure 6: Bets from Los Angeles-based participants. Participants (n=21 from Los Angeles) were surveyed and average dollar amount that participants were willing to bet on home and neutral teams under leading and facing deficit circumstances was determined. Error bars represent the standard deviation which is the average difference between a value of the bet and the mean bet.

circumstances.

When viewing the different cities, we saw that the success of the cities' sporting franchise history could impact the fans confidence in their teams' ability of overcoming a deficit or preserving a lead. Those claiming to be fans of Boston sports teams appeared to be the most confident of their franchises' chances to come back from significant deficits, while having significantly more confidence in their team than a neutral team to comeback from a large deficit. This might be tied to the historical success of Boston franchises, as they collectively have the second highest amount of championship wins for a city and the most championships (twelve) since 2000 (7). Philadelphia bettors, however, appeared to be the least confident when their home teams were facing a significant deficit, which would be reasonable as they have only obtained two championships across all athletic franchises in the city since 2000 (7).

The results of the survey could be explained by several potential theories for why the participants bet much more on their favorite team even when the team was down significantly. Firstly, it was likely that, due to the information avoidance of the participants, they do not wish to accept and receive the potential information of their favorite team losing with a bet on the opposing team. Thus, they would bet on their favorite team to avoid predicting such potential results. Secondly, it was also possible that the bias participants have for their home team was strong enough to remain even in unfavorable circumstances. Lastly, risk aversion could have impacted the results as bettors could have been avoiding the bet that shows potentially negative results in their team losing.

There were several potential sources of errors in this survey. Firstly, the survey takers might not have reflected completely what they would do in a real-life betting circumstance in the survey as the survey involves hypothetical scenarios. When treating a hypothetical circumstance in a survey, it was difficult for the participants to express exactly what they would do in an actual betting circumstance as the pressure of actual loss and gain of property could influence the judgment of the bettors. One potential way to alleviate that error is giving a participant an actual monetary bonus when they win their bets. Also, as $10,000 is a large amount of money, a smaller amount of hypothetical money might be more realistic for the participants. In addition, the survey had a limitation in that it only focused on the fans of five major cities and though this reflects a large amount of people, the survey might not be accurate for the entire country. Even though the five cities represented in the research were cities with major athletic markets and a huge amount of professional sports audience, it was still only five cities out of the hundreds in the United States. Thus, there were bound to be possibilities, as well as decisions and opinions not represented and taken into consideration in this research.

Through the results from the study, it was evident that bettors do have the tendency to bet more on their home teams than neutral teams when facing a large deficit. This trend could contribute to bringing several new pieces of information such as a potential area of marketing for betting companies. In addition, the hypothesis proposed in this study could advance theories in related areas such as the idea of information acceptance bias. Those theories could be applied to many other scenarios that involves similar biases due to information avoidance. Those scenarios, related to sports betting or not, could bring to light multiple potential resolution or predictions. For example, this bias could have implications with marketing and other decision sciences.

METHODS

This study aimed to determine if the bias amongst sports bettors on their home team was strong enough to make them bet irrationally. Thus, a simulation of betting was determined to be the most reliable method for the research. A survey that contained several betting scenarios was sent out via Amazon Mechanical Turk to participants in cities whose teams were mentioned in the survey, and data was collected from those surveys.

We designed the survey so that as much bias involving favoring their home or favorite team in a betting scenario as possible was taken into consideration. When the participants began the survey, they were given a variety of scenarios in the four major American sports: hockey, basketball, football, and baseball. In each category, there were four to six scenarios. Each scenario portrayed a game in which one specific team was faced with a rather large deficit. That deficit would be difficult to overcome given the little time left in the game displayed in the scenario. In each of the scenarios, we prompted the participants to imagine that they were given $10,000 to bet on both or either teams in any combination they wish. To make it at least somewhat favorable to bet on the team that is down, the gamble for a team that was down was 2 to 1. Thus, if they bet $1 on team A, the team facing the deficit, and win, they earn $2. We asked the participants to make betting decisions in each of the scenarios with the $10,000. The participant would make decisions regarding how much to bet on either team if they choose to do so instead of betting everything on one team. After they bet, we could find a difference between the bet on the team that was down versus the team that was ahead. For example, if the participant bet $9,000 on the team that was up and $1,000 on the team that was down, the value would be -8000 (1000-9000 = -8000). We could compare this value to the amount the participant bets in the scenarios with neutral teams. This value could lead to the discovery of whether the bias for their home team overshadows the better judgment of the bettor even when the team they favor was faced with a large deficit. The amount of money the participants bet on the neutral teams would demonstrate how the participant would normally treat such a betting circumstance without any potential bias. To ensure that those neutral scenarios would minimize bias, a scenario involving a team's historical rival(s) would be disregarded.

During the survey, there were multiple ways to ensure

Table 1: Cities and teams involved in the survey.

Home teams	Their respective rival teams	Their respective neutral teams
Boston	New York, Los Angeles	Chicago
New York	Boston, Philadelphia	Chicago, Los Angeles
Chicago	Philadelphia	New York, Los Angeles
Philadelphia	New York, Chicago, Boston	Los Angeles
Los Angeles	Boston	Chicago, Philadelphia

that we take into consideration the home team or favorite team of the participant. Firstly, in the 20 scenarios in the survey, the questions appealed to multiple cities. It targeted Chicago, New York, Philadelphia, Boston, and, Los Angeles. In those 20 scenarios, the neutral scenarios from the perspective of New York fans might be the scenarios that feature other home teams of other cities, ensuring the appeal of the survey in a wide range of cities. In addition, the five cities were chosen to have the most popular teams and strong fan bases. In these cities, even if one participant was only a fan of one team in Philadelphia, it was more likely for that participant to have at least one of his other favorite teams in other sports to appear in other fore-mentioned cities, ensuring the accuracy of the results. At the end of the survey, a question also required participants to list their favorite teams in each of the four major professional sports to ensure that we treated their data in each scenario appropriately. In addition to asking the participants their favorite teams, there were also demographic questions. The first demographic question asked for the annual income of the participants. The annual income of the participants would be able to inform the researchers how much the $10,000 bet meant to the participants as those who were affluent might not care for the $10,000 as much as a less affluent participant. Secondly, the demographic question asked for how comfortable the participant was to take risks which could provide information about how the participant bet and whether the bet was normal or abnormal according to the value. In addition to the previous questions, the demographic questions also inquired how often the participant watched sports and whether the participant had bet on sports before.

We also utilized a two-way ANOVA later in the study to analyze the results. The two factors were whether the team was a home team or neutral team and whether the team was facing the large deficit or leading by that large amount. We created three null hypotheses:

1. The factor of the score (whether the team was leading or behind) does not significantly impacts the bets.
2. The factor of the team (whether it was a home team or neutral team) does not significantly impacts the bets.
3. Score and Team interaction do not have a significant impact on bets.

The score is a significant term in the ANOVA analysis with a *p*-value of less than 0.001, rejecting the first null hypothesis, the team is also a significant term with a *p*-value of less than 0.001, affirming the second null hypothesis. Lastly, the team-score interaction is also a significant term with a *p*-value of less than 0.001 rejecting the third null hypothesis. It would be reasonable to conclude that both factors, whether a participant is betting on a favorite (home) team and a neutral team as well as whether the team is facing a large deficit or leading by a large amount, significantly impact the amount of the bet placed.

Received: March 4, 2019
Accepted: February 14, 2020
Published: March 10, 2020

REFERENCES

1. Brinson, Will. “Super Bowl LI: One Vegas Sportsbook Got Absolutely Destroyed by the Patriots' Victory.” *CBS Sports*, February 6, 2017. https://www.cbssports.com/nfl/news/super-bowl-li-las-vegas-gets-absolutely-destroyed-by-the-patriots-victory/ Accessed 10 Sept. 2019.
2. Stanek, Rostislav. "Home bias in sport betting: evidence from Czech betting market." *Judgment and Decision Making*, vol. 12, no. 2, 2017, pp. 168+. *Gale Academic Onefile*, Accessed 24 Dec. 2018.
3. Krizan, Z., & Windschitl, P. D. “The Influence of Outcome Desirability on Optimism.” *Psychological Bulletin*, vol. 133, no. 1, 2017, pp. 95-121., doi:10.1037/0033-2909.133.1.95. Accessed 10 Dec. 2018.
4. Kerr-Dinean, “Are 'Win Probabilities' useless? ESPN's Director of Sports Analytics explains why they're not.” *USA Today*, 22. Feb. 2017. https://ftw.usatoday.com/2017/02/super-bowl-espn-win-probability-atlanta-falcons-new-england-patriots-stats-tom-brady Accessed 6 Feb. 2020
5. Golman, Russell, et al. “Information Gaps for Risk and Ambiguity.” *SSRN Electronic Journal*, 24 Mar. 2015, doi:10.2139/ssrn.2605495. Accessed 10 Dec. 2018.
6. Golman, Russell, Hagmann, David, Loewenstein, George. "*Information Avoidance*." Journal of Economic Literature, 55 (1): pp. 96-135. March, 2017 *Carnegie Mellon University*, Accessed 10 Dec. 2018.
7. Garvin, Patrick. “Has Your City Won as Many Championships as Boston?” *BostonGlobe.com*, February 4, 2019, apps.bostonglobe.com/sports/graphics/2018/10/boston-sports-history/. Accessed 10 Sept. 2019.
8. Chip, and Amos Tversky. “Preference and Belief: Ambiguity and Competence in Choice under Uncertainty.” *Journal of Risk and Uncertainty*, vol. 4, no. 1, 1991, pp. 5–28., doi:10.1007/bf00057884. Accessed 10 Sept. 2019.
9. Morewedge, Carey K., et al. “Betting Your Favorite to Win: Costly Reluctance to Hedge Desired Outcomes.” *Management Science*, vol. 64, no. 3, 2018, pp. 997–1014., doi:10.1287/mnsc.2016.2656. Accessed 10 Sept. 2019.

JOURNAL OF
EMERGING INVESTIGATORS

Article

Assessing and Improving Machine Learning Model Predictions of Polymer Glass Transition Temperatures

Manav Ramprasad[1], Chiho Kim[2]

[1] Joseph Wheeler High School, Marietta, Georgia

[2] School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia

SUMMARY

The success of the Materials Genome Initiative has led to opportunities for data-driven approaches for materials discovery. The recent development of Polymer Genome (PG), which is a machine learning (ML) based data-driven informatics platform for polymer property prediction, has significantly increased the efficiency of polymer design. Nevertheless, continuous expansion of the 'training data' is necessary to improve the robustness, versatility, and accuracy of the ML predictions. Accurate prediction of polymer properties, such as glass transition temperature (T_g), is advantageous for the design of polymers, particularly for high temperature applications. We hypothesized that by adding more data with increased chemical diversity to the dataset, the predictive capabilities of the PG model would improve. In order to test the performance and transferability of the predictive model for T_g (previously trained on a dataset of 450 polymers), we have carefully collected additional experimental T_g data for 871 polymers from multiple data sources. The T_g values predicted by the present PG models for the polymers in the newly collected dataset were compared directly with the experimental T_g to estimate the accuracy of the present model. Using the full dataset of 1321 polymers, a new ML model for T_g was built following past work. The root mean square error (RMSE) of prediction for the extended dataset, when compared to the earlier one, decreased to 27 K from 57 K, thereby supporting our initial hypothesis that increasing the dataset would improve the predictions. To further improve the performance of the T_g prediction model, we are continuing to accumulate new data and exploring new ML approaches.

INTRODUCTION

A polymer is a large molecular system composed of a chemical repeating unit (monomer). Polymers, displaying a dizzying diversity of physical and chemical properties, constitute an important and ubiquitous class of materials (1). Although they are made up of atomic species found from the periodic table, such as carbon, hydrogen, and oxygen, their limited chemical palette still leads to a rich and diverse spectrum of distinct polymers with a broad range of properties. Thus, it is highly non-trivial to find a suitable polymer for a particular application with desired properties in the practically infinite chemical space. As a result, selection of polymers has hitherto proceeded largely by intuition and trial-and-error efforts, which generally tend to advance the materials discovery landscape in a painstakingly slow manner.

In 2011, the White House unveiled the Materials Genome Initiative (MGI) to accelerate the discovery, manufacture, and deployment of advanced materials to a speed twice as fast as in the past, but at a fraction of the cost (2). One of the central pillars of the MGI is the use of data-driven approaches, such as machine learning (ML), to speed up materials discovery, including in polymer science and engineering. Data-driven ML approaches are complementary to traditional approaches, such as trial-and-error methods involving serendipity, used in materials science and engineering (3). ML approaches utilize prior data, information, and knowledge in an effective and efficient manner, as has been demonstrated in many other domains in the past. Classic examples of ML approaches include facial, fingerprint, or object recognition systems; machines that can play sophisticated games such as chess, Go, or poker; and automation systems such as in robotics or self-driving cars (3, 4).

Within the domain of materials science and engineering, the synthesis and testing process in the laboratory tends to be expensive and time-consuming, especially when handling the polymeric system. In order to utilize the data-driven framework, a dataset of several similar materials and their properties must be first collected. This data constitutes "prior knowledge" on this situation, i.e., the data is obtained from previously performed dedicated experiments or from the literature. Each of the materials in the dataset is then converted to a unique numerical representation, typically referred to as the "fingerprint." Finally, a mapping is established between the fingerprint and its properties using ML algorithms such as Gaussian process regression (GPR), thus leading to a predictive surrogate model (5). Subsequently, this model can be used to make instantaneous predictions of the properties of a new material, by simply following the fingerprinting and mapping procedures. The essential elements of this workflow are portrayed in **Figure 1**.

The efficacy of this method has been recently demonstrated as part of the "Polymer Genome" (PG) Project (6). In order to improve upon the predictive capabilities of the ML models implemented, increased data collection is extremely important. The present work deals with testing the capability

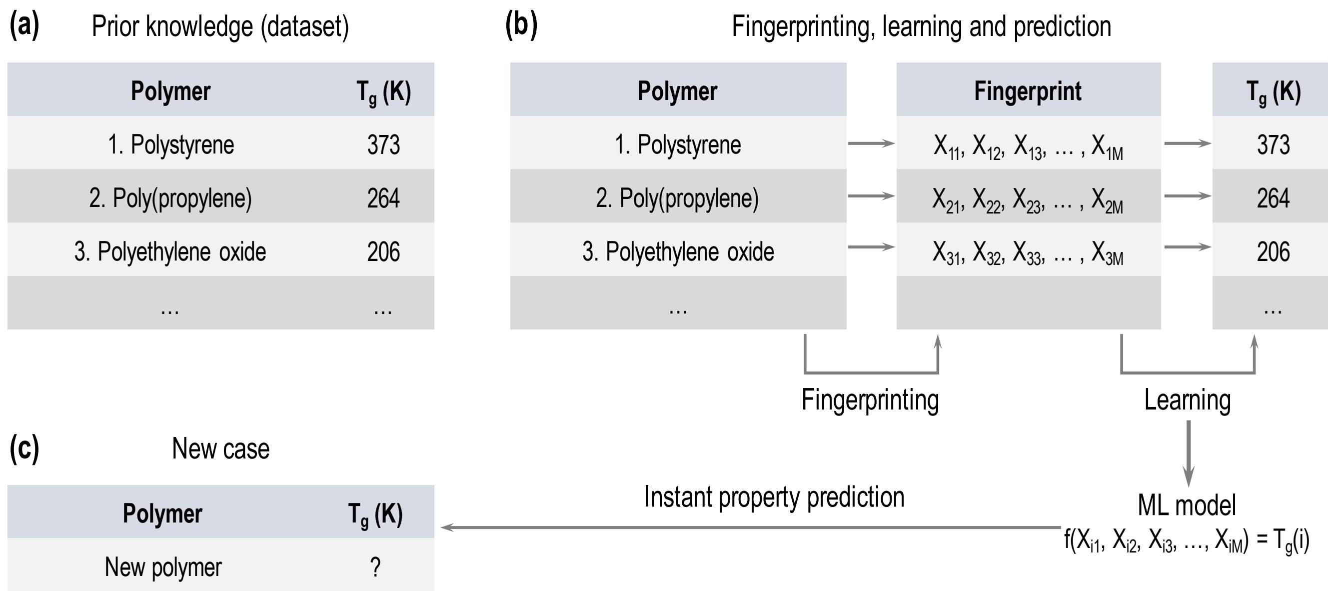


Figure 1: The key elements of machine learning in materials science. (a) Schematic view of an example data set. (b) Creation of a prediction model via the fingerprinting and learning steps. (c) Statement of the problem "What is the T_g of new polymer?"

of PG on new polymers, then using the results of this test to improve the predictive models. The property chosen for this test was the glass transition temperature (T_g) the temperature above which a polymer transitions from brittle and glass-like to viscous and rubber-like. T_g is an important property for many applications, as it determines the temperature ranges at which it is safe to use a polymer. Previously, the model hosted by PG was trained on 450 polymers. Current work demonstrates how expansion of the dataset affects the performance of the ML model. Therefore, we believe that by significantly increasing the size of the dataset, we can improve the predictive capabilities of the PG model in its performance and transferability. We have collected additional experimental T_g data for 871 polymers. The predictions of PG for these new polymers were compared directly with the collected T_g data, and conclusions have been drawn regarding the deficiencies of PG. The original training set was then augmented with this new data, and retraining was performed, ultimately leading to an improvement in the predictive capability of PG.

RESULTS

We refer to the earlier version of PG, which was trained on 450 T_g values, as PG-0. The newer version of PG in which the new T_g data for 871 additional polymers has been incorporated, is referred to as PG-1 (details of data distribution and example polymers in the dataset are shown in the section Methods). Since PG-0 was trained on the original 450 data points, the predictions for those 450 points are fairly accurate. The prediction for the new polymers, on the other hand, is inaccurate, and uncertainty of the prediction is higher. **Figure 2a** shows a parity plot of the performance of PG-0 on both the new dataset of 871 polymers and the initial 450. While many polymers fall closer to the parity line, indicating good agreement between predicted and actual values (values found from the literature), predictions for a certain portion of new polymers are off the parity line.

The poor predictive capabilities for those polymers in the range 300 K - 500 K is mainly due to the difference in fingerprint for the new data points compared to the benchmark (original) data points. In the case of very high T_g values, the PG-0 model performs poorly due to a lack of benchmark data points in the high T_g region (see also: **Figure 3**, showing the distribution of T_g values found in the original and new datasets). In all cases for which the predictions are poor, the uncertainty of the predictions, which is depicted by error bar around data points (**Figure 2**), is relatively higher than those for the original 450 polymers. High uncertainty for a particular case indicates that the polymer is 'not very similar' to the 450 training set polymers of PG-0. Had the scope of the training set been larger, more polymers would have been considered 'more similar' to the training set polymers and would have had more accurate predictions. Overall, the performance in terms of the root mean square error (RMSE) for PG-0 is greater than 50 K for the set of new 871 polymers. This RMSE is higher than desired for T_g predictions (below 30 K). Additionally, of the 871 polymers, 43% have a difference of at least 30 K between the experimental and predicted T_g. This observation indicates that more data points are necessary to improve the predictive performance of the ML model.

Next, we used the 871 new polymers and their corresponding T_g values to augment the original T_g dataset used for PG-0, then retrained to create a new PG-1 GPR predictive model for T_g (**Figure 2b**). As can be seen, a remarkable improvement in predictions emerges. The RMSE in this case is well below 30 K, which is acceptable, as the uncertainties in the actual measurement of T_g is in the

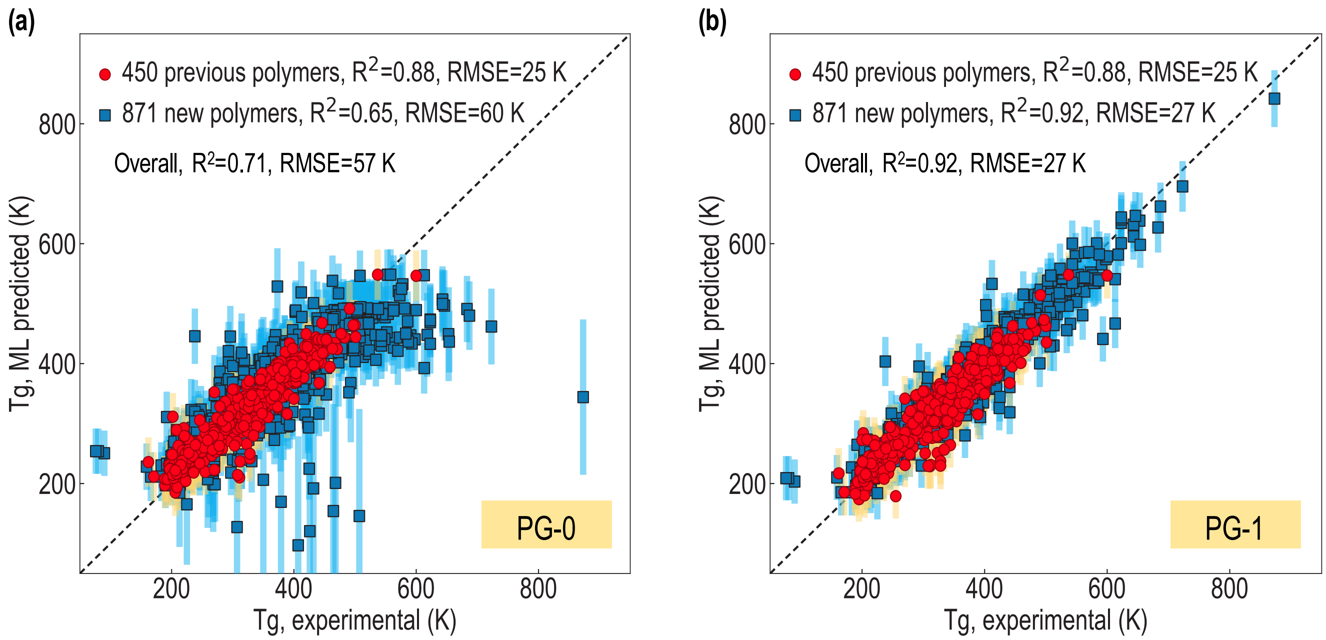


Figure 2: Performance of ML prediction model. Comparison of models trained on (a) 450 previous polymers and (b) 1321 polymers, including 871 new polymers. Error bar represents GPR uncertainty (confidence of prediction).

same range. The uncertainties calculated by GPR, shown by the error bars, have also decreased significantly, again showing an improvement in prediction capabilities. Relative to the original dataset, the new dataset has specifically and purposefully added polymers in new chemical spaces, and has added polymers with high T_g values, i.e., in the 500-700 K range. These aspects have led to a significantly better predictive capability of PG. Further progress can be achieved by systematically adding more diverse data.

DISCUSSION

Although efficient, ML models are accurate and reliable only within the domain of the dataset on which the model was trained. Predictions made for cases that fall outside the domain of the training data (i.e., the dataset originally used to create the models) are not expected to be reliable. In such cases, the new data points that fall outside the original domain of applicability have to be necessarily included in a retraining process to make the predictive model more versatile and transferable.

In summary, to improve upon an existing ML model to predict polymer T_g, a comprehensive dataset of polymer T_g was collected. Machine learning predictions for these new polymers revealed the deficiencies of the previous model. In retraining the machine learning model on the new data, the performance of the predictions dramatically improved, supporting our initial hypothesis that dataset expansion can significantly improve prediction. This work has thus led to a T_g prediction model that has been exposed to a more diverse dataset than before and is hence more versatile. The new model reduced the RMSE for not only new polymers, but also the polymers from the original dataset. The new prediction model presented for T_g, as well as the other polymer properties listed above, is available for free at the PG online platform (6).

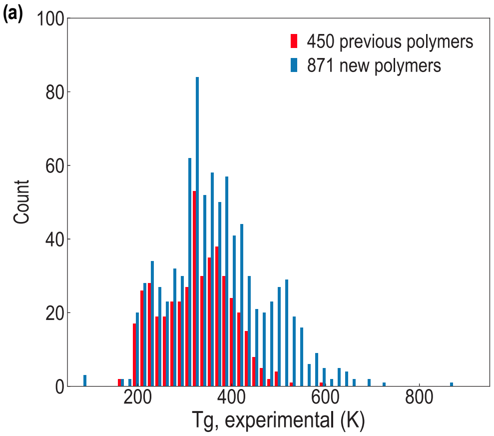
Looking into the future, it would be useful if the prediction pipeline could be inverted, such that polymers could be recommended to meet a specific set of property objectives, such as T_g between 600 K and 650 K. A variety of artificial intelligence-based algorithms (7, 8) may be utilized for such purposes. Solving this inverse problem effectively will significantly accelerate polymer discovery, as inputting requirements for particular polymer properties would result in the suggestion of possible polymers to meet the need.

Besides T_g, many other properties of polymers are important as well. In addition to the T_g prediction, PG also offers predictions of other properties, including 1) electrical properties like bandgap, ionization energy, and electron affinity, 2) dielectric and optical properties such as the dielectric constant and the refractive index, 3) physical and thermodynamic properties like density and atomization energy, 4) solubility properties like Hildebrand solubility parameter and a list of solvents and non-solvents, 5) mechanical properties like tensile strength and Young's modulus, and 6) permeability properties like gas (He, H_2, CO_2, N_2, O_2, and CH_4) permeability. Each of these predictive models within PG could potentially go through an improvement due to new data infusion.

MATERIALS AND METHODS

Data for this work were obtained from publicly available collections of experimental measurements (9, 10) and an online repository of polymer properties (11). The new polymer dataset is highly diverse, and its constituent polymers are composed of nine atomic species: carbon, hydrogen, oxygen,

(a)



(b)

Polymer name	SMILES	$T_{g,experimental}$ (K)	$T_{g,PG-0}$ (K)	$T_{g,PG-1}$ (K)
Polystyrene	CC(C1=CC=CC=C1)	373	368	353
Poly(propylene)	CC(C)CC(C)CC(C)	264	280	256
Poly(caprolactone)	CCCCCOC(=O)CCCCCOC(=O)	208	229	217
Poly(ethylene adipate)	CCCCC(=O)OCCOC(=O)	230	243	235
Poly(ethylene glycol)	CCOCCO	206	185	189
Polyglycolide	CC(=O)OCC(=O)O	312	308	312
Polyacetal	COCO	211	203	204
Poly(phenylene sulfide)	C(C=C1)=CC=C1SC(C=C2)=CC=C2S	391	378	372
Poly(p-xylene)	CC(C=C1)=CC=C1C	343	357	349
...	...	...	...	...

Figure 3: (a) Distribution of the T_g values for all polymers considered in this work. (b) Sample polymer dataset with SMILES representations and experimental T_g values.

nitrogen, sulfur, fluorine, chlorine, bromine and iodine. The T_g of the 1321 polymers (450 polymers from previous work and 871 newly collected polymers) in the dataset varied widely, ranging from 76 K to 873 K with a mean of 354 K (**Figure 3a**). The repeat units of the polymers were represented using the simplified molecular-input line-entry system (SMILES) (12). Examples of SMILES representations are shown in **Figure 3b** with the original name of polymers and T_g.

In order to capture the key features that may control T_g, we utilized the hierarchical polymer fingerprinting scheme (13). The fingerprint building process involves assessing three hierarchical levels of features. The first is at the atomic scale, wherein atomic fragments occur. This set of descriptors captures the type of atoms and atomic connectivity in the polymers. For our 1321 polymers, there are 128 such components. The next level deals with quantitative structure property relationship descriptors (14), such as the estimated surface area of the polymer repeating unit and fraction of rotatable bonds. Such descriptors, 39 in total, form the next set of components of our overall fingerprint. The third level descriptors captured morphological features, such as the topological distance between aromatic rings and the length of sidechains. We include 22 morphological features in the fingerprint.

The ML model was built by mapping the descriptors to the T_g values using GPR with a sum-kernel consisting of a radial basis function kernel and a white-noise kernel. During the model development step, we used an out-of-sample testing scheme to validate the ML model. We randomly partitioned the T_g dataset so that 80% was used for the model training and 20% was used for the validation of the trained model (5-fold cross-validation). Among five models trained on different random choices of training set, the single model delivering the most accurate predictions for the test set was selected as the best cross-validation model. Using the fixed kernel with the hyper-parameters extracted from this model, we obtained the final model which was retrained on our entire dataset. In the prediction step, data points with fingerprints very close to the new fingerprint value are weighted more than data points with fingerprints farther away. This means that if the new polymer is similar in terms of fingerprint to some polymers already in the data set, GPR will give a T_g value close to that of those similar polymers. Details of the approach used may be found in previously published work (13).

Received: August 3, 2019
Accepted: March 4, 2020
Published: March 18, 2019

REFERENCES

1. Krevelen, Dirk Willem van. *Properties of Polymers: Their Correlation with Chemical Structure*. Elsevier, 1997.
2. “Materials Genome Initiative.” www.mgi.gov
3. Ramprasad, R., Batra, R., Pilania, G., Mannodi-Kanakkithodi, A., and Kim, C. “Machine Learning in Materials Informatics: Recent Applications and Prospects.” *npj Computational Materials*, vol. 3, 2017, pp. 54.
4. Jordan, M. I. and Mitchell, T. M. “Machine Learning: Trends, Perspectives, and Prospects.” *Science*, vol. 349, no. 6245, 2015, pp. 255-260.
5. Williams, C. K. I., and Rasmussen, C. E. “Gaussian Processes for Regression.” *Advances in Neural Information Processing Systems 8*. MIT Press: Cambridge, 1996, pp. 514-520.
6. “Polymer Genome: An informatics platform for polymer property prediction and design using machine learning” www.polymergenome.org
7. Shafkat, I. “Intuitively Understanding Variational Autoencoders.” *Towards Data Science*, https://towardsdatascience.com/intuitively-understanding-variational-autoencoders-1bfe67eb5daf.
8. Sadeghi, J., Sadeghi, S., and Niaki, Seyed Taghi A. "Optimizing a hybrid vendor-managed inventory and transportation problem with fuzzy demand: An improved particle swarm optimization algorithm." *Information*

Sciences, vol. 272, no. C, 2014, pp. 126-144.
9. Brandup, J., Immergut, E.H., and Grulke, E.A. *Polymer Handbook*. 4th ed., 1999, John Wiley and Sons, New York.
10. Askadskii, A. A. *Computational Materials Science of Polymers*. 2003, Cambridge: Cambridge Int.
11. "Chemical Retrieval on the Web (CROW)" http://polymerdatabase.com
12. Weininger, D. "SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules." *Journal of Chemical Information and Modeling*, vol. 28, no. 1, 1988, pp. 31-36.
13. Kim, C., Chandrasekaran, A., Huan, T.D., Das, D., and Ramprasad, R. "Polymer Genome: A Data-powered Polymer Informatics Platform for Property Predictions." *Journal of Physical Chemistry C*, vol. 122, no. 31, 2018, pp. 17575-17585.
14. Nantasenamat, C., Isarankura-Na-Ayudhya, C., Naenna, T., and Prachayasittikul, V. A. "Practical Overview of Quantitative Structure-Activity Relationship." *EXCLI Journal*, vol. 8, 2009, pp. 74-88.

JOURNAL OF
EMERGING INVESTIGATORS

Article

The impacts of varying types of light on the growth of five *Arabidopsis* varieties

Brendt Wang, Sarah Connors, Katherine Kaminski
Acton-Boxborough Regional High School, Acton, Massachusetts

SUMMARY

***Arabidopsis* is known as the "fruit fly of plants." It is small and easy to grow, has a short life cycle, and has a small, easy-to-manipulate genome. Using Arabidopsis, we tested the effects of varied light conditions on the plant growth of mutants with dysfunctional light pathways. We tested five different strains: wild type, a phytochrome A mutant (phyA), a phytochrome B mutant (phyB), a phyA/phyB double mutant, and a DET1-1 mutant. With these mutants, we investigated how varied wavelengths and exposure of light affect the growth of the mutants. We found that the phyA mutant, the phyB mutant, and the double mutant all grew well in red light, with high germination rates and the largest average plant size. The phyB mutant grew the best under blue light, with the highest germination rate and the second largest average plant size. Under natural light, every strain grew relatively well, with high germination rates and consistent sizes. Although the DET1-1 mutant had a lower average size compared to the phyA and phyB mutations, it had the highest germination percentage, making it the most successful under no-light conditions.**

INTRODUCTION

Just like animals, plants have developed many mechanisms to respond to the surrounding environment. They have systems of receptors that receive stimuli and activate pathways to create responses. One of these systems is the phytochrome system. Plants rely on photoreceptors to mediate responses to light stimuli. Phytochromes are types of photoreceptors that are sensitive to red light, a main component of natural light, and far-red light, a lower wavelength of red light often produced after light is filtered through the leaves of other plants. Red light activates phytochromes while far-red light (FR) de-activates them. When phytochromes are active, growth is induced, and when inactive, growth is slowed (1). Furthermore, activated phytochromes trigger germination; thus, plants only germinate when exposed to red light. Research has already been done on the effects of phytochrome A and B mutations on *Arabidopsis*. One study found that phyA had a germination defect in FR while the phyB mutant had a germination defect in the dark; however, they also found that the effects of the phyA mutation could be suppressed by the phyB mutation (2). Both the phyB and the phyA mutants grew well under red light but were underdeveloped under FR. In contrast, the phyA phyB double mutant showed underdevelopment in red light but better development in FR, especially in the cotyledons (embryonic leaves) (3). Another study has also found the *det1* gene in *Arabidopsis* to allow the plants to grow like a light-grown plant in the absence of light (4). We wanted to see how these mutants would behave under more varied light conditions, not just under the types of light for which they have mutations. Thus, we tested these mutants not only under red light and natural light, but also under no light and blue light to see how different wavelengths would affect development in plants with nonfunctional phytochromes.

Phytochrome A and B have overlapping but different functions. PhyA is much more sensitive to FR and is responsible for germination (initial growth from seed) and de-etiolation (the greening of plants through the development of chloroplasts). Under the shade of other plants, light is often filtered of red and blue light leaving FR which is lower on the spectrum. FR triggers a light pathway through phyA stimulating germination and de-etiolation, which is an important stage towards plant maturation. PhyA also inhibits responses for avoiding shade, like the elongation of the hypocotyl (stem). When exposed to high levels of red light, phyA degrades. Under shade, plants will accelerate their growth in attempt to outcompete competitors, but excessive elongation growth can inhibit plants from establishing maturity, causing abortion. With a high FR to red light ratio, phyA inhibits elongation while promoting germination and de-etiolation. The null phyA mutation would lead to unchecked etiolation, which includes long hypocotyls and undeveloped leaves (5). PhyB also regulates de-etiolation but in a different manner. Under red light, phyB is activated, suppressing shade avoidance responses. When red light is reduced, the phyB becomes inactive which in turn stimulates shade avoidance. Thus, the phyB mutation would prevent the plant from suppressing excessive elongation (6). The DET1-1 mutant is slightly different because the *det1* gene acts as a transcriptional repressor for genes that are expressed by light stimulus transduction pathways. A nonfunctional *det1* gene would prevent the mediation of plant development in response to light, causing the plant to grow regardless of stimuli— including in the dark (4).

Based on previous research, we hypothesized that the DET1-1 mutant would grow best in the dark since it would grow regardless of stimuli from the lack of light. We also hypothesized that the phyA and phyB mutations would grow best in the red light because the pathway suppressing shade

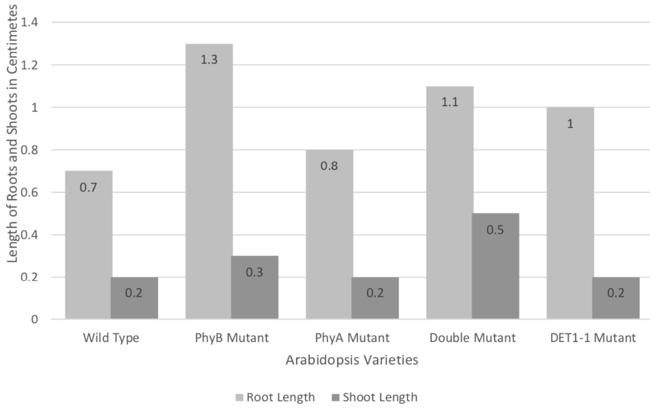


Figure 1: The mean length of five Arabidopsis varieties' roots and shoots in 24-hour light. The graph shows that every variety was successful at producing both shoots and roots in our control experiment. The phyB mutant had the greatest mean root growth at 1.4 cm, and the double mutant had the greatest mean shoot length at 0.5 cm.

avoidance responses, de-etiolation, and elongation under light would be blocked, and the double mutant would grow best under the blue light because with both pathways for red light blocked, the plant would be forced to rely on blue light to initiate a growth response.

RESULTS

We tested how mutant varieties of *Arabidopsis* respond to different colors and amounts of light. The tests were run on five varieties of *Arabidopsis*: wild-type; CS6213, which had a mutation in phyB; CS6219, which had a mutation in phyA; CS6224, which had mutations in both phytochrome A and B; and CS6158, which had a *det1* gene mutation that encouraged growth in the dark. We prepared and plated four agar plates under similar conditions to minimize experimental error; we prepared the same agar solutions on the same day, and we allowed it to sit for the same amount of time before parafilming each plate. Our independent variables were the four different light conditions under which we ran our experiments, and the dependent variable was the resulting growth of each *Arabidopsis* variety. We determined growth success by looking at the average of the total lengths of each plant, including the roots and shoots, and germination ratio under each condition. Plants with longer average lengths and higher germination ratios were considered more successful. We consider germination ratios to be more important than the lengths when determining growth success, so we would consider a variant with a high germination rate and smaller lengths to be more successful than a variant with low germination but longer lengths.

Our baseline showed that all five varieties grew comparably well in 24 hours of direct light. They all had a germination percentage ranging from 75% to 100% and average lengths ranging from 0.9-1.6 cm (**Figure 1**). Based on the parameters gathered from the baseline experiment, we ran a positive control trial in natural light to test each variety's ability to undergo daily functions in a normal growth environment. We also ran trials in no light to test the DET1-1 mutant, and trials in both red and blue light to test the phytochrome mutations. We measured the root and shoot length of each plant after one week and calculated the successful germination percentage.

Under natural sunlight, there were no statistically significant

		NATURAL (SUN)LIGHT		NO LIGHT		RED LIGHT		BLUE LIGHT	
TYPE OF SEED	Root/ Shoot	Mean±SD (cm)	Stan. Err.	Mean±SD (cm)	Stan. Err.	Mean±SD (cm)	Stan. Err.	Mean±SD (cm)	Stan. Err.
WILD TYPE (N=6)	Roots	0.9±0.40	0.07	0.0±0.0	0.00	0.5±0.06	0.02	0.5±0.06	0.02
	Shoots	0.2±0.12	0.02	0.0±0.0	0.00	0.6±0.15	0.05	0.2±0.06	0.02
PHYB MUTANT (N=6)	Roots	0.9±0.18	0.04	0.7±0.00	0.00	0.9±0.13	0.03	1.0±0.24	0.04
	Shoots	0.4±0.05	0.01	1.1±0.00	0.00	1.1±0.13	0.03	0.3±0.12	0.02
PHYA MUTANT (N=6)	Roots	0.6±0.29	0.06	0.4±0.00	0.00	0.6±0.22	0.04	0.7±0.24	0.06
	Shoots	0.2±0.07	0.01	1.2±0.00	0.00	0.6±0.21	0.03	0.2±0.21	0.05
DOUBLE MUTANT (N=6)	Roots	0.4±0.19	0.05	0.0±0.00	0.00	0.6±0.10	0.03	1.4±0.14	0.07
	Shoots	0.4±0.40	0.10	0.0±0.00	0.00	1.5±0.32	0.11	0.6±0.07	0.04
DET1-1 MUTANT (N=6)	Roots	1.0±0.37	0.06	0.7±0.19	0.08	0.8±0.14	0.02	0.9±0.31	0.05
	Shoots	0.1±0.00	0.00	0.4±0.32	0.14	0.1±0.04	0.01	0.1±0.06	0.01

Table 1: The mean root and shoot length with germination percentage of each Arabidopsis variety in four light conditions. Each plate includes an Arabidopsis wild type and four mutant varieties, and all plates were set up identically. The table shows the average root and shoot lengths of each variety, as well as the germination percentage. All plates were placed at the same angle to get qualitative data about phototropism. The mean, standard error, and standard deviation were calculated separately for each variety in each condition. The no-light data for all varieties except the DET1-1 mutant are slightly misleading, because only one plant germinated in each condition, so the mean of the results is only based on one test.

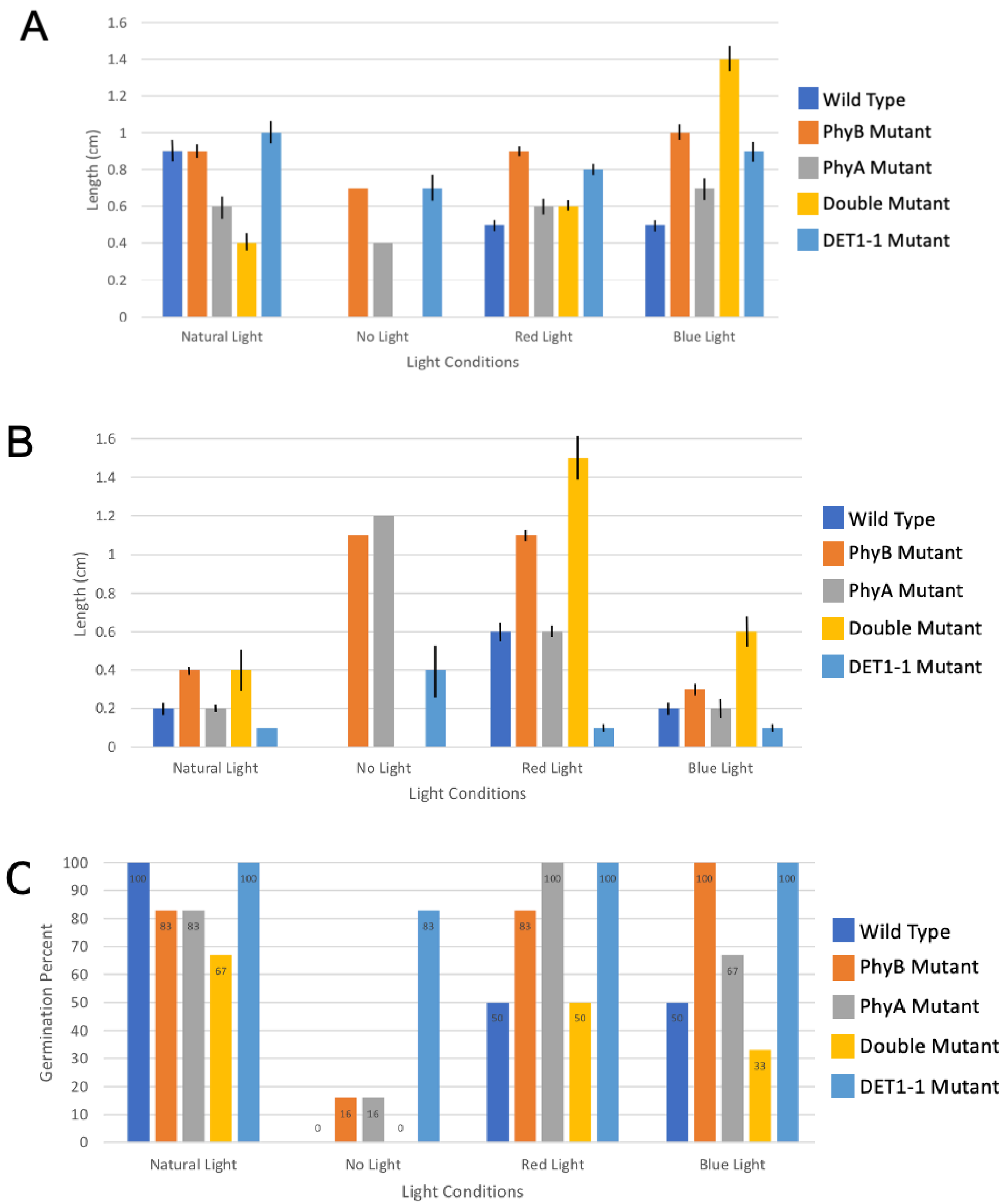


Figure 2: The mean root and shoot lengths of five Arabidopsis varieties grown in sunlight, no light, and red and blue light after one week. A) Mean shoot lengths in all light conditions. **B)** Mean root lengths in all light conditions. The graphs are organized by light condition, so each graph compares the root or shoot lengths of the five varieties under each of the four light conditions. The error bars were created using the standard error values calculated separately for each variety in each condition, as shown in Table 1. **C)** Germination percentage of five Arabidopsis varieties' grown in sunlight, no light, red light, and blue light. In cases where the majority of the seeds did not germinate, the numerical result is based on very few trials and is not a true average.

differences in the shoot lengths of the five varieties. However, the phyA mutant and double mutant had significantly less root growth compared to the others, averaging 0.6 cm and 0.4 cm respectively. Overall, the wild type, phyB, and double mutant all grew well, with high germination percentages from 83%-100% and average lengths from 1.1 cm - 1.3 cm long (**Table 1**). We interpreted overlapping error bars as not showing a statistically significant difference and non-overlapping error bars as suggesting a possible statistical significant between treatments.

Under no light, the phyB mutant, phyA mutant, and DET1-1 mutant were the only mutants to experience any growth. While the phyA mutants had an average length of 1.6 cm and the phyB mutants 1.8 cm, only 16% of the phyA mutants and phyB mutants germinated, while 83% of DET1-1 mutants did so. Thus, although the DET1-1 mutant had a smaller average length of 1.1cm, under no light the DET1-1 mutant grew the best. (**Figure 2**).

In the plate grown with red light, all of the varieties had statistically similar roots lengths of 0.5 cm to 0.9 cm except for the wild type and the phyA mutant which were less statistically significant compared to the phyB mutant. Furthermore, the phyB mutant and double mutant had the greatest shoot lengths, 1.1 cm and 1.5 cm respectively, while the DET1-1 mutant had the shortest shoot growth at only 0.1 cm (**Figure 2**). Overall, the phyB mutant grew best under red light, with an 83% germination rate and an average length of 2.0 cm, and the double mutant grew next best with an average length of 2.1 cm and a 50% germination rate.

The plate grown under blue light produced similar results to the plate grown under red light, with the phyB mutant growing most successfully with an average length of 1.3 cm and a 100% germination rate. The double mutant had the longest length of 2.0 cm, but had a very low germination rate of 30%. The wild type showed diminished growth with 50% germination and an average length of 0.7 cm. According to our data, varying amounts and types of light results in unique growth patterns amongst the five *Arabidopsis* varieties.

DISCUSSION

Our data supports the hypothesis that the DET1-1 mutant grows best in the dark and that there would be variation between the positive control and the sunlight test due to the plants following their normal growth cycle rather than a 24-hour day, which was mimicked by the constant light source in the control. The data also supports our hypothesis that under blue light, the double mutant grows most successfully, since with both red light growth-inducing pathways blocked, the plant relies solely on blue light for energy and growth. However, the data refuted our hypothesis for the tests under red light. In the red light, we expected both of the mutants with one functioning phytochrome, the phyA mutant and the phyB mutant, to grow the best. Contrary to our expectations, the double mutant missing both phytochromes grew the best, along with the phyB mutant instead of the phyA mutant. For the red light, we predicted the double mutant to grow the worst because it has mutations in both phytochromes, which are the pigments that plants use to capture red light. A normal functioning phyA would degrade under red light, inhibiting de-etiolation and plant elongation, while a normal phyB would promote de-etiolation under red light, suppressing shade avoidance responses (5). Mutations in the phytochromes block these light pathways, but their functions cancel each other out. With a dysfunctional phyB pathway, de-etiolation would be less active and shade avoidance would be more active; however, the phyA mutation would leave de-etiolation and elongation unchecked. Thus, the double mutant would be able to de-etiolate and have excessive elongation (6). This is reflected by the increased growth of the double mutant under red light and blue light. Thus, the double mutant was able to grow best under red light due to unregulated elongation. The phyB mutant would also grow well in red and blue light since the pathway suppressing shade avoidance responses under light would be blocked, allowing it to grow longer. The results of the DET1-1 mutant supports the idea that the mutation blocks the *det1* gene, which regulates the transcription of light-mediated pathways for plant development. With a nonfunctional gene, the plant will develop regardless of light stimuli. Therefore, the DET1-1 mutant was able to grow in the dark (4).

Other researchers came to similar conclusions. In a study conducted by Peng Liu and Robert Sharrock, they found that the phytochromes A and B have slightly different functions that are not directly involved in the same pathway, explaining how our double phytochrome mutant could still grow and capture light instead of having the light-capturing pathway shut down (1). They also observed extended growth in their phyA mutants, matching the longer shoots and roots found in the phyB mutant throughout the experiment (1). The increased growth found with a phyA mutation could be further explored, because the mutation enables growth in shady areas, opening new environments to sustain agriculture.

Since our experiment had many steps that took place over the course of a few weeks, there was a lot of room for small errors to build up in our data. The media created was not measured when it was split into the two plates, so one plate may have more agar than others. This difference in nutrients is one of the small inconsistencies in the experiment. Secondly, due to our schedules, it was not always possible to record the data exactly a week after. Instead, the data was collected after a week and one day or a week and two days. In order to finish our data collection on time, we only let the plates sit in the refrigerator for a 72-hour germination period during the follow-up test, compared to the germination period of a week that was used in the baseline. This factor may contribute to the lower germination rates that we observed in the follow-up. Another possible bias is the effect of ambient light on the tests using blue and red light. In the red light test, the plate was placed in a box near the windows, with an opening on the side facing the classroom. The placement of the box near the window may have given the plants some ambient sunlight, as well as some ambient light from the classroom, which may have skewed the data. In the test with the blue light, the lamp was placed in a closet which was dark most of the time, but the light was occasionally turned on, and the plant received some ambient light.

In further investigations, we would explore how a mutant containing mutations in both the *det1* gene and the phytochromes would grow in situations with reduced or no light. This mutant could be compared to plants known to grow well in the shade, to investigate if plants have naturally evolved these mutations in order to grow in new environments.

MATERIALS AND METHODS

The *Arabidopsis* plants were grown in 0.8% agarose plates. The environment was created by mixing a solution of 0.8 grams of agarose with 100 ml of water, which was then boiled by placing the flask in the microwave while it was "sealed" by a paper towel stopper. The flask is placed at room temperature until cool (up to 24 hours) to kill any bacteria spores that may have entered the solution. This boiling and cooling process is repeated two more times to remove all spores. Although a 24 hour cooling period between the boiling steps is preferred, the same result was also reached by boiling the solution twice in one day, allowing the solution to cool completely between boils.

The plants were grown in flat plates, but the plates were gridded on the bottom. This allowed each row to hold six seeds of each type, as long as the seed was placed in the center of one of the grid boxes in that row. To prepare the plates, the agar solution was split equally between two plates. The plates, sealed with parafilm, were placed in the refrigerator to set overnight. This process was completed twice since four plates were needed for the experiment.

The following strains were obtained for this experiment from the Arabidopsis Biological Resource Center at Ohio State University: Stock #CS39005 (wild type), Stock #CS6213 (phyB mutant), Stock #CS6219 (phyA mutant), Stock #CS6224 (double phytochrome mutant), and Stock #CS6158 (DET1-1 mutant). After the plates were set, the rows were labeled from top to bottom with the stock numbers in the order listed above, skipping the first gridded row on the plate, which was labeled as X for clarity. The plates were removed from the refrigerator, and the parafilm was removed to plate the seeds. The seed canisters were opened inside small petri dishes to catch overflow seeds. Using a toothpick, one seed was picked up and placed in one of the gridded boxes in the corresponding row for its seed type. The seed type containers were sealed into the petri dishes when each of the six grid spots had been filled for that seed type. This process was repeated for all seed types. The plates were sealed and put back into the refrigerator for a one-week germination period.

After the germination period, the plates were placed into their experimental environments. The plates were oriented so that they were upright, with the light source above the X row. The no-light environment was created by wrapping the plate in tinfoil. This plate and the plate for natural light were placed adjacent to the same window following the above orientation. Finally, one environment with a blue lamp and one with a red lamp in which the light source is only the colored light were created. One plate was placed in each environment.

Every week, the length of the roots and shoots was measured on each plant. Pictures were taken of the model plant for that test, which was chosen on week one and labeled with a dot in the grid box on the bottom of the plate. The pictures were taken under a microscope. In addition, the number of germinated seeds in each row was recorded.

ACKNOWLEDGEMENTS

We would like to thank Mr. Aaron Mathieu for his guidance on this project and the submission process. We would also like to thank the Arabidopsis Research Center at Ohio State University for supplying the arabidopsis seeds.

Received: July 26, 2019
Accepted: October 19, 2019
Published: March 25th, 2020

REFERENCES

1. "Plant Sensory Systems and Responses." *OpenStax CNX*, Biology. May 27, 2016, http://cnx.org/contents/185cbf87-c72e-48f5-b51e-f14f21b5eabd@10.53.
2. Li, Jigang, *et al*. "Phytochrome Signaling Mechanisms." *The Arabidopsis Book*, American Society of Plant Biologists, 2011, www.ncbi.nlm.nih.gov/pmc/articles/PMC3268501/.
3. Reed, J. W., *et al*. "Phytochrome A and Phytochrome B Have Overlapping but Distinct Functions in *Arabidopsis* Development." *Plant Physiology*, vol. 104, no. 4, Jan. 1994, pp. 1139–1149., doi:10.1104/pp.104.4.1139., www.plantphysiol.org/content/104/4/1139.
4. Chory, J. "*Arabidopsis Thaliana* Mutant That Develops as a Light-Grown Plant in the Absence of Light." *CELL*, vol. 58, iss. 5, Sep. 5, 1989, pp. 991-999 www.arabidopsis.org/servlets/TairObject?type=publication&id=5675.
5. Franklin, Keara A, and Garry C Whitelam. "Phytochrome a Function in Red Light Sensing." *Plant Signaling & Behavior*, Landes Bioscience, vol. 2, iss. 5, 2007, pp. 383–385.www.ncbi.nlm.nih.gov/pmc/articles/PMC2634220/.
6. Klaassen, Pieter. "Roots and Root Hairs." *Roots in Plants | CANNA Gardening USA*, CANNA, www.cannagardening.com/roots_and_root_hairs.

JOURNAL OF
EMERGING INVESTIGATORS

Article

Analysis of technology usage of teens: correlating social media, technology use, participation in sports, and popularity

Vedat Can Düzgezen, Banu Acarlar, and Özgür Öner
Aci High School, Istanbul, Turkey

SUMMARY

This study tests the correlation between technology usage and teens' social lives. The addition of student popularity and the effects of extracurricular activities on technology usage were also examined. A group of 50 students was surveyed (25 males and 25 females; 25 middle schoolers and 25 high schoolers). The survey primarily asked the students to rate the social environment in their school, find the ratio of their in-school to out of school friends, vote for the three most popular students in their grade, and identify their technology usage from one to five (1 representing not dependent at all and 5 representing extremely addicted). A negative correlation was found between participation in extracurricular activities and technology usage (p=0.032), which means that students who participated in extracurricular activities used statistically significantly less technology than the ones who do not. There was no significant difference between the technology usage of middle and high school students. One major finding was that boys used technology mainly for gaming and entertainment (p=0.039), whereas girls mainly used it for social media (p=0.016). Most interestingly, the survey showed that the students who were voted to be more popular by others had higher social media usage on average than those who were not. Unexpectedly, a common answer received in the popularity ranking question was the denial of any presence of popularity in the specified grade. The denying students received significantly fewer "popularity votes" than others. The final results added to an increased understanding of the relationship between technology usage and teens' social lives.

INTRODUCTION

Today's teenagers spend a lot of time on the internet and social media, with 95% of the teens reporting that they have access to or own a smartphone with internet connection and 45% of them saying that they are "constantly online" (1). As a result, experts have identified this problem as internet addiction. According to Young and Rogers, people become addicted to the internet in the same way that they become addicted to drugs, alcohol, and—most similarly—gambling, which result in academic, social, and occupational impairment (2). In addition, previous studies have found a statistically significant correlation between the usage of the internet and depression (3). In the teens' case, this unveils a bigger problem. According to Erik Erikson, teenage years encompass an important developmental stage of life, as teenagers are in search of their unique identities. However, the portrayal of themselves on the internet causes teens to get confused about their emerging new identity (4). With the extended amount of internet use, teens start to create an online personality for themselves. Research done by the Girl Scouts specifies that 74% of girls agree that other girls use social media to make themselves look "cooler" than they are, and 42% say that this statement describes them (5). According to the Pew Research Center, only 25% of teens spend time with their friends after school on a daily basis, and 5% do not meet with their friends outside of school (6). According to the Unified Theory of Adoption and Use of Technology (UTAUT), although it is evident that both genders use technology and social media intensively, males and females do not use it in the same way (7). The intention of this research project is to determine the difference between males' usage and females' usage of technology as well as trying to correlate technology usage and popularity of students at school. In addition, a lot of middle and high school students spend time after school on sports teams and activity clubs. Another factor this study will investigate is whether extracurricular activities after school with which students are occupied with impact their internet usage. Also, this research aims to find the correlation between the thoughts of students about their social life in school and their social media usage. All in all, there are many factors to technology usage among teens, and this study aspires to create a better understanding of the complex social dynamics in school. To clarify, since no intervention is taking place, it is hard to directly infer causal relationships. There may be variables that are not considered in the study, resulting in an illusion of a correlative relationship.

For one of the expected results, whether or not internet usage is affected by extracurricular activities and sports seems clear. We hypothesize that the results would suggest that spending time playing sports and participating in extracurricular activities would decrease technology usage. Since previous studies have proven the beneficial effects of physical exercise and sport participation on self-control (8), although we don't currently have data to support this model, one possibility is that that exercising contributes to solving internet addiction. One can resist the urge of going through the reward and excitement packed social media sites with the discipline exercising regularly brings. On the other hand, preventing boredom by filling free time with extracurricular activities also stops the wanting of using the internet. Whether or not this expected result is correct in our case will be tested through the surveys and data collected from the phones of teenagers. Testing an expected result may also help check the data and the selected subject group.

One hypothesis questions whether there is a correlation between the thoughts of students about social life in school and their social media usage. The 8th and 9th grade students at the surveyed high school spend time socializing in and outside of school. The school surveyed Aci High School, which has one 20-minute, one 50-minute, and several 5-minute breaks between classes. Students have around two hours of free time at school that they can spend socializing. The question in the survey regarding this hypothesis required the students to rate the social environment during these breaks, which are defined to be the immediate physical surroundings, social relationships, and cultural milieus within which defined groups of people function and interact (9). According to Moawad and Ebrahem adolescents' extensive use of electronic communication to interact with their peers may impair their relations with their parents, siblings, and other family members (10). This suggests that adolescents are using technology to profoundly interact and communicate with friends. According to Denworth (11), friendship takes time to develop. The more time two people spend together, the more likely they are to become friends. Since many adolescents use technology to communicate with their peers, one would expect that the more technology they use the better their friendship with others would be, since they would spend more time not necessarily together physically but with each other online. This may signify that students who stay in touch after school are more likely to develop stronger friendship bonds than those who do not. According to Laugeson (12), the lack of social connections and friendship greatly predicts juvenile delinquency, hate towards school, and mental health problems, which can affect someone's liking or disliking of the social environment at school. Thus, the data may unveil that those who have weaker friendship bonds or less usage of social media/communication applications may rate the social environment at school as being worse than those who have stronger friendship bonds or higher usages. Tying it all together, we hypothesized that someone with high social media and technology usage for communication would rate the social environment at school higher than those who use technology primarily for playing games and entertainment purposes.

We found that students who took elective courses after school had a lower technology usage with 24.66 hours per week compared to 30.89. Another major finding was the usage difference between the sexes with females on average spending 54.86% of their technology usage on social media compared to males with 37.62%. Similarly, the findings suggested that male students spent more time on gaming than females with 50.52% of the total technology usage, compared to females' 24.67%.

RESULTS

We asked how the technology usage could be correlated to social life in teenagers with the help of a survey of male and female middle and high school students.

The data collected revealed that the average technology usage (hours per week) for the 8th and 9th graders was 25.86 hr/wk (hours per week) (**Figure 1**). The 9th graders had an average of around three hours more than the 8th graders with 27.03 hr/wk versus 24.25 hr/wk. The maximum time spent on a phone was 46 hr/wk. The average amount of time spent on social media by the 8th and 9th graders was 11.85 hr/wk,

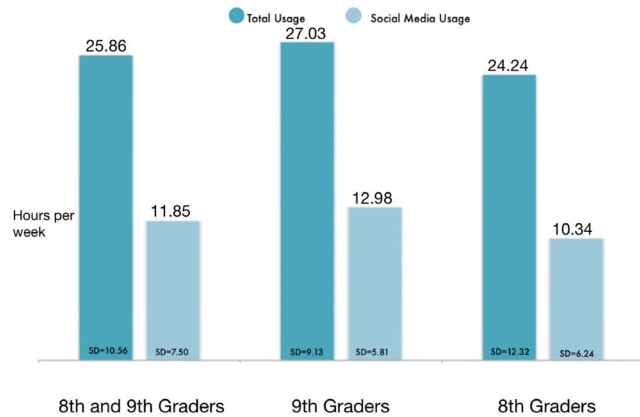


Figure 1: The technology usage averages of 8th , and 9th graders. The total technology usage average is 25.86 hours per week (SD=10.56) and 11.85 hr/wk (SD=7.50) for total social media usage. 9th graders had an average total technology usage of 27.03 hr/wk (SD=9.13) and social media usage of 12.98 hr/wk (SD=5.81). 8th graders had an average total technology usage of 24.24 hr/wk (SD=12.32) and social media usage of 10.34 hr/wk (SD=6.24).

with the maximum being 29.91 hr/wk. The 9th graders led this category with an average of 12.98 hr/wk, while the 8th graders fell back with an average of 10.34 hr/wk. On average, the surveyed teens spent 45.55% of their time on social media. The 9th graders spent on average nearly 5% more time on social media than the 8th graders with 47.97% versus 42.21%.

Out of the 50 students studied, 28% specified that they either did not do sports outside of school or did not participate in any clubs. Only 4% said that they neither participated in clubs nor sports teams outside of school. The average rating on a scale of 0-5 that the students gave to the "social environment" was 3.97. According to the survey, the kids classified themselves as above average technology users by putting up an average of 3.87 to the question: If 5 is extremely addicted to technology and 0 not at all, what would you classify yourself as? Even the students who were clearly under the average technology usage claimed that they were above average internet users. On average, the surveyed students had obtained 66.14% of their friends from school, displaying the importance of the school environment in friendship, with the majority of the friends of surveyed students being in the same school with them. One interesting question on the survey (Appendix A) was the last one: If you are comfortable, name 3 people from your grade that you would consider as the most "popular" (defined as someone or something is liked, enjoyed, or supported by many people) (30). During the surveying process, a considerable amount of people stated that their grades neither used the term "popular" nor had "popular" kids. This seemed to be a controversial topic among teens since many students did not feel comfortable answering this question. Most of them replied with the exact words: "Our grade does not have such a thing as popularity". Upon creating a table from the registered votes, it was clear that the people who received "popularity votes" were not hesitant on voting for others. Those who received 0 votes, without exception, stated that popularity did not exist; however, not all people who denied the existence of popularity received 0 votes. It appears that most of the students who are deemed unpopular deny that the term "popularity" is widely accepted in their grades.

In the intergroup analysis, the rates of social media usage

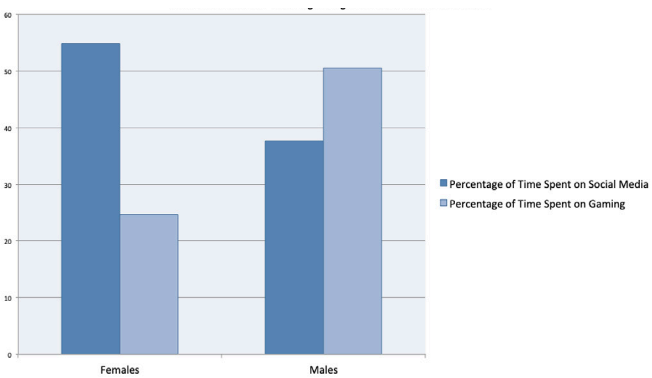


Figure 2: Social media and gaming usages of males and females. On average males spent 37.62% of their technology usage in social media and 50.52% in gaming. Whereas females spent 54.86% of their technology usage in social media and 24.27% in gaming.

between the sexes were examined first. Descriptive statistics of the variables created according to gender are given in the graph below (**Figure 2**). We showed that there was a significant difference between the social media usage rates of male and female students. Especially when the average was examined, we saw that the female students spent 54.86% (SD=13.65) of their free time in social media compared to the 37.62% (SD=21.25) spent by the male students. Similarly, when the time devoted to games and entertainment was examined, we saw that male students spent 50.52% (SD=16.04) of their free time on games and entertainment compared to female students with 24.67% (SD=17.89) (t-test, p=0.03).

The threshold for statistically significant p-values for the t-test was taken as alpha=0.05. When the t-test results were examined, a significant difference was found between the rate of social media use to all technology usage, social media usage, in-school friend rates, and game and entertainment ratio for males and females (t-test, p=0.01). This means that statistically, male (25.17 hours per week, SD=11.08) students used social media less than female (28.70 hk/wk, SD=11.28) students. Similarly, there seemed to be a significant difference in the percentage of in-school friends between females and males, with females reportedly having more in-school friends than males (t-test, p=0.03). The higher number of in-school friends of females may be because of their higher use of social media and communication applications than males. Further research would be necessary to investigate this result.

When we examined the students who take elective courses (including playing sports) and those who do not, the average technology usage values between the two groups appeared to be close to each other. **Figure 3** shows the significant differences between these two groups when we take a critical value of 0.05 for the t-test. For this reason, we can say that there is a significant difference in technology usage amounts between those who take elective courses and those who do not.

When the data above (**Figure 3**) of students who are engaged in sports activities and those who do not were examined, we could show that the students who play sports spent less time on social media than the students who do not play sports (30.89 hr/wk, SD=17.40, compared to 24.66, SD=13.24). Similarly, there was a difference in the scores obtained according to the answers to the questions. It is striking that especially non-student-athletes were also more interested in technology than the student athletes. Because the values for the rate of friends in school (p=0.06) and social environment ratings (p=0.07) were close to the critical value (p=0.05), we notice that they are close to being statistically significant. This means that there might be a real difference of in-school friends of students who participate in sports and students who do not; however, this data failed to demonstrate a statistically significant correlation. Similarly, these tests and data cannot prove any causation between the two variables and should be approached with scepticism.

According to the correlation test, the "Game and Entertainment Ratio Compared to Total Usage, and the Social Media Ratio Compared to Total Usage" variable had a high inverse correlation with an r-value of r=0.67 (p=0.04) for females and r=0.73 (p=0.03) for males. This suggests that, with statistical significance, the students who do not spend their free time on social media spent time on games and entertainment, and vice versa. This was obtained by analyzing social media and gaming application usages of individuals. Although the data signifies an inverse correlation, we could not specify whether the variables have a causal relationship.

When analyzed in female students, there was a weak positive correlation between "Social Media Ratio Compared to Total Usage" and "Social Environment Score" variables r=0.33 (p=0.02). In other words, the social environment score of female students increased slightly as social media usage increased.

To summarize, the data supported both the hypothesis and the expected results. From the analysis, we could see that, in fact, spending time playing sports and participating in extracurricular activities decreased technology usage. However, we could not obtain a causative relationship from the data since, again, no variable was manipulated. We can only infer that, in this specific group, the usage of technology was lower for individuals who spent time doing sports and extracurricular activities. For the hypothesis concerning the relationship between social media usage and social environment ratings of students, results from the correlation test showed that there was a correlation between the two data. Although this suggests that individuals who have higher social media usage ranked the social environment at school higher, we cannot say that one causes another. Another aim of this research was to pinpoint the differences in the technology usage of males and females. From the t-test results, we see

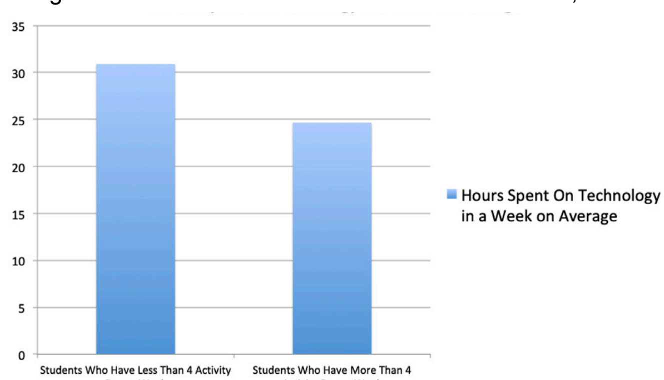


Figure 3: Technology usages of students who have more than four activity days a week and those who do not. The average total usage for the students who have more than four activity days a week was 30.89 and 24.66 for those who do not.

that males and females use technology differently. Females use technology mainly to go through social media while males mainly use technology for entertainment and games. The results did not suggest a significant difference between the technology usages of middle and high school students.

DISCUSSION

When an intergroup analysis was made between the males and females who participated in the study, there was a noticeable statistical difference between the technology usages of the two groups. We calculated that the females spent 54.86% of their total technology usage on social media, compared to 37.62% of usage in males. Similarly, with a p-value equalling 0.03, males spent more time on mobile games than females did. Thus, we can infer that females on average tend to spend their time on social media while males prefer gaming on mobile applications. With this information, we can see that females and males tend to use their time on their mobile phones differently, helping us understand the general usage.

We found that there was a statistically significant difference in the technology usage of students who participated in extracurricular activities and those who did not. Additionally, we showed that students who participated in extracurricular activities spent less time on their technological devices. Although this semblance is only a correlation, not causation, it could increase our understanding of the subject. For example, understanding that participating in clubs and sports teams correlates with decreased internet usage, could help parents who are concerned about their children's social media usage to help their children by encouraging them to participate in those kinds of activities. One interesting finding similar to this one is the social environment ratings and in-school friend numbers of students who participated in sports. Student athletes, with a value trending towards significance (t-test, p=0.06), both rated their social environment higher than non-student-athletes and reported that they had more in-school friends than those who did not do sports. This could also help parents understand the social dynamics of schools, even if the results are not completely statistically significant, which could help children who have problems socializing or making friends at school find common ground with others or overcome their problem.

There are several factors that affect one's technology usage that cannot be accounted for. This research aimed to find correlations rather than causations, since every individual is unique. Although there are several factors that affect an individual's internet usage, correlations can greatly help in understanding the general picture. Since personality, socializing preferences, friend groups, and family influences differ in every individual, it is hard to estimate whether someone's preferences would change when the variables in this study were to change (sex, age, technology usage, and activity and application preferences). Another setback of the study is that the measurement methods were not standardized or taken from psychology literature. Future studies may develop or validate the findings of this study by using measures set by a standardized source such as the the DCT-IA by the DSM-5 (The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition)(31).

Several prior studies were conducted on the correlation between social media usage and clinical depression (3). This topic is worth further discussion, since it is estimated that around 20% of all adolescents get diagnosed with depression (13). Comparing prior research and this experiment's findings, social media usage may be one of the causes. The overwhelming use of social media usage among teens and the manipulation of physical appearance may cause teens to create a false sense of being. With the extended use of the internet, teens today create an online personality for themselves, which is usually different from their real-life identity. Research done by the Girl Scouts specifies that 74% of girls agree that other girls use social media to make themselves look "cooler" than they are, and 42% say that this statement describes them (5). Teens may create or believe in higher standards of beauty given the extreme usage of social media and body manipulation techniques, and get depressed when they cannot reach the increased standards. The creation of a "better" or "cooler" personality online may cause them to devalue their real-life identity. This may be found in any age range as Yang and Brown found the use of an altered Facebook self-representation to obtain a contemporary higher self-esteem in transition to college (32).

Prior to this study, there have been many arguments about the effects of social media usage in an adolescent's well-being and emotional status. Many studies, such as that of Kross, have found that Facebook use predicted a negative shift in both life satisfaction and how people feel moment-to-moment (14). Contrarily, many studies, such as Orben's "Social Media's Enduring Effect on Adolescent Life Satisfaction," have found that social media use is not, in and of itself, a strong predictor of life satisfaction among the adolescents (30). Both of these contrasting studies have been criticized due to their methods of measuring well-being, thus not resolving the debate. Although the results of this study do not definitely prove anything or end the discussion, they support Orben's findings with a significant difference in the numbers of in-school friends and social environment evaluation scores of students with high and low social media usages. An external cause may make social media usage and the user's well-being appear negatively correlative, thus causing the debate. As long as the absence of a perfect well-being metric and ability to survey a large proportion of the users continues, so will the debate.

A term that is worth researching is the term popularity. The fact that around 20 students gave the exact response ("There is no such thing as popularity in our term") seems to be more than a mere coincidence. The purpose of searching for the most "popular" students was to be able to draw a correlation between the popularity of a student and their social media usage or social environment point (the rating that helps us calculate how favourable a school's social environment is from the perspective of the students). However, interestingly there was no correlation. What stood out was that the popular kids (the kids who received votes from others) did not refuse to vote for others, yet the ones with fewer than five votes refused to write down any names. We hypothesize that the ones who are not deemed popular by their peers do not want to admit their unpopularity by stating that there is no such thing as popularity. However, this hypothesis would need more time and data to be resolved with clear reasoning.

METHODS

To test the hypothesis concerning the relationship between technology usage and the social lives of teenagers, 8th and 9th graders of a private high school were surveyed. Every student was asked to give consent for their data to be used anonymously in our research project. The survey requested the amount of time each student spent on their phone, going through apps classified as social media, games and entertainment, and creativity and productivity.

Some apps were more popularly used than others in their categories. According to the survey, Instagram, Snapchat, WhatsApp, Facebook, Twitter, and TikTok were the most used social media apps (unordered). YouTube, 9Gag, Netflix, Hulu, PUBG Mobile, Fortnite Mobile, Growtopia, Minecraft, and Sims were the most popular applications under the category "Games and Entertainment". Finally, the Apple Notes, Notability, Safari, Puffin, PowerPoint, Word, Google Drive, Google Docs, and Google Slides were the most used apps under the category "Creativity and Productivity" (28, 29).

The amount of time each student spent on these apps was received via the new screen time feature that calculates the time spent on each app. The data taken from the screen time feature were matched to a number representing the students' names for privacy. After the data were processed, a survey was made using the questions listed in the appendix.

The students were asked about what extracurricular clubs they were participating in and how much time they spent attending it. They were also questioned about how much time they usually spent during each week on sports. The students were asked to rate the social environment at school, estimate the number of friends attending the same school, and describe their technology usage as above or below average without knowing the data. The survey was complemented by the values taken from the screen time feature on personal devices.

The results of the survey were analysed using the SPSS Program; t-tests and correlation tests were performed to obtain statistically significant (p-value = 0.05) different mean values and correlations from the data.

Received: October 22, 2019
Accepted: February 12, 2020
Published: March 27, 2020

REFERENCES

1. Anderson, Monica, and Jingjing Jiang. "Teens, Social Media & Technology 2018." *Pew Research Center: Internet, Science & Tech*, Pew Research Center, 31 Dec. 2019, www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/.
2. Young, Kimberly S., and Robert C. Rogers. "The Relationship Between Depression and Internet Addiction." *CyberPsychology & Behavior*, vol. 1, no. 1, 1998, pp. 25–28., doi:10.1089/cpb.1998.1.25.
3. Üneri, Özden Şükran, and Canan Tanıdır. "Evaluation of Internet Addiction in a Group of High School Students: a Cross-Sectional Study / Bir Grup Lise Öğrencisinde Internet Bağımlılığı Değerlendirmesi: Kesitsel Bir Çalışma." *Dusunen Adam: The Journal of Psychiatry and Neurological Sciences*, 2011, pp. 265–272., doi:10.5350/dajpn2011240402.
4. Poole, Sarah, and John Snarey. "Erikson's Stages of the Life Cycle." *Erikson's Stages of the Life Cycle*, SpringerLink, 1 Jan. 1970, link.springer.com/referenceworkentry/10.1007/978-0-387-79061-9_1024.
5. "Who's That Girl? Image and Social Media." *Girl Scouts of the USA*, June 2010, www.girlscouts.org/en/about-girl-scouts/research/publications.html.
6. Lenhart, Amanda. "Teens, Technology, and Friendships." *Pew Research Center*, 1 Oct. 2015, www.pewresearch.org/staff/amanda-lenhart/.
7. Goswami, Ananya, and Sraboni Dutta. "Gender Differences in Technology Usage—A Literature Review." *Open Journal of Business and Management*, vol. 04, no. 01, 2016, pp. 51–59., doi:10.4236/ojbm.2016.41006.
8. Park, Jae-Ahm, et al. "Effect of Sports Participation on Internet Addiction Mediated by Self-Control: A Case of Korean Adolescents." *Kasetsart Journal of Social Sciences*, vol. 37, no. 3, 2016, pp. 164–169., doi:10.1016/j.kjss.2016.08.003.
9. Pathak, Elizabeth B, and Michele Casper. "A Definition of 'Social Environment.'" *American Journal of Public Health*, vol. 91, no. 3, Apr. 2001, p. 465., doi:10.2105/AJPH.91.3.465a.
10. Moawad, Gehan E. "The Relationship between the Use of Technology and Parent Adolescents Social Relationship." *Journal of Education and Practice*, vol. 24, no. 4, 2003, doi:10.1016/s0197-4572(03)00224-6.
11. Laugeson, Elizabeth. "The Science of Making Friends: Helping Socially Challenged Teens and Adults." *The Help Group*, UCLA Autism Research Alliance, 21tb7c2uredpvu4mah6gyjmt-wpengine.netdna-ssl.com/assets/3B Laugeson - The Science of Making Friends.pdf.
12. Borchard, Therese J. "Why Are So Many Teens Depressed?" *World of Psychology*, 8 July 2018, psychcentral.com/blog/why-are-so-many-teens-depressed/.
13. Kross, Ethan, et al. "Facebook Use Predicts Declines in Subjective Well-Being in Young Adults." *PLoS ONE*, vol. 8, no. 8, 2013, doi:10.1371/journal.pone.0069841.
14. Arisoy, Ozden. "Internet Bagimliligi." *Turkiye Klinikleri Psikiyatri - Özel Konular*, Turkiye Klinikleri, 1 Jan. 1970, www.turkiyeklinikleri.com/article/tr-internet-bagimliligi-54344.html.
15. Ciairano, Silvia, et al. "Parenting and Adolescent Well-Being in Two European Countries." *Adolescence*, U.S. National Library of Medicine, 2008, www.ncbi.nlm.nih.gov/pubmed/18447083.
16. Correa, Teresa. "Who Interacts on the Web?: The Intersection of Users' Personality and Social Media Use." *American Psychological Association*, APA Psych Net, 2010, psycnet.apa.org/record/2009-24129-014.
17. Currie, Melissa. "Parenting in the Technology Generation: Exploring the Impact of Technology on Adolescents and Their Parents." *UQ ESpace*, The University of Queensland, 10 July 2014, espace.library.uq.edu.au/view/UQ:360600.
18. Akcay, Behiye. "THE RELATIONSHIP BETWEEN TECHNOLOGY AND ETHICS; FROM SOCIETY TO SCHOOLS ." *Turkish Online Journal of Distance Education*, vol. 9, no. 4, Oct. 2008. *Istanbul University* , ISSN: 1302-6488.
19. Halabi, Jehado. "Attitudes of Saudi Nursing Students toward Nursing Research." *Saudi Journal for Health Sciences*, vol. 5, no. 3, 2016, p. 118., DOI:10.4103/2278-0521.195813.
20. Jones, Sydney, and Susannah Fox. "Generations Online in 2009." *Pew Research Center: Internet, Science & Tech*, Pew Research Center, 31 Dec. 2019, www.pewresearch.org/internet/2009/01/28/generations-online-in-2009/.

21. Kuss, Daria J., and Mark D. Griffiths. “Internet Addiction: A Real Addiction?” *Internet Addiction in Psychotherapy*, 2015, pp. 54–104., doi:10.1057/9781137465078_4.
22. Morrison, Catriona M., and Helen Gore. “The Relationship between Excessive Internet Use and Depression: A Questionnaire-Based Study of 1,319 Young People and Adults.” *Psychopathology*, vol. 43, no. 2, 23 Jan. 2010, pp. 121–126. *NCBI*, DOI:10.1159/000277001.
23. Ofcom. *Children and Parents: Media Use and Attitudes Report*. Ofcom UK, 23 Oct. 2012, www.ofcom.org.uk/__data/assets/pdf_file/0020/56324/main.pdf.
24. Moawad, Gehan N, and Gawhara G Ebrahem. “The Relationship between Use of Technology and ParentAdolescents Social Relationship.” *Journal of Education and Practice* , vol. 7, no. 4, 2016, ISSN:2222-288X .
25. Whang, Leo Sang-Min, *et al*. “Internet Over-Users Psychological Profiles: A Behavior Sampling Analysis on Internet Addiction.” *CyberPsychology & Behavior*, vol. 6, no. 2, 2003, pp. 143–150., DOI:10.1089/109493103321640338.
26. Yang, Shu Ching, and Chieh-Ju Tung. “Comparison of Internet Addicts and Non-Addicts in Taiwanese High School.” *Computers in Human Behavior*, vol. 23, no. 1, 2007, pp. 79–96., doi:10.1016/j.chb.2004.03.037.
27. “List of Most-Played Mobile Games by Player Count.” *Wikipedia*, Wikimedia Foundation, 20 Mar. 2020, en.wikipedia.org/wiki/List_of_most-played_mobile_games_by_player_count.
28. Clement, J. “Top U.S. Mobile Social Apps by Users 2019.” *Statista*, 10 Mar. 2020, www.statista.com/statistics/248074/most-popular-us-social-networking-apps-ranked-by-audience/.
29. Orben, Amy, *et al*. “Social Media’s Enduring Effect on Adolescent Life Satisfaction.” *Proceedings of the National Academy of Sciences*, vol. 116, no. 21, June 2019, pp. 10226–10228., DOI:10.1073/pnas.1902058116.
30. “POPULARITY: Meaning in the Cambridge English Dictionary.” *Cambridge Dictionary*, dictionary.cambridge.org/dictionary/english/popularity.
31. Tu, Dongbo, *et al*. “A New Measurement of Internet Addiction Using Diagnostic Classification Models.” *Frontiers in Psychology*, vol. 8, Oct. 2017, DOI:10.3389/fpsyg.2017.01768.
32. Yang, Chia-Chen, and B. Bradford Brown. “Online Self-Presentation on Facebook and Self Development During the College Transition.” *Journal of Youth and Adolescence*, vol. 45, no. 2, Mar. 2015, pp. 402–416., DOI:10.1007/s10964-015-0385-y.

Sponsorship

Editor's Circle $10,000+

ALLOY THERAPEUTICS

Patron $5,000+

NEW ENGLAND BioLabs® Inc.
enabling technologies in the life sciences

PORTFOLIOS WITH PURPOSE®

Institutional Supporters

HARVARD UNIVERSITY

HARVARD MEDICAL SCHOOL

Tufts UNIVERSITY

Charitable Contributions

We need your help to provide mentorship to young scientists everywhere.

JEI is supported by an entirely volunteer staff, and over 90% of our funds go towards providing educational experiences for students. Our costs include manuscript management fees, web hosting, creation of STEM education resources for teachers, and local outreach programs at our affiliate universities. We provide these services to students and teachers entirely free of any cost, and rely on generous benefactors to support our programs.

A donation of $30 will sponsor one student's scientific mentorship, peer review and publication, a six month scientific experience that in one student's words, 're-energized my curiosity towards science', and 'gave me confidence that I could take an idea I had and turn it into something that I could put out into the world'. **If you would like to donate to JEI, please visit https://emerginginvestigators.org/support, or contact us at questions@emerginginvestigators.org.** Thank you for supporting the next generation of scientists!

'Journal of Emerging Investigators, Inc. is a Section 501(c)(3) public charity organization (EIN: 45-2206379). Your donation to JEI is tax-deductible.'