|CURIEUX|

ACADEMIC JOURNAL

November Issue

Issue 33: Part Two

Editing Staff

Chief-In-Editor

Caroline Xue

George Hasnah

Chief of Operations

Anshal Vyas

Assisting Editors

Olivia Li

Luzia Thomas

Madalyn Ramirez

Shefali Awasthi

Soumya Rai

Sarah Strick

Table Of Contents

Page 6: Relationship Between Food Insecurity and Cardiovascular Health Outcomes and Efficacy of Food Banks in Improving Cardiovascular Outcomes Among Those Experiencing Food Insecurity By Meghana Simhadri, Oluwaseyi Adeyinka, MD, MPH

Page 15: Why has the United States threatened to withdraw support to the UNSC Resolution 1820? A Feminist Perspective By Helen Bai

Page 23: The Nonlinear Dynamics of a Ball Bouncing on a Vibrating Ferrite Rod By Ke Li and Kangqi Liao

Page 37: The Neuroscience and Psychology Behind the Joys of Acting By Aditi Shankar

Page 42: Bell's Theorem and Entanglement By Shreyas Dillon

Page 53: An Analysis on the Educational Potential and Challenges of ChatGPT By Jihong Kim

Page 60: Recent Advances in Mesenchymal Stem Cell-based Therapy for Diabetes Mellitus By Lancy ShangGuan

Page 72: From Black Hole Thermodynamics to Holographic Principle By Chuyue Zhang

Page 82: Mechanisms of Cancer Resistance in Mammals By Amari Anderson

Page 105: Economics of Bitcoin and Inflation: Is Bitcoin an Inflation Hedge? By Gokulakrishnan Rathinasamy

Page 117: Assessing gKDR-GMM Model's Performance on *C. elegans* Brain Activity Data By Reia Lee

Page 131: Leaders in Transition: A Comparative Study of Military and Political Figures During Epochal Transformations By Minggang Li

Page 142: Machine Learning for Autonomous Decision-Making and Visual Computing for Robotics By Damla Köse

Page 155: 1995 Rugby World Cup; Illusory Racial Reconciliation in the Rainbow Nation By Ted Huneke

Page 162: Vaping: Comprehensive Review and Health Effects on Youth By Kevin Cao

Page 177: Long-term Health Consequences of Bulimia Nervosa By Catherine Jian Xi Teng and Cindy Xin Wen Zhang

Page 188: The Impact of Data and Analytics on Strategic Decision-Making in the National Basketball Association By Sarp Kologlu

Page 204: Application of CRISPR/Cas 9 by Biotechnology Companies to Treat Genetic Disease By Atiksh Vats

Page 214: Parenting Styles and Academic Outcomes: Combatting Your Negative Learned Behaviors By Cassandra Avallone

Page 221: NLP Sentiment Analysis of News Articles: Uncovering Opinion Dynamics in Contemporary Media by Akshay Shivkumar

Page 238: Comparing Prominent Generative Language Models for Classifying Political Alignment Of Limited Context Bigrams By Sankalp Singh

Page 248: Environmental and Genetic Risk Factors of Breast Cancer and Disparities By Jia Yue (Joyce) Deng

Page 264: Music and Bridging Divide on National and International Scales By Anjali Schoetz

Page 272: Fostering Academic Advancement in São Luís: Enhancing Research Opportunities and Laboratory Infrastructure for High School Students By Camila W. Maalouf

Page 283: Using Machine Learning to Improve the Accuracy of Sentiment Analysis on Amazon Consumer Reviews By Mahatru Guddamsetty

Page 290: Korea's Independence Movements: Influence of International Context in the East Asia-Pacific Region By Yejun Cheon

Page 302: Stratification and Power: A Study on Power and Status in Hierarchical Structures By Michael Zhu

Page 312: Putting the Brakes on T Cell Exhaustion: Platforms for Chimeric Antigen Receptor (CAR) T Cell Regulation Lead the Way By Chungbi Yoon

Page 328: "Mass Shootings in America: The Mistaken Story" An Opinion Piece By Caroline Browning

Page 331: Brain Regions Activated in Different Types of Context-Dependent Memory: A Review By Abigail Bogdanovsky

Page 340: Overexpression of ATP Synthase in E.Coli By Jude Thomas, Meta Foster, and David Crowley

Page 346: Why People Obey the Law: A Critical Review of Existing Research by Makayla P. Wu

Page 353: An Investigation of the Labor Movements in the US and China By Jianing Chen

Page 360: Modified A* Algorithm With Octile Distance By Eunchan Lee

Page 375: Exoplanet Detection with Decision Trees By Sriram Loganathan

Page 383: Performance of OpenCV HOG Model for Various Image Quality Changes By Soumik Sinha

Page 389: The Comorbidities of Iron Deficiency Anemia and Psychiatric Illness: Analyzing Global Data to Bridge the Gap in the Literature By Sophia Smallwood

Page 393: In the Name of Profit and Compassion: Increasing Employee Engagement By Serena Wu

Page 399: WWI and Women's Fashion By Emma Lu

Page 405: Evaluating the Correlation of Different NBA Factors on Team Success By Ishan Katpally

Page 415: The Effect and Solutions of Climate Change in Orange County By Amudhan Gurumoorthy

Page 427: Using AI Models to Predict the Occurrence of Invasive Species Based on Habitat Preferences By JanetLou Guan

Page 440: Borderline Personality Disorder: Dissecting Gendered Stigmas, Historical Context, and Media Impact By Victoria Lai

Page 456: The Influence of the Shape of the Cam Pulley on Leg Extension Machines By Junghyun Koh

Page 466: Using Regulatory T-Cells to Treat Type 1 Diabetes in Humans By Jaisal Vohra

Page 470: Identifying Invasive Plant Species with Convolutional Neural Networks By Kwanhee Lee

Page 477: Evolutionary Aviation Network Flow Optimization Using Topological and Demographic Features By Rahul Rao

Page 486: Type 1 Diabetes in Children: Affect on Cognition and Biotechnological Solutions By Shiksha Anand

Relationship Between Food Insecurity and Cardiovascular Health Outcomes and Efficacy of Food Banks in Improving Cardiovascular Outcomes Among Those Experiencing Food Insecurity By Meghana Simhadri, Oluwaseyi Adeyinka, MD, MPH

Abstract

Food insecurity, an issue characterized by limited or uncertain access to sufficient and nutritious food, has wide-ranging consequences. Dietary choices significantly impact cardiovascular health, including the development of conditions like diabetes. However, individuals grappling with food insecurity often encounter constraints in their food options and struggle to obtain high-quality, nutritious meals. Consequently, they face an elevated risk of experiencing adverse cardiovascular outcomes. To address this critical concern, food banks have emerged as establishments dedicated to meeting the nutritional needs of individuals confronted with food insecurity. Traditionally, food banks have provided emergency food assistance during times of crisis. More recently, researchers have explored the potential of food banks to serve as platforms for implementing nutritional interventions aimed at improving cardiovascular health outcomes and addressing risk factors. This review aims to assess different interventions that have been implemented through food banks in the United States to address the interconnected issues of food insecurity and increased risk of cardiovascular outcomes and risk factors, especially diabetes. By synthesizing the existing evidence, this review seeks to identify effective strategies for improving cardiovascular health among food-insecure populations and inform future interventions in this field.

Introduction

Food insecurity, a condition in which households lack access to adequate food because of limited money or other resources, is a leading health and nutrition issue in the United States (Gundersen and Ziliak). The Food and Agriculture Organization of the United Nations has developed a scale for food insecurity, defined from food security to severe food insecurity ("Hunger"). Those experiencing moderate food insecurity consume a lower quality and/or quantity of food and are not certain about their ability to obtain food ("Hunger"). This can cause malnutrition, stunted growth in children, micronutrient deficiencies, and obesity ("Hunger"). Those experiencing severe food insecurity do not have food to eat, even going days without eating.

Food insecurity remains a relevant and pervasive issue. In 2021, approximately 10.2 percent (13.5 million) of households experienced food insecurity at some point during the year (*USDA ERS - Key Statistics & Graphics*). These households lacked sufficient money or resources to acquire enough food for all their members (*USDA ERS - Key Statistics & Graphics*). Among the food-insecure households, 6.4 percent (8.4 million) had low food security, meaning they managed to avoid significant disruptions to their eating patterns by adopting coping strategies such as relying on federal food assistance programs or community food pantries (*USDA ERS - Key Statistics & Graphics*). On the other hand, 3.8 percent (5.1 million) of households faced very low food security, resulting in disrupted eating patterns and reduced food

intake for one or more household members due to insufficient resources (*USDA ERS - Key Statistics & Graphics*). In terms of population, 33.8 million individuals lived in food-insecure households in 2021 (*USDA ERS - Key Statistics & Graphics*). Within this group, 8.6 million adults resided in households with very low food security (*USDA ERS - Key Statistics & Graphics*). Additionally, there were 5.0 million children living in food-insecure households where both adults and children faced food insecurity (*USDA ERS - Key Statistics & Graphics*). Among the nation's children, 521,000 (0.7 percent) lived in households where at least one child experienced very low food security (*USDA ERS - Key Statistics & Graphics*).

One way food insecurity can manifest is an insufficient quantity of food. Those experiencing food insecurity may not have enough food to meet their dietary needs, leading to hunger and malnutrition (*Hunger and Poverty in America | Feeding America*). Food insecurity can also result in the consumption of poor-quality food due to the lack of access to nutritious foods. Individuals affected by this may have to rely "on low-cost, energy-dense foods that are often high in unhealthy fats, added sugars, and sodium," leading to an increased risk of chronic diseases (Gundersen and Ziliak). Further, food insecurity can include uncertainty about consistent and reliable access to food, causing individuals stress and anxiety (*Hunger and Poverty in America | Feeding America*). Lastly, there are several coping strategies individuals experiencing food insecurity may employ, including eating less food, foregoing some meals, and consuming a lower-quality diet. (*Hunger and Poverty in America | Feeding America*).

Several specific groups had rates of food insecurity that surpassed the national average of 10.2 percent (*USDA ERS - Key Statistics & Graphics*). These groups included all households with children, with a rate of 12.5 percent, and households with children under the age of 6, with a rate of 12.9 percent (*USDA ERS - Key Statistics & Graphics*). Furthermore, households headed by a single woman experienced a higher rate of 24.3 percent, while households headed by a single man had a rate of 16.2 percent (*USDA ERS - Key Statistics & Graphics*). Women living alone had a food insecurity rate of 13.2 percent, and men living alone had a rate of 12.3 percent (*USDA ERS - Key Statistics & Graphics*). Moreover, food insecurity was more prevalent in households with Black, non-Hispanic reference persons (19.8 percent) and Hispanic reference persons (16.2 percent) (*USDA ERS - Key Statistics & Graphics*). A household reference person is an adult household member in whose name the housing unit is owned or rented (*USDA ERS - Key Statistics & Graphics*). Additionally, households with incomes below 185 percent of the poverty threshold had a significantly higher rate of 26.5 percent (*USDA ERS - Key Statistics & Graphics*). In 2021, the Federal poverty line for a family of four was $27,479 (*USDA ERS - Key Statistics & Graphics*). These findings highlight the disparities in food insecurity rates among specific demographic and economic groups, emphasizing the need for targeted support and interventions to address these challenges.

According to the Universal Declaration of Human Rights, "Everyone has the right to a standard of living adequate for the health and well-being of himself and of his family, including food, clothing, housing and medical care and necessary social services, and the right to security in the event of unemployment, sickness, disability, widowhood, old age or other lack of

livelihood in circumstances beyond his control" (Nations). To respect and ensure the protection of this human right, it is essential to address the issue of food insecurity.
The consequences of food insecurity include reduced productivity and earnings, increased healthcare costs, economic losses for communities, and perpetuation of poverty cycles with inter-generational impacts (*Hunger and Poverty in America | Feeding America*). Malnutrition and hunger associated with food insecurity can lead to decreased energy levels, difficulty concentrating, and impaired cognitive function, all of which hinder work performance and limit educational attainment (*Hunger and Poverty in America | Feeding America*). Furthermore, individuals experiencing food insecurity are more likely to have chronic health conditions, resulting in higher healthcare expenses (Seligman, Laraia, et al.). Inadequate nutrition exacerbates health problems, leading to additional healthcare needs and costs (Seligman, Laraia, et al.).

The impact of food insecurity extends beyond individuals to communities (*USDA ERS - Key Statistics & Graphics*). Insufficient access to food affects local economies by reducing workforce productivity and constraining consumer spending (*USDA ERS - Key Statistics & Graphics*).

One of the most significant long-term effects of food insecurity is its contribution to poverty cycles (*Hunger and Poverty in America | Feeding America*). Limited access to nutritious food impedes individuals' ability to thrive, ultimately restricting their educational attainment and future earning potential (*Hunger and Poverty in America | Feeding America*). Consequently, children growing up in food-insecure households may face persistent challenges in education, employment, and overall economic well-being, perpetuating the cycle of poverty across generations (*Hunger and Poverty in America | Feeding America*).

Health Outcomes of Concern

The primary cardiovascular outcomes of concern in this review are heart attack and stroke. A stroke occurs when a blood vessel bursts in the brain or when the blood supply to the brain is blocked ("About Stroke | Cdc.Gov"). A heart attack occurs when enough blood does not reach part of the heart muscle ("Heart Attack Symptoms, Risk Factors, and Recovery | Cdc.Gov"). Individuals with diabetes have an increased likelihood of developing heart disease. Further, they are more prone to possessing specific risk factors, such as high blood pressure or high cholesterol, which elevate the risk of experiencing heart attacks or strokes (*Diabetes, Heart Disease, & Stroke - NIDDK*). The elevated blood glucose levels that are associated with diabetes can inflict damage on blood vessels and nerves that regulate heart function and blood flow (*Diabetes, Heart Disease, & Stroke - NIDDK*). This damage accumulates and contributes to the development of heart disease (*Diabetes, Heart Disease, & Stroke - NIDDK*). Diabetics often develop heart disease earlier in life than those without diabetes. Compared to adults without diabetes, adults with diabetes are almost twice as likely to experience heart disease or stroke (*Diabetes, Heart Disease, & Stroke - NIDDK*).

Hypertension significantly increases the risk of heart disease and stroke, two leading causes of death in the United States ("Facts About Hypertension | Cdc.Gov"). In 2021 alone, hypertension was a primary or contributing cause of 691,095 deaths in the United States ("Facts About Hypertension | Cdc.Gov"). Alarmingly, nearly half the adults in the United States, approximately 116 million individuals, have hypertension ("Facts About Hypertension | Cdc.Gov").

Strokes and heart attacks also pose significant health risks. In 2021, approximately 1 in 6 deaths attributed to cardiovascular disease was due to a stroke ("About Stroke | Cdc.Gov"). Someone in the United States experiences a stroke every 40 seconds, and strokes claim a life, on average, every 3 minutes and 14 seconds ("About Stroke | Cdc.Gov"). Annually, over 795,000 individuals in the United States suffer from a stroke ("About Stroke | Cdc.Gov").

Diet plays a crucial role in cardiovascular health. Research suggests that a higher-quality diet is associated with a reduced risk of recurrent cardiovascular events in individuals aged 55 and above with cardiovascular disease or diabetes (Dehghan et al.). This shows the importance of promoting healthy eating habits to reduce the recurrence of cardiovascular events.

The impact of food insecurity on cardiovascular health is notable. When people do not have adequate access to nutritious food, they are unable to consume the nutritious food necessary to prevent cardiovascular health issues. Notably, Wang et al. showed that between 2011 and 2017, a rise in food insecurity correlated with elevated cardiovascular mortality rates among nonelderly adults. This connection remained significant after accounting for shifts in demographics, economic conditions, and healthcare accessibility (Wang et al.). The link between food insecurity and increased cardiovascular mortality could be the effect food insecurity has on the prevalence of cardiovascular risk factors like diabetes and hypertension (Wang et al.). Further, it may also have made medication adherence more challenging (Wang et al.). Between 2011 and 2017, Wang et al. demonstrated there was an association between food insecurity's increase in US counties and nonelderly adults' increase in cardiovascular mortality rates after accounting for demographic, economic, and healthcare access factors. Food insecurity may have an association with higher cardiovascular mortality through increasing the prevalence of diabetes and hypertension, which are cardiovascular risk factors; it may also complicate medication adherence (Wang et al.).

In summary, diet plays a crucial role in preventing cardiovascular events, emphasizing the importance of a high-quality diet. Because food insecurity limits access to nutritious food, it can lead to the development of cardiovascular risk factors and exacerbate cardiovascular health outcomes. Promoting healthy eating habits and addressing food insecurity are essential in mitigating the risks associated with cardiovascular disease.

Consolidation of Key Findings

A major finding across the included literature was that food bank users were often not consuming the required amounts of food and nutrients. *The Role of Food Banks in Addressing Food Insecurity: A Systematic Review* stated studies using 24-hour diet recall saw the

underconsumption of required amounts of fruits, vegetables, meat or meat alternatives, and milk among almost all participants, while several also saw underconsumption of legumes (Bazerghi et al.) The same review also stated that through reviewing food-bank-provided food, only one study concluded the food was adequate, whereas others found the food to be inadequate (Bazerghi et al.). Among the studies that found food provided by food banks to be inadequate, requirements for micronutrients, such as vitamins A and C, milk, meat, and meat alternatives were not met, and a sufficient quantity of food for one week was not found to be provided, revealing that in many cases food banks were not able to meet the nutritional needs of their users (Bazerghi et al.).

Similarly, *The nutritional quality of food parcels provided by food banks and the effectiveness of food banks at reducing food insecurity in developed countries: a mixed‐method systematic review* mentions that one study that evaluated the nutritional quality of food parcels "reported a mean Nutrient Density Score of 0.3 (SD 0.3)," suggesting food parcels are deficient in important nutrients (Oldroyd et al.). This review also states that one study found that only 25 percent of food bank parcels met the requirements for fruits and vegetables, which is notable because fruits and vegetables are necessary components for heart-healthy diets (Oldroyd et al.). Food banks did, however, increase the fruit and vegetable intake of their users, with this same review stating that average fruit and vegetable servings per person increased from 0.22 to 3.33 in the pre-post study (Oldroyd et al.).

This review also included that one study stated carbohydrates contributed the greatest proportion of energy in the parcels (62.2%); this proportion is considerably higher than the recommended proportion of carbohydrates (50%) (Oldroyd et al.). Furthermore, between the two studies that reported the sugar content of the parcels, both found sugar recommendations to be exceeded (Oldroyd et al.). These points are notable because disproportionate consumption of carbohydrates and sugar may lead to an increased risk of heart disease, diabetes, and poor cardiovascular outcomes.

This review was unable to determine a decisive conclusion about the impact of food banks on health outcomes. However, it was found food banks could address food insecurity, improve dietary intake, and make a slight impact on specific health markers. *The nutritional quality of food parcels provided by food banks and the effectiveness of food banks at reducing food insecurity in developed countries: a mixed‐method systematic review* stated studies that evaluated the effect of food banks on insecurity saw improvements in food insecurity after six months, showing that over time, food banks were able to ameliorate some food insecurity. Further, this review found a positive correlation between the quality of food parcels and the quality of users' diets, meaning that if food banks provided high-quality food to their users, the quality of users' diets was also high, demonstrating how food banks could help improve clients' diets. For example, by improving food parcels' nutritional quality through increasing fruits and vegetables and removing nutrient-poor snacks, the intake of vitamin C, potassium, and fruits and vegetables was improved significantly (Oldroyd et al.).

Moreover, food banks' diabetes interventions were able to improve dietary intake, showing that food bank initiatives targeted specifically at diabetics were also successful. In

summary, after assessing dietary quality compared to no intervention in controls, a considerable improvement was observed through interventions (Oldroyd et al.). Also importantly, two studies in Figure 3 of *Interventions to Address Food Insecurity Among Adults in Canada and the US A Systematic Review and Meta-analysis,* show that food insecurity interventions led to significant changes in glucose levels, highlighting that food banks are able to affect positive health outcomes in their users (Ferrer et al.; Seligman, Lyles, et al.).

Conclusions

Based on the literature, food banks have the potential to address and improve specific cardiovascular health markers but are not currently able to reduce the risk of long-term cardiovascular diseases. One review stated that food banks play a significant role in providing short-term solutions to exacerbated food insecurity, but they have limitations in their ability to improve overall food security outcomes because of lacking amounts of nutrient-dense foods, especially vegetables, fruits, and dairy, indicating that food banks are able to help keep people fed in emergencies, but the quality and nutritional value of the available foods limit their ability to serve their users (Bazerghi et al.).

This study also suggests that because food banks cannot completely ameliorate short or long-term food insecurity and cannot meet nutritional requirements for their users, they may actually contribute to users' poorer health outcomes. (Bazerghi et al.).

However, there is evidence that food banks have the potential to improve health markers, including Hb1Ac. One included paper showed food banks' diabetes interventions improved food insecurity and diet, and improved health outcomes such as HbA1c levels, although it is also stated that the studies' findings cannot be generalized to all food bank users and that findings did not always reach significance (Oldroyd et al.). Further, a separate review stated among the 15 studies encompassing clinically significant health outcomes, the ones focusing on correlations with HbA1c exhibited the most robust evidence, and while the association was not statistically significant, a possible association with HbA1c reduction was indicated by the pooled analysis, further suggesting that food banks may have the ability to improve health outcomes, especially diabetes (Oronce et al.).

It has been established that food banks have a strong potential to improve cardiovascular health markers and reduce the cardiovascular risk factors of their users, most notably through improving diabetes markers. However, there are some factors preventing food banks from best serving their clients and most effectively improving their health outcomes. For example, there are some groups that are often overlooked. One such group is those who cannot cook food; a user stated, "Well they give ya a lot of rice and they give ya a lot of vegetables… but if you're homeless like me it is not necessarily easy to cook those types of food" (Douglas et al.). This shows that food banks sometimes do not consider those who are unable to cook their food, causing even nutritious food that must be cooked to be unhelpful. Another group that is sometimes overlooked is those with religious restrictions. One user stated, "I found some foods that I didn't really use because I am Hindu and we didn't eat beef and sometimes like canned

food or the noodles or whatever they have beef flavours," showing that the food provided by food banks may not be suitable for all religious groups (Lee et al.). A third group that is also often not considered is those with medical dietary needs, including those with diabetes. One food bank client stated "Since I became diabetic I can't eat 90% of the stuff they have there so it's not been that much help to me. The stuff they give out, I can't have" (Enns et al.). Thus, food banks are not always catering to their clients' medical needs, with some being unable to use the provided resources.

These situations highlight some of the overarching areas of improvement for food banks. One review remarks that food banks can improve food security outcomes when perishable food groups are available, when operational resources are sufficient, and when the needs of users are recognized and addressed (Bazerghi et al.) In addressing these issues, food banks will improve the food security of their clients, therefore helping improve their health outcomes. The same review identifies three main hurdles food banks face in meeting the needs of their users: that food bank clients are increasing in number, that donations are not increasing accordingly or are not appropriate for the food bank's needs, and that enough nutritional training to provide education and advice to users is not provided to food bank staff (Bazerghi et al.).

To address these issues, this paper proposes some possible solutions. First, food banks can provide clients with increased food choices. This will allow users to select the foods that are best suited to their needs and preferences. Continuing on this idea, food banks can tailor their offerings to better suit their clients' needs. They can consider dietary restrictions, whether for religious, medical, or other reasons and provide food that accommodates these needs, allowing all users to be served effectively by food banks. Similarly, they can provide options that do not need to be cooked to better serve those who are unable to cook their meals. They can also increase staff training to ensure staff are equipped to give advice on nutrition and diet. Such an initiative would help educate food bank users about eating food that is best for their heart and overall health, which would help improve cardiovascular health outcomes and reduce the risk factors of food bank clients. Finally, to make the mentioned suggestions possible, it is imperative that more funds are allocated to food banks. This would ensure they have the resources to best serve their populations.

Works Cited

"About Stroke | Cdc.Gov." *Centers for Disease Control and Prevention*, 4 May 2023, https://www.cdc.gov/stroke/about.htm.

Bazerghi, Chantelle, et al. "The Role of Food Banks in Addressing Food Insecurity: A Systematic Review." *Journal of Community Health*, vol. 41, no. 4, Aug. 2016, pp. 732–40. *DOI.org (Crossref)*, https://doi.org/10.1007/s10900-015-0147-5.

Dehghan, Mahshid, et al. "Relationship Between Healthy Diet and Risk of Cardiovascular Disease Among Patients on Drug Therapies for Secondary Prevention: A Prospective Cohort Study of 31 546 High-Risk Individuals From 40 Countries." *Circulation*, vol. 126, no. 23, Dec. 2012, pp. 2705–12. *DOI.org (Crossref)*, https://doi.org/10.1161/CIRCULATIONAHA.112.103234.

Diabetes, Heart Disease, & Stroke - NIDDK. https://www.niddk.nih.gov/health-information/diabetes/overview/preventing-problems/heart-disease-stroke. Accessed 7 Aug. 2023.

Douglas, Flora, et al. "Resourcefulness, Desperation, Shame, Gratitude and Powerlessness: Common Themes Emerging from A Study of Food Bank Use in Northeast Scotland." *AIMS Public Health*, vol. 2, no. 3, 2015, pp. 297–317. *PubMed*, https://doi.org/10.3934/publichealth.2015.3.297.

Enns, Aganeta, et al. "Experiences of Food Bank Access and Food Insecurity in Ottawa, Canada." *Journal of Hunger & Environmental Nutrition*, vol. 15, May 2020, pp. 1–17. *ResearchGate*, https://doi.org/10.1080/19320248.2020.1761502.

"Facts About Hypertension | Cdc.Gov." *Centers for Disease Control and Prevention*, 6 July 2023, https://www.cdc.gov/bloodpressure/facts.htm.

Ferrer, Robert L., et al. "Primary Care and Food Bank Collaboration to Address Food Insecurity: A Pilot Randomized Trial." *Nutrition and Metabolic Insights*, vol. 12, 2019, p. 1178638819866434. *PubMed*, https://doi.org/10.1177/1178638819866434.

Gundersen, Craig, and James P. Ziliak. "Food Insecurity And Health Outcomes." *Health Affairs*, vol. 34, no. 11, Nov. 2015, pp. 1830–39. *DOI.org (Crossref)*, https://doi.org/10.1377/hlthaff.2015.0645.

"Heart Attack Symptoms, Risk Factors, and Recovery | Cdc.Gov." *Centers for Disease Control and Prevention*, 12 July 2022, https://www.cdc.gov/heartdisease/heart_attack.htm.

"Hunger." *Food and Agriculture Organization of the United Nations*, http://www.fao.org/hunger/en/. Accessed 6 Aug. 2023.

Hunger and Poverty in America | Feeding America. https://www.feedingamerica.org/hunger-in-america/poverty. Accessed 7 Aug. 2023.

Lee, Sarah, et al. "Exploring the Experience of Food Insecurity among University Students Caring for Children: A Qualitative Descriptive Study." *Journal of Hunger & Environmental Nutrition*, vol. 15, Dec. 2018, pp. 1–12. *ResearchGate*, https://doi.org/10.1080/19320248.2018.1557093.

Nations, United. "Universal Declaration of Human Rights." *United Nations*, United Nations,

https://www.un.org/en/about-us/universal-declaration-of-human-rights. Accessed 7 Aug. 2023.

Oldroyd, Lucy, et al. “The Nutritional Quality of Food Parcels Provided by Food Banks and the Effectiveness of Food Banks at Reducing Food Insecurity in Developed Countries: A Mixed-Method Systematic Review.” *Journal of Human Nutrition and Dietetics*, vol. 35, no. 6, 2022, pp. 1202–29. *Wiley Online Library*, https://doi.org/10.1111/jhn.12994.

Oronce, Carlos Irwin A., et al. “Interventions to Address Food Insecurity Among Adults in Canada and the US.” *JAMA Health Forum*, vol. 2, no. 8, Aug. 2021, p. e212001. *PubMed Central*, https://doi.org/10.1001/jamahealthforum.2021.2001.

Seligman, Hilary K., Courtney Lyles, et al. “A Pilot Food Bank Intervention Featuring Diabetes-Appropriate Food Improved Glycemic Control Among Clients In Three States.” *Health Affairs (Project Hope)*, vol. 34, no. 11, Nov. 2015, pp. 1956–63. *PubMed*, https://doi.org/10.1377/hlthaff.2015.0641.

Seligman, Hilary K., Barbara A. Laraia, et al. “Food Insecurity Is Associated with Chronic Disease among Low-Income NHANES Participants.” *The Journal of Nutrition*, vol. 140, no. 2, Feb. 2010, pp. 304–10. *PubMed Central*, https://doi.org/10.3945/jn.109.112573.

USDA ERS - Key Statistics & Graphics. https://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-u-s/key-statistics-graphics/. Accessed 7 Aug. 2023.

Wang, Stephen Y., et al. “Food Insecurity and Cardiovascular Mortality for Nonelderly Adults in the United States From 2011 to 2017: A County-Level Longitudinal Analysis.” *Circulation: Cardiovascular Quality and Outcomes*, vol. 14, no. 1, Jan. 2021, p. e007473. *DOI.org (Crossref)*, https://doi.org/10.1161/CIRCOUTCOMES.120.007473.

Why has the United States threatened to withdraw support to the UNSC Resolution 1820? A Feminist Perspective By Helen Bai

Abstract

On June 19, 2008, the UN Security Council Resolution 1820 condemning sexual violence as an act of war was passed unanimously with 15 votes; it declared that rape and other forms of sexual violence can constitute war crimes or a constitutive act with respect to genocide, marking a large step forward for female representation on an international stage. However, the US threatened opposition to the resolution, causing it to lose a key element of its structure: the establishment of a formal regulation system. The results of the debate raises many questions on the intent of the US opposition against such a strongly supported resolution and the extent of its power within the UN. The foregoing begs the question how do we understand the US threat to withdraw its support from this resolution. How we understand this question is critical to our perspectives on the future of Resolution 1820 and subsequent similar resolutions.

Introduction

United Nations Security Council (UNSC) Resolution 1820 condemning sexual violence as an act of war was passed unanimously with 15 votes. The Security Council demanded an immediate and complete cessation of sexual violence against civillians by all parties. As an extension of the previous statement on sexual violence in times of war, Resolution 1325, the resolution served as reinforcement, intending to strengthen the already existing resolution. Resolution 1820 declared that rape and other forms of sexual violence can constitute war crimes or a constitutive act with respect to genocide, marking a large step forward for female representation on an international stage. The fifteen-member Council adopted the resolution without conflict, affirming the Council's intention to establish state-specific regimes, imposing targetted measures against warring factors who has committed sexual violence against female civilians.

However, the United States (US) threatened opposition to the resolution, causing it to lose a key element of its structure: the establishment of a formal regulation system. Previous discussion within the Council affirmed the Council's readiness to address systematic sexual violence directed at female civilians as part of widespread campaign against civilians. According to a press release by the United Nations Security Council, the Council demands immediate actions to protect civilians, including but not limited to military disciplinary measures. Of the permanent members, the US was the sole state to raise opposition to the resolution, claiming that the resolution carried insensitive language relating to gender and sexual health. As a result of the threat of a veto, the Security Council was forced to limit regulation policies to nothing more than a stern warning. The results of the debate raises many questions on the intent of the US opposition against such a strongly supported resolution and the extent of its power within the UN. Although Resolution 1820 was successfully passed, the US opposition raises questions abouts its intentions in such an action and how domestic American policies may reflect the same

idea. In light of the recent overturn of Roe vs. Wade, a Supreme Court case protecting abortion as a federal right, it is clear that US policy surrounding women's right is shifting to a conservative stance, as opposed to progressive liberalism.

The US publicly announced that the withdrawal of support was a result of the 'insensitive' language reflected in the resolution. Zalmay Khalilzad, US ambassador, stated that Resolution 1820's use of language surrounding reproductive health was "largely concerning", hence the US opposition. However, one of the conditions that the US posed for the resolution was to eliminate the establishment of a formal regulation system. This action raises further questioning of US intent, as officials failed to offer explanation for the opposition of regulation. In 2019, the Trump administration has drawn a harsh line on using phrases that could be seen as a cover-up for 'liberal promotion'. Such ideals may have largely contributed to the overturn of Supreme Court case Roe v Wade, demolishing the federal right to abortion. For a country so keen on protecting the rights and liberties of all, the US has largely watered-down a resolution meant to provide services and "support for survivors of sexual violence". The US portrays itself as the 'protector' of life, liberty, and property through its Constitution and the 'American democracy'. Throughout history, the US has illustrated itself in a progressive light, advertising its protection of civil liberties for global citizens. The withdrawal of US support for Resolution 1820 exhibits a complete turnaround in its foreign policy on the protection of women's rights in times of war, turning liberty into a weapon to force other states into political compliance. This intervention during Trump's presidency precedes the overturning of Roe vs Wade, but there is little work examining what this interference might reveal about the place of women in domestic politics. Beyond Trump, there seems to be a close connection between the US opposition of Resolution 1820 and the overturn of Roe vs Wade. Both cases involve what authorities describe to be 'indecent language' surrounding female reproduction health, and conclude with the limitation of women's rights in exchange for political leverage. The foregoing begs the question: How do we understand the US threat to withdraw its support from Resolution 1820, and what that might mean for women beyond the stage of global politics. How we understand this question is critical to our perspectives on the future of Resolution 1820 and subsequent similar resolutions.

This paper explores this question in three main sections. It will begin with accounting for the place of women in American foreign policy and how that correlates with the position of women within American domestic politics. An analysis of this connection will proceed, highlighting how American performance on the international stage reflects its domestic policies. Lastly, the implications this holds for women will be investigated and a conclusion will be drawn. My analysis will draw mainly on feminist international studies and security studies frameworks.

Accounting for the place of women in American foreign policy

America's withdrawn support highlights an over-debated occurrence while distracting the public eyes from women's rights on the global scale. This opposition shifts political attention

from an under-engaged set of issues surrounding women's rights to the over-engaged theme of US dominance in global politics. Furthermore, this withdrawn support continues to be engaged as different to the US's domestic policies surrounding women's rights when it is largely the same. The resolution on sexual violence in armed conflict was also forced to omit references to sexual and reproductive health after the US threatened to exercise its veto power, illustrated in Adamczyk's policy brief (2019). Adamczyk describes the US opposition to stem from a protective perspective over reproductive health. However, questions have been raised about what exactly about a regulation system triggers such opposition.

The US opposition might have come from various intentions, but the consequences of such actions are the same. Puechgulrbal (2010) argues that the perpetuation of stereotyping language in these documents removes women's agency and maintains them in the subordinated position of victims. The institution of the UN leaves the male monopoly of power unchallenged and presents gender mainstreaming as a non-political activity. However, while this patriarchal monopoly claims to view gender mainstreaming as "non-political", it is often used on a political stage for leverage and power, as demonstrated by this resolution. The Council is beginning to recognize and constitute women as agential subjects and suggest that the fragmented and mutable representations of women in Council resolutions offer a unique opportunity for critical engagement with what 'women' might be, do, or want in the field of gender and security (Shepherd 2011). This usage of women's rights as a chip on a bargaining table reveal's the Council and US's intention of objectifying protection against sexual violence.

Utilizing women's rights in international politics is a problem that eclipses the challenge with UN resolutions and women's inclusion in contexts of violence in war and peacebuilding. These already under-examined issues run the risk of disappearing completely with the rapid shift of attention from policy to resistance. The experience since passage of UNSC Resolution 1325 has shown that UN resolutions alone do not guarantee women's effective inclusion in peacebuilding (Sharp 2015). Rather than helping the advocacy of women's rights, the UNSC has enabled other nations to use these rights as 'sticks' to impose political power. Instead of including women's voices in the debate, women have been added as a bargaining chip to the poker table. Not only so, but Macfarlane (2021) argues that Security Council Resolutions are not an effective method to prevent conflict-related sexual violence. In fact, it may do more harm than good to the efforts to end wartime sexual violence. Once the notion of sexual violence being used for political power has been raised, more countries will abuse this policy in order to gain political power on an international stage. UNSC Resolutions are highly limited in their regulation policies. Though these promises sound good in theory, there is not much done in practice. Ahmed (2011) argues that violence against women is embedded in structural factors, which exist in peacetime and only become magnified during periods of armed conflict.

Furthermore, such a resolution has framed sexual violence as a threat to international security. Stabel (2013) utilizes Finnemore & Sikkink's theory on normative change in her research to provide insight in how and why sexual violence would be framed as a matter of international security. Issues failing within the Security Council's mandate as threats to women,

to state, and to the international community. Rather than choosing to resolve the threat of sexual violence against women in times of war, leading nations of the UN reflect this attention to the fight for political power. By ignoring outcries for protection against sexual violence, the US is able to twist the battle for women's rights into a threat to the rest of the UNSC. This issue demonstrates the US perspective on women's rights, as it directly reflects current US domestic policies. The abuse of power by the US not only threatens the realization of women's rights, but extends that threat to the rest of the international community. US officials state that the reason for heavy opposition on matters of women's rights is a result of language used on reproductive health. However, instead of choosing to resolve this minor issue, the US instead turns Resolution 1820 and 1325 into a political weapon. The UNSC's inability to respond escalates this into an international crisis, taking away what little attention was previous focused on the matters of women's protection against sexual violence.

Women in American Policy

The foregoing demonstrates the dominant strategies governing women's rights in international politics and domestic American policy. However, the false contrast between American foreign and domestic policies are even more apparent when women's issues are foregrounded. The logics underpinning American foreign policy are the very same ones informing domestic policies. Both advocate strongly for women's rights, but lack any progressive action. Domestic strategies to maintain a 'home-front' focus for activists means that these activists are blinded to processes marginalizing women in international politics. Failure to comprehend these continuities between domestic and foreign policies limits action purely to the domestic sphere.

American values on women's rights typically take a liberal stance. The US has long prided itself on its advocacy for equality and freedom, but how much of that advocation has come to fruition? O'Conner (1996) discusses the Women's Suffrage Movement, and how it fought for women's right to vote in the US. In her paper, O'Conner explored the history of this movement, tracing it all the way back to the Seneca Falls Convention in 1848. This movement concluded with the passing of the 19th Amendment to the US Constitution, granting women the right to vote. However, for many years after that, women were severely criticized for dabbing in politics, and even led to the second rise of the Cult of Domesticity, the belief that a woman's job was in the home, taking care of her husband and children. What seemed like a step forward in women's rights actually led to more social pressure for women to leave the workforce and become housewives. Another social rights movement that led to the expansion of women's rights was the Second Wave of Feminism. This was a sexual and political revolution, where the presence of women in workplaces and social establishments became more common and accepted. The Second Wave of Feminism movement laid a foundation for women to be accepted as full citizens of the US, not just legally, but socially as well. The US was a country built on protests, and women's rights are not exempt from that continuation.

The history of activism geared toward the recognition and acceptance of American women as full citizens with full participation in politics is often referred to when engaging with threats to these political gains in contemporary politics. However, this limited approach focuses entirely on domestic politics following the disjunction between the domestic and the international. Failure to engage with both 'fronts' when facing the issue of women's rights, means that activists and scholars are restricted in the extent to which they can track the continuities in American policy regarding women's rights. The US's withdrawal of support for UNSC Resolution 1820 happened more than three years before the overturning of Roe vs Wade. This occurrence heavily foreshadowed the impending loss on women's rights policies, but did not receive nearly as much media attention as it should have.

The false contrast between American foreign and domestic policies becomes even more apparent when women's issues are brought to the forefront. While there may be rhetoric and advocacy for women's rights in both realms, the underlying logics that inform American foreign policy are the same ones shaping domestic policies. Both domains claim to advocate strongly for women's rights, but in practice, they often lack progressive action. Domestic strategies aimed at maintaining a "home-front" focus for activists can inadvertently blind them to the processes marginalizing women in international politics. By failing to comprehend the continuities between domestic and foreign policies, activists primarily concentrate their efforts within the domestic sphere. This limited perspective hampers their ability to fully understand the challenges faced by women in international contexts and to address the broader implications of policies that undermine women's rights globally.

American values on women's rights have typically taken a liberal stance, and the country has long prided itself on its advocacy for equality and freedom. The Women's Suffrage Movement, discussed by O'Conner (1996), fought for women's right to vote in the US, culminating in the passage of the 19th Amendment to the US Constitution, granting women suffrage. However, despite this apparent progress, women continued to face criticism for their involvement in politics for many years afterward. In fact, the movement led to the second rise of the Cult of Domesticity, reinforcing the belief that a woman's role was limited to the home, caring for her husband and children. Thus, what initially seemed like a step forward in women's rights actually resulted in increased social pressure for women to withdraw from the workforce and assume traditional domestic roles.

Another significant social rights movement that contributed to the expansion of women's rights was the Second Wave of Feminism. This movement, characterized by a sexual and political revolution, facilitated the increasing presence and acceptance of women in workplaces and social establishments. The Second Wave of Feminism laid a foundation for women to be recognized as full citizens of the United States, not only in a legal sense but also socially. It was a testament to the fact that the US, as a country built on protests and social change, has continuously seen activism as a means to advance the rights of various marginalized groups, including women. However, when engaging with contemporary threats to these political gains, the history of activism often focuses solely on domestic politics, overlooking the disjunction

between the domestic and international spheres. By failing to engage with both fronts, activists and scholars restrict their understanding of the continuities and intricacies of American policy concerning women's rights. This limited approach hinders their ability to track the implications of policies across both the domestic and international arenas.

The US's withdrawal of support for UNSC Resolution 1820, which occurred more than three years prior to the overturning of Roe v. Wade, serves as a significant precursor to the erosion of women's rights. Unfortunately, this event did not receive the attention it deserved from the media and broader public discourse. Recognizing the interconnectedness of domestic and international policies on women's rights is crucial for anticipating and addressing potential threats to these rights. It requires a comprehensive understanding of the continuities in American policy, beyond a narrow focus on domestic politics alone.

Conclusion

Had activists within the US focused some of their attention on global politics rather than strictly following the ‘home-front’, the overturning of Roe vs. Wade could have been predicted. What transpires is that chipping away at women’s rights on an international stage is a space for rehearsing similar political decisions at home. The US had allowed for the overturn of Roe vs. Wade because similar action on the international level did not raise much concern. Had more resistance risen at the news of US opposition to a resolution on the protection of sexual violence, perhaps Roe vs Wade would not have the capacity to be overturned.

The United States' opposition to United Nations Security Council Resolution 1820, which condemns sexual violence as an act of war, raises significant questions about its intentions and the correlation between its foreign and domestic policies on women's rights. Despite the resolution being passed unanimously by the Security Council, the US threatened to veto it, leading to the removal of a formal regulation system from the resolution. The US cited concerns about language related to gender and sexual health, but the conditions it posed for the resolution, including the elimination of a regulation system, raise doubts about its true intentions. The opposition by the US reflects a shift in its domestic policies regarding women's rights, as evidenced by the recent overturn of the Supreme Court case Roe v. Wade, which protected abortion as a federal right. The US, traditionally seen as a progressive advocate for civil liberties, has shifted towards a conservative stance on women's rights. This raises concerns about the US's commitment to protecting women's rights, both domestically and internationally. Analyzing the place of women in American foreign policy reveals a problematic trend of using women's rights as a bargaining chip for political power on the global stage. The US opposition to Resolution 1820 undermines the efforts to address systematic sexual violence against women in times of war and peacebuilding. Instead of promoting women's agency and protection, the US turns women's rights into a political weapon, jeopardizing the realization of these rights and threatening the international community's commitment to addressing sexual violence.

The connection between American foreign and domestic policies on women's rights becomes more apparent when focusing on women's issues. American values, which espouse

equality and freedom, often do not translate into tangible progress for women's rights. Historical movements such as the Women's Suffrage Movement and the Second Wave of Feminism have brought important changes, but the pressure to conform to traditional gender roles and the recent erosion of women's rights demonstrate the ongoing challenges faced by women in American society. The limited engagement of activists and scholars with both domestic and international arenas hinders their understanding of the continuities in American policy on women's rights. The US's withdrawal of support for Resolution 1820 served as a precursor to the overturning of Roe v. Wade, but received insufficient attention. By neglecting the international dimension, activists missed an opportunity to predict and resist the erosion of women's rights domestically.

In conclusion, the US opposition to Resolution 1820 and its subsequent actions highlight the interplay between American foreign and domestic policies on women's rights. The use of women's rights as a bargaining chip for political power on the international stage reflects a troubling trend. By undermining the global efforts to protect women from sexual violence, the US raises concerns about its commitment to women's rights. It is crucial to understand the connections between the US's actions in international politics and its domestic policies, as this understanding shapes perspectives on the future of resolutions like 1820 and their implications for women's rights worldwide. Activists and scholars must broaden their focus to encompass both domestic and global dimensions to effectively advocate for women's rights and address the challenges faced by women in contemporary society.

Works Cited

Black, Renee. "Mainstreaming Resolution 1325 on Women, Peace, and Security." *Potentia: Journal of International Affairs*, vol. 1, 2009, pp. 75–89, https://doi.org/10.18192/potentia.v1i1.4369.

Chomsky, Noam. *Hegemony or Survival?* Hamish Hamilton, 2003.

"THE EDIBLE WOMAN: THE PSYCHOLOGY OF EARLY SECOND-WAVE FEMINISM." *Margaret Atwood*, 2007, pp. 9–34, https://doi.org/10.1163/9789401204545_003.

Finnemore, Martha, and Kathryn Sikkink. "International Norm Dynamics and Political Change." *International Organization*, vol. 52, no. 4, 1998, pp. 887–917, https://doi.org/10.1162/002081898550789.

Fitzpatrick, Brenda. "After Security Council resolution 1325." *Tactical Rape in War and Conflict*, 2016, pp. 181–208, https://doi.org/10.1332/policypress/9781447326694.003.0008.

Gordon, Linda. "Socialist Feminism." *New Labor Forum*, vol. 22, no. 3, 2013, pp. 20–28, https://doi.org/10.1177/1095796013499736.

Herman, Edward S. *Manufacturing Consent: The Political Economy of the Mass Media*. Createspace Independent Publishing Platform, 2017.

"II.B.17 United Nations Security Council Resolution 1820 (On Acts of Sexual Violence against Civilians in Armed Conflicts) (19 June 2008)." *International Law & World Order*, pp. 1–4, https://doi.org/10.1163/ilwo-iib17.

Kosinski, Michelle, and Eli Watkins. "US Threatens to Veto UN Resolution on Helping Rape Victims, over Mentions of Reproductive Health | CNN Politics." *CNN*, Cable News Network, 23 Apr. 2019, www.cnn.com/2019/04/22/politics/us-united-nations-sexual-violence-resolution/index.ht

Larson, T.A. "The Woman Suffrage Movement in Washington." *Women Suffrage*, 1994, pp. 318–351, https://doi.org/10.1515/9783110971064.318.

Macfarlane, Emma. "Resolutions Without Resolve: Turning Away from UN Security Council Resolutions to Address Conflict-Related Sexual Violence." *Michigan Journal of Gender & Law*, no. 27.2, 2021, p. 435, https://doi.org/10.36641/mjgl.27.2.resolutions.

Pratt, Nicola. "Reconceptualizing Gender, Reinscribing Racial-Sexual Boundaries in International Security: The Case of UN Security Council Resolution 1325 on 'Women, Peace and Security'1." *International Studies Quarterly*, vol. 57, no. 4, 2013, pp. 772–783, https://doi.org/10.1111/isqu.12032.

Ruby, Felicity. "Security Council Resolution 1325." *Rethinking Peacekeeping, Gender Equality and Collective Security*, https://doi.org/10.1057/9781137400215.0018.

Sharp, Ingrid. "Feminist Peace Activism 1915 and 2010: Are We Nearly There Yet?" *Peace & Change*, vol. 38, no. 2, 2013, pp. 155–180, https://doi.org/10.1111/pech.12010.

Twenty Years of Protection of Civilians at the UN Security Council - ODI, cdn.odi.org/media/documents/12709.pdf. Accessed 20 Sept. 2023

The Nonlinear Dynamics of a Ball Bouncing on a Vibrating Ferrite Rod

By Ke Li and Kangqi Liao

Abstract

This paper investigates the dynamics of a ball bouncing on a ferrite rod subjected to an alternating magnetic field. Employing linear magnetostriction theory, we establish the equation governing the vibration of the ferrite rod. Subsequently, we scrutinize the impact of frequency-doubling phenomena on the ball's motion and analytically deduce the probability distribution function (PDF) of the ball's rebound velocity after collision. To testify the theoretical analysis, numerical simulations are executed to simulate the ball's trajectory. Furthermore, Neural Network curve fitting algorithm is employed to approximate the PDF of the rebound velocity based on the simulation results. Our analytical PDF consistently predicts the peak of the rebound velocity distribution, albeit exhibiting slight disparities with the PDF obtained via numerical simulations. Computation results also show that the ball would eventually fall into "Locking Region", exhibiting periodic motion. Experiments are also conducted, and the resulting PDF follows the general pattern of the analytically deduced function at high velocities but deviates slightly from the theoretical analysis at low velocities.

1. Introduction

Applying an alternating current (AC) to a coil wrapped at the lower end of a ferrite rod, the rod will evince deformation in alignment with the magnetic field—a phenomenon recognized as the magnetostriction effect. When the AC current frequency approaches the natural frequencies of the ferrite rod, if a ball is placed at the upper end of it, it would start to bounce. Subsequent interactions between the ball and the rod engender an intricate rebounding pattern.

Despite the apparent simplicity of the dynamic system, it evinces an extraordinary susceptibility to initial conditions and certain randomness, which serve as indicators of chaotic motion, thereby necessitating the utilization of statistical methodologies for a quantitative characterization of this hopping dynamic. This interaction between a ball and a vibrating platform has been studied thoroughly [1–4], including the fixed points, strange attractors, Lyapunov exponents and so on. Previous works [5,6] have shown both experimentally and theoretically that models of thermodynamic and statistical physics could be employed to study the behavior of granular materials subjected to vertical vibration. The work of Warr et al [7] investigates the motion of a particle bouncing on a sinusoidally vibrating plane. They established that the probability distribution of the rebound velocity of the ball after each impact follows a Boltzmann distribution using generalized Langevin Equation. However, the previous works only focused on sinusoidally vibrating platforms. When addressing the scenario of a ball rebounding on a vibrating ferrite rod, the anisotropic characteristics of the ferrite material engender a frequency doubling effect. This effect introduces additional frequency components, which is often just one component twice as large as the excitation frequency, alongside the fundamental excitation frequency arising from the alternating current.

This study investigates the dynamics of a ball bouncing on a ferrite rod undergoing vibration under the influence of frequency doubling effects. The study derives the probability distribution function (PDF) of the ball's rebound velocity and testifies it with computer simulation and Neural Network curve fitting. Interestingly, the computational simulations presented in this paper also show that the ball would eventually enter the "Locking Region" [4], after several hundred to several thousand collisions, thereby manifesting periodic motion, whose period is an infinite series of bounces in a finite amount of time. It is also found experimentally that the general pattern of the analytically deduced PDF suits the experimental results well at high velocities but deviates slightly from experiments at low ones.

2. Experimental Apparatus

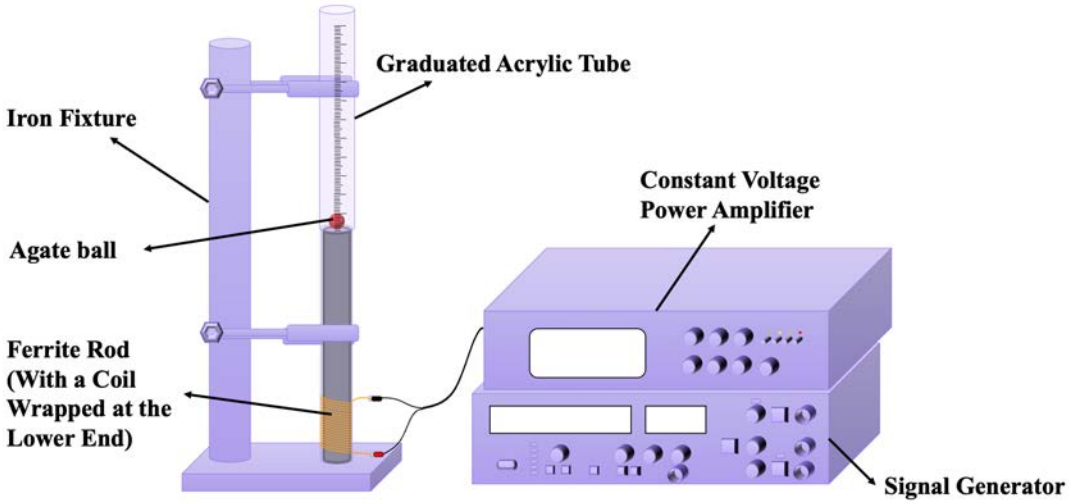


Fig 1: Experimental Apparatus

In the experiment, a 1631A function signal generator was used to output a sinusoidal wave, which was then amplified using the SAST Xianke constant voltage power amplifier (240W) and inputted to the coil wound around the lower end of the ferrite rod. A small ball was initially placed at the upper end of the ferrite rod and its motion was confined by a graduated acrylic tube. When the frequency of the AC in the coil matches the natural frequencies of the ferrite rod, the ball would start to bounce.

The entire setup was positioned directly in front of a CCD camera and the motion of the small ball was recorded by it. The rebound velocity of the ball after each collision could be obtained using the relationship $v = \sqrt{2gh}$.

3. The Vibration of the Ferrite Rod

3.1 Magnetostriction

Ferrite is a ceramic material made by mixing and firing large proportions of iron (III) oxide (Fe2O3, rust) blended with small proportions of one or more additional metallic elements, such

as strontium, barium, manganese, nickel, and zinc. Magnetostriction [8] is a property of some ferromagnetic materials which causes them to expand or contract in response to a magnetic field. This effect allows the material to convert electromagnetic energy into mechanical energy. As a magnetic field is applied to the material, its molecular
dipoles and magnetic field boundaries rotate to align with the field. This causes the material to strain and elongate.

Under the magnetic field B in z direction, a stress S will be produced in the ferrite rod in the same direction according to the following equation:

$$S = \lambda H = \frac{\partial u\,(z,t)}{\partial z} \tag{1}$$

where H is the magnetic field strength, u is the displacement of each point on the rod in z direction at position z and time t, and λ the magnetostriction coefficient.

3.2 Vibration Equation

The origin of the coordinate is set at the lower end of the ferrite rod, which is wrapped by a coil with a sinusoidally alternating current in it. After ignoring the vibration in the radial direction, the equation of vibration of the ferrite rod is governed by:

$$\frac{\partial^2 u}{\partial z^2} + \lambda \frac{\mathrm{d}\overline{H}}{\mathrm{d}z} = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} \tag{2}$$

where $\overline{H}$ stands for the magnetic field in z direction averaged over the cross section of the ferrite at z which can be calculated using Biot-Savart's law, as shown in Fig. 2, and $c = \sqrt{E/\rho}$ stands for the sound velocity (E and ρ stands for the Young's modulus and density of the rod)

Generally, the equation governing the vibration of the rod under arbitrary external force is:

$$E\frac{\partial^2 u}{\partial z^2} + f(z,t) = \rho\frac{\partial^2 u}{\partial t^2} \tag{3}$$

For steady vibration under sinusoidal external forces, f and u could be reduced to

$$f(z,t) = F(z)\mathrm{e}^{\mathrm{i}\omega t} \tag{4}$$

$$u = U(z)\mathrm{e}^{\mathrm{i}\omega t} \tag{5}$$

Eq. (3) thus becomes

$$E\frac{\mathrm{d}^2U}{\mathrm{d}z^2} + F + \rho\omega^2 U = 0 \tag{6}$$

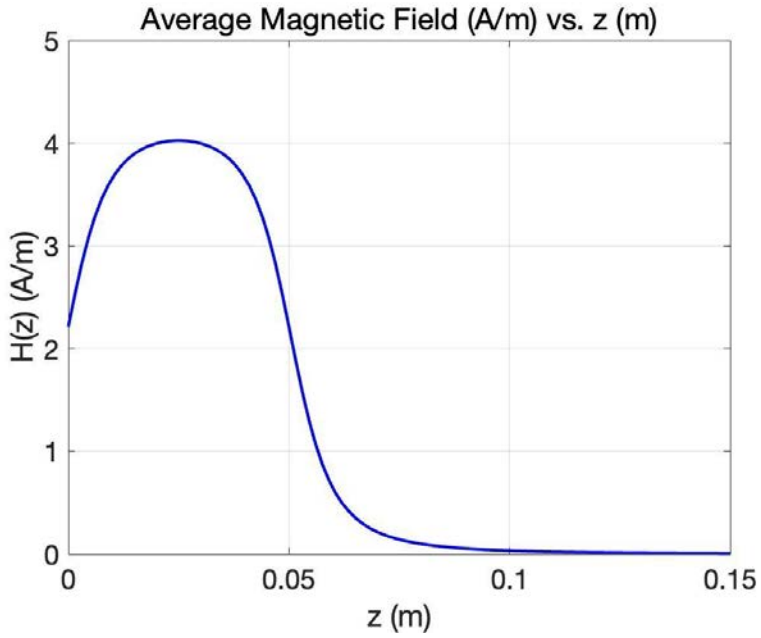


Fig.2: $\overline{H}(z)$ obtained through Biot-Savart's Law for a ferrite rod and a coil circular with circular cross sections, with magnetic permeability $\mu_r = 100$, current $I = 2.2A$, number of turns $N = 60$ radius of the rod and coil $R = 0.01$

For a ferrite rod with two free ends, the Green's function is:

$$G_0 = \begin{cases} -\frac{\cos(kz)\cos(kl-kz')}{kc^2\sin(kl)}, & 0 \le z < z' \le l \\ -\frac{\cos(kz')\cos(kl-kz)}{kc^2\sin(kl)}, & 0 \le z' < z \le l \end{cases} \tag{7}$$

where l stands for the length of the ferrite rod. By the definition of Green's function, the solution of Eq. (3) could be expressed as:

$$U = \frac{1}{\rho}\int_0^l G_0(z,z')F(z)\mathrm{d}z' \tag{8}$$

Inserting $F(z) = E\lambda\frac{d\overline{H}}{dz}$ into Eq. (8) and taking the imaginary part of the solution u, the solution could be obtained as:

$$u(z,t) = \left(\int_0^z \frac{\lambda \cos(kz') \cos(kl - kz) \frac{\mathrm{d}\overline{H(z)}}{\mathrm{d}z}}{k \sin(kl)} \mathrm{d}z' + \int_z^l \frac{\lambda \cos(kz) \cos(kl - kz') \frac{\mathrm{d}\overline{H(z)}}{\mathrm{d}z}}{k \sin(kl)} \mathrm{d}z' \right) \sin(\omega t) \tag{9}$$

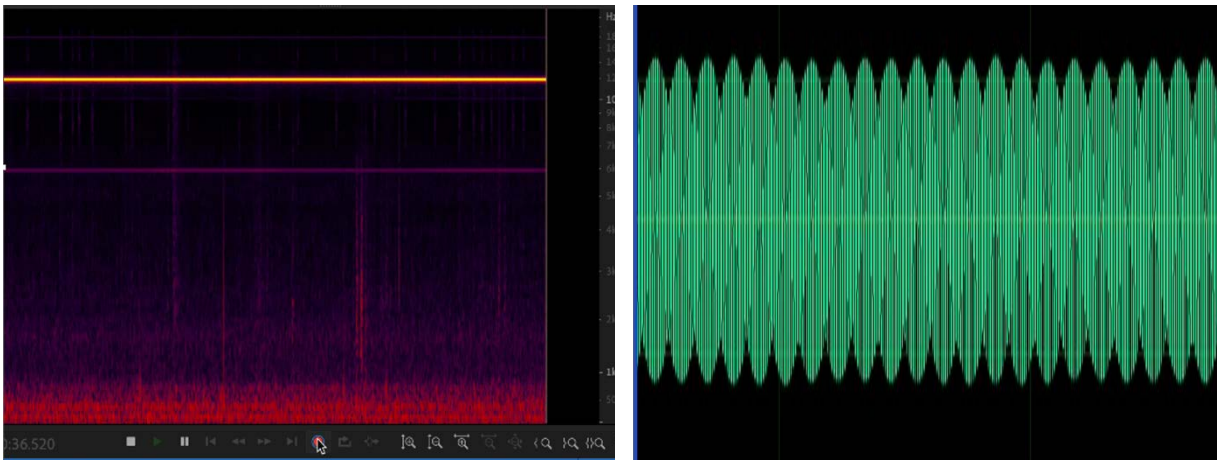
Frequencies $f_n = nc/2l$ are called natural frequencies. As the frequency approaches f_n, $kl \to n\pi$, $u \to \infty$, which is recognized as resonance. When the AC frequency approaches natural frequencies, the velocity of the vibration of the rod would be comparable to that of the falling ball.

3.3 Frequency Doubling

In Eq. (2), the vibration of a ferrite rod under alternating magnetic field is discussed under the assumption of isotropy and linearity. As indicated by Eq. (9), the vibration can be described by a sine function of form:

$$u(l,t) = A \sin(\omega t) \tag{8}$$

where $A = max\left(u|_{z=l}\right)$. However, S is often not a linear function of H (or the magnetization M) but an even function of it.



(a) (b)

Fig.3: (a) is the spectrogram of the sound produced, (b) is the sound recorded

By analyzing the sound produced by the ferrite rod recorded by microphones in experiments, it is observed that the sound emitted by the vibrating ferrite rod primarily consists of the AC frequency and even multiples of that frequency (this is also analyzed theoretically in [8]). But in most cases, only the excitation frequency and a frequency twice as

high are included, as shown by Fig.3. In these situations, the phenomenon of "beats" can be clearly observed. In this paper, we only discuss the cases in which:

$$u(l,t) = A_1 \sin(\omega t) + A_2 \sin(2\omega t) \tag{11}$$

4. The Dynamics of the Ball

4.1 The Ball's Equation of Motion

Now consider the collision between the vibrating ferrite rod and a ball in free fall. In this section, we use dimensionless units. Time is measured in units of $T = 2\pi/\omega$, acceleration in units of $g/2$, velocity in units of $\pi g/\omega$ and position in units of $2\pi^2 g/\omega^2$. Then the dimensionless height of the upper end of the ferrite rod $u(l,t)$ is:

$$h(\tau) = \frac{\Gamma_1}{2\pi}\sin(2\pi\tau) + \frac{\Gamma_2}{2\pi}\sin(4\pi\tau) \tag{12}$$

where $\Gamma_1 = A_1\omega^2/\pi g$ and $\Gamma_2 = A_2\omega^2/\pi g$. Let τ_n denote the time of the n *th* collision, $W_n > 0$ the relative velocity between the ball and the ferrite rod immediately after the n *th* collision and Δ_n the distance between the ball and the upper end of the ferrite rod after the n *th* collision. After elementary algebra, it could be obtained that;

$$\begin{aligned}\Delta_n(\tau) = {} & \frac{\Gamma_1}{2\pi}\left[\sin(2\pi\tau_n) - \sin(2\pi\tau)\right] + \frac{\Gamma_2}{2\pi}\left[\sin(4\pi\tau_n) - \sin(4\pi\tau)\right] \\ & + \left[W_0 + \Gamma_1\cos(2\pi\tau_n) + 2\Gamma_2\cos(2\pi\tau_n)\right](\tau - \tau_n) - (\tau - \tau_n)^2\end{aligned} \tag{13}$$

The next collision happens when $\Delta_n(\tau) = 0$, which means that τ_{n+1} is the smallest value greater than τ_n that satisfies $\Delta_n(\tau_{n+1}) = 0$. Once τ_{n+1} is calculated using τ_n and W_n, W_{n+1} could be obtained through the following equation:

$$\begin{aligned}W_{n+1} &= -\alpha\frac{\mathrm{d}\Delta_n}{\mathrm{d}\tau_{n+1}} \\ &= -\alpha\{W_n - 2(\tau_{n+1} - \tau_n) + \Gamma_1\left[\cos(2\pi\tau_n) - \cos(2\pi\tau_{n+1})\right] \\ &\quad + 2\Gamma_2\left[\cos(4\pi\tau_n) - \cos(4\pi\tau_{n+1})\right]\}\end{aligned} \tag{14}$$

where α is the restitution coefficient.

4.2 High Bounce Approximation

When the relative velocity W is large enough, the velocity of the vibrating platform could be ignored. This is called the High Bounce Approximation [4]. From Eq. (13), and ignoring the phase of the platform, τ_{n+1} could be approximated by:

$$\tau_{n+1} \approx W_n + \Gamma_1 \cos(2\pi\tau_n) + 2\Gamma_2 \cos(2\pi\tau_n) \tag{15}$$

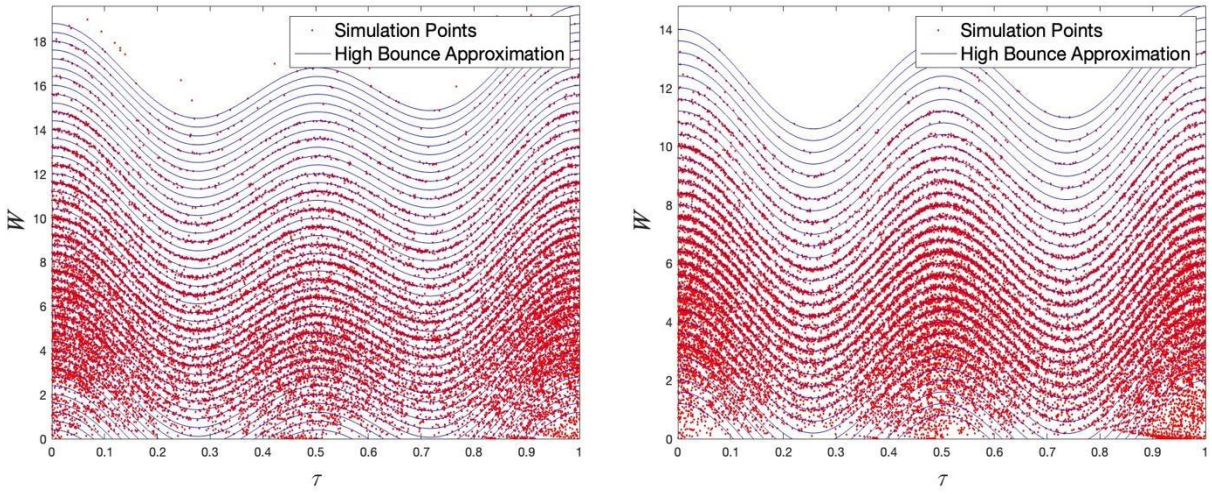
Substituting this equation to Eq. (14), W_{n+1} could be approximated by

$$W_{n+1} \approx \alpha[\tau_{n+1} - \tau_n + \Gamma_1 \cos(2\pi\tau_{n+1}) + 2\Gamma_2 \cos(4\pi\tau_{n+1})] \tag{16}$$

The High bounce approximation is valid when the velocity of the upper end of the vibrating ferrite rod is sufficiently large, so the (τ, W) for each collision would distribute around the spirals

$$W = \alpha[\frac{k}{2} + \Gamma_1 \cos(2\pi\tau) + 2\Gamma_2 \cos(4\pi\tau)] \tag{17}$$

because the rod's velocity reaches its maximum at around $\tau = \frac{k}{2}$, where k is an arbitrary integer. This pattern is clearly shown in Fig.4.



(a) $\Gamma_1 = 1.5,\ \Gamma_2 = 1$ (b) $\Gamma_1 = 0.5,\ \Gamma_2 = 1$

Fig. 4: The $(\tau\ mod\ 1,\ W)$ of 30000 collisions in each simulation is plotted on the graph above. The solid lines are the results of the high bounce approximation, the dots represent simulated collisions.According to the high bounce theory, the dots should be distributed around the solid lines.

4.3 The Probability Distribution Function of the Rebound Velocity

In this section, time and position are measured in usual SI units, seconds and meters. From Fig.4, it could be clearly seen that the dynamics of the ball bouncing on the vibrating ferrite rod exhibits certain "randomness": the red dots generated during simulations are distributed *around* the blue spirals described by Eq.(17) but not exactly *on* them. It is then reasonable According to Warr et al [7], the absolute rebound velocity of the ball after the n *th* collision, $v_n = \frac{\pi g}{\omega}\left[W_n + \Gamma_1 \cos(2\pi\omega t) + 2\Gamma_2 \cos(4\pi\omega t)\right]$, satisfies the generalized discrete Langevin equation:

$$\frac{\mathrm{d}v_n}{\mathrm{d}n} = (\alpha - 1)\, v_n + \frac{\beta}{v_n} + \phi \tag{18}$$

In analogue with the original Langevin equation that describes the movements of an ensemble of a particle, the term on the left represents the friction force, the first two terms on the right stands for a conservative force and the last term represents the white noise force. The rebound velocity v thus follows a Boltzmann distribution under the potential field $\frac{1}{2}(\alpha - 1)v^2 + \beta ln(v)$:

$$P(v) \propto v^{\frac{\beta}{kT}} \exp\left(\frac{\alpha - 1}{2k_B T} v^2\right) \tag{19}$$

where $k_B T$ and β are two constants to be determined.

In the context of a ball rebounding on a vibrating ferrite rod, the white noise ϕ stands for the completely random component of the rod's contribution to the rebounding velocity of the ball, and β/v_n is a bias term, reflecting that the velocity of the ferrite rod at collision is truly random only when v_n is infinite.

Warr et al [7] calculated that for a ball bouncing on a sinusoidally vibrating platform,

$$P(v) = \frac{2\left(\frac{2(1-\alpha)}{(1+\alpha)^2 (A\omega)^2}\right)^{\frac{3+\alpha}{2(1+\alpha)}}}{\Gamma\left(\frac{3+\alpha}{2(1+\alpha)}\right)} v^{\frac{2}{1+\alpha}} \exp\left(-\frac{2(1-\alpha)}{(1+\alpha)^2 (A\omega)^2} v^2\right) \tag{20}$$

We now expand this distribution function to the vibrating ferrite rod whose equation of vibration is $u(l,t) = A_1 \sin(\omega t) + A_2 \sin(2\omega t)$ by replacing the term $A\omega$ with $A'\omega'$.

First, consider the case where $A_1/A_2 \geq 2$, as shown in Fig.5.

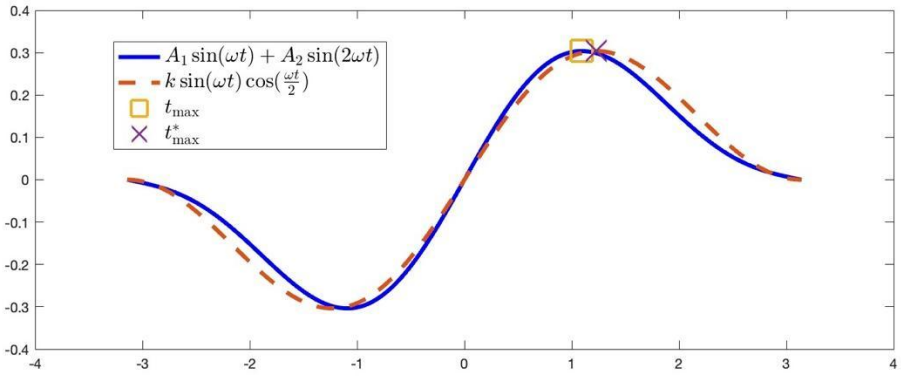


Fig.5: The wave $A_1\sin(\omega t)+A_2\sin(2\omega t)$ for $A_1/A_2 \geq 2$ within the interval $\left[-\frac{T}{2},\frac{T}{2}\right]$

The wave $A_1\sin(\omega t)+A_2\sin(2\omega t)$ could be approximated by the wave $k\sin(\omega t)\cos\left(\frac{\omega t}{2}\right)$ within the interval $[-T/2,\ T/2]$. The maximums of the two waves are the same. We now define t_{max}, $t^*_{max} \in [-T/2, T/2]$ as :

$$t_{\max}=\arg\max_t(A_1\sin(\omega t)+A_2\sin(2\omega t))=\frac{1}{\omega}\arccos(\frac{-A_1+\sqrt{A_1^2+32A_2^2}}{8A_2})$$
$$t^*_{\max}=\arg\max_t(\sin(\omega t)\cos(\frac{\omega t}{2}))=\frac{1}{\omega}\arcsin(\sqrt{\frac{1}{3}}) \quad (21)$$

If the maximums of the two waves are the same. it could be obtained that:

$$k=\frac{A_1\sin(\omega t_{\max})+A_2\sin(2\omega t_{\max})}{\sin(\omega t^*_{\max})\cos(\frac{1}{2}\omega t^*_{\max})} \quad (22)$$

It is reasonable to calculate A' and ω' as

$$A'=\frac{1}{2}\int_{\frac{-T}{2}}^{\frac{T}{2}}k\cos(\frac{1}{2}\omega t)dt=k$$
$$\omega'=\omega \quad (23)$$

For cases where $A_1/A_2 < 2$, the frequency doubling vibration of the ferrite rod could be approximated by two segments of vibration, one within $\left[-\frac{1}{4}T,\frac{1}{4}T\right]$, the other within $\left[\frac{1}{4}T,\frac{3}{4}T\right]$.

Within the interval $\left[-\frac{1}{4}T, \frac{1}{4}T\right]$, the function $A_1 \sin sin\,(\omega t) + A_2 \sin(2\omega t)$ could be approximated by $k_1 \sin(2\omega t)$, as shown by Fig.6.

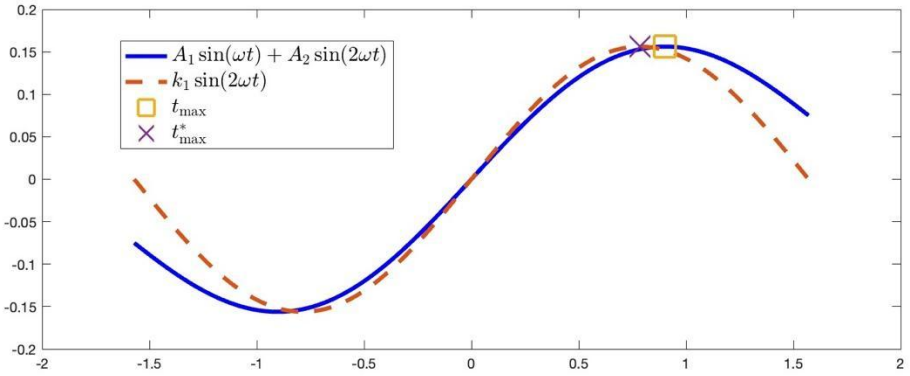


Fig.6: The wave $A_1 \sin(\omega t) + A_2 \sin(2\omega t)$ for $A_1/A_2 < 2$ within the interval $\left[-\frac{T}{4}, \frac{T}{4}\right]$

Similar to the cases where $A_1/A_2 \geq 2$, the maximum values of the two functions should be the same, then k_1 could be calculated as:

$$
\begin{aligned}
t_{\max} &= \arg\max_t(A_1 \sin(\omega t) + A_2 \sin(2\omega t)) = \frac{1}{\omega}\arccos\left(\frac{-A_1 + \sqrt{A_1^2 + 32A_2^2}}{8A_2}\right) \\
t^{*(2)}_{\max} &= \arg\max_t(\sin(2\omega t) = \frac{\pi}{4\omega} \\
k_1 &= \frac{A_1 \sin(\omega t_{max}) + A_2 \sin(2\omega t_{max})}{\sin(2\omega t^*_{max})}
\end{aligned}
\tag{24}
$$

Within the interval $\left[\frac{1}{4}T, \frac{3}{4}T\right]$, the function could also be approximated by $k_2 \sin(2\omega t)$, as shown in Fig.7. k_2 could be calculated using the condition that the minimum value of the two functions are the same:

$$
\begin{aligned}
t_{\max} &= \arg\max_t(A_1 \sin(\omega t) + A_2 \sin(2\omega t)) = \frac{1}{\omega}\arccos\left(\frac{-A_1 + \sqrt{A_1^2 + 32A_2^2}}{8A_2}\right) \\
t^{*(2)}_{\max} &= \arg\max_t(\sin(2\omega t) = \frac{\pi}{4\omega} \\
k_1 &= \frac{A_1 \sin(\omega t_{max}) + A_2 \sin(2\omega t_{max})}{\sin(2\omega t^*_{max})}
\end{aligned}
\tag{25}
$$

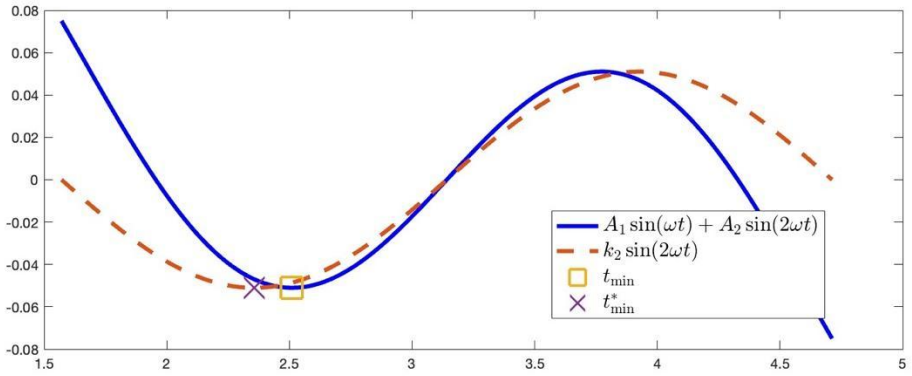


Fig.7: The wave $A_1\sin(\omega t) + A_2\sin(2\omega t)$ for $A_1/A_2 < 2$ within the interval $\left[\frac{T}{4}, \frac{3T}{4}\right]$

The A' and ω' of frequency doubling vibration could be calculated as:

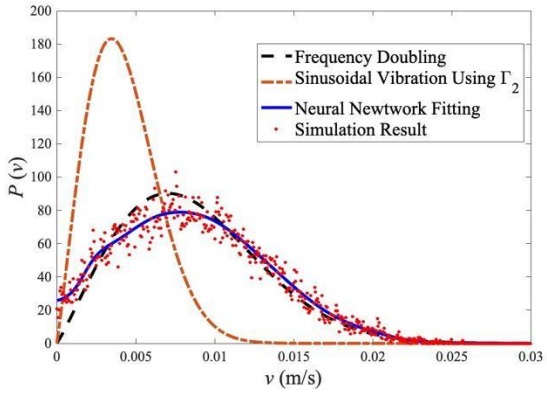
$$A' = (k_1 + k_2)/2$$
$$\omega' = 2\omega \tag{23}$$

4.4 The Simulation Results

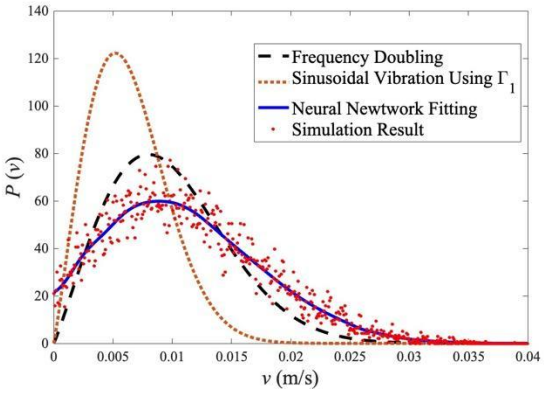
We simulated the collisions between the falling ball and the vibrating ferrite rod by solving Eq. (13) and (14) numerically. All collisions with negative absolute velocity v were ignored.

For each pair of Γ_1 and Γ_2, we simulated 30000 collisions and generated 500 red dots to approximate the PDF of the absolute velocity of the ball after collision, as shown in the figures above. ω is assumed to be $\omega = 3000 \times 2\pi\, s^{-1}$ in all simulations.

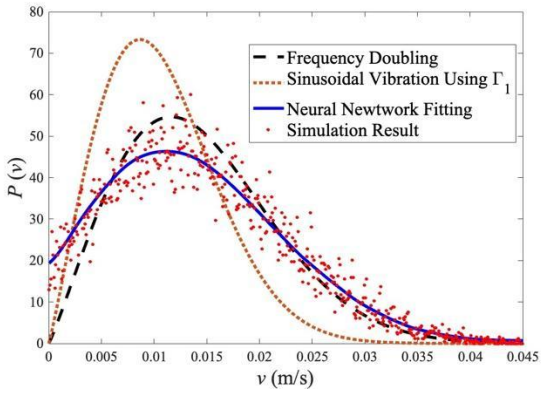
In simulations for each pair of Γ_1 and Γ_2, "locking region" was always detected. If the ball enters the "locking region", it "sticks" to the ferrite rod and performs an infinite series of collisions of decreasing velocities within a finite length of time, ultimately performing periodic motion.



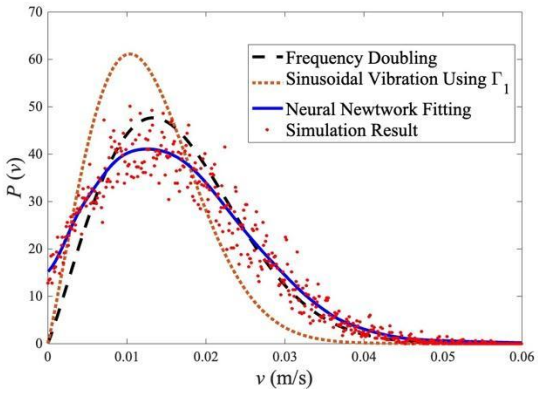
(a) $\Gamma_1 = 0.5,\ \Gamma_2 = 1$



(b) $\Gamma_1 = 1.5,\ \Gamma_2 = 1$



(c) $\Gamma_1 = 2.5,\ \Gamma_2 = 1$



(d) $\Gamma_1 = 3,\ \Gamma_2 = 1$

Fig.8: (a) - (d) show the probability density distributions of the bouncing velocity of the ball under Γ_1 and Γ_2. The orange dashed line represents the case where the ferrite rod exhibits sinusoidal vibration, while the black dashed line represents modulated vibration. The red data points represent the simulated results, and the blue line represents the fitted curve from the NN.

The PDFs obtained from the simulations do not pass through the origin. In this way they are significantly different from Boltzmann distribution. Therefore, we employed a Neural Network algorithm to generate the approximated PDF based on the 500 red dots generated by simulation to better compare the simulation results with our analytically deduced PDF in Section 3.3. Neural Network (NN) is a deep learning algorithm used to fit a curve based on data points without giving the curve any analytical form. In this paper, the tool TensorFlow is used to train the NN. We create 6 hidden layers, each with 60 nodes.

From the simulation results above, it is apparent that our PDF for the vibrating platform is more accurate than Warr's original analysis on the sinusoidally vibrating platform and is basically correct in predicting the peak of the distribution, but its general pattern still deviates slightly from the simulation results.

4.5 The Experimental Results

Using the apparatus in Section 2 and AC of frequency 5850 hz, 157 collisions are recorded and analyzed. The restitution coefficient α in the experiment is 0.87. The A_1 and A_2 are extremely small and they could not be measured subjected to our experimental instruments. However, it could still be observed that the experimental results could be well fitted by a Boltzman distribution in Eq.(20) as shown in Fig.9. The A' of the best fitting curve is $5.03 \times 10^{-6}\ m$, which is indeed a reasonable value.

The collision with high velocities is better approximated by the Bolztman distribution compared to those with lower velocities. This is probably because the relative rebound velocity is a nonlinear function of incoming relative velocity at low velocities. Furthermore,

the collision duration at low velocities might be comparable to the rod's period, which would make the theoretical analysis invalid.

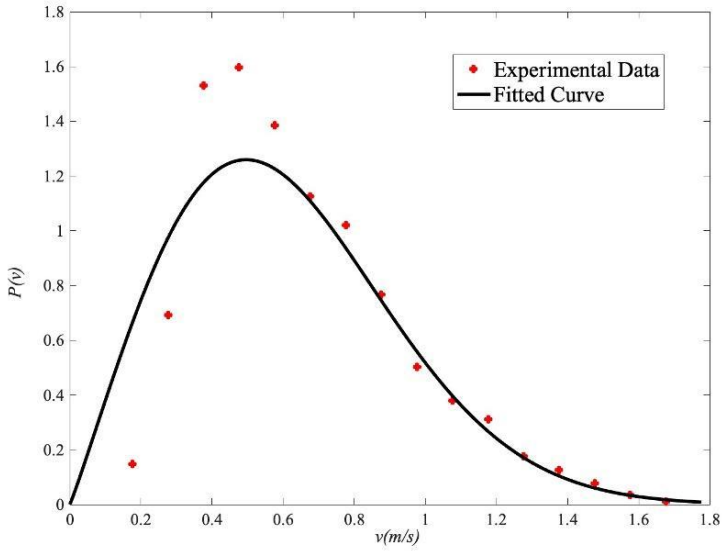


Fig.9: Experimental Results

5. Conclusion

This paper discusses the dynamics of a ball bouncing on a ferrite rod under an alternating magnetic field created by the AC in the coil at its end. In Section 3, the vibration equation of the ferrite rod under the assumption of linearity and isotropy is derived, revealing that the height of the rod's upper end is a sine function of time $u(l,t) = A\sin sin\,(\omega t)$. However, since the ferrite rod is generally heterogenous in real experiments, the frequency domain of $u(l,t)$ should be a combination of the AC frequency and a frequency twice as high (which is also observed in simulations and experiments), given by:
$u(l,t) = A_1 \sin(\omega t) + A_2 \sin(2\omega t)$.

In Section 4, the ferrite rod is treated as a platform vibrating beneath the ball, and the ball's trajectory is investigated. Neural Network curve fitting algorithm is used to approximate the PDF of the ball's rebound velocity by iteratively solving Eq.(13) and Eq.(14), and it is found that the rebound velocity v of the ball approximately follows a Boltzmann distribution in Eq.(20), with $A\omega$ rescaled by a factor k or $k_1 + k_2$ according to Eq.(23) and Eq.(26) compared to Warr's analysis on sinusoidally vibrating platforms. Additionally, it is observed during the numeric simulations that regardless of the initial conditions, the ball eventually enters the "locking region", exhibiting periodic motion in the end. In real experiments, it was observed that the PDF is well approximated by a Boltzmann distribution of the form of Eq.(20) at high velocities but deviates slightly from it at low velocities. This could be explained by the non-linear relationship between the incoming and rebounding relative velocity and the relatively long collision duration at low velocities.

Works Cited

1. J.Y. Chastaing. Dynamics of a bouncing ball. *American Journal of Physics*, 83(518):518–524, 2015.
2. T.Gilet. Completely inelastic ball. *Physical Review E*, 79(055201):1–3, 2009.
3. A. Mehta. Novel temporal behavior of a nonlinear dynamics system: The completely inelastic bouncing ball. *Physical Review Letters*, 65(4):391–396, 1990.
4. J.M.Luck. Bouncing ball with a finite restitution: chattering, locking, and chaos. *Physical Review E*, 48(5):3988–3998, 1993.
5. S.Warr. Energy input and scaling laws for a single particle vibrating in one dimension. *Physical Review E* ,52(5):5596–5601, 1995.
6. J. Lee. Scaling behavior of granular particles in a vibrating box. *Physica A*, 219:305–326, 1995.
7. S.Warr. Probability distribution functions for a single particle vibrating in one dimension: experimental study and theoretical analysis. *Physical A: Statistical Mechanics and its Applications*, 231(4):551–574, 1996.
8. K. Shoemaker. The mathematical modeling of magnetostriction, 2018.
9. V. Balakrishnan. *Elements of nonequilibrium statistical mechanics*. ANE Books, 2008.
10. J. Rajchenbach. Fluidization of a bidimensional powder. *Europhysics Letters*, 16(2), 1991.
11. H.J. Herrmann. Simulations of two-dimensional arrays of beads under external vibrations: Scaling behavior. *Physical Review E*, 50(4), 1994.

The Neuroscience and Psychology Behind the Joys of Acting By Aditi Shankar

Introduction

As a shy child, I always marveled at the confidence and passion that stage actors exuded. With the encouragement of my parents, I began my performing arts journey in first grade as an ensemble member in the musical *Frog and Toad*. The part that I most vividly remember playing was a squirrel. There I was, dressed all in black, my outfit completed by a brown vest and bushy brown tail as long as my torso. On my first performance night, I tentatively stepped out onto the stage, holding fake cloth leaves in my trembling hands. At first, the leaves fell shyly from my fingers, but then I led my fellow squirrels, tossing leaves exuberantly and creating an autumn flurry. The concoction of sheer joy, nervousness, excitement, and adrenaline that I felt on stage with the lights shining on me was absolutely priceless. As I carried this passion with me through high school, a question began to form: What was it about acting that allowed me to feel so immersed and joyful? This essay investigates that question by exploring the psychology and neuroscience behind the joys of acting and how stage fright and stress impact one's absorption in their performance.

Acting and Flow

When one performance ends, I cannot wait for the next. During each show, I become completely immersed in the happiness, adrenaline, focus on my role, and perception that time seems to fly. Psychologists commonly refer to this state of mind as "flow" (Hruska). Flow is a psychological state during which one feels completely engrossed in an activity. Before a person experiences flow, there must be: 1) a perceived challenge-skill balance, such that the activity is neither boring nor overly grueling for the participant, 2) meaningful steps to reach concrete goals, and 3) "clear, unambiguous feedback" to fulfill changing demands and maintain flow (Hruska). During a flow-inducing activity, people experience the following: 1) total concentration, 2) a sense of deep enjoyment of the activity itself, rather than the outcome, 3) a deep immersion in the task, 4) a sense of control, 5) a lack of self-consciousness, and 6) a distorted sense of time (e.g., time feels faster; Hruska). But what happens at the neurological level when experiencing flow?

In the brain, the "spikey" and "sharp" beta brain waves that arise when one is multitasking, experiencing anxiety, and shifting their attention become smooth, freeflowing alpha waves at the onset of flow (brainbiz.com.au). This slowed, less erratic brain wave activity curates a headspace that allows people to make fewer, deeper neural connections. As one enters the rewarding flow state, their neurons begin to fire a variety of excitatory and inhibitory neurotransmitters, or electrical signals that stimulate or block certain neurological functions. Some of the neurotransmitters released during the flow state include dopamine ("feel good" signal), anandamide ("bliss" signal), norepinephrine (energy and attentiveness signal), serotonin (calm signal), and endorphins (pain blockers and stress relievers) (brainbiz.com.au). During flow,

the prefrontal cortex, a part of the brain that controls logic and thinking, becomes less active as the limbic system (i.e., the "emotional" brain) activates, creating a joyous experience.

Neuroscience of Acting

Researchers have not studied the neuroscience of acting extensively. Nonetheless, one recent study investigated the neuroscience behind actors playing the roles of Romeo and Juliet. This study revealed that when fully embodying the characters, study participants engaged in less self-processing (Brown et al., 1-20). Self-processing is a term to describe reflecting on one's personal thoughts or beliefs. This reduction in self-processing is consistent with the qualities of flow, where the person becomes less focused on their own ego and more so on the activity at hand. The researchers also reported that when actors responded to prompts in character, their brains had reduced activity in the midline network of the frontal lobe. Thus, researchers concluded that acting is a "deactivation-driven process," which results in a "loss of self" (Brown et al., 1-20). External pressure and self-consciousness can negatively affect a person's mindset, actions, and the quality of the activity. So, removing these obstacles is crucial for deriving enjoyment.

Emotions and Acting

Emotions are also a key part of the psychological experience of acting (Trueman). An article from the Guardian notes that actors often use their own emotions to enhance the credibility of their characters' emotions onstage. There are two different categories of emotions involved in acting: personal emotions that one experiences in response to the plot and fabricated emotional responses as the character responds to another actor's behavior (Trueman). To make the characters' emotions appear genuine, actors must internalize their characters' intentions and reactions to different plot events. When in character, the actor learns the emotional significance of the stimulus to their character according to the script. The learned emotional response is stored in the amygdala, the center for emotions and emotional memory, and in the hippocampus for recall (Tyng et al., 1-22). These reactions can then be voluntarily recalled during the next performance. Many actors enjoy their craft because it can feel emotionally liberating to play another role. They can see acting as a process of self-discovery or use it as an outlet to express themselves in a different way from their everyday selves (Trueman).

Causes of Stage Fright

While this paper has discussed several factors that predict performance enjoyment, a common obstacle to enjoying performance is stage fright (psychologytoday.com). According to Psychology Today, freezing onstage, feeling nauseous, suffering from dry mouth, and stumbling over words are all common symptoms of stage anixety. Stage fright inhibits a person from diving fully into their role because it hinders flow (Spahn et al., 1-12). *Frontiers in Psychology* released a study that investigated how musicians experience flow. When these musicians were surveyed about their musical performance anxiety (MPA) levels, they reported that the more performance

anxiety they experienced, the less flow they encountered onstage. The findings also stated that stronger coping skills (i.e., the ability to reassure oneself when experiencing stage fright) reduced participants' symptoms of MPA (i.e., diminished nervousness and stress) and increased self-efficacy (i.e., self-confidence and belief in one's own abilites). These psychological changes led participants to experience enhanced flow (Spahn et al., 1-12). However, because the relationship between performance anxiety and flow has has not been studied specifically in theater, future research should investigate these potential linkages.

Stage anxiety may also be influenced by a person's personality and beliefs about stress (Bentham, McGonigal). For example, those with autotelic personalities are charaterized as people who continuously seek out activities with an appropriate skill-challenge balance and enjoy setting goals and receiving clear, real-time feedback. In other words, people who naturally have autotelic personalities are more likely to experience flow and, likely, less stage fright (Bentham). On the other hand, people who have negative beliefs about the stress that they expect to experience while on stage may enjoy performing less because they are fixated on the negative outcomes of stress, such as lower success and poor mental health consequences. According to psychologist Kelly McGonigal, those who believe that their stress responses are helpful and energizing are more confident and less anxious during high pressure situations, such as performing onstage (McGonigal). In addition, people who have positive beliefs about stress are able to change their bodies' physiological reactions to stressful situations to be less taxing. However, those who believe stress is harmful have an intensely negative reaction to stress, harming their ability to enjoy perfoming, feel confident, and perform excellently.

Tips to Reduce Performance Anxiety and Enhance Flow

Getting into flow during performance is crucial to feeling fulfilled and finding enjoyment (Hruska). When actors express joy and passion through their performances, the audiences can feel it too. To reduce performance anxiety, research and my personal experiences suggest using tools such as meditation, listening to energetic music, and talking to fellow cast mates to get focused. Calming yourself down before a seemingly daunting task is a skill that actors must be able to use for any performance (psychologytoday.com). Although stage fright never vanishes, actors must reframe their stress mindset about performance to make their experience on stage more enjoyable. The more experience a person has, the more manageable stage fright becomes over time (psychologytoday.com).

Conclusion

This article investigated the reasons why some people deeply enjoy acting, while others face heightened anxiety in the process. Studies have explored brain activity during performance, showing that musical performance anxiety and flow are inversely related (Spahn et al., 1-12). Meanwhile, during flow, the brain releases calming, "feel-good" neurotransmitters, and the brain becomes more focused and less frenetic (brainbiz.com.au). Additionally, this paper explained that while acting, actors often enhance their emotions by adopting the perspectives of their

characters and enjoy the art of this emotional expression. From research about the psychological and neurological experiences of acting, I now know that waiting behind the curtains with a positive stress mindset and focused state of mind will help me tackle any challenges the stage presents me with, so that I can prance around the stage without a care in the world, and you can too.

Works Cited

Bentham, Nicola. "Understanding the Autotelic Personality." *Motivation*, Motivation, 2018, vocal.media/motivation/understanding-the-autotelic-personality.

Brown, Steven, et al. "The neuroscience of Romeo and Juliet: an fMRI study of acting." *Royal Society Open Science*, vol. 6, no. 3, 13 Mar. 2019, p. 181908, https://doi.org/10.1098/rsos.181908.

Hruska, Emese. "What Musicians Should Know about Flow and Why." *PerfActionist*, PerfActionist, 27 Feb. 2021, perfactionist.org/articles/what-musicians-should-know-about-flow-and-why/.

Martin, Pete. "What Happens to an Actor's Brain When Acting?" *Backstage*, Backstage, 3 Mar. 2021, www.backstage.com/uk/magazine/article/what-happens-to-an-actors-brain-when-acting-67811/.

Performance by Kelly McGonigal, *How to Make Stress Your Friend*, June 2013, https://www.ted.com/talks/kelly_mcgonigal_how_to_make_stress_your_friend?language=en.+Accessed+20+Sept.+2023. Accessed 20 Sept. 2023.

Spahn, Claudia, et al. "Live music performance: The relationship between flow and Music Performance Anxiety." *Frontiers in Psychology*, vol. 12, 2021, https://doi.org/10.3389/fpsyg.2021.725569.

"Stage Fright." Edited by Psychology Today Staff, *Psychology Today*, Psychology Today, www.psychologytoday.com/us/basics/stage-fright. Accessed 20 Sept. 2023.

"The Neuroscience of Flow." *Brain Biz*, Brain Biz, brainbiz.com.au/the-neuroscience-of-flow/#:~:text=FLOW%20Conditions&text=In%20flow%2C%20brain%20changes%20occur,effortless%20execution%20of%20the%20task. Accessed 20 Sept. 2023.

Trueman, Matt. "'When You Cry, You Really Cry': The Emotional Toll of Stage Acting." *The Guardian*, Guardian News and Media, 11 Aug. 2016, www.theguardian.com/stage/2016/aug/11/actors-emotions-on-stage-kate-fleetwood-michelle-terry-ben-miles.

Tyng, Chai M., et al. "The Influences of Emotion on Learning and Memory." *Frontiers in Psychology*, vol. 8, 24 Aug. 2017, pp. 1–22, https://doi.org/10.3389/fpsyg.2017.01454.

Bell's Theorem and Entanglement By Shreyas Dillon

Abstract

Entanglement is one of the defining features of quantum mechanics. Physicists and philosophers, including Einstein, have been bothered by the implications of joint measurements of entangled particles for some time. One such implication is the apparent break from locality. Our universe looks local but measurements of entangled particles suggest it may be non-local. This led to many theories being formulated in order to create a more complete picture during measurements of entangled particles. One of the most prominent theories was a class known as "local hidden variable" theories, which assigned an additional, unknowable quantity to a particle's state. However, John Bell outlined a theory, which has since gained considerable experimental support, that showed performing joint measurements of entangled particles contradicts any "local hidden variable" theory. In this paper, we will discuss John Bell's theorem and how it shows the differences between quantum mechanics and local hidden variable theories. We will explain the impact this theorem had on the quantum world and how they were proved in an experimental setting, including the closure of several "loopholes."

Introduction

The quantum world is riddled with mysteries and perplexing phenomena that are distinct from classical physics. For centuries we have tried to make sense of these behaviors, with physicists such as Albert Einstein, Niels Bohr, Max Planck, and Erwin Schrödinger paving the way. Despite many breakthroughs, the foundation of quantum theory still remains puzzling. One example of this is the indeterminacy surrounding measurement. In classical physics, all measurements are deterministic and there is no fundamental limit to how precise a classical measurement can be. In the quantum world, however, there is a limit to the precision of measurements (governed by the uncertainty principle) and outcomes are probabilistic.

Many theories have been postulated surrounding the "complete" definition of quantum states and why quantum measurement differs from the classical world but one primary group is the local hidden variable theories. The famous EPR (Einstein Poldosky Rosen) paradox speculated about one of these theories in an attempt to mesh classical and quantum physics. The EPR paper questioned quantum mechanics' ability to describe reality and concluded that additional variables were necessary to describe reality completely [5]. They believed that all particles have an extra quantity attached to their states that determines the measurement outcome prior to actual measurement. This theory seemed to reconcile one of the main gaps separating classical and quantum measurement and was the primarily held theory for some years but soon fell out of favor with the release of John Stewart Bell's theorem. This theorem utilized quantum entanglement in its establishment and highlighted the differences between predictions made by quantum mechanics and local hidden variable theories. A uniquely quantum behavior, entanglement is the linking of particles that remains regardless of the distance between them. Einstein described it as "spooky action at a distance" and Bell used these particles' unique

correlations to contradict the EPR hidden variable predictions. This created a highly important finding known as Bell's theorem. This theorem is a fundamental one in quantum physics and it showcases a main difference between classical and quantum mechanics. Bell's theorem breaks Einstein's attempted meshing of classical and quantum physics and further elucidates just how unique the quantum world is.

Background

To fully understand Bell's theorem, it is necessary to understand some more technical definitions of quantum mechanical topics such as Dirac notation, superposition, and Bell states. In the appendix, we provide detailed descriptions of these terms. In addition to the technical terms, it is important we understand the history before Bell's theorem, specifically the EPR group's argument. The EPR group's thought experiment was conducted with pairs of entangled particles that appeared to be communicating faster than the speed of light, in direct contradiction with relativity and locality. Their initial argument utilized "squeezed states", which are states where uncertainty for one of the measurable variables is altered ("squeezed") in such a way that can make the particles entangled on that variable. However, an analogous experiment can be conducted with the spins of particles, specifically spin ½ particles in the "singlet" spin state. If you measure components of these particles' spin, quantum mechanics dictates that if a measurement of the first particle yields 1 (spin up), the second particle will yield -1 (spin down) and vice versa. You are then able to predict the state of the second particle after measurement of the first.

In attempts to understand this break from relativity the EPR group posited that there must be some other variable, which we will denote as λ attached to a particle that determines its state prior to measurement. It is important to note that λ is not strictly defined, it can be a discrete or continuous variable or even a set of variables. In our case, it takes the form of a single continuous parameter. Additionally, the angle of the measurement detectors changed the outcome of measurement. With λ and the detector angles in mind, the measurement of one particle in an outcome we'll call A is dependent on $\vec{a}$ and λ where $\vec{a}$ is the unit vector that represents the angle of the first particle detector. Similarly, a result B depends on $\vec{b}$ (unit vector of second detector) and λ . They assumed that B is independent of the setting, $\vec{a}$, and A on$\vec{b}$. If $p(\lambda)$ is the probability distribution of λ, then the expectation value of the products of spin are:

$P(\vec{a}, \vec{b}) = \int p(\lambda) A(\vec{a}, \lambda) B(\vec{b}, \lambda) d\lambda$. It was expected that this would equal the

quantum mechanical expectation value of the product of the spins for the singlet state. This was not the case, however, and these differences were the foundation for Bell's theorem.

Bell's theorem

What Bell identified is that this quantum mechanical prediction is inconsistent with any local hidden variable theory. Bell based his theorem upon a generalization of the EPR experiment. The particle detectors in the EPR case were oriented in the same direction, but Bell

allowed them to differ and rotate independently. Additionally, instead of analyzing the direct measurements, he calculated the average value of the product of the spins. The expectation value for this product (in the singlet state) is $P(a\rightarrow, b\rightarrow) = -\ a\rightarrow \cdot\ b\rightarrow$ where $a\rightarrow$ and $b\rightarrow$ are the arbitrary orientations for the two particle detectors. He then utilized several components of local hidden variable theories and employed them to create a mathematical inequality that could highlight this disparity between classical and quantum mechanics. The first of these components was the hidden variable λ itself, which was viewed as a necessity to define the "complete" state of the two particles. Secondly, locality was assumed, i.e. the measurement of the second particle was considered independent of the angle of the first detector. Thus, there are two functions that represent the results of measurement of the two particles: $A(a\rightarrow, \lambda)$, $B(b\rightarrow, \lambda)$. These two functions can only take on the values of ± 1. From this, along with the probability distribution of λ, we get the prediction for this average of spins (shown in the last section), as dictated by a local hidden variable theory:$P(a\rightarrow, b\rightarrow) = \int p(\lambda)A(a\rightarrow, \lambda)B(b\rightarrow, \lambda)d\lambda$. When the detectors are aligned (as they were in the EPR experiment) the results of measurement are perfectly anti-parallel, so $A(a\rightarrow, \lambda) = -\ B(a\rightarrow, \lambda)$. We then can eliminate B and change our equation to:

$P(a\rightarrow, b\rightarrow) = -\int p(\lambda)A(a\rightarrow, \lambda)A(b\rightarrow, \lambda)d\lambda$. Now if $c\rightarrow$ is any other unit vector (this step can be confusing if you try to apply it in the experimental context but it is necessary from a mathematical standpoint to move forward) then

$$P(a\rightarrow, b\rightarrow) - P(a\rightarrow, c\rightarrow) = -\int p(\lambda)[A(a\rightarrow, \lambda)A(b\rightarrow, \lambda) - A(a\rightarrow, \lambda)A(c\rightarrow, \lambda)]d\lambda.$$

Additionally, if we note that $[A(b\rightarrow, \lambda)]^2 = 1$ then we can change our equation to:

$$P(a\rightarrow, b\rightarrow) - P(a\rightarrow, c\rightarrow) = -\int p(\lambda)[1 - A(b\rightarrow, \lambda)A(c\rightarrow, \lambda)]A(a\rightarrow, \lambda)A(b\rightarrow, \lambda)d\lambda$$

Now we take the absolute value of both sides of the equation. $A(a\rightarrow, \lambda) \times A(b\rightarrow, \lambda) = \pm 1$ so if we take the absolute value of the product, $|A(a\rightarrow, \lambda)A(b\rightarrow, \lambda)|$, we get that the product has to be 1 or: $|A(a\rightarrow, \lambda)A(b\rightarrow, \lambda)| = 1$. Additionally $p(\lambda)[1 - A(b\rightarrow, \lambda)A(c\rightarrow, \lambda)] \geq 0$ so putting it in absolute value has no effect on it and our equation becomes an inequality:

$$|P(a\rightarrow, b\rightarrow) - P(a\rightarrow, c\rightarrow)| \leq \int p(\lambda)[1 - A(b\rightarrow, \lambda)A(c\rightarrow, \lambda)]d\lambda.$$

Now if we distribute $p(\lambda)$ we get

$$|P(a\rightarrow, b\rightarrow) - P(a\rightarrow, c\rightarrow)| \leq \int p(\lambda) - p(\lambda)A(b\rightarrow, \lambda)A(c\rightarrow, \lambda)d\lambda.$$

The right side becomes $\int p(\lambda)d\lambda - \int p(\lambda)A(b\rightarrow, \lambda)A(c\rightarrow, \lambda)d\lambda$

$p(\lambda)$ follows the normalization condition, which means $\int p(\lambda)d\lambda = 1$ and

$-\int p(\lambda)A(b\rightarrow,\lambda)A(c\rightarrow,\lambda)d\lambda = P(b\rightarrow,c\rightarrow)$. This is the same process from when we

eliminated B and got $P(a\rightarrow,b\rightarrow) = -\int p(\lambda)A(a\rightarrow,\lambda)A(b\rightarrow,\lambda)d\lambda$. With these

simplifications, we now acquire the famous Bell inequality:
$|P(a\rightarrow,b\rightarrow) - P(a\rightarrow,c\rightarrow)| \leq 1 + P(b\rightarrow,c\rightarrow)$[7]. This inequality is consistent with any local hidden variable theory, but clearly highlights the incompatibility of local hidden variable theories (and thus classical mechanics) with quantum mechanics [3]. We will demonstrate this through a worked example.

Imagine we have two spin ½ particles in the singlet state, S^1 and S^2. We measure a component of their spin angular momentum, measuring S^1 in direction $a\rightarrow$ and S^2 in direction $b\rightarrow$. The quantum mechanical prediction for the product of their spins is $<S^1{}_a S^2{}_b> = -\hbar^2/4\,cos(\theta)$ (refer to the appendix for a derivation of this value), where the brackets represent finding the expectation value and θ is the difference of $a\rightarrow$ and $b\rightarrow$. We stick to the convention and normalize the values of measuring a spin to ± 1. This changes the expectation value from $\hbar/2 \times -\hbar/2\,cos(\theta)$ to $1 \times -1\,cos(\theta)$, or just $-cos(\theta)$. Now we have the ability to graph both sides of Bell's inequality (and the differences between them) to see which angle of the detector will provide the maximal violation of this inequality.

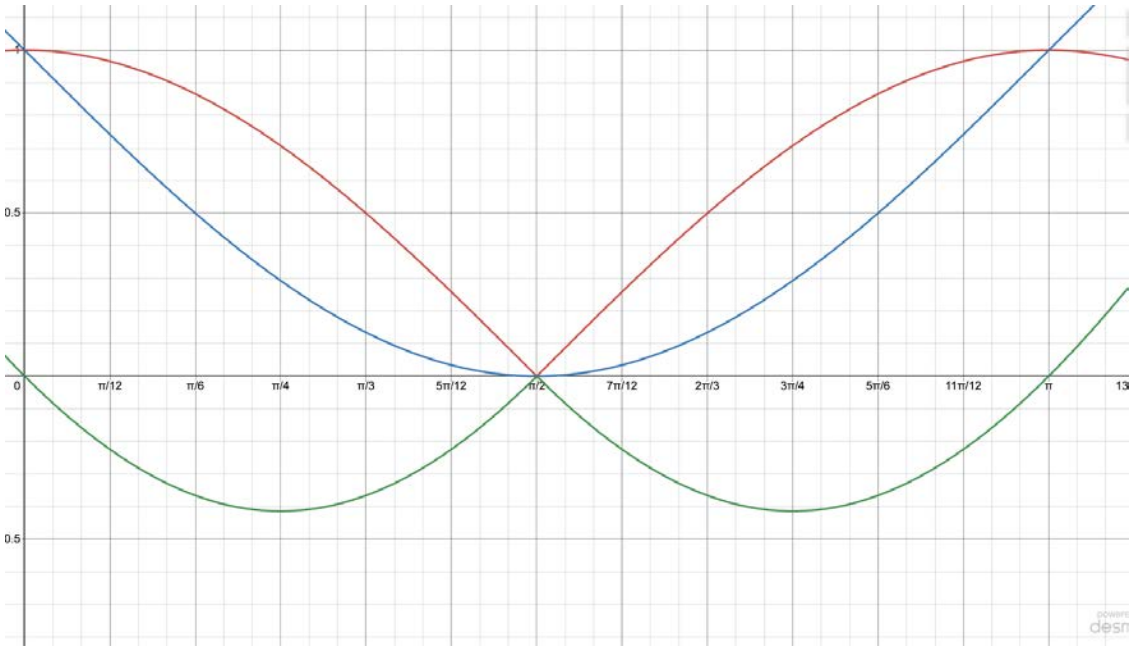


Fig. 1: Graph of Bell's inequality and the differences between each side.

The red graph is the left side of the inequality (the quantum mechanical prediction), blue is the right (the maximum allowed in a local hidden variable theorem), and green is the difference. The negative values given by the difference are important to note as it means that the maximum value allowed in a local hidden variable theory is surpassed by quantum mechanics, causing a significant contradiction. Continuing, the next thing we need is a value of θ. The value that returns the clearest violation of Bell's inequality is $\theta = \pi/2$, or $a \rightarrow - b \rightarrow = \pi/2$. This is the angle that provides the maximal difference in the inequality because, as evidenced by our graph, quantum mechanics predicts 0 for this angle. This means that it is the result for which we have the least amount of knowledge about either of the particles, $P(a \rightarrow, b \rightarrow) = 0$. It is easy to check this by plugging in $\pi/2$ for θ in our expectation value for the product of spins. Now for the reference vector $c \rightarrow$. We will have it make a 45-degree angle with $a \rightarrow$ and $b \rightarrow$. (The reference vector doesn't have a strict place to be; we can pretty much pick its location). Additionally, π/4 is the minimum value for the graph of the difference and thus the maximal violation of the inequality. This setup of $a \rightarrow$, $b \rightarrow$, and $c \rightarrow$ can be represented graphically by having $a \rightarrow$ and $b \rightarrow$ form axes and $c \rightarrow$ be exactly between them:

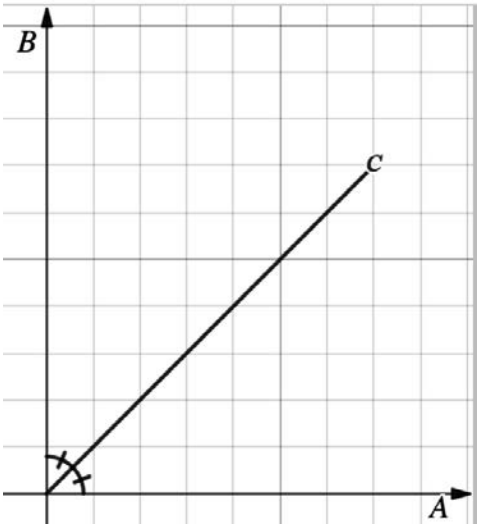


Fig 2: Graphical representation of orientation vectors

With this, we can obtain the result of Bell's inequality for these particles. Taking Bell's inequality: $|P(a \rightarrow, b \rightarrow) - P(a \rightarrow, c \rightarrow)| \leq 1 + P(b \rightarrow, c \rightarrow)$, and using our 3 vector graph to visualize, we get:

$P(a \rightarrow, b \rightarrow) = - cos(0 - 90) = 0$

$P(a \rightarrow, c \rightarrow) = - cos(0 - 45) = - \frac{\sqrt{2}}{2}$

$P(b \rightarrow, c \rightarrow) = - cos(90 - 45) = - \frac{\sqrt{2}}{2}$

We will take $-\frac{\sqrt{2}}{2} \approx -.707$.

Plugging all this into our inequality we get

$|0 + .707| \leq 1 - .707 \Rightarrow .707 \leq .293$ which is clearly false.

Local realism disproved

Bell's contradictions of local hidden variable theories had major implications on the quantum world. It meant that either quantum mechanics is fundamentally wrong or nonlocality is an intrinsic property of nature. Both of these options had considerable consequences if proved true, so experimental tests were performed [1]. Many of the first experiments were carried out by J.F Clauser, an experimental physicist working at Columbia University. He and several colleagues entangled the polarization of two optical photons (light particles) and sent them to two separate polarizers, with the only outputs being ±1 [6]. These polarizers were set to rotate and measure coincidence rates. Coincidence rates are the amount of time both photons were detected by the polarizers. As the polarizers were rotated, the coincidence rates changed because the angle of the polarizer directly affected the particles' detection rates. Using his experimental setup, and principles of hidden variable theories, Clauser was able to create constraints for the coincidence rates, which quantum mechanics violated. This result was hugely significant as it was the first time Bell's ideas were tested experimentally. Clauser's experiments represent the critical step of transitioning from theory to physical reality and laid the experimental foundation for the disprovement of local realism in nature.

Even though this foundation was in favor of quantum mechanics, there were two primary loopholes that needed closure in order to have confidence in these results. "Loopholes" refer to ways the experiments were conducted that left room for a steadfast believer in local hidden variable theories to argue against Bell's theorem. The first of these two loopholes was the locality loophole. In both the EPR and Bell's paper, it was assumed that the orientation of one polarizer had no effect on the measurement of a particle at the other polarizer. In order to have this be the case and close this loophole in an experimental setting, Bell realized that the orientations of the polarizers needed to be chosen after the photons were released. This would close this loophole because of the principle of relativistic causality, which states that the cause of something must come before its effect. This effect cannot influence anything that would require movement faster than the speed of light to reach. If the orientations are chosen mid-flight and the detectors are positioned far enough away so that they are out of the "light cone" of the entangled photons, then it cannot have any effect on the initial state of the particles, as that would signify causality that was faster than light. To fulfill this experimental necessity Alain Aspect and his team based at the Institut d'Optique created their own experiment. Their setup was very similar to Clauser's; however, instead of predetermining the angle of the detectors and keeping that angle for the whole experiment, Aspect added an optical switch at each detector that was able to change the angle of the polarizers mid-flight [2]. The results of this experiment, even including this optical switch, agreed with quantum mechanical predictions and violated Bell's inequality by 6 standard deviations. Aspect's experiment further cemented this idea of nonlocality in nature but there was still another loophole present in experiments.

The detection loophole was a loophole in the previously conducted experiments based on the heralding efficiency of the technology used. Heralding efficiency is the probability of detecting one photon when its entangled partner has also been detected. In both Clauser and

Aspect's experiments, the violations of Bell's inequality were found from a very small sample of total emitted pairs as the heralding efficiency was poor for technology of the time. Utilizing a small sample of the entangled pairs meant that the results might not be representative of the total pairs. To close this loophole, the heralding efficiency had to be more than 66%. This issue was solved in 2013, as technology had advanced to the point where new photodetectors had intrinsic detection efficiencies of over 90% [8]. This was more than enough to close the detection loophole and both Bradley Christensen and Mary Rowe conducted experiments utilizing these new technologies that produced results in favor of quantum mechanics [4].

Both loopholes had now been closed, but separately. This was until Anton Zeilinger, at the University of Vienna, created an experiment that closed both these loopholes simultaneously. His experiment utilized the idea of rapidly changing polarizers that Aspect added but employed a type of random number generator that served as the mechanism to close the locality loophole. Additionally, their heralding efficiency was 75%, more than the 66.6% necessary to sufficiently conclude that lack of detection was yielding results in favor of quantum mechanics. The results of this experiment were in agreement with quantum mechanics and through definitive violations of Bell's inequality and closures of experimental loopholes it was confirmed that there is nonlocality in nature and no local hidden variable theory can be used to describe reality.

Conclusion

With this mathematical result and subsequent experiments, we have successfully proved that quantum mechanics is incompatible with any local hidden variable theory. Bell strategically utilized entanglement, a uniquely quantum phenomenon, to demonstrate that we can not account for this seeming nonlocality demonstrated by entangled particles with local hidden variable theories. We are left with the unsettling fact that nature itself is nonlocal. Einstein and the EPR group thought this nonlocal behavior that quantum mechanics demonstrated was unsettling and they began efforts to understand it. Their attempts to create a "complete" depiction of quantum systems and physical reality through the combination of classical and quantum physics led to yet another solely quantum property. This further highlights the uniqueness of quantum mechanics/physics as a field, as along with its probabilistic nature, nonlocality distinguishes it from classical physics. The individuality that characterizes quantum mechanics is what makes it such an intriguing field of study for many. Understanding the microscopic interactions that govern our world is something we have yet to even scratch the surface of, but through physicists such as Bell and the products of their life's work, we come closer to fathoming just how the enigmatic universe we live in functions.

Appendix

Dirac notation and superposition measurement. Dirac notation primarily makes use of bras ($< |$) and kets ($| >$). A ket is a vector and is representative of the state of a particle. A bra can be viewed as a function of row vectors that when multiplied with a ket yields a (complex) number. In order to mathematically represent the eigenstates of a particle prior to a measurement,

Dirac notation is often utilized. A very basic, 2 eigenstate example of possible spin states for a particle would look something like this: $|\psi >= c_0|\uparrow > \ + c_1|\downarrow >$.
Here we can see the use of kets to represent the particle's possible states. The constants c_0 and c_1 represent the probability amplitude of each state. For a certain constant c0 the probability that state will be measured is $|c_0|^2$. Additionally, $\sum_{n=0}^{\infty} |c_n|^2 = 1$. When measuring, we also utilize the projection operator $\widehat{P} = |\alpha >< \alpha|$. This operator picks out only the portions of other vectors that lie along $|\alpha >$.

As you can see, the measurement procedure is vastly different from the classical case. Classically, measurement outcomes can be perfectly predicted (if we have full knowledge of the system). Quantum mechanics differs due to its probabilistic nature, and only after an observation does this uncertainty collapse and one state is randomly chosen. Even if we know the particle is in a known superposition of two states before measurement we don't know which state will be the one measured, only the probabilities associated with each state. There is no way to know what state the particle will end up in prior to measurement, except in the case of entangled particles. Entangled particles are correlated in a way that allows us to be able to gain information on the state of the second particles after only observing the first. This is best shown through the case(s) of maximally entangled particles, or Bell states. The four Bell states are:

$|\Phi +>= \frac{1}{\sqrt{2}}[|\uparrow a > |\uparrow b >+ |\downarrow a > \ |\downarrow b >$

$|\Phi ->= \frac{1}{\sqrt{2}}[|\uparrow a > |\uparrow b >- |\downarrow a > \ |\downarrow b >$

$|\Psi +>= \frac{1}{\sqrt{2}}[|\uparrow a > |\downarrow b >+ |\downarrow a > \ |\uparrow b >$

$|\Psi ->= \frac{1}{\sqrt{2}}[|\uparrow a > |\downarrow b >- |\downarrow a > \ |\uparrow b >$

We will walk through an example measurement on one of these Bell states to better understand this. We have two maximally entangled particles in spin ½ states that can be represented as:

$|\Psi ->= \frac{1}{\sqrt{2}}[|\uparrow a >\downarrow |b >- |\downarrow a > \ |\uparrow b >$

We measure the first particle (particle a), and it is spin up. We apply the spin-up projector to get $|\uparrow a >< \uparrow a|\Psi ->$. Expanding, we get

$\frac{1}{\sqrt{2}}[|\uparrow a >< \uparrow a|\uparrow a > |\downarrow b >- |\uparrow a >< \uparrow a|\downarrow a > \ |\uparrow b >$

$< \uparrow a|\uparrow a >= 1$ and $< \uparrow a|\downarrow a >= 0$ so simplifying gives us

$\frac{1}{\sqrt{2}}[|\uparrow a > (1)|\downarrow b >- (0)|\uparrow a > \ |\uparrow b >$

$=\frac{1}{\sqrt{2}}|\uparrow a > |\downarrow b >$. We have figured out the state of particle b without explicitly measuring it, aided by knowledge of the initial states of the particles.

<Ψ-$|\uparrow a >< \uparrow a|\Psi ->$ = $|< \uparrow a|\Psi ->|$^2 = probability of finding a particle in eigenstate $\uparrow a$

<u>Quantum mechanical prediction for expectation product of spins value</u>:

The singlet state of two spin ½ particles can be represented in Dirac notation as $|00>=\frac{1}{\sqrt{2}}(|\uparrow\downarrow>-|\downarrow\uparrow>)$. What we need to find is the expectation value for the product of the spins of these 2 particles in this singlet state. We represent this as $<S^1{}_a S^2{}_b>$, with the subscripts a and b representing the direction we measure the spins. We can choose our measurement axes, in this case, we will have it so that a lies on the z-plane and b on the xz-plane. These axes are our base for measurement. There are many other possible axes we could have chosen to represent our spins; this is just one way we can express it. Thus $S^1{}_a = S^1{}_z$ and $S^2{}_b = cos(\theta)S^2{}_z + sin(\theta)S^2{}_x$ (the cosines and sines are necessary to represent the angle from the z-axis. Additionally theta is the angular distance between a and b). $<S^1{}_a S^2{}_b> = <00|S^1{}_a S^2{}_b|00>$. We start with the multiplication of the ket: $S^1{}_a S^2{}_b|00>$.

$$S^1{}_a S^2{}_b|00> = S^1{}_a S^2{}_b\left(\frac{1}{\sqrt{2}}(|\uparrow\downarrow>-|\downarrow\uparrow>)\right).$$

$$S^1{}_a S^2{}_b\left(\frac{1}{\sqrt{2}}(|\uparrow\downarrow>-|\downarrow\uparrow>)\right)=\frac{1}{\sqrt{2}}(S^1{}_a S^2{}_b)(|\uparrow\downarrow>-|\downarrow\uparrow>)$$

$$\frac{1}{\sqrt{2}}(S^1{}_a S^2{}_b)(|\uparrow\downarrow>-|\downarrow\uparrow>)=\frac{1}{\sqrt{2}}[S^1{}_z(cos(\theta)S^2{}_z + sin(\theta)S^2{}_x)](|\uparrow\downarrow>-|\downarrow\uparrow>)$$

Distributing the kets we get

$$\frac{1}{\sqrt{2}}[(S_z\uparrow)(cos(\theta)S_z\downarrow + sin(\theta)S_x\downarrow) - (S_z\downarrow)(cos(\theta)S_z\uparrow + sin(\theta)S_x\uparrow)]$$

Using the pauli matrices for spin ½ particles in the z and x axes we get

$$\frac{1}{\sqrt{2}}[(\frac{\hbar}{2}\uparrow)(cos(\theta)(-\frac{\hbar}{2}\downarrow) + sin(\theta)(\frac{\hbar}{2}\uparrow) - (-\frac{\hbar}{2}\downarrow)(cos(\theta)\frac{\hbar}{2}\uparrow + sin(\theta)\frac{\hbar}{2}\downarrow)]$$

$$\Rightarrow\frac{\hbar^2}{4}(cos(\theta)\frac{1}{\sqrt{2}}-\uparrow\downarrow + sin(\theta)\frac{1}{\sqrt{2}}\uparrow\uparrow + cos(\theta)\frac{1}{\sqrt{2}}\downarrow\uparrow + sin(\theta)\frac{1}{\sqrt{2}}\downarrow\downarrow)$$

$$\Rightarrow\frac{\hbar^2}{4}(cos(\theta)\frac{1}{\sqrt{2}}(-\uparrow\downarrow+\downarrow\uparrow) + sin(\theta)\frac{1}{\sqrt{2}}(\uparrow\uparrow+\downarrow\downarrow)$$

Along with the singlet state for spin ½ particles there are the triplet states, which can be represented as:

$$|11>=|\uparrow\uparrow>$$

$$|10>=\frac{1}{\sqrt{2}}(|\uparrow\downarrow>+|\downarrow\uparrow>)$$

$$|1-1>=|\downarrow\downarrow>$$

We can use these to now have

$$\frac{\hbar^2}{4}(-cos(\theta)|00> + sin(\theta)|11> + |1-1>$$

With this expression, we are ready to multiply with the bra giving us

$\frac{\hbar^2}{4} < 00|(-\cos(\theta)|00> + \sin(\theta)(|11> + |1-1>)$

$\Rightarrow \frac{\hbar^2}{4}(-\cos(\theta) < 00|00> + \sin(\theta) < 00|11> + < 00|1-1>)$

Both the $|11> \text{ and } |1-1>$ states are orthogonal to the singlet state($|00>$, so their inner product will yield 0, giving us

$\frac{\hbar^2}{4}(-\cos(\theta) < 00|00>$. The inner product of the singlet state with itself is 1 so our final answer is $< S^1_a S^2_b >= -\frac{\hbar^2}{4}\cos(\theta)$ [7].

Works Cited

[1]Aspect, Alain. “Closing the Door on Einstein and Bohr’s Quantum Debate.” *Physics* 8 (December 16, 2015): 123. https://doi.org/10.1103/PhysRevLett.115.250401.

[2] Aspect, Alain, Jean Dalibard, and Gérard Roger. “Experimental Test of Bell’s Inequalities Using Time- Varying Analyzers.” *Physical Review Letters* 49, no. 25 (December 20, 1982): 1804–7. https://doi.org/10.1103/PhysRevLett.49.1804.

[3]Bell, J.S. “ON THE EINSTEIN PODOLSKY ROSEN PARADOX.” *Physics* Vol. 1, no. No.3 (November 4, 1964): 195–200.

[4]Christensen, B. G., K. T. McCusker, J. Altepeter, B. Calkins, T. Gerrits, A. Lita, A. Miller, et al. “Detection-Loophole-Free Test of Quantum Nonlocality, and Applications.” *Physical Review Letters* 111, no. 13 (September 26, 2013): 130406. https://doi.org/10.1103/PhysRevLett.111.130406.

[5]Einstein, A., B. Podolsky, and N. Rosen. “Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?” *Physical Review* 47, no. 10 (May 15, 1935): 777–80. https://doi.org/10.1103/PhysRev.47.777.

[6]Freedman, Stuart J., and John F. Clauser. “Experimental Test of Local Hidden-Variable Theories.” *Physical Review Letters* 28, no. 14 (April 3, 1972): 938–41. https://doi.org/10.1103/PhysRevLett.28.938.

[7]Griffiths, D.J. (2018) Introduction to Quantum Mechanics. 3rd Edition, Prentice Hall, New York. https://doi.org/10.1017/9781316995433

[8]Rowe, M. A., D. Kielpinski, V. Meyer, C. A. Sackett, W. M. Itano, C. Monroe, and D. J. Wineland. “Experimental Violation of a Bell’s Inequality with Efficient Detection.” *Nature* 409, no. 6822 (February 15, 2001): 791–94. https://doi.org/10.1038/35057215.

An Analysis on the Educational Potential and Challenges of ChatGPT By Jihong Kim

Abstract

ChatGPT, a product of OpenAI, stands as one of the most advanced conversational artificial intelligence models in the world. It is built on the GPT (Generative Pre-trained Transformer) architecture and is intended to comprehend and produce text that resembles human speech in response to input. Its extensive training data, which includes a sizable percentage of the internet, enables it to respond to a variety of requests, from straightforward factual enquiries to intricate philosophical discussions. Its ability to synthesize and create knowledge has made it quite controversial in the realm of education because it has the potential to undermine academic integrity. As a result, institutions around the world, such as the University of Hong Kong and Sciences Po in France, have banned its use outright. However, some groups of scholars and educators recognize its potential to streamline the learning process for students and argue for its integration, rather than exclusion, from schools. In the midst of this controversy, this review paper draws on recent scholarship to outline appropriate and inappropriate uses for ChatGPT in education. Based on technological advancements throughout the history of education, such as the calculator and Google, this paper advocates for a nuanced framework that would help students benefit from AI tools without sacrificing critical thinking skills. Practical suggestions for educational institutions will also be discussed. This research has implications for the future of technology in education, especially AI, which is quickly becoming an essential part of modern society.

Introduction

Technological advancement has always been a catalyst for societal evolution, reshaping the way people live, work, and interact. As technology develops, it introduces efficiencies that can streamlight processes, improve productivity, and foster connectivity over a larger group of people. The Internet, for instance, has democratized access to information by removing or lowering barriers to education and communication. However, rapid societal change inevitably causes confusion and controversy. ChatGPT, therefore, also has been met with widespread social debate. ChatGPT, a product of OpenAI, stands as one of the most advanced conversational artificial models in the world. It is capable of comprehending and producing text that resembles human speech in response to input. There already are a number of articles discussing its uses in different fields, especially in education. Since its first release in late 2022, there have been around 24,500 research papers discussing its application in education; this is a particularly rich topic of discussion since contemporary education is easily exploited through the utilization of this new technology that is ChatGPT. In light of this situation, through exploring the situations in which the use of ChatGPT is appropriate or inappropriate, this research aims to gain insight towards an applicable determinant that can provide guidelines where it can clarify the right and wrong uses of ChatGPT within the educational field.

The Controversy Around ChatGPT

OpenAI launched ChatGPT, a subsequent development in their GPT (Generative Pre-trained Transformer) series, on November 30, 2020. Its release came after the success and learnings from its predecessor, GPT-3. ChatGPT was designed to interact with users in a conversational manner by generating human-like text based on the prompts it received. With the vast amount of data it was trained on, ChatGPT could produce coherent and contextually relevant content, answer questions, create narratives, assist in problem-solving, and even emulate specific writing styles when prompted (ChatGPT). ChatGPT was immensely popular, surpassing 1 million users in 5 days and earning hundred million users in two months. Nowadays, it has 1.6 billion visits every month and is the second fastest growing platform in history (Milmo).

The initial reaction to ChatGPT's release was primarily one of trepidation and caution. Many educational institutions promptly classified ChatGPT as a possible cheating tool given the ability of the AI to produce lengthy and quite logical text passages. Such unprecedented ability was considered to possibly undermine conventional educational testing and assessment methods. Major U.S. school districts and eminent UK colleges even went so far as to forbid it or issue warnings against its use. For example, the Los Angeles Unified School District, ranking second in size in the U.S., promptly restricted access to OpenAI's site on its educational networks. This move was quickly mirrored by other educational districts; by January, numerous educational authorities across English-speaking regions, including states like Washington, New York, Alabama, and Virginia in the U.S., as well as Australian states like Queensland and New South Wales, had prohibited the use of the software (Heaven). This mistrust wasn't limited to institutional decisions; similar concerns were expressed in the broader public conversation, which was mainly concerned with questions of academic integrity, plagiarism, and the veracity of student work.

Media played a significant role in shaping these perceptions, with initial reports emphasizing the challenges posed by ChatGPT to academic integrity. Concerns about students misusing the tool for dishonest purposes overshadowed potential educational benefits, leading to widespread skepticism about ChatGPT's role in educational settings. The technological prowess of ChatGPT, especially its proficiency in generating text that could effortlessly pass off as student-generated content, intensified these concerns.

However, paranoia did not last long thanks to people who started to discover ChatGPT's potential as an educational assistant. Multiple sources support this claim; Hemsley et al. asserts that ChatGPT can interpret unclear commands and provide easily understandable summaries for students with communication and literacy challenges. Crawford contends that ChatGPT has the potential "to support students to be more immersed in their learning, it is possible that the use of an AI tool – like ChatGPT – could free students from the demand for cognitive resources to reach a state of flow and become fully immersed in the task." On the same note, Emily Donahoe, a writing tutor and educational developer at the University of Mississippi, realized that ChatGPT helps students focus on specific skills without getting sidetracked (Heaven). In this way, experts are finding ways that ChatGPT can support student performance during the learning process.

Based on ChatGPT's remarkable capabilities and its potential to foster better education, banning ChatGPT in fear of cheating does not seem to be the best choice. Schools have been challenged with multiple tools that "facilitate cheating," such as calculators, Google, Wikipedia, etc. At the same time, allowing ChatGPT without any restrictions is also a risky move given that, unlike the aforementioned tools, it can generate hunan-like content with ease. Most, if not all, of the research papers discussing ChatGPT's usefulness reach a similar conclusion: a new educational policy or paradigm is required to incorporate ChatGPT or, in fact, any AI. Tlili highlights the necessity for a new teaching paradigm in response to the educational shift towards using chatbots. Similarly, Anderes recommends changing the definition of cheating to accommodate AI usage: "Another idea would be to add to the definition of cheating by incorporating text such as ''cheating includes the use of advanced original text creation AI when it is specifically not allowed by the instructor for a given assignment.'' Such suggestions call for greater nuance in schools' attitudes towards ChatGPT.

ChatGPT and Other Educational Tools in Context

For the educational industry to undergo a proper paradigm shift, it is imperative that we explore the inappropriate and appropriate uses of ChatGPT. The primary concern educational institutions have regarding ChatGPT is its potential to threaten academic integrity. Unlike traditional search engines or databases that simply retrieve and display pre-existing information, ChatGPT operates on a more dynamic principle. Built on the GPT architecture, It constructs a new response based on its understanding of the topic, the context provided, and the patterns it has observed in similar contexts during its training. This extensive pre-training allows it to have a broad base of knowledge from which it can draw upon, making its generated content rich and varied. ChatGPT does not simply retrieve knowledge; it can synthesize and create knowledge. Therefore, students can obtain original essays or responses in a matter of seconds without much effort without critical thinking. This convenient cycle of instant gratification would further motivate a student to opt for immediate answers provided by ChatGPT rather than taking time to understand complex concepts themselves. A student's ability to think critically can atrophy due to overreliance on ChatGPT. Essentially, ChatGPT not only deteriorates skill but also detracts a student from achieving the goal of education itself.

However, it is not the first time education has been met with the advent of new tools that supposedly "harm" education. The calculator, for instance, was forbidden initially after its introduction in the late 1960s. Many educators and parents believed that it would undermine the foundational arithmetic skills students were supposed to develop. The fear was that students would become overly reliant on this tool, losing their ability to perform basic calculations without it. However, calculators were then allowed to students over grade 8 starting from 1975, and were allowed to all grade levels five years later (Watters). Educators realized the calculator's potential as a convenient yet powerful tool. By automating simple calculations, students could focus on more complex mathematical concepts and problem-solving strategies. This shift in perspective allowed for a more holistic approach to mathematics, where the emphasis was not

just on rote calculations but on understanding the underlying principles and applications. Once students understood the fundamental mechanisms of arithmetic calculation or any other concept that could be automated by the calculator, calculators saved a considerable amount of time.

The same applies to a more recent example: search engines. The late 1990s and early 2000s witnessed the rapid proliferation of the internet and search engines. Just like calculators, these tools were initially met with skepticism in educational settings. The internet, with its vast reservoir of information, was seen as a potential distraction for students. There were concerns about the credibility of online sources, the potential for plagiarism, and the fear that students would rely too heavily on search engines, bypassing traditional research methods and critical thinking. However, as with the calculator, educators soon began to recognize the potential benefits of the internet and search engines in the learning process. Although "traditional" research skills were no longer needed, the advent of the internet called for new skills, such as the ability to discern inaccurate and accurate information or to choose appropriate sources from a myriad of them. Other than that, utilizing the search engine was no different from using encyclopedias or books. The fundamental skills of research still held, namely detecting bias, evaluating the credibility of sources, and synthesizing information from various resources to form a coherent understanding or argument. Thanks to the advent of search engines, students could minimize time waste during research; the act of rummaging through bodies of text without guarantee of finding desired information was no more.

Essentially ChatGPT is the next generation of search engines. Traditional search engine gives users a list of sources of information to choose from. From searching and obtaining information, there is a mediary stage of comparing sources and examining them. ChatGPT gets rid of the mediary stage and immediately provides the reader with whatever information is requested. As aforementioned, utilization of ChatGPT does not render skills completely useless nor bring dramatic change. ChatGPT isn't flawless; it "hallucinates" occasionally, providing factually wrong responses as if they were true. Therefore, certain skills like cross-checking will continue to be relevant while development of new skills are necessitated, like prompt engineering, the ability to craft specific and effective prompts to extract accurate and relevant information from ChatGPT. Consequently, rejecting ChatGPT entirely seems like a gross overreaction akin to forbidding the use of Google or calculators.

However, it is important to remember that ChatGPT has the remarkable capability to generate original content, which is a technology completely unprecedented. In that aspect, ChatGPT must be considered a tool *too* convenient. Google and calculator aids one's attempt to solve a problem, while ChatGPT can directly solve it, letting a person skip the whole thought process and instantly reach an answer. It eliminates the cognitive process which is supposed to be vitalized. Therefore, when used improperly, it is indeed a threat to academic integrity.

Suggestions for Educational Institutions

Using ChatGPT to streamline the cognitive process is encouraged, but using it to skip the whole cognitive process renders education meaningless. Unfortunately, at the current stage it is

hard to partially disable the features of ChatGPT; therefore, it is up to the student to manage oneself regarding the use of ChatGPT, and the educational system can provide instructions to help students use it appropriately.. The same applies to Google; a student can use Google to look up complete essays to plagiarize or paraphrase, while another can use Google to gather information for the assignment. Appropriate use largely depends on the user instead of the tool. The essence of education is not just the acquisition of knowledge but the development of critical thinking, problem-solving abilities, and a lifelong love for learning. When students use ChatGPT as a crutch, bypassing the cognitive processes essential for genuine understanding, they do themselves a disservice. However, when used as a supplementary tool – to clarify doubts, explore topics in depth, or even challenge one's understanding – ChatGPT can be invaluable.

Currently the appropriate use of ChatGPT is loosely defined or left to be decided by educators on an individual level. The appropriate use of ChatGPT, according to UC Berkeley Office of Ethics is limited to those "consistent with UC Berkeley's Principles of Community" while utilizing "Publicly-available information" (UC Berkeley). These principles of community, however, are broad and can be interpreted in various ways, making it a potentially ineffective guideline for regulating the use of a specific tool like ChatGPT. Harvard, similarly, has not devised guidelines specific to ChatGPT, but is "leveraging existing University policies" (Harvard University et al.) to determine its use. Applying existing standards to a revolutionary tool could potentially stifle its applications or fail to address its unique challenges. The challenge with ChatGPT, and AI in general, is to determine how to use it to enhance the learning experience rather than diminish it.

The key lies in fostering a sense of responsibility and discernment among students. Educators can guide students on how to use ChatGPT effectively, emphasizing its role as a supplementary tool rather than a primary source of information. For instance, as aforementioned, ChatGPT can be used to clarify challenging concepts given its excellent ability to explain high-level ideas in layman's terms. Instead of depending solely on ChatGPT, students can use it as a secondary source of information to further their understanding. Since teachers can answer a limited number of questions, and must prioritize the questions of underperforming students, students who desire to explore beyond the curriculum can take advantage of the knowledge of ChatGPT. ChatGPT, when utilized appropriately, is the equivalent to a teacher that can answer an unlimited number of questions regardless of time and location. As students demand more sophisticated knowledge from ChatGPT, they are also presented with an opportunity to practice cross-checking and validating information; given ChatGPT's occasional "hallucinations," students will be encouraged to cross-reference the information with trusted sources such as textbooks or scholarly articles. We are in a transitional phase in which we are challenged to deal with revolutionary tech that is expected to change various aspects of our lives. As we navigate this transitional phase, it becomes evident that our challenge is not just to integrate revolutionary tech, but to do so in a way that enriches the learning experience.

Conclusion

The introduction of ChatGPT into the educational sphere has undeniably sparked a whirlwind of discussions, debates, and deliberations. Its prowess offers a tantalizing glimpse into the future of education but also brings forth a slew of concerns. This research embarked on a journey to discern the appropriate and inappropriate applications of ChatGPT in education, striving to shed light on the nuanced interplay between this advanced technology and the foundational principles of learning.

The trajectory of ChatGPT, from its meteoric rise in popularity to the subsequent reservations from educational institutions, mirrors the broader narrative of technological integration into established systems. The societal pendulum, swinging between caution and curiosity, underscores the intricate dance of progress and preservation. History reminds us of the initial skepticism surrounding tools like calculators and search engines, which eventually found their harmonious place in the educational tapestry, augmenting rather than overshadowing the learning process. With ChatGPT, the challenge is multifaceted: ensuring it complements the educational journey, while safeguarding the essence of cognitive exploration and critical thinking. Yet, as we navigate this intricate landscape, broader ethical quandaries emerge. The potential for ChatGPT's misuse casts shadows over academic integrity and fairness. If ChatGPT becomes an ally for some students and remains elusive for others, are we inadvertently creating disparities in the educational arena? Moreover, the current discourse seems disproportionately skewed towards higher education, leaving the implications for pre-college education relatively uncharted. This oversight beckons a more holistic exploration, especially given the malleability and impressionability of younger minds. Such topics are beyond the scope of this paper, but it is certain that future research addressing ChatGPT’s impact on primary and secondary schools and the digital gap will be an important addition to the growing scholarship about the new AI tool.

Works Cited

Milmo, Dan. “CHATGPT Reaches 100 Million Users Two Months after Launch.” The Guardian, Guardian News and Media, 2 Feb. 2023, www.theguardian.com/technology/2023/feb/02/chatgpt-100-million-users-open-ai-fastest-growing-app.

Heaven, Will Douglas. “ChatGPT Is Going to Change Education, Not Destroy It.” MIT Technology Review, MIT Technology Review, 7 Apr. 2023, www.technologyreview.com/2023/04/06/1071059/chatgpt-change-not-destroy-education-openai/.

Watters, Audrey. “A Brief History of Calculators in the Classroom.” Hack Education, 12 Mar. 2015, hackeducation.com/2015/03/12/calculators.

UC Berkeley. Appropriate Use of ChatGPT and Similar AI Tools | Office of Ethics. ethics.berkeley.edu/privacy/appropriate-use-chatgpt-and-similar-ai-tools.

Harvard University, et al. “Guidelines for Using ChatGPT and Other Generative AI Tools at Harvard.” Office of the Provost, provost.harvard.edu/guidelines-using-chatgpt-and-other-generative-ai-tools-harvard.

Tlili, A., Shehata, B., Adarkwah, M.A. et al. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. Smart Learn. Environ. 10, 15 (2023). https://doi.org/10.1186/s40561-023-00237-x

Anders, Brent A. “Is using ChatGPT cheating, plagiarism, both, neither, or forward thinking?” Patterns, vol. 4, no. 3, 2023, p. 100694, https://doi.org/10.1016/j.patter.2023.100694.

Crawford, Joseph, et al. “Leadership is needed for ethical chatgpt: Character, assessment, and learning using Artificial Intelligence (AI).” Journal of University Teaching and Learning Practice, vol. 20, no. 3, 2023, https://doi.org/10.53761/1.20.3.02.

Hemsley, Bronwyn, et al. “Will AI Tech like ChatGPT Improve Inclusion for People with Communication Disability?” The Conversation, 19 May 2023, theconversation.com/will-ai-tech-like-chatgpt-improve-inclusion-for-people-with-communication-disability-196481.

Recent Advances in Mesenchymal Stem Cell-based Therapy for Diabetes Mellitus By Lancy ShangGuan

Abstract

Diabetes mellitus (DM) is one of the most devastating chronic autoimmune diseases and the leading cause of death in the world. It is characterized by impaired glucose homeostasis and insulin deficiency which can result in multi-systemic chronic complications like retinopathy and peripheral vascular diseases. Conventional therapies, including insulin replacement and oral hypoglycaemic agents, often fail to provide long-term glycaemic control and prevent diabetic complications. Therefore, an effective therapy that can preserve residual β-cell mass and protect newly generated β-cell from autoimmune destruction is desired. Mesenchymal stem cell (MSC)--based therapy has emerged as a promising avenue for the treatment of diabetes due to its immunomodulatory, regenerative, and differentiation capabilities. It can potentially produce, store and supply insulin to maintain glucose homeostasis in the body. This review focuses on the recent advances in MSC-based therapy for diabetes mellitus, including the development of stem cell therapy for diabetes and molecular mechanisms of action of MSCs. This review will explore the current challenges, limitations and potential future directions of MSC therapies in DM.

Keywords

Mesenchymal stem cells, Diabetes Mellitus, Diabetic complications, stem cell Therapy

Introduction

Diabetes Mellitus is a global health concern, around 537 millions adults are living with diabetes in the world in 2021 (“IDF Diabetes Atlas | Tenth Edition”). It is characterized by chronic hyperglycaemia resulting from defects in insulin secretion, insulin action, or both (Sheriff Sheik Abdulazeez). Hyperglycaemia causes multiple complications including nephropathy, neuropathy and retinopathy in diabetes mellitus patients (Rahim et al.).

The main strategy of the current treatment for decreasing the blood glucose level for diabetes patients is the exogenous supply of insulin. Although it successfully decreased the blood glucose level, it is not capable of completely mimicking endogenously secreted insulin released from pancreatic β-cells, which makes it hard to maintain the optimum level of blood glucose and often causes hypoglycaemic coma (Rahim et al.). As a result, there is a growing interest in exploring alternative therapeutic strategies, including regenerative medicine approaches like mesenchymal stem cell (MSC) therapy.

Mesenchymal stem cells are a type of adult stem cell found in various tissues, including bone marrow, adipose tissue, and umbilical cord tissue (M. Ezquer et al.). Beyond their differentiation capacity, mesenchymal stem cells possess immunomodulatory properties and secrete paracrine factors that influence neighbouring cells. These characteristics make MSC-based therapy an attractive option for diabetes treatment, as they can potentially regulate

immune responses, promote beta-cell regeneration, and mitigate inflammation (Chen, Du, et al.), (Dean & McEntyre), (M. Ezquer et al.), (Akash Kumar Singh et al.).

Diabetes mellitus

Diabetes mellitus is a metabolic, chronic disease that is characterized by abnormal levels of blood glucose which ultimately leads to serious damage to the blood vessels, heart, kidneys and nerves (World). The aetiology of diabetes mellitus is multifaceted, with genetic, environmental, and lifestyle factors playing crucial roles in its pathogenesis.

DM can be categorized into three major types – Type 1 diabetes, type 2 diabetes and Gestational diabetes (Paras Kumar Mishra et al.). Type 1 diabetes is characterized by insulin-producing beta-cell destruction due to the immune complex-mediated attack, which usually leads to a deficiency of insulin in the body that is required for glucose metabolism (Sheriff Sheik Abdulazeez), (Paras Kumar Mishra et al.). Type 2 diabetes is characterized by the body's inability to respond to secreted insulin or insufficient insulin synthesis (Sheriff Sheik Abdulazeez).

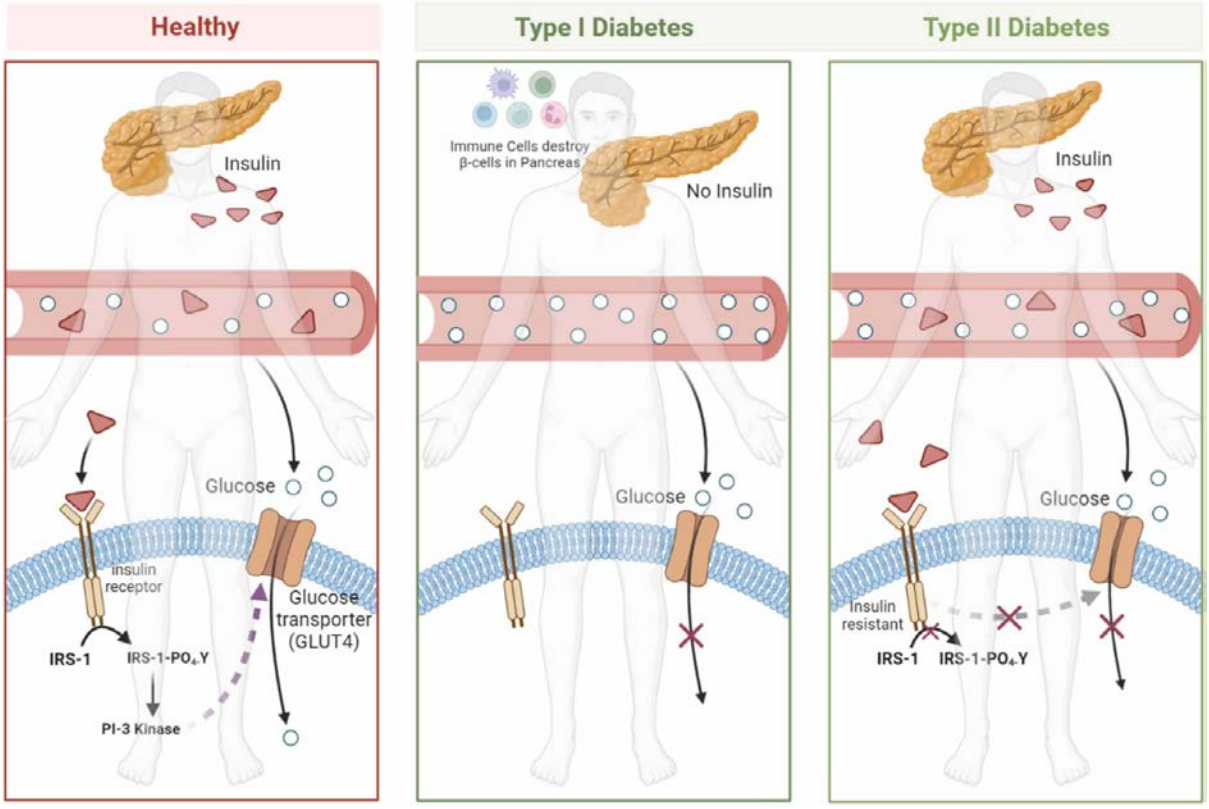


Figure 1. Comparison of Type 1 and Type 2 diabetes; Type 1 diabetes's pancreas does not produce insulin, Type 2 diabetes is the body unable to respond to insulin properly which results in impaired glucose metabolism. (Akash Kumar Singh et al.)

Gestational DM is a type of diabetes that appears during the second trimester of gestation causing high blood glucose levels and disappears after the birth of the baby (Paras Kumar Mishra et al.). The complications of DM are diverse and encompass cardiovascular disease, neuropathy, nephropathy, and retinopathy, among others. Management of DM is multifaceted and involves lifestyle modifications, pharmacotherapy, and insulin administration when necessary. The global prevalence of DM has been steadily rising, posing significant challenges to healthcare systems worldwide. Therefore, a comprehensive understanding of the underlying mechanisms, risk factors, and effective management strategies is essential for mitigating the burden of this disease on public health (Sheriff Sheik Abdulazeez), (Paras Kumar Mishra et al.), (World).

Current Treatment

Traditional therapeutic methods such as insulin injection and pancreas transplant are the most prominent treatments for diabetes mellitus, although they can successfully decrease the blood glucose level in hyperglycaemic patients, there are still many difficulties and concerns about them (Rahim et al.), (Paras Kumar Mishra et al.).

Exogenous insulin through daily injections is frequently associated with concerns like a failure in glucose metabolism control due to the inability to completely mimic endogenously secreted insulin from beta cells, therefore it does not represent a definitive cure for DM patients (M. Ezquer et al.), (Paras Kumar Mishra et al.).

Around 1,000 DM patients received pancreas transplants from cadaveric donors in the USA in 2021 as an alternative to insulin administration (1Scientific Registry of Transplant Recipients, Hennepin Healthcare Research Institute, Minneapolis, MN et al.). Although islet transplantation from cadaveric subjects was proposed to be a permanent cure, the increased risk of infections, the need for immunosuppressive drugs, and their unavailability restricted its use (M. Ezquer et al.).

Oral medications are primarily used to treat type 2 diabetes. They may not be suitable for individuals with type 1 diabetes, as type 1 diabetes requires insulin replacement therapy due to the absence of insulin production. It may become less effective over time as diabetes progresses. Oral medications can cause side effects, which may vary depending on the specific medication. Common side effects include gastrointestinal discomfort, vomiting, stomach upset etc (*Side Effects of Diabetes Medication | London Diabetes Centre*).

These treatments reverse neither disease progression nor cellular dysfunction (Huang et al.). Therefore, methods to promote either the development of the body's existing beta cells or the supply of stem cell-derived insulin-producing cells would offer future therapy possibilities for patients with complicated diabetes (Paras Kumar Mishra et al.).

Development of MSC-Based Therapy for DM

The development of stem cell therapy for diabetes mellitus represents a promising avenue in the pursuit of novel treatment modalities for this complex metabolic disorder. Stem cells are

unspecialized, self-renewing cells that can develop into many different cell types, which can range from muscle cells to brain cells (*Default - Stanford Medicine Children's Health*)(Figure 2).

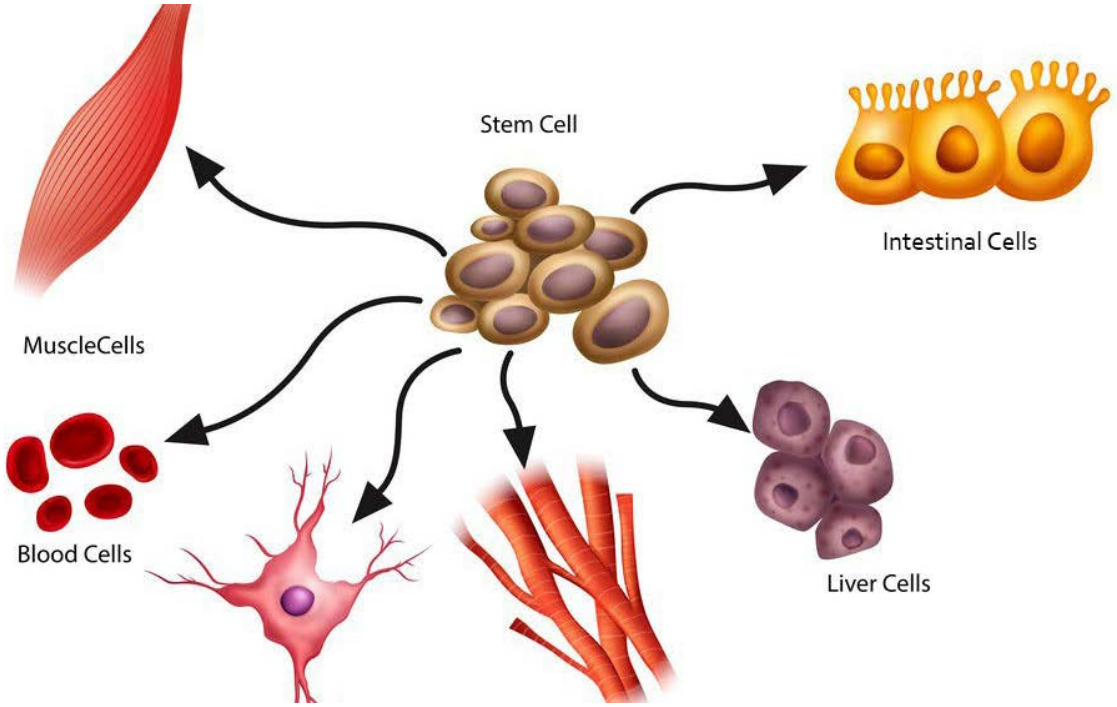


Figure 2. Stem cells can differentiate into many different types of cells. Including muscle cells, blood cells, liver cells etc. (Cade Hildreth (CEO)

Stem cell therapy is a form of regenerative medicine that aims to repair damaged cells in the body by modulating the immune system and reducing inflammation (Cona). Recent advances in stem cell research have provided avenues to generate insulin-producing cells in vitro, holding great promise for both type 1 and type 2 diabetes patients. It is expected that this therapeutic method should produce, store, and supply insulin to maintain glucose homeostasis (Rahim et al.).

Over the past few decades, significant progress has been made in this field, with various sources of stem cells, including embryonic, induced pluripotent, and adult stem cells, explored for their therapeutic potential. The differentiation of these stem cells into insulin-producing β -like cells has been achieved through directed differentiation protocols, often involving the manipulation of key signalling pathways and transcription factors. Preclinical studies in animal models have demonstrated encouraging results in terms of glycaemic control and insulin secretion following stem cell transplantation (Rahim et al.), (Paras Kumar Mishra et al.), (Cona).

Mesenchymal stem cells (MSCs) have emerged as a compelling candidate in the field of regenerative medicine for the treatment of DM, they are currently the most widely used stem cells in clinical trials for diabetes (Eleonora de Klerk & Matthias Hebrok). MSCs are multipotent adult stem cells that can be extracted from a variety of sources, including bone marrow, adipose tissue, dental pulp, liver, kidney, umbilical cord blood, Wharton's jelly etc. (M. Ezquer et al., 2014), (Jayasinghe et al.), (Figure 3).

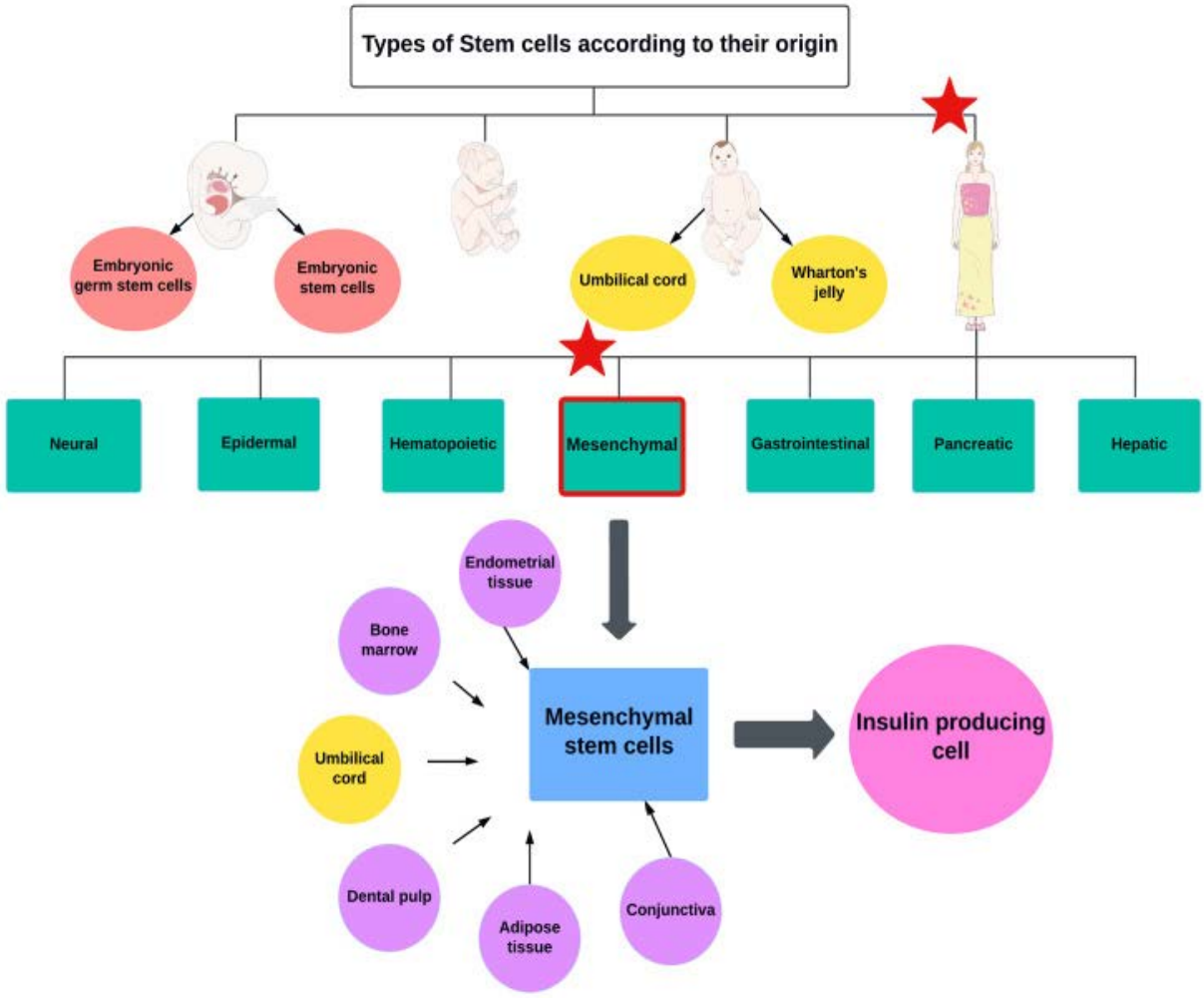


Figure 3. Types of stem cells according to their origin and source of mesenchymal stem cells include bone marrow, adipose tissue, umbilical cord tissue etc. (Jayasinghe et al.)

These cells possess unique immunomodulatory properties and the capacity for differentiation into multiple cell lineages, making them particularly attractive for diabetes therapy. Their immunomodulatory and proangiogenic roles can be used to help arrest β-cell destruction, facilitate endogenous β-cell regeneration, preserve residual β-cell mass and prevent disease recurrence, making them the ideal candidates for the comprehensive treatment of diabetic patients (M. Ezquer et al.). Furthermore, their paracrine signalling and secretion of trophic factors have been shown to exert anti-inflammatory and anti-apoptotic effects on damaged pancreatic tissue. Clinical trials have demonstrated its ability in lowering levels of fasting blood sugar, haemoglobin A1c, and C-peptide, and in treating microvascular complications (Jayasinghe et al.). While preclinical studies and early-phase clinical trials have demonstrated the safety and potential efficacy of MSC-based therapies, challenges such as optimization of MSC sourcing,

standardization of protocols, and long-term safety assessments remain pertinent. The evolving landscape of MSC-based treatments for DM holds promise for addressing the underlying pathophysiological mechanisms of the disease and offers a potential avenue to improve glycaemic control and reduce diabetes-related complications.

Action Mechanism of MSCs

Differentiation into Induced Pluripotent Cells

One key mechanism involves the differentiation of MSCs into induced pluripotent cells (iPSCs), which subsequently contribute to the regeneration of pancreatic islet cells. MSCs possess the inherent capability to differentiate into multiple cell lineages, including insulin-producing β-cells (Dave), (Walczak et al.). Through a series of complex molecular signalling pathways, such as the Wnt and Notch pathways, MSCs can be induced to undergo differentiation into iPSCs, which exhibit a greater degree of plasticity and can differentiate into various cell types (Wang et al.), (Zhang et al.). Upon transplantation into the diabetic host, these iPSCs can further differentiate into functional β-like cells, leading to enhanced insulin production and glucose homeostasis (Walczak et al.). Furthermore, MSCs exert their therapeutic effects through paracrine signalling, immunomodulation, and anti-inflammatory properties, which collectively contribute to the amelioration of diabetes-related complications (Walczak et al.), (Zang et al.).

Regeneration of pancreatic islet β-cells

MSC-based therapy capitalizes on a multifaceted molecular mechanism for the regeneration of pancreatic islet β-cells in the context of diabetes treatment. The fundamental element of this mechanism lies in the promotion of β-cell regeneration. MSCs, owing to their plasticity, can differentiate into β-like cells through the activation of crucial signalling pathways, including Hedgehog, PDX-1, and Ngn3 (Zang et al.), (Barachini et al.). This transdifferentiating process leads to the emergence of functional insulin-producing cells, thereby augmenting the insulin reservoir within the pancreatic islets. Furthermore,(e.g., VEGF, HGF), cytokines, and chemokines (Zang et al.), (Khatri et al.). These molecules collaboratively foster a conducive microenvironment within the diabetic pancreas, facilitating β-cell survival, proliferation, and functional restoration. Importantly, MSCs also exhibit immunomodulatory properties, effectively suppressing inflammatory responses and autoimmune reactions that pose a threat to β-cell integrity. In this manner, MSC-based therapy not only replenishes the depleted β-cell population but also shields these newly generated β- cells from immune-mediated destruction (Barachini et al.).

Protection of endogenous pancreatic islet β-cells

Beyond their potential for differentiation into insulin-producing β-like cells, MSCs exhibit an inherent ability to preserve existing β-cells. This protective effect is mediated through paracrine signalling, where MSCs secrete a repertoire of bioactive factors, including

anti-inflammatory cytokines (e.g., IL-10, TGF-β), chemokines, and growth factors (e.g., IGF-1) (Khatri et al.), (Barachini et al.). These molecules collectively mitigate local inflammation and dampen the autoimmune responses that commonly lead to β-cell destruction in diabetes. By fostering an immunomodulatory milieu and curtailing detrimental immune reactions, MSC-based therapy not only replenishes the β-cell population but also shields these cells from immune-mediated harm. This dual action—enhanced regeneration and protection of endogenous β-cells—constitutes a comprehensive molecular approach that holds promise for the restoration of normoglycemia in DM(Jayasinghe et al.), (Khatri et al.), (Barachini et al.).

Amelioration of insulin resistance

MSCs exert a significant influence on insulin sensitivity. This effect is mediated through the secretion of paracrine factors, such as adiponectin, interleukin-6 (IL-6), and tumour necrosis factor-alpha (TNF-α), which play pivotal roles in modulating insulin signalling pathways (Seng Kar Yap et al.), (Chen, Fan, et al.). By enhancing insulin sensitivity in peripheral tissues, particularly in adipose, muscle, and liver tissues, MSC-based therapy effectively mitigates the insulin resistance commonly associated with type 2 DM (Huang et al.). Furthermore, MSCs' anti-inflammatory properties contribute to the reduction of chronic low-grade inflammation, a hallmark of insulin resistance. Through these intricate molecular mechanisms, MSC-based therapy holds promise for not only replenishing β-cell populations but also improving systemic insulin responsiveness, thus providing a comprehensive approach toward achieving glycaemic control in diabetes (Seng Kar Yap et al.), (Chen, Fan, et al.), (Huang et al.), (Deng et al.).

Discussion

The recent advances in MSC-based therapy for DM represents a notable leap forward in the pursuit of innovative treatments for this debilitating condition. These advancements underscore the therapeutic potential of MSCs in diabetes management, with compelling preclinical and early-phase clinical data demonstrating their ability to address key pathophysiological aspects of the disease. MSCs, with their immunomodulatory properties and capacity to stimulate tissue repair and regeneration, have exhibited promise in both type 1 and type 2 diabetes contexts. In the past decade, several notable studies have underscored the multifaceted roles of MSCs in diabetes management. For instance, a recent investigation by Barachini et al in 2023 (Barachini et al.) demonstrated that the transplantation of MSCs in type 1 diabetic murine models led to a substantial reduction in autoimmune responses against pancreatic islet cells, resulting in improved glycaemic control. Furthermore, the study by Zang et al. in 2017 (Zang et al.) and Gao et al. in 2022 (Gao et al.) shed light on the reparative effects of MSCs on beta-cell function in type 2 diabetes patients. Their findings revealed that MSC-based therapies not only attenuated insulin resistance but also promoted the proliferation and survival of insulin-secreting beta-cells. Furthermore, the ongoing exploration of alternative MSC sources and optimization of their characteristics, such as genetic engineering to enhance functionality, showcase the dynamic evolution of this field. Nevertheless, challenges remain, including the

need for standardized protocols, long-term safety assessments, and comprehensive mechanistic insights. The clinical translation of MSC-based therapies into widespread practice hinges on addressing these concerns and further refining their applications. Overall, recent advances in MSC-based therapy for diabetes mark a significant stride toward more effective and potentially curative approaches for this prevalent metabolic disorder.

The future directions of MSC-based therapy in the treatment of DM are poised to shape the landscape of regenerative medicine. These advancements cover a broad strategy targeted at maximizing the therapeutic potential of MSCs. Genetic engineering of MSCs for heightened functionality, encompassing enhanced differentiation into insulin-producing cells and augmented secretion of paracrine factors, holds substantial promise. Combination therapies, involving MSCs alongside other regenerative or immunomodulatory interventions, may emerge as a potent strategy, particularly when integrated with β-cell replacement approaches (Jayasinghe et al.) (M. Ezquer et al.). Finally, endeavours to enhance cost-effectiveness and accessibility, and the establishment of stem cell banking repositories will solidify MSC-based therapies as a pivotal force in the diabetes treatment paradigm.

Conclusion

MSC-based therapy represents a promising avenue for the treatment of DM. While current treatments effectively manage the disease, they fall short of providing a cure and often come with unwanted side effects. MSCs offer a unique opportunity to address the underlying pathology of diabetes by promoting β-cell regeneration, modulating the immune response, and reducing inflammation. Early results from clinical trials evaluating the security and effectiveness of MSC-based treatment are promising. MSC-based therapy may hold the key to a more efficient and long-lasting therapeutic option for DM with further study and protocol improvement.

Acknowledgements

The author would like to acknowledge and give her warmest thanks to Dr. Hazal Gezmis and Dr. Hakan Coskun who made this work possible. The author would also like to thank the Cambridge Centre for International Research (CCIR) for the support throughout this process.

Works Cited

1Scientific Registry of Transplant Recipients, Hennepin Healthcare Research Institute, Minneapolis, MN, et al. “Pancreas.” *Hrsa.gov*, 2021, srtr.transplant.hrsa.gov/annual_reports/2021/Pancreas.aspx. Accessed 31 Aug. 2023.

Aguayo-Mazzucato, Cristina, and Susan Bonner-Weir. “Stem Cell Therapy for Type 1 Diabetes Mellitus.” *Nature Reviews Endocrinology*, vol. 6, no. 3, Nature Portfolio, Mar. 2010, pp. 139–48, https://doi.org/10.1038/nrendo.2009.274. Accessed 6 Sept. 2023.

Akash Kumar Singh, et al. “Recent Trends and Advances in Type 1 Diabetes Therapeutics: A Comprehensive Review.” *European Journal of Cell Biology*, vol. 102, no. 2, Elsevier BV, June 2023, pp. 151329–29, https://doi.org/10.1016/j.ejcb.2023.151329. Accessed 20 Aug. 2023.

Bani, R., et al. “Efficacy of Stem Cell Application in Diabetes Mellitus: Promising Future Therapy for Diabetes and Its Complications.” *Cureus*, Cureus, Inc., Feb. 2021, https://doi.org/10.7759/cureus.13563. Accessed 20 Aug. 2023.

Barachini, Serena, et al. “Mesenchymal Stem Cell in Pancreatic Islet Transplantation.” *Biomedicines*, vol. 11, no. 5, Multidisciplinary Digital Publishing Institute, May 2023, pp. 1426–26, https://doi.org/10.3390/biomedicines11051426. Accessed 6 Sept. 2023.

---. “Mesenchymal Stem Cell in Pancreatic Islet Transplantation.” *Biomedicines*, vol. 11, no. 5, Multidisciplinary Digital Publishing Institute, May 2023, pp. 1426–26, https://doi.org/10.3390/biomedicines11051426. Accessed 21 Sept. 2023.

Cade Hildreth (CEO. “What Are Stem Cells? What They Are and What They Do.” *BioInformant*, BioInformant, 27 Sept. 2022, bioinformant.com/what-are-stem-cells/. Accessed 1 Sept. 2023.

Chen, Guang, et al. “Human Umbilical Cord-Derived Mesenchymal Stem Cells Ameliorate Insulin Resistance via PTEN-Mediated Crosstalk between the PI3K/Akt and Erk/MAPKs Signaling Pathways in the Skeletal Muscles of Db/Db Mice.” *Stem Cell Research & Therapy*, vol. 11, no. 1, Springer Science+Business Media, Sept. 2020, https://doi.org/10.1186/s13287-020-01865-7. Accessed 6 Sept. 2023.

Chen, Shuai, et al. “Current Progress in Stem Cell Therapy for Type 1 Diabetes Mellitus.” *Stem Cell Research & Therapy*, vol. 11, no. 1, Springer Science+Business Media, July 2020, https://doi.org/10.1186/s13287-020-01793-6. Accessed 16 Aug. 2023.

Cona, Louis A. “Mesenchymal Stem Cells (MSCs): A Comprehensive Overview of Their Properties and Uses.” *Dvcstem.com*, 25 Aug. 2023, www.dvcstem.com/post/what-are-mesenchymal-stem-cells. Accessed 2 Sept. 2023.

---. “Stem Cell Therapy: A Comprehensive Overview (2023).” *Dvcstem.com*, 25 Aug. 2023, www.dvcstem.com/post/stem-cell-therapy. Accessed 1 Sept. 2023.

Dave, Shruti D. “Mesenchymal Stem Cells Derived in Vitro Transdifferentiated Insulin-Producing Cells: A New Approach to Treat Type 1 Diabetes.” *Advanced Biomedical Research*, vol. 3, no. 1, Medknow, Jan. 2014, pp. 266–66, https://doi.org/10.4103/2277-9175.148247. Accessed 6 Sept. 2023.

---. “Mesenchymal Stem Cells Derived in Vitro Transdifferentiated Insulin-Producing Cells: A New Approach to Treat Type 1 Diabetes.” *Advanced Biomedical Research*, vol. 3, no. 1, Medknow, Jan. 2014, pp. 266–66, https://doi.org/10.4103/2277-9175.148247. Accessed 6 Sept. 2023.

Dean, Laura, and Jo McEntyre. “Introduction to Diabetes.” *Nih.gov*, National Center for Biotechnology Information (US), 7 July 2004, www.ncbi.nlm.nih.gov/books/NBK1671/. Accessed 13 Aug. 2023.

“Default - Stanford Medicine Children’s Health.” *Stanfordchildrens.org*, 2019, www.stanfordchildrens.org/en/topic/default?id=what-are-stem-cells-160-38. Accessed 1 Sept. 2023.

Deng, Zihui, et al. “Infusion of Adipose-Derived Mesenchymal Stem Cells Inhibits Skeletal Muscle Mitsugumin 53 Elevation and Thereby Alleviates Insulin Resistance in Type 2 Diabetic Rats.” *Molecular Medicine Reports*, Spandidos Publishing, Apr. 2018, https://doi.org/10.3892/mmr.2018.8901. Accessed 6 Sept. 2023.

Eleonora de Klerk, and Matthias Hebrok. “Stem Cell-Based Clinical Trials for Diabetes Mellitus.” *Frontiers in Endocrinology*, vol. 12, Frontiers Media, Feb. 2021, https://doi.org/10.3389/fendo.2021.631463. Accessed 1 Sept. 2023.

Endocrine Society. “Diabetes Treatments.” *Endocrine.org*, Endocrine Society, 24 Jan. 2022, www.endocrine.org/patient-engagement/endocrine-library/diabetes-treatments#:~:text=Diabetes%20can%20be%20treated%20with,insulin%20to%20maintain%20glucose%20control. Accessed 16 Aug. 2023.

Figliuzzi, Marina. “Mesenchymal Stem Cells Help Pancreatic Islet Transplantation to Control Type 1 Diabetes.” *World Journal of Stem Cells*, vol. 6, no. 2, Baishideng Publishing Group, Jan. 2014, pp. 163–63, https://doi.org/10.4252/wjsc.v6.i2.163. Accessed 21 Sept. 2023.

Gao, Shuang, et al. “Mesenchymal Stem Cells (MSCs): A Novel Therapy for Type 2 Diabetes.” *Stem Cells International*, vol. 2022, Hindawi Publishing Corporation, Aug. 2022, pp. 1–17, https://doi.org/10.1155/2022/8637493. Accessed 6 Sept. 2023.

Huang, Qiulan, et al. “Mesenchymal Stem Cells: An Excellent Candidate for the Treatment of Diabetes Mellitus.” *International Journal of Endocrinology*, vol. 2021, Hindawi Publishing Corporation, May 2021, pp. 1–11, https://doi.org/10.1155/2021/9938658. Accessed 5 Sept. 2023.

---. “Mesenchymal Stem Cells: An Excellent Candidate for the Treatment of Diabetes Mellitus.” *International Journal of Endocrinology*, vol. 2021, Hindawi Publishing Corporation, May 2021, pp. 1–11, https://doi.org/10.1155/2021/9938658. Accessed 6 Sept. 2023.

“IDF Diabetes Atlas | Tenth Edition.” *Diabetesatlas.org*, 2022, diabetesatlas.org/#:~:text=Diabetes%20around%20the%20world%20in%202021%3A,and%20783%20million%20by%202045. Accessed 13 Aug. 2023.

Jayasinghe, Maleesha, et al. “The Role of Mesenchymal Stem Cells in the Treatment of Type 1 Diabetes.” *Cureus*, Cureus, Inc., July 2022, https://doi.org/10.7759/cureus.27337. Accessed 5 Sept. 2023.

Khatri, Rahul, et al. “Mesenchymal Stem Cells Promote Pancreatic β-Cell Regeneration through Downregulation of FoxO1 Pathway.” *Stem Cell Research & Therapy*, vol. 11, no. 1, Springer Science+Business Media, Nov. 2020, https://doi.org/10.1186/s13287-020-02007-9. Accessed 6 Sept. 2023.

Li, Ming, and Susumu Ikehara. “Stem Cell Treatment for Type 1 Diabetes.” *Frontiers in Cell and Developmental Biology*, vol. 2, Frontiers Media, Mar. 2014, https://doi.org/10.3389/fcell.2014.00009. Accessed 26 Aug. 2023.

Li, Yanju, et al. “Efficacy of Mesenchymal Stem Cell Transplantation Therapy for Type 1 and Type 2 Diabetes Mellitus: A Meta-Analysis.” *Stem Cell Research & Therapy*, vol. 12, no. 1, Springer Science+Business Media, May 2021, https://doi.org/10.1186/s13287-021-02342-5. Accessed 6 Sept. 2023.

M. Ezquer, et al. “Mesenchymal Stem Cell Therapy in Type 1 Diabetes Mellitus and Its Main Complications: From Experimental Findings to Clinical Practice.” *Journal of Stem Cell Research & Therapy*, 2014, www.semanticscholar.org/paper/Mesenchymal-Stem-Cell-Therapy-in-Type-1-Diabetes-to-Ezquer-Arango-Rodr%C3%ADguez/df6acef00f07a88654f0f87957c575cf8dbfebe3. Accessed 20 Aug. 2023.

Mathur, Aanchal, et al. “The Safety and Efficacy of Mesenchymal Stem Cells in the Treatment of Type 2 Diabetes- a Literature Review.” *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*, vol. Volume 16, Dove Medical Press, Mar. 2023, pp. 769–77, https://doi.org/10.2147/dmso.s392161. Accessed 6 Sept. 2023.

Moreira, Alvaro, et al. “Therapeutic Potential of Mesenchymal Stem Cells for Diabetes.” *Journal of Molecular Endocrinology*, vol. 59, no. 3, Bioscientifica, Oct. 2017, pp. R109–20, https://doi.org/10.1530/jme-17-0117. Accessed 6 Sept. 2023.

Nemati, Marzieh, et al. “Efficiency of Stem Cell (SC) Differentiation into Insulin-Producing Cells for Treating Diabetes: A Systematic Review.” *Stem Cells International*, vol. 2021, Hindawi Publishing Corporation, Feb. 2021, pp. 1–9, https://doi.org/10.1155/2021/6652915. Accessed 6 Sept. 2023.

Paras Kumar Mishra, et al. “Stem Cells as a Therapeutic Target for Diabetes.” *Frontiers in Bioscience*, vol. 15, no. 1, Frontiers in Bioscience, Jan. 2010, pp. 461–61, https://doi.org/10.2741/3630. Accessed 26 Aug. 2023.

Rahim, Fakher, et al. “Stem Cell Therapy for Patients with Diabetes: A Systematic Review and Meta-Analysis of Metabolomics-Based Risks and Benefits.” *Non-Coding RNA Investigation*, vol. 2, AME Publishing Company, Nov. 2018, pp. 40–40, https://doi.org/10.21037/sci.2018.11.01. Accessed 23 Aug. 2023.

Seng Kar Yap, et al. “Human Umbilical Cord Mesenchymal Stem Cell-Derived Small Extracellular Vesicles Ameliorated Insulin Resistance in Type 2 Diabetes Mellitus Rats.”

Pharmaceutics, vol. 14, no. 3, Multidisciplinary Digital Publishing Institute, Mar. 2022, pp. 649–49, https://doi.org/10.3390/pharmaceutics14030649. Accessed 6 Sept. 2023.

Sheriff Sheik Abdulazeez. "Diabetes Treatment: A Rapid Review of the Current and Future Scope of Stem Cell Research." *Journal of the Saudi Pharmaceutical Society*, vol. 23, no. 4, Elsevier BV, Sept. 2015, pp. 333–40, https://doi.org/10.1016/j.jsps.2013.12.012. Accessed 13 Aug. 2023.

"Side Effects of Diabetes Medication | London Diabetes Centre." *London Diabetes Centre*, 3 Dec. 2021, londondiabetes.com/news-and-events/common-side-effects-of-diabetes-medication/. Accessed 5 Sept. 2023.

Song, Na, et al. "Mesenchymal Stem Cell Immunomodulation: Mechanisms and Therapeutic Potential." *Trends in Pharmacological Sciences*, vol. 41, no. 9, Elsevier BV, Sept. 2020, pp. 653–64, https://doi.org/10.1016/j.tips.2020.06.009. Accessed 21 Sept. 2023.

Walczak, Maciej, et al. "Directed Differentiation of Human IPSC into Insulin Producing Cells Is Improved by Induced Expression of PDX1 and NKX6.1 Factors in IPC Progenitors." *Journal of Translational Medicine*, vol. 14, no. 1, BioMed Central, Dec. 2016, https://doi.org/10.1186/s12967-016-1097-0. Accessed 6 Sept. 2023.

---. "Directed Differentiation of Human IPSC into Insulin Producing Cells Is Improved by Induced Expression of PDX1 and NKX6.1 Factors in IPC Progenitors." *Journal of Translational Medicine*, vol. 14, no. 1, BioMed Central, Dec. 2016, https://doi.org/10.1186/s12967-016-1097-0. Accessed 6 Sept. 2023.

Wang, Haiying, et al. "The Wnt Signaling Pathway in Diabetic Nephropathy." *Frontiers in Cell and Developmental Biology*, vol. 9, Frontiers Media, Jan. 2022, https://doi.org/10.3389/fcell.2021.701547. Accessed 6 Sept. 2023.

Wang, Mengyuan, et al. "Mesenchymal Stem Cell-Based Immunomodulation: Properties and Clinical Application." *Stem Cells International*, vol. 2018, Hindawi Publishing Corporation, June 2018, pp. 1–12, https://doi.org/10.1155/2018/3057624. Accessed 21 Sept. 2023.

World. "Diabetes." *Who.int*, World Health Organization: WHO, 13 May 2019, www.who.int/health-topics/diabetes#tab=tab_1. Accessed 31 Aug. 2023.

Wu, Xiaomo, et al. "Mesenchymal Stromal Cell Therapies: Immunomodulatory Properties and Clinical Progress." *Stem Cell Research & Therapy*, vol. 11, no. 1, Springer Science+Business Media, Aug. 2020, https://doi.org/10.1186/s13287-020-01855-9. Accessed 21 Sept. 2023.

Zang, Li, et al. "Mesenchymal Stem Cell Therapy in Type 2 Diabetes Mellitus." *Diabetology & Metabolic Syndrome*, vol. 9, no. 1, Springer Science+Business Media, May 2017, https://doi.org/10.1186/s13098-017-0233-1. Accessed 6 Sept. 2023.

Zhang, Han, et al. "Regulatory Mechanisms of the Wnt/β-Catenin Pathway in Diabetic Cutaneous Ulcers." *Frontiers in Pharmacology*, vol. 9, Frontiers Media, Oct. 2018, https://doi.org/10.3389/fphar.2018.01114. Accessed 6 Sept. 2023.

From Black Hole Thermodynamics to Holographic Principle By Chuyue Zhang

Abstract

The Hawking radiation leads to the rise of the information paradox as the universe loses information about matter inside black holes as the black hole thermally evaporates. This paradox underscores the conflict between general relativity and quantum mechanics. Quantum gravity seeks to unify general relativity and quantum mechanics to make predictions at the Planck scale where standard model, quantum field theory, and general relativity are predicted to break down. This reconciliation has an interesting manifestation in the correspondence between black hole mechanics and laws of thermodynamics. This paper uses this correspondence as a starting point to analyze its implications for degrees of freedom in quantum gravity. With reference to research articles, including Susskind's "The World as a Hologram" and G.'t Hooft's "Dimensional Reduction in Quantum Gravity," this paper introduces the development of intriguing speculations about a novel framework for quantum gravity—the holographic principle, which posits that the information contained within a three-dimensional system can be adequately encoded on a two-dimensional surface.

Introduction

Black holes are predicted by Einstein's field equation, the pillar of general relativity. It is composed of a singularity and an event horizon. Anything that crosses the event horizon will be unable to escape the black hole's gravitational pull, being reduced to mass, charge, and momentum—black holes are "hairless." Initially, the black hole's event horizon which is directly proportional to mass is supposed to be unable to decrease. However, the discovery of Hawking's radiation suggested otherwise. The black hole can actually "radiate" by pulling apart anti-particles that form spontaneously due to vacuum fluctuation at the edge of the event horizon. Particles with positive energy that are released constitute the radiation and those with negative energy that are captured decrease the mass of the black hole. However, now that the black hole is found to have the potential of eventually "evaporating", a problem dubbed the information loss paradox arises. If the radiation has a thermal nature, then the pure state which the matters are in before falling into the black hole has been reduced to a mixed state when they are radiated out thermally—the information describing the state of the in-falling matters has been lost during the process. This paradox represents the insufficiency of general relativity because the quantum mechanics description of a black hole implies that the event horizon is not able to completely veil the black hole's interior from outside observers due to the quantization of matters (S.D. Mathur 3). Since in the vicinity of Black hole, both general relativity and quantum mechanics effect need to be taken into account, a field named quantum gravity comes in, attempting to unify general relativity and quantum mechanics in order to extrapolate predictions at the Planck scale such as at the singularity. This potential combination of general relativity and quantum mechanics has an interesting manifestation in the correspondence between black hole mechanics and laws of thermodynamics. Therefore, using black holes as the subject of examination, this

paper seeks to use this correspondence to analyze its implication of degrees of freedom, which is positively related to entropy, an important measure of information, in quantum gravity.

Method

The research method this paper will employ is a literature review. Since the field of focus is theoretical physics, the research would not be based on data analysis. The research will draw information from two articles: The World as a Hologram and Dimensional Reduction in Quantum Gravity, by Susskind and G.'t Hooft, respectively. The arguments made in both papers will be summarized and analyzed to provide support for this paper's thesis. Information will also be extracted from papers like "What Exactly is the Information Loss Paradox" and videos like "World as a Hologram Explained" to supplement understanding.

Laws of Thermodynamics and Black Hole Mechanics

The Zeroth Law of Thermodynamics

Thermodynamics primarily regards how work, energy, temperature, and heat relate to each other. There are four laws that dictate how these parameters interact.

The zeroth laws indicates that if thermodynamic system A is in thermal equilibrium with system C, and thermodynamic system B is also in thermal equilibrium with system C, then system A and system B must be in thermal equilibrium with each other. Being in thermal equilibrium means that the systems do not exchange heat over time. In other words, the systems have the same temperature. Considering the aspect of temperature will make this law more understandable. Temperature is the average kinetic energy, a quantity proportional to mass and velocity squared, of a system of particles. If two systems have particles with the same average kinetic energy, statistically, the overall change in kinetic energy of a system due to collisions between particles in the two systems will be 0, thus no net exchange of heat.

The First Law of Thermodynamics

The first law is built upon the equation: $dU = Q - W$: the change in internal energy (d U) of a closed system is equal to the difference between energy inputted as heat $(Q)$and the work done by the system to the surroundings(W). The internal energy U is the total potential energy and kinetic energy stored between atoms and molecules in the system which, under ideal conditions, the contribution of potential energy is negligible. U, constituted primarily by kinetic energy, is thus dependent on pressure, temperature, number of molecules, and volume, whose relation is governed by the ideal gas equation: $PV = nRT$. Note that by doing work on the surrounding, the energy of the system decreases as it is transferred out, therefore there's a negative sign in front of W. For instance, when a water bottle is put on a fire, heat is added to the water in the bottle, and the water's temperature increases. When a cylinder of gas is capped with a movable piston, pressing the piston inward would be doing positive work on the gas, and the gas's internal energy increases while allowing the piston to move outward means doing positive work on the environment thus negative work on the gas, so internal energy decreases. In essence,

this law concerns that energy is neither created nor destroyed, it can only be transferred or converted.

The Second Law of Thermodynamics

The second law regards the direction of natural processes. It states that without external energy inputted, heat flows naturally from hot bodies to cold bodies, but not the reverse. Entropy S can be calculated by Q/T. If heat Q has been transferred between two systems, one with higher T and one with lower T, the change in S would be $\frac{Q}{T_{low}} - \frac{Q}{T_{high}}$, making change in S positive. If two temperatures are the same, then the change in S is 0. Therefore, the theorem that the total entropy can never decrease is established. (Drake) This law also indicates that when the change in entropy is zero, a process could be considered reversible since an infinitesimal change could reverse the direction of heat flow.

The Third Law of Thermodynamics

The third law suggests that there is a constant value that the entropy of a system will eventually approach as the temperature tends to absolute zero (0 Kelvin) when the system is in a ground state with minimal thermal energy. If the ground state is a unique state, meaning there's only one micro-state possible for the system, then the residual entropy would be zero.

The Zeroth Law of Black Hole Mechanics

The laws of black hole mechanics seek to describe the properties of black holes. The zeroth law says that, for stationary black holes, the surface gravity is constant across the black hole's event horizon, meaning that on every location of the event horizon, the gravitational pull by the black hole is constant. For spherical black holes, the value of surface gravity is: $\frac{c^2}{d}$, where d is the diameter of the event horizon. The event horizon is the area of the black hole since entering the event horizon means an inescapable destiny of falling into the singularity. Since diameter d is related to the black hole radius: $\frac{2GM}{c^2}$, surface gravity is solely determined by the mass of the black hole.

The First Law of Black Hole Mechanics

The first law operates on the equation $dE = \kappa \div (8G\pi) \times dA + \Omega dJ + \Phi dQ$ which in words means that the change in energy of black hole is related to the change in area, angular momentum, and electric charge. The last two terms refer to energy due to rotation and energy due to electromagnetism, respectively. Since black holes are hairless, meaning that they only have mass, charge, and angular momentum, these three quantities would be the legitimate source of energy storage. The interesting part is how the factor of area comes in, and this will be discussed in the later section.

The Second Law of Black Hole Mechanics

The second law states that the change in the area of the event horizon over time is never smaller than zero. This law has originally been formulated based on the assumption that due to immense gravitational attraction, the black hole will release nothing, neither matter nor radiation, and that it only accretes, which adds to the mass and thus the area. Generalized Second Law in order for it to hold true. After identifying the area to be representative of black holes' entropy, the GSL reasserts that the total entropy of black holes and the matter and radiation field exterior to the black hole together will not decrease with time (Bekenstein 2339).

The Third Law of Black Hole Mechanics

The third law indicates that the surface gravity of a black hole cannot go to zero. Although the surface area of the event horizon can be potentially increased by adding in particles that increase the black hole's momentum, the rate of increase in the area decreased gradually along with the decrease in surface gravity as angular momentum and mass approaches the relation of $\frac{J}{M^2} = 1$ where J is momentum and M is mass of the black hole. This relation can only be actualized under the ideal condition of having infinite time and the infinite divisibility of matters (Bardeen 169-170).

Relation between Black Hole Mechanics and Thermodynamics

There seems to exist a certain relationship between black hole mechanics and thermodynamics. This relation gives us more than a mathematical analogy thanks to quantum mechanics which toppled over the statement that black holes have zero temperature, zero entropy, and no thermal radiation.

The Zeroth law in black hole mechanics indicates that the surface gravity is constant over the event horizon of a stationary black hole. The phenomenon of constancy reminds me of the Zeroth law of thermodynamics, which indicates that the temperature is uniform over a body in thermal equilibrium. In this case, surface gravity seems to have played the role of temperature. Going further, the first law of black hole mechanics in vacuum case relates mass, area, and angular momentum of two nearby stationary black holes through the equation of $dM = (\kappa \div (8G\pi))dA + \Omega dJ$. The format of this equation resembled that of the first law of thermodynamics stating that dE is equal to TdS + state parameters of two nearby thermal equilibrium state of a system. Bekenstein pointed out that the statement that the area of a black hole can never decrease is analogous to the statement that a close system's total entropy can never decrease in the second law of thermodynamics. This correspondence suggests a relationship between the black hole's event horizon's area and entropy. Thus, we can establish from the first law that E is equivalent to M, T is correlated with surface gravity, and entropy is correlated with A.

Hawking's temperature further solidified the correlation by discovering that a black hole, originally thought to have zero temperature, actually has a temperature of $T = \frac{\kappa}{2\pi}$ due to Hawking radiation. And Hawking-Bekenstein entropy of a black hole is calculated to be

$S_{bek} = \frac{A}{4}$, confirming the correlation established in the previous paragraph. These relationships hint to us that general relativity and quantum mechanics might be reconcilable as many quantities that describe black holes in general relativity turn out to be a measure of their thermodynamic properties and quantum behaviors (Wald 2-3).

Different Avatars of Entropy

In the thermodynamic sense, entropy is about the behavior of a thermodynamic system. It is calculated as the thermal energy per unit temperature. It is how spread out and in-concentrated energy is so that it is unavailable for doing work (Muller). It measures a system's level of disorderedness. It dictates the direction of natural processes: in a closed system, heat flows spontaneously from the hot to the cold. In the information perspective, the higher the entropy, the less information about a system an external observer has. In a statistical sense, Boltzmann entropy can be defined as the number of micro-states, that are equally likely, a system can be in which means that it has taken into account the atomic nature of matters and their constant random motion. Boltzmann's theory allows for exceptions in the Second Law of Thermodynamics such as having a spontaneous flow from hot to cold. However, due to the extremely small probability of such an occurrence, the possibility of this happening can be deemed insignificant. Boltzmann's entropy is mathematically calculated by setting $S = k_b ln\Omega$ where Ω represents the number of micro-states a system can possibly be in given a set of macroscopic parameters, or measurable quantities. For instance, a cup of water's volume, mass, density, and temperature are the macroscopic parameters that describe the state of this particular system of water. The molecules that constitute this cup of water, however, can be in multiple different configurations that still yield a cup of water with the same volume, mass, density, and temperature. And entropy is a count of such micro-states. The more the entropy, the more information a system contains. The information definition of entropy can be related to the statistical sense by considering the example of hard drive storage. The information in the computer is stored in the binary system which the binary digits can be represented by a certain state of small magnets. For instance, a magnet with a North Pole on the top can represent 1 and a magnet with a South Pole on top can represent 0. The more magnets there are, the more information that can be stored. As the configuration of micro-states in a system can be considered the "magnets", the more micro-states there are the more information a system contains. Normally, entropy for bodies is expressed as correlated to the volume which makes sense because the particles roam in a three-dimensional world and thus the information about them should be dependent on the dimension. But the fact that a black hole's entropy is actually correlated with area inspired some new understanding regarding the storage of information in a system.

The Holographic Theory

In previous sections, we talked about Hawking-Bekenstein entropy as helping to define the correspondence between the two sets of laws, however, this discovery has further

implications. The mathematical definition of entropy is $S = \ln ln\, N$ and N is the number of a system's microscopic states, which is the configuration of the system's particles' position and velocity. But for a black hole, having $N = e^{S_{bek}}$, would result in an enormous number of micro-states that could be attributed to a description nowhere of the black hole since no "deformation modes" can be found on the event horizon (Mathur 2).

Bekenstein has established a bound for entropy $S \leq \frac{kA}{4l_p^2}$ where S represents the maximum quantum information a volume of space can store, A is the area of the event horizon and l_p is the Planck length. But it occurs to us that this maximum is related to A, rather than V, meaning that the information about a three-dimensional space is associated to a two-dimensional area instead of volume. This odd relation is the first sign of what is going to be called the holographic theory.

In an attempt to solve the information loss paradox, t'Hooft proposed that the quantum information of matters that fall into the black hole is encoded in Hawking radiation and on the event horizon. Assuming this statement is true, although all the information is preserved on the event horizon, the interior of the black hole still exists as an observer crossing the event horizon still has the experience of falling through. Therefore, the area of the event horizon contains information that describes a black hole's interior space. Inspired by black holes, t'Hooft and Susskind extended this claim further, arguing that not only for black holes, closed two-dimensional surface can encode the degrees of freedom of a three-dimensional space as if the information of it is projected onto a screen, hence like a hologram (O'Dowd).

Holographic Principle Interpreted by Quantum Gravity

In order to construct a Hilbert space, the generalized form of Euclidean vector space equipped with the inner product that can go to infinite dimension, a system should have discrete degrees of freedom at tiny distance scales which allows for the state to correspond with an element of Hilbert space and time-reversible laws of evolutions at local scale to maintain properties of quantum mechanics.

Degrees of freedom help define the state of a system and to have a discrete degree of freedom means that the system's configuration can be described by a finite set of parameters. Reversible laws of evolution mean that the evolution of a system operating under the laws is time-symmetric. Black holes do not seem to be reversible as quantum information seems to have been lost through the process of "evaporating", but it has been suggested that black holes do leave signals behind. Then the problem lies in approximating degrees of freedom in black holes.

Since an arbitrary amount of matter can be "absorbed" due to the black hole's immense gravitational pull and an arbitrary amount of matter can be "radiated" due to Hawking radiation, the entropy diverges near the event horizon as the degree of freedom tends to infinity. This infinite amount contradicts the finite value of the black hole's entropy in reality. As a result, black holes' entropy should be attributed to quantized fields instead of only the space-time metric. One way to estimate it is by using thermodynamics. The entropy of a black hole is given by

$$S = 4\pi M^2 + C$$

There exists a relationship between entropy and degree of freedom such that

$$e^S = N$$

where N is the dimension of Hilbert space. In a system that can only be described by Boolean variables such as those having n number of spins in a discrete theory,

$$N = 2^n = e^S$$

As the number of spins is equivalent to the number of degrees of freedom, degrees of freedom can be expressed in terms of entropy and in terms of area A:

$$n = \frac{S}{ln2} = \frac{4\pi M^2}{ln2} = \frac{A}{4ln2}$$

When calculating the amount of mutually orthogonal states (states whose inner product is 0 implying that the probability of finding other states within one state is 0) without creating micro-black holes, it is found that the entropy is maximized when a single black hole has the maximum allowed size. Therefore, S_{max} will be equal to 1/4 A, and the number of possible states will be n. This result shows peculiarity in that the maximum entropy is related to the area rather than the volume. If one approaches from regularized fermion theory with a cut-off at Planck distance there will be more states calculated than there actually are in reality because many collapsed into black holes, suggesting that a stricter cut-off than plank length is needed. The implication of this result is that it is possible to represent all events inside a closed surface by the degrees of freedom on the surface, one Boolean variable per Planckian surface element, thus establishing the principles of holographic theory ('t Hooft 2-6).

Holographic Principles Interpreted by Susskind

Consider the world as a three-dimensional lattice with Planck length spacing and spin-like degrees of freedom, the distinct orthogonal quantum state in a space of volume V will be

$$N(V) = 2^n$$

where n is the number of sites. $logN$ is the entropy which will be equal to $nlog2$. By substituting

$$n = \frac{V}{l_p^3}$$

we get

$$S = \frac{Vlog2}{l_p^3}$$

thus giving the impression that entropy is proportional to volume. However, according to Bekenstein, states in the above-mentioned equation could have too much energy that a black hole could form. And the entropy derived for a black hole is $S = \frac{Alog2}{l_p^2} = \frac{A}{4}$. The reason why entropy as a function of V cannot be admitted is that it could violate the second law of thermodynamics. If a space of V has energy less than a black hole of the same size, by throwing in matter with infinitesimal mass and making such space a black hole, the created black hole's entropy will be smaller than the original entropy of the space, thus leading to the phenomenon of

entropy decreasing. This phenomenon again pointed toward the necessity of the holographic principle.

Susskind assumed matters to be composed of elementary structureless constituents called partons. Such partons can be projected onto a screen by a light ray intersecting the screen at the right angle. Their presence will be represented by a pixel light at the position of the light ray intersection. The situation of concern is what will happen when an infinitely dense object is mapped. Take the example of a black hole. When a black hole is mapped to the screen, the image is of the same area as that of the event horizon. The area element on the screen will not be smaller than that at the horizon due to the focusing theory of general relativity, which states that under the assumption that matters have positive energy and gravity attracts, null geodesic congruence, or bundles of light rays, will converge and eventually meet together in curved spacetime, thus ensuring that the information density, the amount of information per unit area, will not exceed one bit per Planck area. Trying to increase the information density by putting another black hole behind would be futile since the lensing effect of the first black hole would project the black hole behind its image as a ring. Lensing happens when the light rays behind a massive object are bent around it due to the gravitational effect. The image of a parton is not unique because the image of a parton can be at various possible locations on the screen as multiple light rays coming out from the parton satisfy the rule of presence defining. Due to this non-uniqueness, the state of the screen is said to be at quantum superposition, which means the system maintains the properties of quantum mechanics.

The above mechanism concerns encoding static images, but it is also possible to factor in the time element to code non-static events. An event (x,t) can be coded onto the screen by considering all the light rays that hit the screen at a time instant as the “light front”. In light front quantization of quantum gravity, there’s no direct reference of longitudinal direction needed, and transverse space can be seen as a discrete lattice made of binary pixels of Planckian length spacing. As a result, the obscuring of longitudinal direction can be interpreted as a dimensional reduction, which is the nature of the holographic principle (Susskind 2-8).

Hologram

The mechanism by which the hologram works can be conceived by imagining a two-dimensional lattice. The lattice is composed of cells, and the whole system is considered the field where the rules that govern how the cells interact with each other are considered the field theory. The sizes of the cells can be changed as the viewers zoom in and zoom out. In other words, the cell sizes are dependent on the observation tools: the more accurate the tool the smaller the grid sizes. First, assume that there are rules for cell interactions of different scales, that is, there’s a set of rules acting on microscopic cells when the observer zooms in and there’s also a set of rules on macroscopic cells that contains the microscopic cells when the observer zooms out. To say the field is scale-invariant is to say that the rules governing microscopic cells and macroscopic cells are identical. Such a field theory is called the conformal field theory in which the interaction rules are preserved despite local changes in scale. As a result, entities

existing at different scales do not influence each other. New degrees of freedom are thus added to the lattice constituting an extra dimension. If the entities' values on a different scale are conceived as a number line extending vertically out of a grid point, the two-dimensional lattice plane would exist just like a three-dimensional lattice box. Thus, in this sense, a two-dimensional surface can encode the degrees of freedom in a three-dimensional space (O'Dowd).

Conclusion

Inspired by the similarity between black hole's behaviors and thermodynamics, a correlation between black hole's properties and thermodynamic properties is established. The discovery that for black holes, a place where both gravitational and quantum effects are not negligible, entropy is related to the area of the event horizon led to further speculation of a new formulation of quantum gravity. The resulting holographic principle proposed that information from a three-dimensional system can be sufficiently encoded on a two-dimensional surface. As assistant professor Shaghoulian has stated, "Because of black holes, quantum gravity should be formulated as a non-gravitational quantum system in fewer dimensions (Shaghoulian)."

Works Cited

Bardeen, J. M., et al. “The Four Laws of Black Hole Mechanics - Communications in Mathematical Physics.” *SpringerLink*, Springer-Verlag, 1973. https://link.springer.com/article/10.1007/BF01645742

Bekenstein, Jacob D. “Black Holes and Entropy.” *Physical Review D*, American Physical Society, 15 Apr. 1973. https://doi.org/10.1103/PhysRevD.7.2333

Drake, D.W.F. “Second Law of Thermodynamics.” *Encyclopædia Britannica*, Encyclopædia Britannica, inc., 9 Aug. 2023. www.britannica.com/science/second-law-of-thermodynamics

Mathur, S.D. “What Exactly Is the Information Paradox?” *SpringerLink*, Springer Berlin Heidelberg, 1 Jan. 1970. https://link.springer.com/chapter/10.1007/978-3-540-88460- 6_1

Muller, D. “The Most Misunderstood Concept in Physics.” *YouTube*, 1 Jul. 2023. www.youtube.com/watch?v=DxL2HoqLbyA

O’Dowd, Matthew. “The Holographic Universe Explained.” *YouTube*, 10 Apr. 2019. www.youtube.com/watch?v=klpDHn8viX8

Shaghoulian, Edgar. “Black Hole Discovery Helps to Explain Quantum Nature of the Cosmos.” *Scientific American*, Scientific American, 1 Sept. 2022. www.scientificamerican.com/article/black-hole-discovery-helps-to-explain-quantum- nature-of-the-cosmos/

Susskind, Leonard. “The World as a Hologram.” *AIP Publishing*, AIP Publishing, 1 Nov. 1995. https://doi.org/10.1063/1.531249

Wald, Robert. “The Thermodynamics of Black Holes.” *Living Reviews in Relativity*, U.S. National Library of Medicine, 9 July 2001. https://pubmed.ncbi.nlm.nih.gov/28163633/

't Hooft, G. “Dimensional Reduction in Quantum Gravity.” *arXiv.Org*, 20 Mar. 2009, https://doi.org/10.48550/arXiv.gr-qc/9310026

Mechanisms of Cancer Resistance in Mammals By Amari Anderson

Abstract

Cancer is a complex disease that threatens the health of a broad range of species. Despite the fact that cancer is the second leading cause of death worldwide, there are a number of species that are remarkably resistant to cancer, including the elephant, naked mole rat, blind mole rat, whale, horse, cow, and microbat. These animals exhibit a natural resistance to cancer through various mechanisms, including enhanced DNA repair, hypersensitivity to genotoxic stress, resistance to metastatic invasion, and duplicate copies of tumor suppressor genes. These mechanisms can provide insight into preventing, as well as treating, cancer in less resistant species. This paper reviews potential mechanisms for cancer resistance across a variety of mammals and provides insight as to why some mammals are more resistant to cancer than others.

Introduction

Cancer is characterized by the uncontrolled proliferation of abnormal cells. As the second leading cause of death worldwide (World Health Organization: WHO), this complex disease threatens the health of people globally. While treatments for cancer exist, treatment options are often ineffective, and access to treatment is a barrier in many countries. Research on cancer prevention is becoming an important focus, as prevention is not only more cost-effective but also circumvents the need to go through onerous treatment regimens.

Previously, cancer prevention research has mainly focused on conscious choices humans can make to decrease their risk of cancer. Lifestyle changes, such as being physically active, having a healthy diet, and limiting alcohol consumption have all been identified as ways to prevent cancer (American Institute for Cancer Research). Although these strategies don't entirely prevent cancer, they can empower people to decrease their risk of cancer by making conscious lifestyle choices. Discovering strategies to more effectively prevent cancer in humans may be possible by examining the characteristics of naturally cancer-resistant mammals.

Body mass is thought to be a factor contributing to cancer resistance. Statistically, animals with a higher body mass should have a higher rate of developing cancer, as they have more cells than smaller animals, meaning they have a higher chance of mutagenesis which can result in cancer. However, some large animals actually have a much lower rate of cancer compared to humans and mice. Peto's Paradox states that there is no real correlation between body mass and cancer incidence, suggesting that larger animals have developed natural tumor-suppressing mechanisms to counteract their increased cancer risk. In support of this theory, research has found that mammals larger than 5-10 kg possess an important anticancer mechanism known as replicative senescence. During replicative senescence, cells enter a state of permanent growth cycle arrest when telomeres, the regions of repetitive DNA sequences found at the end of the chromosomes, reach a critically short length (Chadwick; Seluanov et al. "Telomerase Activity"). This is due to the repression of telomerase activity, which maintains the length of telomeres, in somatic tissues, causing telomeres to continue shortening as cells divide (Seluanov et al. "Telomerase Activity"). Replicative senescence is thought to be an anticancer mechanism that limits cell proliferation, thus working to prevent cancer. Telomerase must be

active to form human tumors and is used to extend the lifespan of cancer cells. Since larger mammals have inactive telomerase, they have an extra defense mechanism against tumorigenesis.

The lifespan of an animal also plays a large role in the animal's resistance to cancer. Long-lived mammals tend to display cancer resistance due to evolutionary adaptations that keep them cancer-free until post-reproductive age (Seluanov et al. "Mechanisms"). Therefore, long-lived mammals capable of reproducing later in life may have evolved cancer resistance to support their prolonged reproductive period. On the other hand, it is possible that because these animals developed mechanisms that decrease their risk of cancer, they now have longer lifespans. In either case, lifespan and cancer resistance are both linked.

Diets can also impact cancer resistance. In humans, a healthy diet that limits red or processed meat can decrease cancer risk (Gonzales and Riboli; Key et al.). Similarly, herbivorous mammals have a lower cancer mortality rate compared to carnivorous mammals (Vincze et al.). While the diet of a mammal is not the sole determining factor on the lifespan or cancer resistance, analyzing characteristics that cancer-resistant mammals possess, such as herbivorous diets, could allow for a better understanding of cancer resistance across mammals.

Mice and rats are commonly used as research models for various fields, including cancer. Information from these mammals has allowed for extraordinary breakthroughs in the development of treatment for human cancers (Anisimov et al.). However, these short-lived, cancer-prone mammals may not be the ideal research models to uncover novel mechanisms of cancer prevention that could be translated to human cancer prevention. Using inherently cancer-resistant mammals, such as elephants, whales, naked mole rats, blind mole rats, horses, cattle, and bats, as models for cancer research may allow for breakthroughs in the field of cancer research, resulting in treatment strategies and methods for cancer prevention. This paper explores the mechanisms that contribute to cancer resistance in mammals and proposes potential focuses of future research to advance our understanding of cancer prevention and resistance.

Elephants

Elephants, of the family Elephantidae and the order Proboscidea, are large land mammals of three species: the African forest elephant, the Asian elephant, and the African bush elephant, which is the largest living land mammal. Found in parts of Africa, South Asia, and Southeast Asia, elephants are herbivorous mammals that live in deserts, marshes, and forests.

The size of an elephant mainly depends on the sex and species of the individual but can vary from 2000 kg to 7000 kg (Shoshani). Similarly, the lifespan of elephants varies depending on the captivity status of the elephant. Captive elephants tend to have an average lifespan of around 17-20 years (Clubb et al.), while those that live in the wild may live anywhere from 40-80 years old (Shoshani; Clubb et al.).

The majority of research surrounding elephants of the Loxodanta genus and cancer resistance focuses on African bush elephants. When results are specific to African forest elephants, a distinction will be noted. Otherwise, African bush elephants may simply be referred to as "African elephants."

Cancer in Elephants

Elephants are large, long-lived mammals. Thus, they have developed anticancer mechanisms to decrease their cancer risk until they are past their reproductive age. Elephants are thought to be extremely resistant to cancer and have a significantly lower cancer incidence rate compared to other mammals, such as humans, cats, and dogs.

Cancer rates in elephants differ between species. Asian elephants are much more prone to cancer than African bush elephants. One study found that 41% of Asian elephants (17/41) displayed evidence of neoplasia, compared to 6% of African elephants (2/35) ("Elephant" Tollis et al.). Out of the 41 Asian elephants, 6 were diagnosed with malignant cancer, making the malignant cancer incidence rate in Asian elephants 14%. There were no cases of malignant cancer found in the African elephants ("Elephant" Tollis et al.). Another study reported a cancer incidence rate of over 40% in elephants, with a malignancy rate of 24% (Boddy et al.). However, this study combined Asian and African bush elephants and had a small sample size of 17, which can impact the results.

Asian elephants display a higher cancer incidence rate compared to African bush elephants because neoplasia in mature female Asian elephants is extremely common. Neoplasia of the reproductive tract seems to be common in adult, female Asian elephants and was observed in 80% of 80 adult female Asian elephants (Landolfi et al.). Myometrial leiomyomas were the most common tumor observed and have a reported prevalence of anywhere from 71-100% in mature, female Asian elephants (Landolfi et al.; Montali et al.). Malignant tumors accounted for 12 of the 63 reported uterine neoplasms (19%), and 67% of those malignant tumors were found to have metastasized (Landolfi et al.). It's important to note that this high rate of neoplasia only seems to be prevalent in the reproductive tract of mature, female Asian elephants, and does not appear to be a prevalent issue in other elephant subcategories.

Prior to studies reporting a high rate of neoplasia in female Asian elephants, incidences of neoplasia across elephants were thought to be extremely uncommon. While elephants do have a low cancer mortality rate of less than 5% (Abegglen et al.), researchers now know that cancer incidence in elephants varies depending on the sex, age, and species of the elephant, and neoplasia is common in some groups of elephants. While this may seem to challenge the notion of elephants' cancer resistance, it's important to note that only a specific subset of elephants is susceptible to only a single type of cancer. Overall, elephants are remarkably resistant to cancer due to their remarkable anticancer mechanisms.

Cancer Resistance in Elephants: The TP53 Gene

Although female Asian elephants have a higher rate of neoplasia than previously thought, these mammals still possess remarkable strategies that protect themselves against the devastating disease that is cancer.

Elephants have multiple pseudogene copies of a tumor suppressor gene known as TP53 (Abegglen et al.). TP53 triggers apoptosis, or programmed cell death, in response to a variety of signals sent by sensors that track stressors such as DNA damage or radiation (Abegglen et al.). If

the degree of damage to the cell is excessive, TP53 can halt the cell cycle until conditions become normalized within the cell. If the cell is irreparably damaged, TP53 can trigger apoptosis, causing the cell to die (Hanahan and Weinburg). This makes it an essential tumor-suppressor gene across mammals.

African bush elephants tend to have anywhere from 19-23 extra copies of the gene, while the Asian elephant genome indicates that Asian elephants harbor anywhere from 10-37 copies of the gene("Elephant" Tollis et al.; Abegglen et al.; Sulak et al.). For comparison, humans have only one copy of TP53 (Abegglen et al.). The extra copies that elephants harbor appear to be all pseudogenes and house a variety of changes in the sequence of the gene but are still associated with a heightened apoptotic response in elephants.

Elephants are extremely sensitive to cellular stress because of their extra copies of TP53. This causes them to demonstrate apoptosis at much lower rates of DNA damage compared to humans and other species (Abegglen et al.). While this early-apoptosis strategy is efficient, it's also possibly very draining. Elephants likely have to deal with constant cell death, which increases their need for cell replacement. Elephants have most likely evolved adaptations that counteract the damage caused by frequent apoptosis ("Mechanisms" Seluanov et al.). Studies done on mice support this theory, as mice with active p53 display symptoms such as accelerated aging, reduced fertility, shorter lifespan, reduced size, atrophy across organ systems, loss of stem cells, and tissue cellularity (Tyler et al.; Maier et al.). However, mice that were engineered to carry extra copies of wild-type Trp53 transgenic alleles did not display similar symptoms and instead were protected from cancer (Garcia-Cao et al.).

Additionally, older Asian elephants were found to display a lower rate of apoptosis, suggesting an age-related decrease in hypersensitivity and apoptosis across Asian elephants (Abegglen et al.). This may provide reasoning as to why neoplasia is common in adult, female Asian elephants. However, further research would have to be done to confirm this, as well as determine why it is so prevalent in female Asian elephants only.

Naked Mole Rats

The naked mole rat (*Heterocephalus glaber*) is a small, subterranean rodent commonly found in the drier areas of East Africa. Mature naked mole rats typically weigh around 35 grams but can reach a mass of over 110 grams (Buffeinstein et al.). They are the longest-lived rodents, with an average lifespan of 28 years in captivity (Buffeinstein).

Cancer and Cancer Resistance

Although mammals have some of the highest cancer rates, cases of neoplasia in naked mole rats are rare. Multi-year-long observations of naked mole rat colonies have recorded no incidences of cancer (Buffeinstein; "Spontaneous" Delaney et al.). In fact, only a few cases of spontaneous neoplasia in naked mole rats have been reported ("Initial Case" Delaney et al.). Not only are they inherently resistant to natural cancer development, but they are also highly resistant to induced tumorigenesis ("Hypersensitivity" Seluanov et al.; Liang et al.; Manov et al.).

Naked mole rats use an early-acting anticancer mechanism known as early contact inhibition (“Hypersensitivity” Seluanov et al.). Regular contact inhibition, a characteristic of normal cells, is a natural process that arrests cell division once cells reach a high density or come into close contact with each other, preventing hyperproliferation. Contact inhibition is often lost in cancer cells. Unlike most other species, naked mole rats exhibit early contact inhibition, which causes cells to proliferate very slowly in culture and arrest cell division earlier than human and mouse cells(“Hypersensitivity” Seluanov et al.; “Distinct” Seluanov et al.). Early contact inhibition requires less cell contact to inhibit growth compared to regular contact inhibition and acts as a barrier defense against tumorigenesis and malignant growth by restricting cell proliferation.

Early contact inhibition is a result of the production of a large amount of the extracellular signal high molecular mass hyaluronan (HMM-HA), a substance that is secreted by naked mole rat cells at a five times larger rate than human or mouse cells (Tian et al.). The oversecretion of HMM-HA is likely an adaptation that initially served as a way to improve the skin elasticity of naked mole rats for their subterranean lifestyle (Tian et al.). However, the larger amount of HMM-HA secreted serves as a critical anticancer mechanism in naked mole rat cells. When the signaling pathway that leads from HMM-HA to p16, the gene that triggers the activation of early contact inhibition, is removed, naked mole rat cells are more susceptible to hyperplasia and cancer development (Tian et al.). The signaling pathway also requires the presence of the CD44 receptor, which is a cell surface adhesion receptor that is expressed in many cancers and plays a role in regulating metastasis (Tian et al.).

Early contact inhibition is strongest when both tumor suppressor pathways p53, which triggers apoptosis, and pRb, which controls cell division, are expressed (Hanahan and Weinburg). Naked mole rat cells only lose early contact inhibition when both p53 and pRb are inactive, thus causing their cells to proliferate beyond the standard rate and undergo regular contact inhibition (“Hypersensitivity” Seluanov et al.). When naked mole rat cells undergo regular contact inhibition upon deletion of p53 and pRb, it is triggered by the activation of the gene p27 instead of p16, similar to human and mouse cells (“Hypersensitivity” Seluanov et al.). When only one of the two tumor suppressor genes, p53 and pRb, in the naked mole rat is active, cells may proliferate beyond the naked mole rat’s standard rate but undergo apoptosis when the cells become too close in proximity (“Hypersensitivity” Seluanov et al.). As a result, cell density remains low. Similarly, when alternative reading frame (ARF), a gene specific to naked mole rats that plays a role in tumor suppression, is inactivated, naked mole rat cells induce senescence to protect the cells against oncogenic transformation (Miyawaki et al.). This indicates that naked mole rats can sense the loss of one of their tumor suppressor pathways and use backup methods to protect themselves against cancer and hyperproliferation.

Aging and cancer are both the result of an accumulation of cell damage (Lopez-Otin et al.). The two processes share many similarities and are linked in a variety of ways. When a species exhibits resistance to aging, it may also possess notable adaptations that affect cancer development. Naked mole rats are an example of a species that does not display evident signs of

aging (Buffeinstein). They have excellent anti-aging defenses, display adequate reproductive potential well into their third decade of living, and have minimal age-related physiological changes (Buffeinstein; O'Connor et al.). It is possible that the mechanisms that contribute to the naked mole rat's lack of aging may play a role in its cancer resistance. For example, the accuracy of protein translation is thought to play an important part in aging. Naked mole rats have more precise protein translation, which decreases the risk of anomalous proteins that can cause cancer (Azpurua et al.). It is possible that the naked mole rat's accurate protein translation may contribute to both its cancer resistance and lack of aging.

Naked mole rats are also capable of fructose-driven glycolysis (Park et al.). Similarly to their overproduction of HMM-HA, this adaptation evolved as a way for naked mole rats to be better suited to life underground. The ability to use fructose instead of glucose to fuel glycolysis is an example of an adaptation that allows naked mole rats to be more resistant to hypoxic environments. Interestingly, fructose-mediated actions are thought to be a possible risk factor for cancer, and fructose has been found to induce the proliferation of tumors (Liu et al.). However, it is possible that the potential increase in cancer risk due to fructose utilization has been counteracted by the various anticancer strategies employed by the naked mole rat.

Blind Mole Rats

Blind mole rats (*Nannospalax ehrenbergi* superspecies) are subterranean mammals commonly found in mountain valleys and forests in the Middle East. They are long-lived rodents that weigh 100-200 grams and have a documented maximum lifespan of 21 years (Edrey et al.). Although blind mole rats are closer to mice than to naked mole rats, the two both possess long lifespans relative to their body mass and remarkable abilities, such as inherent cancer resistance.

Cancer and Cancer Resistance

Blind mole rats are incredibly resistant to cancer. A 40-year-long observation of thousands of blind mole rats reported zero spontaneous tumors in the animals (Gorbunova et al.). There have been zero confirmed cases of cancer in blind mole rats, suggesting these animals are one of the most cancer-resistant mammals. They are also incredibly resistant to induced tumorigenesis and were shown to kill cancer cells of other species, proving they possess incredibly efficient anticancer mechanisms (Manov et al.).

Blind mole rats employ an anticancer mechanism known as concerted cell death. Concerted cell death is triggered by the release of IFN-β (Gorbunova et al.). Interferons (IFNs) are proteins that play a role in the immune response by signaling that pathogens, such as viruses, are present. They also play crucial roles in tumor suppression. Unlike naked mole rat cells, blind mole rat cells proliferate rapidly for around 7-20 population doublings. Due to their sensitivity to hyperplasia, likely evolving as a way to recognize specific preludes to cancer, blind mole rat cells release IFN-β once they've reached a high density and then undergo concerted cell death. *In vitro* cells arrest proliferation for around 3 days, after which all cells on the plate detach and die within 4 days by a mix of both necrotic and apoptotic responses (Gorbunova et al.). Concerted cell death primarily uses necrosis, a process that is usually associated with the uncontrolled death

of a large group of cells which is often harmful to the organism. On the other hand, apoptosis is a programmed process of cells that usually has beneficial, healing effects because it does not lead to widespread cell death (D'Arcy). Due to the major differences between the two cell death pathways, necrosis is seen as an inefficient way of eliminating unwanted cells compared to apoptosis. However, the response of necrosis in blind mole rats virtually eliminates all of the possibly oncogenic cells by not only killing off the possibly pre-malignant cells but also the cells surrounding them, making it a strong anticancer mechanism. The reasoning as to why blind mole rats primarily employ necrotic responses instead of apoptotic responses in concerted cell death may lie in their unusual p53 sequence, which was found to be unable to induce apoptosis, yet was more sensitive to inducing growth arrest (Ashur-Fabian et al.). This unique p53 gene substituted lysine for arginine, a substitution that has been found in human tumors that display a resistance to hypoxia. This suggests that this adaptation evolved to better suit blind mole rats in their subterranean habitat (Ashur-Fabian et al.). Additionally, both tumor suppressor pathways, p53 and pRb, must be inactivated to abrogate concerted cell death (Gorbunova et al.).

Concerted cell death is not the only cancer resistance mechanism used by the blind mole rat. These animals possess a higher rate of DNA/RNA editing, reduced chromosome arrangements, and an over-representation of short interspersed elements (SINEs), non-coding sequences that play a role in cell survival against stressors such as DNA damage, compared to mice and rats, which is most likely linked to the rodent's resistance to cancer and tolerance of hypoxia (Fang et al.).

Blind mole rats also produce a large amount of HMM-HA (Tian et al.). The large amount of HMM-HA in blind mole rats is likely an adaptation that protects the rodents from oxidative damage. This adaptation may contribute to cancer resistance. However, blind mole rats do not exhibit early contact inhibition, and HMM-HA has not been identified as a property of concerted cell death (Manov et al.). Therefore, there is no substantial evidence that points to HMM-HA playing a role in the cancer resistance of the blind mole rat. Further research is required to determine whether HMM-HA has anticancer mechanisms in the blind mole rat.

Another factor that may contribute to cancer resistance in the blind mole rat is its uniquely modified variant of heparanase, an enzyme that degrades heparan sulfate on the cell surface and extracellular matrix (Nasser et al.). Heparanase is mainly overexpressed in tumors, and its overexpression is associated with increased metastasis. The uniquely spliced variant found in the blind mole rat inhibits heparan sulfate degradation and suppresses glioma tumor growth, which may play a role in cancer resistance (Nasser et al.).

Bowhead Whales: Large Mammals with No Cancer

The bowhead whale (*Balaena mysticetus*) is a large aquatic mammal that resides in Arctic and subarctic waters. Weighing in at 75-100 tons, bowhead whales are estimated to have a natural lifespan of over 200 years and are one of the longest-living mammals (NOAA Fisheries).

The bowhead whale possesses many genes that result in enhanced DNA repair, immune response, and cell-cycle regulation (Table 1) (Keane et al.). Genes under positive selection in the bowhead whale include the excision repair cross-complementing group 1 (ERCC1) and histone

deacetylase (HDAC1 & HDAC2). ERCC1 is involved in DNA repair, while HDAC1 and HDAC2 both play roles in the regulation of chromatin structure and have been associated with longevity in fruit flies (Keane et al.). These genes are also involved in death receptor signaling, which detects death outside the cell and triggers apoptosis (Keane et al.; "Death Receptor Pathway").

The bowhead whale also possesses multiple copies of various genes, such as PCNA, LAMTOR1, and UVRAG which contribute to cancer resistance (Keane et al.; "Return" Tollis et al.). Proliferating cell nuclear antigen, or PCNA, is involved in DNA repair. In the bowhead whale, LAMTOR1 is not only copied multiple times, but it also has unique changes in its amino acids. LAMTOR1 is involved in activating the mTORC1 gene, which is associated with cancer and aging (Keane et al.; "Return" Tollis et al.). It also plays a role in controlling cellular growth. UVRAG genes are tumor suppressor genes. Other duplicated genes found in the bowhead whale that have roles in the cell cycle, DNA repair, and cellular stress response include PSMD4, UCHL3, ARPP19, and STOML2 (Keane et al.). Additionally, the bowhead whale has unique substitutions in the POLE gene, which is a cancer-related gene that is involved in DNA repair and replication ("Return" Tollis et al.).

Genes associated with tumor suppression are overexpressed in the kidney of the bowhead whale ("Transcriptome" Seim et al.). An example of one of these genes is E4F transcription factor 1 (E4f1), which directly interacts with p53 and plays a role in growth arrest and apoptotic responses. Interestingly, the bowhead whale's liver exhibited a reduced expression of tumor suppressors, yet a higher expression of proto-oncogenes such as WSB1 and PDGFRB ("Transcriptome" Seim et al.). Additionally, the bowhead whale genome is virtually devoid of short interspersed nuclear elements (SINEs), which are short, repetitive non-coding sequences that are involved in cell survival when facing stressors such as DNA damage and infection (Keane et al.). On the other hand, other cancer-resistant mammals, such as the blind mole rat, have a numerous amount of SINEs. This indicates that the bowhead whale relies on mechanisms other than SINEs for their cancer resistance.

Gene(s)	Mutation(s)	Role(s)
Excision repair cross-complementing group 1 (ERCC 1)	Under positive selection	- DNA Repair - Death receptor signaling
Histone deacetylase (HDAC 1 & 2)	Under positive selection	- Regulation of chromatin structure - Longevity - Death Receptor Signaling

Proliferating cell nuclear antigen (PCNA)	Duplicate copies	- DNA Repair
LAMTOR1	Multiple copies; Unique amino acid changes	- Activates the mTORC1 gene (associated with cancer and aging) - Controls cellular growth
UVRAG	Multiple copies	- Tumor suppressor
PSMD4, UCHL3, ARPP19, and STOML2	Multiple copies	- DNA repair - Cellular stress response - Plays a role in mitosis
POLE	Unique substitutions	- Cancer-related - DNA repair - DNA replication
E4F transcription factor 1 (E4f1)	Overexpressed	- Directly interacts with p53 - Growth arrest - Apoptotic responses
WSB1 and PDGFRB	Overexpressed	- Proto-oncogenes (promotes cancer).

Table 1: Genetic mutations in the bowhead whale genome that contribute to cancer resistance. Data from. Keane et al.; "Return" Tollis et al.; "Transcriptome" Seim et al.

Currently, there have been no reports of cancer in bowhead whales. However, it is extremely likely that cancer in bowhead whales goes undetected due to the fact that we do not know much about their natural causes of death. Moreover, it is very rare to find a stranded bowhead whale. When samples are collected from dead bowhead whales, there is usually a delay between when the whale died and when the samples are collected and analyzed, increasing the chance of compromised data. Further research on cancer resistance in the bowhead whale should focus on sample collection of dead bowhead whales to gather comprehensive data surrounding bowhead whales and cancer incidence. Nonetheless, bowhead whales are thought to be resistant to cancer due to their unique genetic changes and slow mutation rate ("Return" Tollis et al.).

Cattle and Horses

Cattle (*Bos taurus*) are large, herbivorous, hoofed mammals. Although the weight of cattle can vary from 300 kg to 1500 kg, the size of these animals primarily depends on the breed and sex of the individual. Bulls tend to weigh more than cows of the same breed. Since cattle are mainly used in food production, these animals are typically slaughtered before they reach the natural end of their life. However, the average natural lifespan of cattle is around 20-25 years.

Horses (*Equus ferus caballus*) are domesticated, herbivorous, hoofed mammals that vary in size, shape, and color, depending on various factors. The size of a horse is largely determined by its breed, ranging from under 400 kg to over 1000 kg. The lifespan of a horse also varies depending on factors such as breed, care, and environment. Most horses can live anywhere from 25-30 years, with some horses living well into their 30s and 40s, and rarely into their 50s and 60s.

Loss of Placental Invasion = Loss of Metastatic Cancers?

Cows and horses are thought to have a lower rate of malignant cancers due to their method of placentation (D'Souza and Wagner). The placenta is a necessary part of pregnancy in mammals, as it allows maternal and fetal blood to come into close proximity with each other to exchange molecules. Many species, including humans, dogs, and cats, accomplish this task by placental invasion, whether endotheliochorial or hemochorial (Moffett and Loke). During placental invasion, the trophoblast cells of the placenta invade the uterus and maternal blood vessels (Moffett and Loke; Graham and Lala). Hemochorial placentation, employed by humans and rodents, is the most invasive placentation in mammals. Hoofed mammals, such as horses and cows, as well as their relatives, like whales, use epitheliochorial placentation (D'Souza and Wagner). Epitheliochorial placentation has no trophoblast invasion at all and instead exchanges molecules between maternal and fetal blood via noninvasive mechanisms. Species with no placental invasion may have evolved mechanisms to suppress trophoblast invasion, in turn also increasing the body's ability to suppress metastatic invasion in certain locations (D'Souza and Wagner). One study reviewed the rate of malignant cancer in animals with differing levels of placental invasion (D'Souza and Wagner). The study found that horses and cows, animals with a loss of invasive placentation, had a much lower rate of malignant cancers in certain locations compared to dogs and cats, animals with invasive placentation (D'Souza and Wagner) (Table 2). Bovine stroma has been proven to be much more resistant to cell invasion when tested *in vitro* (Kshitiz et al.). On the other hand, human stroma is more permissive to invasion (Kshitiz et al.). While research on the permissiveness of stroma of all mammals that employ epitheliochorial placentation has not been done, it is theorized that mammals that use non-invasive placentation, including horses, are better equipped to resist metastatic invasion (Kshitiz et al.).

One study reported that out of 4407 cows examined, only 98 showed evidence of neoplasia (2.22%) (Ramos et al.). The most common tumors in cattle were lymphomas and epithelial tumors, namely squamous cell carcinoma. Lymphomas in bovine have been associated with the bovine leukemia virus, as well as the size of the herd, which may indicate a genetic predisposition to the cancer (Ramos et al.).

Another study found that out of 1.3 million cattle examined, only 302 tumors were found (0.02%) (Anderson and Sandison). However, the cattle examined were confirmed to be a part of the dairy/meat industries and were slaughtered before reaching the natural end of their life, possibly skewing results. Cancer risk increases along with age, making it possible for it to appear as if cattle (that have been slaughtered at a relatively young age) have a lower rate of cancer than normal. Additionally, the study was done in 1969, making it possible that the technology and/or methods used may have made it so that some incidences of neoplasia were not identified.

A study of 774 horses examined, found that only 65 possessed neoplasms (8.39%) (Ramos et al.). Skin cancers are typically the most common type of cancer found in horses, which varies in frequency depending on the breed and pigmentation of the horse. Pale-pigmented horses develop skin cancers such as melanoma more frequently than other horses with darker pigmentation. In addition to melanomas, other common cancers found in horses include equine sarcoids, squamous cell carcinomas, and granulosa cell tumors (Cotchin). Equine sarcoids are usually benign and non-metastatic but can become locally aggressive under certain circumstances and have a high chance of recurrence due to their resistance to therapy (Ogluszka et al.). While these tumors aren't typically life-threatening, they can negatively affect the horse's quality of life (Ogluszka et al.).

Location	Species	Malignant Tumors	Total Tumors	Malignancy Rate
Skin	Cow	11	105	10%
	Horse	26	437	6%
	Dog	338	1691	20%
	Cat	35	51	41%
Glandular Epithelium	Cow	10	45	22%
	Horse	15	175	8%
	Dog	553	1184	47%
	Cat	45	57	79%
Non-glandular epithelium	Cow	187	200	94%
	Horse	119	136	88%
	Dog	263	316	83%
	Cat	46	50	93%
Connective Tissue	Cow	0	22	0%
	Horse	13	127	10%
	Dog	323	626	52%
	Cat	18	31	59%

Table 2: Frequency of malignant tumors in the cow, dog, horse, and cat at four anatomical locations. Table from D'Souza and Wagner.

It's important to remember that although there is a correlation between lower malignancy rates and loss of placental invasion, this may not be causal. Other factors that may contribute to

these results include body mass, age, environment, and diet. It is also possible that cows and horses possess other anticancer mechanisms that contribute to their cancer resistance. However, the data is consistent with the theory that only certain anatomical locations of animals with a loss of placental invasion are resistant to malignant tumors, as cancer malignancy rates are significantly higher in the non-glandular epithelium and lower in other areas, such as the skin and connective tissues.

Bats

Bats, mammals of the order Chiroptera, live relatively long lives compared to what is expected for mammals of their size. Bats are largely diverse mammals, with over 1300 species across 20 families, and comprise 20% of all classified mammalian species, second only to rodents. Most bats have an average lifespan of 10-20 years, depending on the species. Certain microbat species, such as the little brown bat (*Myotis lucifugus*), have been reported to live for over 30 years. The longest-living individual bat was a Siberian bat, and was determined to be around 41 years old. Bats can be found virtually anywhere in the world and have a diet that ranges from fruits to blood depending on the species.

Bats are reservoirs for a variety of viruses that are damaging to multiple species, including humans. However, these viruses appear to cause no clinical disease in bats. Experimentally infecting laboratory bats with a virus only led to, at most, a subclinical infection with few symptoms (Middleton et al.). This indicates that bats possess a variety of effective antiviral mechanisms, some of which may contribute to cancer resistance in the bat.

Cancer and Cancer Resistance

Only a few cases of cancer have been reported in bats, despite their prevalence in scientific research ("Mechanisms" Seluanov et al.; Olds et al.). Mutations that may contribute to cancer resistance in bats are only present in some species and, due to the diversity of bats, there is no confirmation that adaptations found in some species of bats are similar across all bat species. It is likely that different species and lineages of bats have evolved their own cancer-resistance mechanisms, a theory that should be tested with further research into the subject.

It is possible that factors that affect the exceptional longevity of bats may have anticancer effects. For example, bats possess an altered growth hormone/insulin-like growth factor 1 axis ("Genome" Seim et al.). Some species of bat, such as the Brandt's bat (*Myotis brandtii*) and the big brown bat (*Eptesicus fuscus*), harbor a deletion in the growth hormone receptor transmembrane ("Genome" Seim et al.). GHR mutations and dysfunction have been associated with a decreased risk of diabetes and cancer in humans and mice ("Genome" Seim et al.). Multiple insulin signaling-associated genes were also expressed at a higher level in Brandt's bats ("Genome" Seim et al.). This was similarly expressed in GHR dysfunctional long-lived mice, supporting the idea that the GHR in *M. brandtii* is dysfunctional and may cause these bats to exhibit a longer lifespan. Various other aging-related changes in bats may also contribute to cancer resistance.

One study showed that long-lived bats share similarities with the naked mole rat (Ricci et al.). These shared similarities may contribute to these mammals' unique cancer resistance and long lifespans. Both long-lived bats and naked mole rats showed reduced non-LTR retrotransposon, a type of transposable element (TE), accumulation when compared to short-lived species of bats and rodents (Ricci et al.). TEs play a major role in eukaryotic life and are key properties of gene duplication, gene expression regulation, and eukaryotic evolution (Ricci et al.). Furthermore, they also play a role in many devastating human diseases and can ultimately lead to cancer in certain conditions (Ricci et al.; Callinan and Batzer). Therefore, the reduced non-LTR retrotransposon accumulation may enhance cancer resistance (Ricci et al.). Bats may also possess mechanisms to dampen potential health issues related to transposable elements, thus contributing to the antiviral properties of bats (Ricci et al.). Bats and naked mole rats also share an adaptive sequence convergence of ADAMTS9, a novel tumor suppressor that inhibits tumor growth by blocking the IGF1/mTOR pathway, which is associated with cancer, aging, and various diseases (Lambert and Portfors). Additionally, the gene has been indicated in many other age-related conditions (Lambert and Portfors). This shared adaptation may be a reason why both naked mole rats and certain bat species have an apparent resistance to cancer and aging.

Long-lived bat species may also have developed superior mechanisms against cellular injury and damage. For example, long-lived bat species produce more 53BP1 foci compared to short-lived species, even when the same amount of DNA damage occurs (Croco et al.). This DNA damage response pathway results in the activation of genes associated with cell cycle arrest, apoptosis, or senescence. Certain species of bats, such as the *Myotis davidii*, have DNA repair and tumor suppressor genes such as TP53 and BRCA2 under positive selection (Zhang et al.). Myotis bats also possess upregulated telomere maintenance genes, many of which play a role in DNA repair (Foley et al.). However, they are hypersensitive to ultraviolet light (Harper et al.). Another reason why bats may not develop cancer as frequently as other mammals is because of the fact that they possess a lower amount of direct repeats in their mitochondrial DNA (mtDNA), lowering the chance for mutagenesis (Khaidakov et al.). Direct repeats in the mitochondrial DNA are known to contribute to harmful mitochondrial mutations that are suspected to have a significant impact on the aging process (Khaidakov et al.). Mutations in the mitochondrial DNA have also been associated with cancer (Cavalcante et al.). Species with lower levels of direct mtDNA repeats have greater mtDNA stability, allowing for a longer lifespan and a decreased risk of cancer (Khaidakov et al.). This is consistent with the theory that certain species of bats, particularly long-lived ones, have advanced DNA repair strategies that may contribute to cancer resistance.

Greater mouse-eared bats (*Myotis myotis*) have numerous amounts of unique, bat-specific miRNA that are associated with the immune system, aging, and tumor formation pathways. miRNA such as the miR-101-3p and miR-155-5p were upregulated in bats and are associated with apoptosis and immunity (Huang et al.). *M. myotis* bats also downregulate the oncogenic miR-221-5p, which promotes tumor formation in human breast and pancreatic cancers. Over 100

upregulated genes in *M. myotis* bats were enhanced in the mitotic cell cycle, while over 300 down-regulated genes were primarily involved in mitochondrial activity (Huang et al.).

Certain bat species also possess multiple antiviral and anticancer mutations (Scheben et al.). DNA repair genes, including PALB2, IFIT2, POLA1, and POLK, were under positive selection in bats. PALB2 is a part of the BRCA complex and plays a role in tumor suppression. Bats also possess multiple copies of IFIT2, which plays a role in apoptosis and cancer progression (Scheben et al.). CDH1 and CAT were also found to be under positive selection. However, these genes have been linked to cancer, and their evolutionary value to bats is not yet clear.

Other potential reasons for cancer resistance in bats include resistance to oxidative damage (Shen et al.), better maintenance of proteostasis (Pride et al; Salmon et al.), and diet, as frugivorous bats tend to live shorter lifespans than insectivorous bats, likely due to high glucose consumption (Wilkinson and South).

Cancer-resistant Mammals as Models for Cancer Research

Cancer is a disease that has existed long before humans. While the term “cancer” first appeared in around 400 BCE by the Greek physician Hippocrates, the oldest description of the disease dates back to an Egyptian textbook on trauma surgery in around 3000 BCE (American Cancer Society). In this textbook, cancer was described as having no treatment other than removal by cauterization (American Cancer Society). However, due to our enhanced understanding of the disease, there are now a variety of treatment options for cancer, including surgery, hormone therapy, chemotherapy, and radiation therapy.

Since cancer existed before humans, it’s not surprising that cancer is not limited to humans. In fact, cancer can behave quite similarly in humans and other animals. As a result, certain animals are commonly used as cancer research models as they allow for the study of the biology of cancer in a more physiological context *in vivo* instead of in individual cells or tissues *in vitro* (Cancer Research UK). The use of animal models in cancer research allows researchers to better understand cancer as a whole. Using animal models for cancer research has several benefits, including studying how cancer cells grow and metastasize, how effective treatment options are, how certain factors impact risk, and how cancer interacts with other cell types (Pharma Models). Due to this, animal models of cancer have contributed largely to cancer research and the increasing survival rates of cancer patients.

Cancer research primarily uses short-lived, cancer-prone mammals such as mice and rats as research models, both of which have biological and anatomical similarities to humans. These animals play valuable roles in our understanding of cancer, and their use in laboratories has led to various discoveries for cancer treatment, such as the breast cancer drug tamoxifen and the leukemia medication imatinib (Pharma Models). However, compared to mammals that have remarkable characteristics related to cancer, mice and rats may offer limited information regarding cancer resistance, due to their proneness to the disease and lack of genetic variation (de Magalhães).

The field of biomimicry, or biologically inspired engineering, focuses on mimicking elements of nature to invent solutions to human problems (Detanico). The Wright Brothers, for example, observed the flight of birds to invent human flight, leading to the creation of the first airplanes. Similarly, mammals that are resistant to cancer can provide valuable insight into preventing cancer in humans. Cancer-resistant mammals, such as the blind mole rat, have evolved unique and efficient mechanisms to decrease their risk of cancer. These mammals tend to possess a form of efficient and sensitive surveillance of hyperplasia, which allows them to detect and eliminate possible cancer cells in an effective manner. Research that focuses on studying the immune system of these mammals could allow us to emulate their resistance in humans through vaccines or other therapies that enhance the human's immune response to cancer. Other research may focus on using the elephant as a model to engineer extra p53 copies in humans or using the bat as a model to engineer bat-specific gene mutations (BRCA2, IFIT2, and PALB2) that increase cancer resistance.

While cancer-resistant mammals could prove to be invaluable research models, various challenges arise in working with non-traditional animal models. Housing some cancer-resistant mammals, such as elephants, in a laboratory setting is impractical. However, elephants and other cancer-resistant mammals do not have to be housed in a laboratory to be used as models for cancer research. Genes from these mammals can be studied and researchers can partner with zoos to study cancer-resistant mammals such as the elephant and naked mole rat. Additionally, non-traditional animals that are not commonly used as research models have much fewer species-specific resources. Compared to traditional research models, using non-traditional animals as research models can be more difficult, time-consuming, and expensive due to the lack of knowledge surrounding these animals in a research context. To facilitate the research process, methods, protocols, and databases should be further developed for these species. Expanding pre-existing research programs to study how cancer development and resistance work in these remarkably cancer-resistant mammals could lead to extraordinary discoveries that could be applied to cancer prevention and treatment in humans, as well as advance our understanding of cancer across species.

Conclusion

Cancer-resistant mammals tend to exhibit enhanced DNA repair and unique responses to cell proliferation and tumorigenesis. The reasons behind these remarkable adaptations that contribute to cancer resistance vary. Certain species evolved novel adaptations due to their unique lifestyles, and these adaptations have subsequently led to an increased lifespan and decreased cancer risk. The naked mole secretes a high amount of HMM-HA, an adaptation that initially allowed these mammals to live comfortably underground. Now, HMM-HA plays an essential role in early contact inhibition in the naked mole rat, a mechanism that has led to cancer resistance. Similarly, blind mole rats possess a unique p53 sequence which initially evolved to increase their hypoxia resistance. This p53 sequence now contributes to concerted cell death in the blind mole rat, a remarkable anticancer mechanism. Elephants evolved to have multiple copies of TP53, likely as a way to counteract the increased cancer risk associated with their large

size. Bowhead whales, animals that are much larger than elephants, do not possess multiple copies of TP53 but instead possess many genetic changes that result in advanced DNA repair and immune response. Cows and horses are likely resistant to specific metastatic cancers due to their lack of placental invasion, which leads to a less permissive stroma. Bats are diverse mammals that have unique resistance to viral infections, and these mechanisms, along with genetic changes, are thought to have anticancer effects. Analyzing these mechanisms of cancer resistance found in cancer-resistant mammals, uncovering novel strategies, and discovering how these mechanisms can be utilized for cancer treatment and prevention in humans may lead to key breakthroughs in the fight against cancer.

Acknowledgments

Huge thanks to my mentor, Victoria Tobin, for her invaluable guidance throughout the creation of this paper. This paper truly would not have been possible without her feedback and guidance. Thank you.

Works Cited

Cancer. World Health Organization, https://www.who.int/news-room/fact-sheets/detail/cancer. Accessed 30 Aug. 2023.

"How to Prevent Cancer." *American Institute for Cancer Research*, https://www.aicr.org/cancer-prevention/how-to-prevent-cancer/. Accessed 30 Aug. 2023.

Chadwick, Lisa H. "Telomere." *Genome.gov*, https://www.genome.gov/genetics-glossary/Telomere. Accessed 30 Aug. 2023.

Seluanov, Andrei, et al. "Telomerase Activity Coevolves with Body Mass Not Lifespan." *Aging Cell*, vol. 6, no. 1, Feb. 2007, pp. 45–52, https://doi.org/10.1111/j.1474-9726.2006.00262.x.

Seluanov, Andrei, et al. "Mechanisms of Cancer Resistance in Long-Lived Mammals." *Nature Reviews. Cancer*, vol. 18, no. 7, July 2018, pp. 433–41, https://doi.org/10.1038/s41568-018-0004-9.

Gonzalez, Carlos A., and Elio Riboli. "Diet and Cancer Prevention: Contributions from the European Prospective Investigation into Cancer and Nutrition (EPIC) Study." *European Journal of Cancer*, vol. 46, no. 14, Sept. 2010, pp. 2555–62, https://doi.org/10.1016/j.ejca.2010.07.025.

Key, Timothy J., et al. "Diet, Nutrition and the Prevention of Cancer." *Public Health Nutrition*, vol. 7, no. 1A, Feb. 2004, pp. 187–200, https://doi.org/10.1079/phn2003588.

Vincze, Orsolya, et al. "Cancer Risk across Mammals." *Nature*, vol. 601, no. 7892, Jan. 2022, pp. 263–67, https://doi.org/10.1038/s41586-021-04224-5.

Anisimov, Vladimir N., et al. "Cancer in Rodents: Does It Tell Us about Cancer in Humans?" *Nature Reviews. Cancer*, vol. 5, no. 10, Oct. 2005, pp. 807–19, https://doi.org/10.1038/nrc1715.

Shoshani, Jeheskel. "Proboscidea (Elephants)." *eLS*, John Wiley & Sons, Ltd, 25 July 2001, https://doi.org/10.1038/npg.els.0001575.

Clubb, Ros, et al. "Compromised Survivorship in Zoo Elephants." *Science*, vol. 322, no. 5908, Dec. 2008, p. 1649, https://doi.org/10.1126/science.1164298.

Tollis, Marc, et al. "Elephant Genomes Reveal Accelerated Evolution in Mechanisms Underlying Disease Defenses." *Molecular Biology and Evolution*, vol. 38, no. 9, Aug. 2021, pp. 3606–20, https://doi.org/10.1093/molbev/msab127.

Boddy, Amy M., et al. "Lifetime Cancer Prevalence and Life History Traits in Mammals." *Evolution, Medicine, and Public Health*, vol. 2020, no. 1, May 2020, pp. 187–95, https://doi.org/10.1093/emph/eoaa015.

Landolfi, Jennifer A., et al. "Reproductive Tract Neoplasia in Adult Female Asian Elephants (Elephas Maximus)." *Veterinary Pathology*, vol. 58, no. 6, Nov. 2021, pp. 1131–41, https://doi.org/10.1177/03009858211031843.

Montali, R.J., et al. *Ultrasonography and Pathology of Genital Tract Leiomymoas in Captive Asian Elephants: Implications for Reproductive Soundness*.

https://repository.si.edu/bitstream/handle/10088/6159/1e76ac5f-81ac-48f9-a0e2-6e098f1f5337.pdf. Accessed 30 Aug. 2023.

Abegglen, Lisa M., et al. "Potential Mechanisms for Cancer Resistance in Elephants and Comparative Cellular Response to DNA Damage in Humans." *JAMA: The Journal of the American Medical Association*, vol. 314, no. 17, Nov. 2015, pp. 1850–60, https://doi.org/10.1001/jama.2015.13134.

Hanahan, Douglas, and Robert A. Weinberg. "Hallmarks of Cancer: The next Generation." *Cell*, vol. 144, no. 5, Mar. 2011, pp. 646–74, https://doi.org/10.1016/j.cell.2011.02.013.

Sulak, Michael, et al. "TP53 Copy Number Expansion Is Associated with the Evolution of Increased Body Size and an Enhanced DNA Damage Response in Elephants." *eLife*, vol. 5, Sept. 2016, https://doi.org/10.7554/eLife.11994.

Tyner, Stuart D., et al. "p53 Mutant Mice That Display Early Ageing-Associated Phenotypes." *Nature*, vol. 415, no. 6867, Jan. 2002, pp. 45–53, https://doi.org/10.1038/415045a.

Maier, Bernhard, et al. "Modulation of Mammalian Life Span by the Short Isoform of p53." *Genes & Development*, vol. 18, no. 3, Feb. 2004, pp. 306–19, https://doi.org/10.1101/gad.1162404.

García-Cao, Isabel, et al. "'Super p53' Mice Exhibit Enhanced DNA Damage Response, Are Tumor Resistant and Age Normally." *The EMBO Journal*, vol. 21, no. 22, Nov. 2002, pp. 6225–35, https://doi.org/10.1093/emboj/cdf595.

Buffenstein, Rochelle, et al. "Naked Mole Rat." *The Laboratory Rabbit, Guinea Pig, Hamster, and Other Rodents*, 2012, p. 1055, https://doi.org/10.1016/B978-0-12-380920-9.00045-6.

Buffenstein, Rochelle. "Negligible Senescence in the Longest Living Rodent, the Naked Mole-Rat: Insights from a Successfully Aging Species." *Journal of Comparative Physiology. B, Biochemical, Systemic, and Environmental Physiology*, vol. 178, no. 4, May 2008, pp. 439–45, https://doi.org/10.1007/s00360-007-0237-5.

Delaney, M. A., et al. "Spontaneous Histologic Lesions of the Adult Naked Mole Rat (Heterocephalus Glaber): A Retrospective Survey of Lesions in a Zoo Population." *Veterinary Pathology*, vol. 50, no. 4, July 2013, pp. 607–21, https://doi.org/10.1177/0300985812471543.

Delaney, M. A., et al. "Initial Case Reports of Cancer in Naked Mole-Rats (Heterocephalus Glaber)." *Veterinary Pathology*, vol. 53, no. 3, May 2016, pp. 691–96, https://doi.org/10.1177/0300985816630796.

Seluanov, Andrei, et al. "Hypersensitivity to Contact Inhibition Provides a Clue to Cancer Resistance of Naked Mole-Rat." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 46, Nov. 2009, pp. 19352–57, https://doi.org/10.1073/pnas.0905252106.

Liang, Sitai, et al. "Resistance to Experimental Tumorigenesis in Cells of a Long-Lived Mammal, the Naked Mole-Rat (Heterocephalus Glaber)." *Aging Cell*, vol. 9, no. 4, Aug. 2010, pp. 626–35, https://doi.org/10.1111/j.1474-9726.2010.00588.x.

Manov, Irena, et al. “Pronounced Cancer Resistance in a Subterranean Rodent, the Blind Mole-Rat, Spalax: In Vivo and in Vitro Evidence.” *BMC Biology*, vol. 11, Aug. 2013, p. 91, https://doi.org/10.1186/1741-7007-11-91.

Seluanov, Andrei, et al. “Distinct Tumor Suppressor Mechanisms Evolve in Rodent Species That Differ in Size and Lifespan.” *Aging Cell*, vol. 7, no. 6, Dec. 2008, pp. 813–23, https://doi.org/10.1111/j.1474-9726.2008.00431.x.

Tian, Xiao, et al. “High-Molecular-Mass Hyaluronan Mediates the Cancer Resistance of the Naked Mole Rat.” *Nature*, vol. 499, no. 7458, July 2013, pp. 346–49, https://doi.org/10.1038/nature12234.

Miyawaki, Shingo, et al. “Tumour Resistance in Induced Pluripotent Stem Cells Derived from Naked Mole-Rats.” *Nature Communications*, vol. 7, May 2016, p. 11471, https://doi.org/10.1038/ncomms11471.

López-Otín, Carlos, et al. “The Hallmarks of Aging.” *Cell*, vol. 153, no. 6, June 2013, pp. 1194–217, https://doi.org/10.1016/j.cell.2013.05.039.

O’Connor, Timothy P., et al. “Prolonged Longevity in Naked Mole-Rats: Age-Related Changes in Metabolism, Body Composition and Gastrointestinal Function.” *Comparative Biochemistry and Physiology. Part A, Molecular & Integrative Physiology*, vol. 133, no. 3, Nov. 2002, pp. 835–42, https://doi.org/10.1016/s1095-6433(02)00198-8.

Azpurua, Jorge, et al. “Naked Mole-Rat Has Increased Translational Fidelity Compared with the Mouse, as Well as a Unique 28S Ribosomal RNA Cleavage.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 43, Oct. 2013, pp. 17350–55, https://doi.org/10.1073/pnas.1313473110.

Park, Thomas J., et al. “Fructose-Driven Glycolysis Supports Anoxia Resistance in the Naked Mole-Rat.” *Science*, vol. 356, no. 6335, Apr. 2017, pp. 307–11, https://doi.org/10.1126/science.aab3896.

Liu, Haibo, et al. “Fructose Induces Transketolase Flux to Promote Pancreatic Cancer Growth.” *Cancer Research*, vol. 70, no. 15, Aug. 2010, pp. 6368–76, https://doi.org/10.1158/0008-5472.CAN-09-4615.

Edrey, Yael H., et al. “Sustained High Levels of Neuregulin-1 in the Longest-Lived Rodents; a Key Determinant of Rodent Longevity.” *Aging Cell*, vol. 11, no. 2, Apr. 2012, pp. 213–22, https://doi.org/10.1111/j.1474-9726.2011.00772.x.

Gorbunova, Vera, et al. “Cancer Resistance in the Blind Mole Rat Is Mediated by Concerted Necrotic Cell Death Mechanism.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 47, Nov. 2012, pp. 19392–96, https://doi.org/10.1073/pnas.1217211109.

D’Arcy, Mark S. “Cell Death: A Review of the Major Forms of Apoptosis, Necrosis and Autophagy.” *Cell Biology International*, vol. 43, no. 6, June 2019, pp. 582–92, https://doi.org/10.1002/cbin.11137.

Ashur-Fabian, Osnat, et al. “Evolution of p53 in Hypoxia-Stressed Spalax Mimics Human Tumor Mutation.” *Proceedings of the National Academy of Sciences of the United States*

of America, vol. 101, no. 33, Aug. 2004, pp. 12236–41, https://doi.org/10.1073/pnas.0404998101.

Fang, Xiaodong, et al. "Genome-Wide Adaptive Complexes to Underground Stresses in Blind Mole Rats Spalax." *Nature Communications*, vol. 5, June 2014, p. 3966, https://doi.org/10.1038/ncomms4966.

Nasser, Nicola J., et al. "Alternatively Spliced Spalax Heparanase Inhibits Extracellular Matrix Degradation, Tumor Growth, and Metastasis." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 7, Feb. 2009, pp. 2253–58, https://doi.org/10.1073/pnas.0812846106.

Fisheries, Noaa. "Bowhead Whale." *NOAA Fisheries*, 8 Mar. 2023, https://www.fisheries.noaa.gov/species/bowhead-whale.

Keane, Michael, et al. "Insights into the Evolution of Longevity from the Bowhead Whale Genome." *Cell Reports*, vol. 10, no. 1, Jan. 2015, pp. 112–22, https://doi.org/10.1016/j.celrep.2014.12.008.

"Death Receptor Pathway." *Thermo Fisher Scientific*, https://www.thermofisher.com/us/en/home/life-science/antibodies/antibodies-learning-center/antibodies-resource-library/cell-signaling-pathways/apoptosis-death-receptors.html. Accessed 30 Aug. 2023.

Tollis, Marc, et al. "Return to the Sea, Get Huge, Beat Cancer: An Analysis of Cetacean Genomes Including an Assembly for the Humpback Whale (Megaptera Novaeangliae)." *Molecular Biology and Evolution*, vol. 36, no. 8, Aug. 2019, pp. 1746–63, https://doi.org/10.1093/molbev/msz099.

Seim, Inge, et al. "The Transcriptome of the Bowhead Whale Balaena Mysticetus Reveals Adaptations of the Longest-Lived Mammal." *Aging*, vol. 6, no. 10, Oct. 2014, pp. 879–99, https://doi.org/10.18632/aging.100699.

D'Souza, Alaric W. and Günter P. Wagner "Malignant Cancer and Invasive placentation A Case for Positive Pleiotropy between Endometrial and Malignancy Phenotypes." *Evolution, Medicine, and Public Health*, vol. 2014, no. 1, Jan. 2014, pp. 136–45, https://doi.org/10.1093/emph/eou022.

Moffett, Ashley, and Charlie Loke. "Immunology of Placentation in Eutherian Mammals." *Nature Reviews. Immunology*, vol. 6, no. 8, Aug. 2006, pp. 584–94, https://doi.org/10.1038/nri1897.

Graham, Charles H., and Peeyush K. Lala. "Mechanisms of Placental Invasion of the Uterus and Their Control." *Biochemistry and Cell Biology = Biochimie et Biologie Cellulaire*, Jan. 2011, https://doi.org/10.1139/o92-135.

Kshitiz, et al. "Evolution of Placental Invasion and Cancer Metastasis Are Causally Linked." *Nature Ecology & Evolution*, vol. 3, no. 12, Dec. 2019, pp. 1743–53, https://doi.org/10.1038/s41559-019-1046-4.

Ramos, Adriano Tony, et al. “Tumores em animais de produção: aspectos comparativos.” *Ciencia rural* , vol. 38, no. 1, Feb. 2008, pp. 148–54, https://doi.org/10.1590/S0103-84782008000100024.

Anderson, L. J., and A. T. Sandison. “Tumours of Connective Tissues in Cattle, Sheep and Pigs.” *The Journal of Pathology*, vol. 98, no. 4, Aug. 1969, pp. 253–63, https://doi.org/10.1002/path.1710980405.

Cotchin, E. “A General Survey of Tumours in the Horse.” *Equine Veterinary Journal*, vol. 9, no. 1, Jan. 1977, pp. 16–21, https://doi.org/10.1111/j.2042-3306.1977.tb03964.x.

Ogłuszka, Magdalena, et al. “Equine Sarcoids-Causes, Molecular Changes, and Clinicopathologic Features: A Review.” *Veterinary Pathology*, vol. 58, no. 3, May 2021, pp. 472–82, https://doi.org/10.1177/0300985820985114.

Middleton, D. J., et al. “Experimental Nipah Virus Infection in Pteropid Bats (Pteropus Poliocephalus).” *Journal of Comparative Pathology*, vol. 136, no. 4, May 2007, pp. 266–72, https://doi.org/10.1016/j.jcpa.2007.03.002.

Olds, June E., et al. “RETROSPECTIVE EVALUATION OF CASES OF NEOPLASIA IN A CAPTIVE POPULATION OF EGYPTIAN FRUIT BATS (ROUSETTUS AEGYPTIACUS).” *Journal of Zoo and Wildlife Medicine: Official Publication of the American Association of Zoo Veterinarians*, vol. 46, no. 2, June 2015, pp. 325–32, https://doi.org/10.1638/2014-0069R2.1.

Seim, Inge, et al. “Genome Analysis Reveals Insights into Physiology and Longevity of the Brandt’s Bat Myotis Brandtii.” *Nature Communications*, vol. 4, 2013, p. 2212, https://doi.org/10.1038/ncomms3212.

Ricci, Marco, et al. “Comparative Analysis of Bats and Rodents’ Genomes Suggests a Relation between Non-LTR Retrotransposons, Cancer Incidence, and Ageing.” *Scientific Reports*, vol. 13, no. 1, June 2023, p. 9039, https://doi.org/10.1038/s41598-023-36006-6.

Callinan, P. A., and M. A. Batzer. “Retrotransposable Elements and Human Disease.” *Genome Dynamics*, vol. 1, 2006, pp. 104–15, https://doi.org/10.1159/000092503.

Lambert, Matthew J., and Christine V. Portfors. “Adaptive Sequence Convergence of the Tumor Suppressor ADAMTS9 between Small-Bodied Mammals Displaying Exceptional Longevity.” *Aging*, vol. 9, no. 2, Feb. 2017, pp. 573–82, https://doi.org/10.18632/aging.101180.

Croco, Eleonora, et al. “DNA Damage Detection by 53BP1: Relationship to Species Longevity.” *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, vol. 72, no. 6, June 2017, pp. 763–70, https://doi.org/10.1093/gerona/glw170.

Zhang, Guojie, et al. “Comparative Analysis of Bat Genomes Provides Insight into the Evolution of Flight and Immunity.” *Science*, vol. 339, no. 6118, Jan. 2013, pp. 456–60, https://doi.org/10.1126/science.1230835.

Foley, Nicole M., et al. “Growing Old, yet Staying Young: The Role of Telomeres in Bats’ Exceptional Longevity.” *Science Advances*, vol. 4, no. 2, Feb. 2018, p. eaao0926, https://doi.org/10.1126/sciadv.aao0926.

Harper, James M., et al. “Skin-Derived Fibroblasts from Long-Lived Species Are Resistant to Some, but Not All, Lethal Stresses and to the Mitochondrial Inhibitor Rotenone.” *Aging Cell*, vol. 6, no. 1, Feb. 2007, pp. 1–13, https://doi.org/10.1111/j.1474-9726.2006.00255.x.

Khaidakov, Magomed, et al. “Direct Repeats in Mitochondrial DNA and Mammalian Lifespan.” *Mechanisms of Ageing and Development*, vol. 127, no. 10, Oct. 2006, pp. 808–12, https://doi.org/10.1016/j.mad.2006.07.008.

Huang, Zixia, et al. “Blood miRNomes and Transcriptomes Reveal Novel Longevity Mechanisms in the Long-Lived Bat, Myotis Myotis.” *BMC Genomics*, vol. 17, no. 1, Nov. 2016, p. 906, https://doi.org/10.1186/s12864-016-3227-8.

Scheben, Armin, et al. “Long-Read Sequencing Reveals Rapid Evolution of Immunity- and Cancer-Related Genes in Bats.” *bioRxiv*, 19 Aug. 2021, p. 2020.09.09.290502, https://doi.org/10.1101/2020.09.09.290502.

Shen, Yong-Yi, et al. “Adaptive Evolution of Energy Metabolism Genes and the Origin of Flight in Bats.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 19, May 2010, pp. 8666–71, https://doi.org/10.1073/pnas.0912613107.

Pride, Harrison, et al. “Long-Lived Species Have Improved Proteostasis Compared to Phylogenetically-Related Shorter-Lived Species.” *Biochemical and Biophysical Research Communications*, vol. 457, no. 4, Feb. 2015, pp. 669–75, https://doi.org/10.1016/j.bbrc.2015.01.046.

Salmon, Adam B., et al. “The Long Lifespan of Two Bat Species Is Correlated with Resistance to Protein Oxidation and Enhanced Protein Homeostasis.” *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, vol. 23, no. 7, July 2009, pp. 2317–26, https://doi.org/10.1096/fj.08-122523.

Wilkinson, Gerald S., and Jason M. South. “Life History, Ecology and Longevity in Bats.” *Aging Cell*, vol. 1, no. 2, Dec. 2002, pp. 124–31, https://doi.org/10.1046/j.1474-9728.2002.00020.x.

“Understanding What Cancer Is: Ancient Times to Present.” *American Cancer Society*, https://www.cancer.org/cancer/understanding-cancer/history-of-cancer/what-is-cancer.html. Accessed 30 Aug. 2023.

“Involving Animals in Research.” *Cancer Research UK*, 15 Dec. 2016, https://www.cancerresearchuk.org/our-research/involving-animals-in-research.

“Animal Testing Significantly Advances Cancer Research.” *Pharma Models*, 13 May 2014, https://www.pharmamodels.net/blog/animal-testing-significantly-advances-cancer-research/.

de Magalhães, João Pedro. “The Big, the Bad and the Ugly: Extreme Animals as Inspiration for Biomedical Research.” *EMBO Reports*, vol. 16, no. 7, July 2015, pp. 771–76, https://doi.org/10.15252/embr.201540606.

Detanico, Flora Bittencourt. “Biomimicry: Human Creation Inspired By Nature.” *New Acropolis Library*, 18 May 2015, https://library.acropolis.org/biomimicry-human-creation-inspired-by-nature/.

Cavalcante, Giovanna C., et al. “Mitochondria in Tumour Progression: A Network of mtDNA Variants in Different Types of Cancer.” *BMC Genetics*, vol. 23, no. 1, Springer Nature, Feb. 2022, https://doi.org/10.1186/s12863-022-01032-2.

Economics of Bitcoin and Inflation: Is Bitcoin an Inflation Hedge? By Gokulakrishnan Rathinasamy

Abstract

In 2022, inflation has surged to levels not witnessed since the 1980s. The U.S. Consumer Price Index measured inflation of 3% in June 2023 and the highest inflation of 9.1% in June 2022. Cryptocurrency (crypto) is a relatively new asset class and has grown significantly over the past few years; however, the study between inflation and the price of crypto has not been well studied. This research paper examines the correlation between the price of crypto and inflation, highlighting both positive and negative effects. The research employs comparative analysis and market trends to investigate the trends of the volatility of the crypto market. This study will address many questions, such as whether the crypto Bitcoin is an inflation hedge or a safe haven and the economics of Bitcoin and inflation.

Keywords: inflation, inflation hedge, Bitcoin, safe haven, DXY, CCI, Gold, money supply

Introduction

In the modern financial landscape, cryptocurrencies have emerged as decentralized digital assets leveraging blockchain technology, presenting a novel avenue for conducting financial transactions and allocating assets. Despite the growing prominence of cryptocurrencies, their susceptibility to inflation and the implications for startups and companies operating in this domain have yet to be comprehensively examined. The intricate relationship between inflation and cryptocurrencies poses important inquiries, including the viability of Bitcoin as an inflation hedge and the potential volatility that might impede cryptocurrencies from assuming this role.

1.2 Federal Reserve

Inflation and rapidly changing interest rates have posed challenges for many investors, companies, and governmental entities. The impact of inflation across various sectors of the U.S. economy includes the stock market, the housing market, and the treasury sectors. To address inflation, the Federal Reserve has taken measures such as increasing interest rates, which can tamper economic activity to combat inflation: “Additional fiscal support could be costly, but worth it if it helps avoid long-term economic damage and leaves [the economy] with a stronger recovery” (Powell). These actions aim to ensure that consumers can maintain their purchasing power and avoid facing rapidly increasing prices, such as paying double the amount for everyday goods like a carton of eggs.

1.3 Gold and Bitcoin Similarities

Gold is a tangible asset known for its ability to hedge inflation: “Gold is a high liquid asset, which is no one’s liability, carries no credit risk, and is scarce, historically preserving its value over time” (World Gold Council). Other investors, such as Larry Fink, the CEO of

Blackrock, have speculated about whether Bitcoin is the “digital gold” (Goldstein). Gold has been used as an inflation hedge due to its limited supply and its physical characteristics. For instance, when inflation rises, which devalues fiat currency, people look for tangible assets with intrinsic value: “holding paper money will be at risk of devaluation, and investors will invest relatively more in gold to preserve its value” (Borui Xiao). Gold and Bitcoin have similar characteristics, including limited supply and “negative correlation with orthodox paper assets like bonds and stocks” (Ankit Som). Bitcoin's remarkable volatility stems from its emerging nature within the financial landscape. However, many investors remain cautious despite Bitcoin’s potential, believing that Bitcoin's limited historical data makes it risky for individuals and businesses. One prominent skeptic of Bitcoin's prospects is Warren Buffet, the CEO of Berkshire Hathaway, who is known as one of the world’s greatest investors. According to Buffet, Bitcoin doesn’t produce any value that would benefit the economy: “If you told me you own all of the Bitcoin in the world and you offered it to me for $25, I wouldn’t take it because what would I do with it? I’d have to sell it back to you one way or another. It isn’t going to do anything. The apartments will produce rent, and the farms will produce food” (Warren Buffet). Historically, Bitcoin has never existed during an economic period where inflation was over 5%. However, a notable anomaly occurred in June 2022 when Bitcoin experienced a decline and inflation peaked at 9.1%. This led investors to question Bitcoin's viability as an effective hedge against inflation. Nevertheless, it is essential to consider various external factors that might have contributed to Bitcoin's price drop during that period.

1.4 Objective

This research paper has a dual objective. First, it seeks to assess the effectiveness of Bitcoin as a hedge against inflation. Second, it aims to give crypto-related companies valuable insights into the economic dynamics surrounding the correlation between Bitcoin pricing and inflation. By addressing these two key aspects, this study aims to contribute to a better understanding of the role Bitcoin plays in relation to inflation and its implications for businesses in the cryptocurrency sector.

Literature Review

As defined by the Federal Reserve, inflation is the rise of the prices of goods and services, which can lead to substantial economic problems. Monetarists believe inflation occurs due to the expansion of the money supply above the size of an economy (see Table 1). In 2020, the Fed significantly increased the production of money, which could induce inflation. According to the St. Louis reserve bank, “inflation occurs when the money supply in an economy grows faster than the economy’s ability to produce goods and services” (see Table 2). Suppose the economy has so much money circulating. In that case, it creates excessive demand for products and services, which leads to an increase in prices: “More people [have] money to buy things, but there was only so much supply, so prices increased. The newly-printed money they found lost value and did not allow everyone to buy goods” (Liberty University).

Table 1: M2 Money Supply

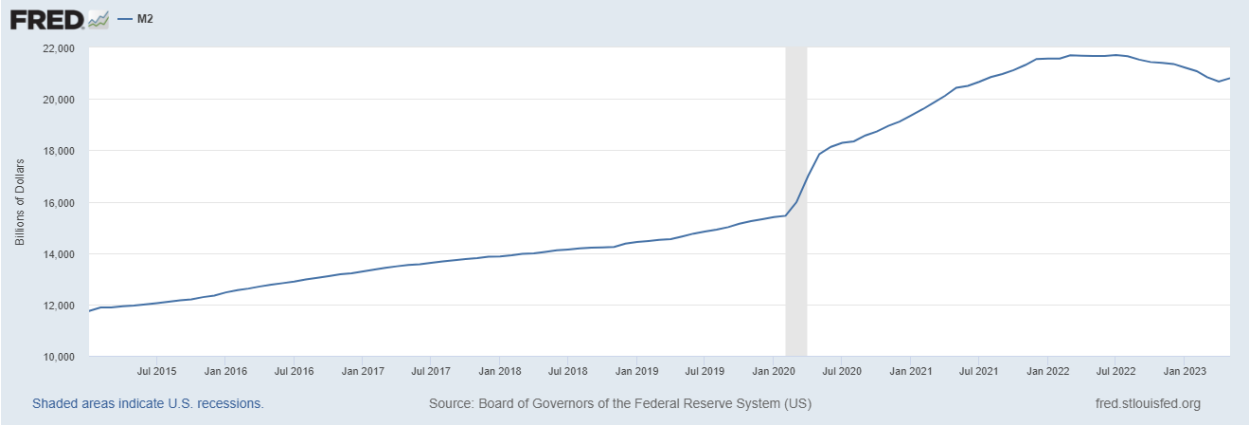
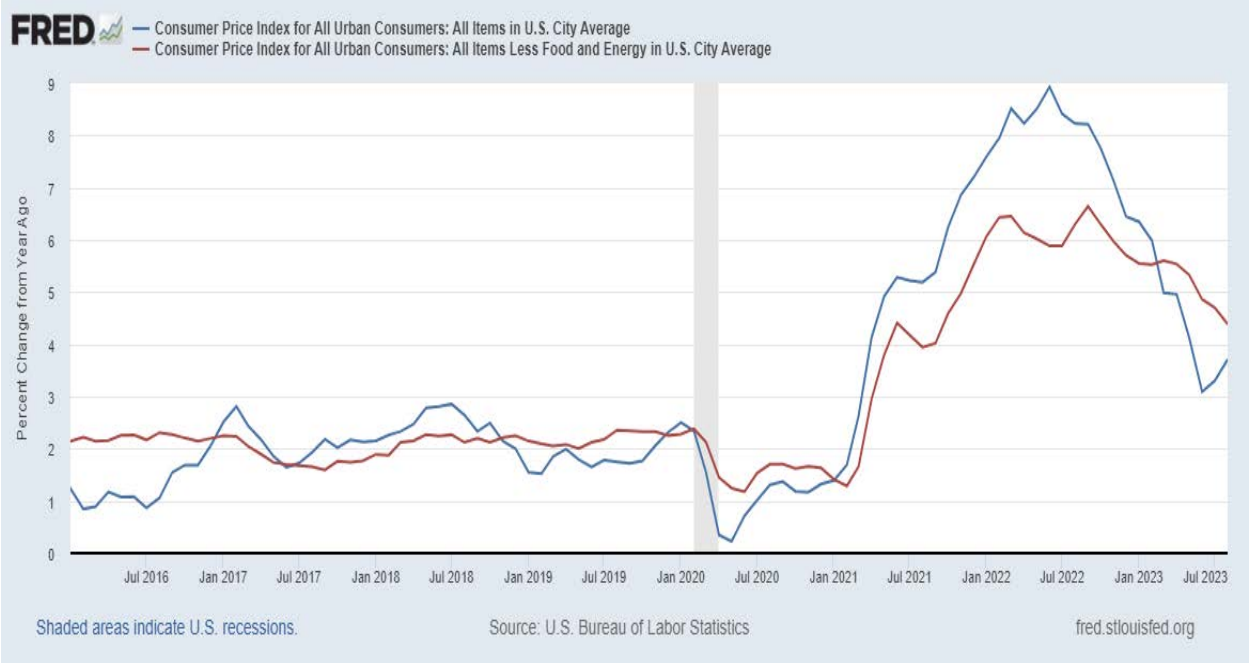


Table 2: Consumer Prices Indexes:

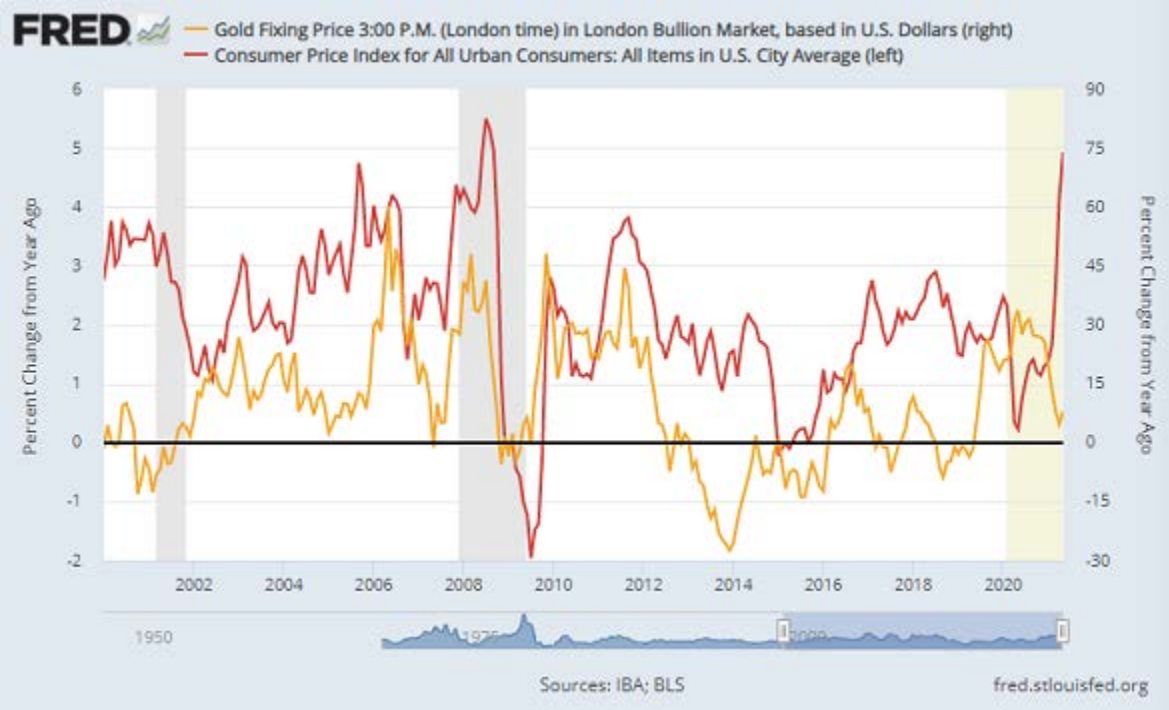


2.2 Inflationary Psychology

Grep Ip, the author of *The Little Book of Economics,* stated, "There are two competing schools of thought on the causes of inflation. Listening to them is like listening to creationists and Darwinians argue over how life began." First, as I have mentioned, are monetarists who believe that increasing money causes inflation. Second, some neo-Keynesians blame both excess spending and inflationary psychology, "in which expectations of high inflation drive high inflation" (RMIT University). Both theories are possible and synergistic; the market is technically and emotionally driven. When investors predict inflation in the coming months, the

market reacts to those pessimistic emotions from the investors. When inflation rises, people are in a dilemma of whether to hold or sell to look for a tangible or hedging asset. Intuit, one of the biggest fintech companies, defined tangible assets as investments that have a physical form so that they can be touched, such as gold or silver.

Table 3: Gold & Consumer Price Index



2.1 Bitcoin's Creation

One of the most significant recessions in U.S. history was the 2008 financial crisis. It was the most severe crisis since the Great Depression. It led to people losing their houses, the failure of major businesses, and a 48% decline in the S&P 500. Consequently, people needed to figure out what to do with their money. Following many bank failures, consumers were skeptical about putting their money in banks. According to the Federal Deposit Insurance Corporation, 25 banks failed. This also caused the Consumer Sentiment Index, or individuals' confidence in an economy, to reach its all-time low.



Table 4: Consumer Sentiment

However, right after the crisis, a software engineer named Satoshi Nakamoto created a cryptocurrency called Bitcoin. His exact identity is still unknown, but many theories about why he invented Bitcoin are due to the 2008 recession since people needed a safe currency. OSWEGO University defined cryptocurrency as a digital currency that is an alternative form of payment utilizing encryption algorithms. In addition, "cryptocurrency's transactions are recorded. The Records are maintained by a decentralized system using cryptography, rather than by a centralized authority" (AFS finance group). Companies that invest heavily in Bitcoin or Crypto startups need clarification and regulation because this is the first time Bitcoin has seen an exceptionally high inflation rate. However, many significant crypto companies concluded it might be an inflation hedge due to its similarities to Gold. The Nasdaq defines an inflation hedge as an "investment designed to protect against inflation and the associated loss of purchasing power." In addition, "Bitcoin behaves similarly to commodities like crude oil and gold, as Bitcoin shares characteristics such as the 'fixed supply' with exhaustible resource commodities" (Sangyup Choi). Binance also said, "Bitcoin's supply will never exceed 21 million, which makes it an attractive store of value resistant to inflation." There are 19,276,325 Bitcoins in existence as of Jan. 29, 2023.

Other Perspectives

Sangyup Choi, a professor at the School of Economics of Yonsei University, found that Bitcoin prices decrease due to financial uncertainty when "an aggregate variable, such as GDP, becoming less predictable" (Columbia University). He also found that Bitcoin prices do not decline in correlation to uncertain policy shocks or changes in government policy. His findings have shown that Bitcoin is *not* a safe haven, meaning assets "negatively correlated with the financial cycle, so safe haven assets tend to provide positive returns during crises" (Rahma Chemkha).

On the other side, many news companies have disagreed with the theory that Bitcoin is an inflation hedge. For instance, Forbes has said, "Bitcoin is not digital gold, and the investment

case of Bitcoin being an inflation hedge is dead. They found that while the S&P 500 was significantly down in September 2022, it was still doing much better than Bitcoin." They even said that the "entire cryptocurrency market is a dud." However, our results show that Bitcoin negatively correlates with the Dollar Index. In addition, we found a decrease in consumer confidence, leading to increased Bitcoin prices. David Hollerith, a senior reporter at Yahoo, found individuals acquiring more cryptocurrency during a decline in their local currency: "Those living in countries with local currencies that have declined 50% or more in value against the U.S. dollar over the past decade — such as South Africa, Mexico, India, and Brazil — were five times more likely to say they planned to purchase cryptocurrency in the coming year." His findings show that a decline in confidence can lead consumers to look into alternative investments to put their money.

Methodology

To analyze if Bitcoin is an inflation hedge or a safe haven, our methodology was a comparison technique with both the Dollar Index (DXY) and the Consumer Confidence Index. Wayne Duggan, a Forbes Advisor contributor, described the Dollar Index as "a market index benchmark used to measure the value of the U.S. dollar relative to other widely traded international currencies." In addition, "The Federal Reserve established the Dollar Index in 1973 to track the value of the U.S. dollar. Two years earlier, President Richard Nixon abandoned the gold standard, which allowed the dollar's value to float freely in foreign exchange (forex) markets" (Wayne Duggan). Economists use the Dollar Index to analyze the economy's performance through exports and imports. For instance, "The more goods the U.S. exports, the more international demand for U.S. dollars to purchase those goods. When demand for the dollar is high, USDX rises" (Wayne Duggan).

In addition, we compared Bitcoin price to the Consumer Confidence Index, which "details consumer attitudes, buying intentions, vacation plans, and consumer expectations for inflation, stock prices, and interest rates" (Conference Board). The Conference Board administers the Consumer Confidence Index, which is based on the Consumer Confidence Survey with "a responding sample size of 3,000 questionnaires" (AKHILESH GANTI). Using the data collected, the Conference Board can analyze the optimism throughout the economy. This predicts the economy because it can show where it is heading in the long term. For instance, OECD states, "An indicator above 100 signals a boost in the consumers' confidence towards the future economic situation, as a consequence of which they are less prone to save and more inclined to spend money on major purchases in the next 12 months." We also used the inflation rate from the consumer price index, which "measures the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services" (U.S. Bureau of Labor Statistics).

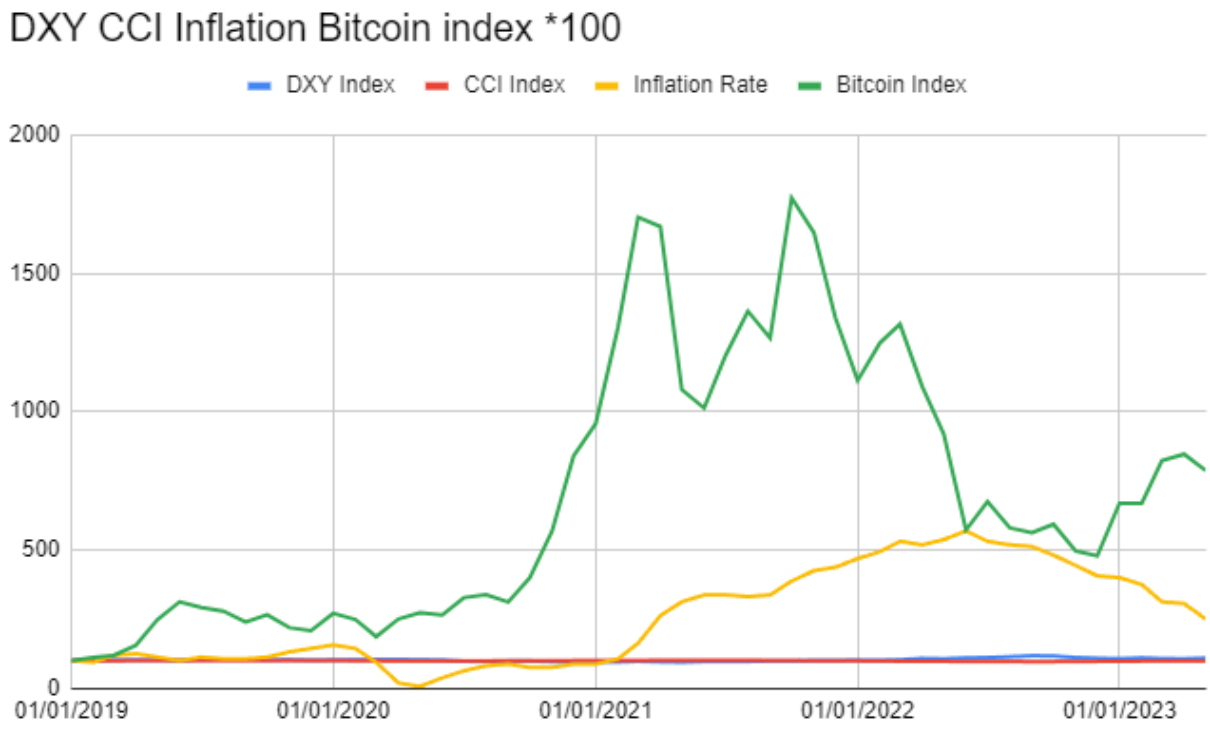
We used the Dollar and Consumer Confidence Indexes because there is a 4-way correlation between all the economic indicators we used. For example, consumer confidence is driven by the rate of inflation. When inflation rises, the dollar's purchasing power decreases,

meaning each dollar buys fewer goods and services. This can lead to a decrease in the value of the Dollar Index as it becomes relatively weaker against other currencies. Investors and consumers notice the dollar devaluation, which decreases their confidence. This leads to how this would impact cryptocurrency and Bitcoin, specifically. According to Robert Kiyosaki, "If the dollar collapses, it might create or stimulate the rise of crypto as a global currency, and in turn, that rise would spell the end for the power of the U.S. dollar globally."

In this research study, we conducted a comparative analysis utilizing four pivotal economic indicators. Specifically, we investigated the correlation between the Dollar Index and the price of Bitcoin and examined the relationship between the Consumer Confidence Index and Bitcoin's price. Furthermore, we investigated the inflation rate and its association with the Consumer Confidence Index and the Dollar Index. These comparisons have yielded profound insights into the intricate interplay between inflation and the dynamic behavior of Bitcoin in the economic landscape. Our research revealed an interdependent relationship among all four economic indicators, highlighting a synergistic relationship. This synergy provides valuable insights into the dynamic interactions between inflation and Bitcoin's behavior in the economic landscape.

Results

Table 5: Comparison of Dollar Index | Consumer Confidence Index | Bitcoin | Inflation | Total Index*100



We found that Bitcoin is an inflation hedge. This analysis shows a negative correlation between Bitcoin and the Dollar Index. During the dates 07/01/21- 11/01/21, we see both an increase in Bitcoin and a substantial increase in inflation. However, we see Bitcoin decrease during 05/01/22. This shows Bitcoin is extremely volatile (unlike Gold), but due to its negative correlation with the Dollar Index, we can conclude that it has hedging characteristics.

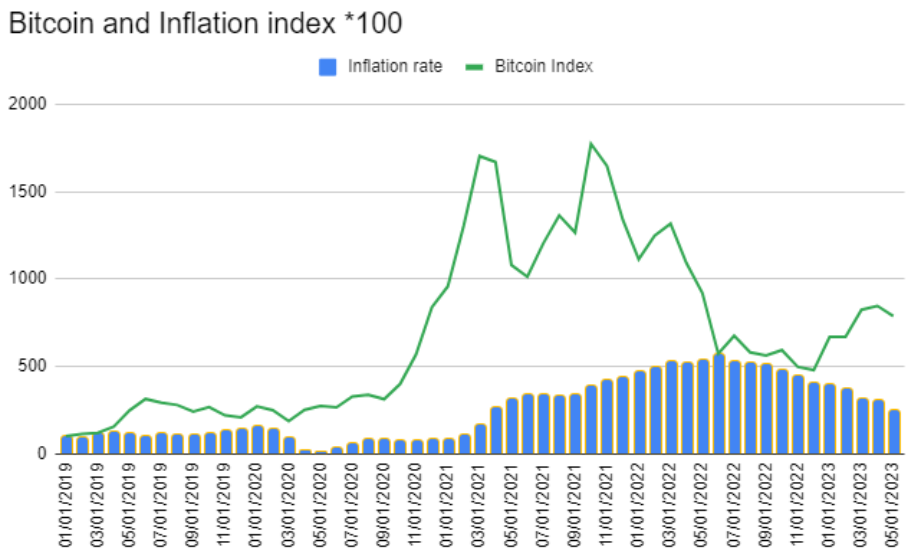


The Consumer Confidence Index shows people's optimism levels each year. We can see that Consumer Confidence during 07/01/21-11/01/21 has a slight decrease while Bitcoin has an increase. When consumers are low in happiness, they look into other assets to put their money in. David Hollerith, a senior reporter at Yahoo, noted a surge in cryptocurrency adoption among individuals seeking protection from declining local currencies. Historically, Gold has served as a reliable hedge against inflation due to its limited supply and tangibility. With its finite supply of 21 million coins, Bitcoin exhibits similar hedging characteristics, making it a potential option to safeguard against inflation.



Additionally, Bitcoin's decentralized authority further bolsters its appeal, as no entity can manipulate its price fluctuations. However, it is essential to note that Bitcoin has only been in the market since 2008, leaving us with no historical background to see how it behaves during periods of inflation. With no historical background, the data is limited to Bitcoin's reaction from 2021-2023 inflation. Many questions still need to be answered, such as what will be the future Bitcoin reactions to inflation? How will Bitcoin react to deflation? There are still many setbacks to this research paper. For instance, many external factors could have caused inflation and the decrease of Bitcoin, such as the 2022 collapse of FTX and the SEC lawsuit against Binance concerning illegal crypto exchanges. Additionally, the Ukraine–Russia war has caused many economic concerns. Our research concludes that Bitcoin is an inflation hedge, but it may not be considered a safe haven asset due to its inherent volatility during economic downturns. The price fluctuations and unpredictability of Bitcoin render it less suitable as a stable and secure investment option.

Our findings demonstrate that Bitcoin's status as an inflation hedge arises from its negative correlation with the Dollar and Consumer Confidence Indexes.



Conclusion

Startups or companies in the cryptocurrency market need to understand that Bitcoin is an inflation hedge. Therefore, they can feel at ease when dealing with cryptocurrency when inflation rises. This research paper answers the question of whether Bitcoin is an inflation hedge and how it correlates with inflation. We found that Bitcoin is indeed an inflation hedge due to its similar characteristics to the price of Gold and its negative correlation with the Dollar Index. However, we did see a significant decrease in Bitcoin during inflationary periods but concluded that we do not have enough historical data because Bitcoin was created after the 2008 recession. Without enough data, we can not pinpoint its specific patterns with inflation due to its first time feeling a 9% inflation rate. This research question impacts the world by allowing startups to feel calm because we concluded that Bitcoin is an inflation hedge; however, many are still skeptical about whether it is a good investment. Many predict crypto to be the powerhouse for payments and currency, so this paper provides valuable information on Bitcoin's reactions toward inflation: "Cryptocurrencies have grown from digital novelties to trillion-dollar technologies with the potential to disrupt the global financial system" (Anshu Siripurapu). On the other hand, "cryptocurrencies have no legislated or intrinsic value; they are simply worth what people are willing to pay for them in the market" (Reserve Bank of Australia

Works Cited

AFS financial group. “What Is Cryptocurrency?” *Www.afsfinancialgroup.com*, 8 Sept. 2021, www.afsfinancialgroup.com/blog/cryptocurrency-9-2-21. Accessed 28 Aug. 2023.

BOARD OF GOVERNORS of the FEDERAL RESERVE SYSTEM. “The Fed - What Is Inflation and How Does the Federal Reserve Evaluate Changes in the Rate of Inflation?” *Board of Governors of the Federal Reserve System*, 2016, www.federalreserve.gov/faqs/economy_14419.htm.

Board of Governors of the Federal Reserve System (US). “M2 Money Stock.” *Stlouisfed.org*, 2020, fred.stlouisfed.org/series/M2SL.

Browder, Keaton. “Opinion: Printing More Currency Will Not Fix Inflation.” *The Liberty Champion*, 15 Nov. 2021, www.liberty.edu/champion/2021/11/opinion-printing-more-currency-will-not-fix-inflation/.

Chemkha, Rahma, et al. “Hedge and Safe Haven Properties during COVID-19: Evidence from Bitcoin and Gold.” *The Quarterly Review of Economics and Finance*, Aug. 2021, https://doi.org/10.1016/j.qref.2021.07.006.

CoinBase. “What Is Inflation?” *Www.coinbase.com*, www.coinbase.com/learn/crypto-basics/what-is-inflation.

Council, World Gold. “The Case for a Strategic Allocation to Gold.” *World Gold Council*, World Gold Council, 31 Mar. 2023, www.gold.org/goldhub/research/relevance-of-gold-as-a-strategic-asset.

Duggan, Wayne. “What Is the U.S. Dollar Index? – Forbes Advisor.” *Www.forbes.com*, 31 May 2023, www.forbes.com/advisor/investing/dollar-index/.

Elkins, Meg , and Robert Hoffmann. “Inflationary Psychology Could Make Things Worse, but for Now It’s in Check.” *Www.rmit.edu.au*, 23 Aug. 2022, www.rmit.edu.au/news/acumen/inflationary_psychology.

Federal Reserve Bank of St. Louis. “Money and Inflation - the Feducation Video Series.” *Www.stlouisfed.org*, 22 Jan. 2013, www.stlouisfed.org/en/education/feducation-video-series/episode-1-money-and-inflation.

GANTI, AKHILESH . “Consumer Confidence Index (CCI): Definition and What It Indicates.” *Investopedia*, 8 Aug. 2023, www.investopedia.com/terms/c/cci.asp#:~:text=The%20CCI%20provides%20insight%20into.

Goldstein, Steve. “BlackRock CEO Larry Fink Compares Bitcoin to Digital Gold as ETF Applications at SEC Stack Up.” *MarketWatch*, 6 July 2023, www.marketwatch.com/story/blackrock-ceo-larry-fink-compares-bitcoin-to-digital-gold-as-etf-applications-at-sec-stack-up-395eb12. Accessed 27 Aug. 2023.

Hollerith, David. “Crypto: Bitcoin Is Seen as Inflation Hedge for Newcomers Outside the US.” *Yahoo News*, 4 Apr. 2022,

news.yahoo.com/crypto-bitcoin-inflation-hedge-231009445.html? Accessed 28 Aug. 2023.

Kozeniauskas, Nicholas, et al. “What Are Uncertainty Shocks?” *Journal of Monetary Economics*, vol. 100, Dec. 2018, pp. 1–15, https://doi.org/10.1016/j.jmoneco.2018.06.004. Accessed 10 Jan. 2021.

OECD. “Leading Indicators - Consumer Confidence Index (CCI) - OECD Data.” *OECD*, 2023, data.oecd.org/leadind/consumer-confidence-index-cci.htm.

OSWEGO. “The Basics about Cryptocurrency | CTS.” *Www.oswego.edu*, www.oswego.edu/cts/basics-about-cryptocurrency.

Powell, Jerome . *Current Economic Issues*. Jerome Powell Federal Reserve Speech. Conference.

---. “Speech by Chair Powell on Current Economic Issues.” *Board of Governors of the Federal Reserve System*, 13 May 2020, www.federalreserve.gov/newsevents/speech/powell20200513a.htm.

Reserve Bank of Australia. “Digital Currencies.” *Reserve Bank of Australia*, 2019, www.rba.gov.au/education/resources/explainers/cryptocurrencies.html.

Siripurapu, Anshu, and Noah Berman. “Cryptocurrencies, Digital Dollars, and the Future of Money.” *Council on Foreign Relations*, 24 Sept. 2021, www.cfr.org/backgrounder/cryptocurrencies-digital-dollars-and-future-money.

Skiadopoulos, George. “Why Gold Prices Go up and down – Five Charts.” *The Conversation*, theconversation.com/why-gold-prices-go-up-and-down-five-charts-138918.

Som, Ankit, and Parthajit Kayal. “A Multicountry Comparison of Cryptocurrency vs Gold: Portfolio Optimization through Generalized Simulated Annealing.” *Blockchain: Research and Applications*, Mar. 2022, p. 100075, https://doi.org/10.1016/j.bcra.2022.100075.

Team, Rich Dad Cryptocurrency. “How the Collapse of the U.S. Dollar Could Fuel a Cryptocurrency Revolution.” *Rich Dad | Financial Education & Coaching for Everyone*, 14 June 2023, www.richdad.com/collapse-of-us-dollar#:~:text=If%20the%20dollar%20collapses%2C%20it. Accessed 28 Aug. 2023.

The Conference Board. “US Consumer Confidence.” *The Conference Board*, 2021, www.conference-board.org/topics/consumer-confidence.

U.S. Bureau of Labor Statistics. “CPI Home : U.S. Bureau of Labor Statistics.” *Bls.gov*, 2022, www.bls.gov/cpi/.

---. “FRED Graph.” *Fred.stlouisfed.org*, 28 Aug. 2023, fred.stlouisfed.org/graph/?g=8dGq.

Xiao, Borui. “A Study on the Hedging Function of Gold in the Light of Economic Policy Uncertainty.” *Computational Intelligence and Neuroscience*, vol. 2022, 16 June 2022, p. e1556475, www.hindawi.com/journals/cin/2022/1556475/, https://doi.org/10.1155/2022/1556475.

Youtube. “Buffett: I Wouldn’t Pay $25 for All the Bitcoin in the World.” *Buffett.cnbc.com*, 2 May 2022,

Assessing gKDR-GMM Model's Performance on *C. elegans* Brain Activity Data By Reia Lee

Abstract

Analyzing the activity and structure of neural circuits is important for understanding the mechanisms of information processing in our brain and nervous system. As experimental data on whole-brain activity at the single neuron level and connectome information on neural circuits have been accumulated, computational methods to integrate these data and reproduce the dynamics of neural activity have been demanded. The recently developed gKDR-GMM model is one such method, but its performance has not yet been fully validated. Here, we applied the gKDR-GMM model to publicly available whole-brain neural activity data of *C. elegans*. We compared the time series simulated by the gKDR-GMM model with the experimental data from multiple perspectives including correlation analysis and principal component analysis. We found that the gKDR-GMM model reasonably reproduces the dynamic features of real whole-brain activity. The stable performance of the gKDR-GMM model for different data indicates that this method is promising as a foundation for integrated analysis of the activity and structure of neural circuits.

Introduction

Analysis of whole brain activity dynamics is crucial for understanding the principles underlying brain function. By capturing the patterns of activity among neurons, one can gain insights into how information is processed and how complex behaviors emerge. The recent advancements in large-scale activity imaging techniques offer new opportunities to explore the mechanisms underlying activity pattern generation in the brain and the transmission of information between neurons or neuronal ensembles. Furthermore, simulations and computational neuroscience allow for comprehensive analysis of these imaging data which lead to deeper insights into mechanisms of neuron behavior and patterns.

Caenorhabditis elegans (*C. elegans*), more commonly known as the roundworm, is a commonly used tool in neuroscience because of its simple and well-characterized nervous system. It serves as a highly valuable model organism for comprehensively studying neural circuits and behavior at the systems level. Recent progress in quantitative behavior analysis, neuronal activity assessment, and novel methods for precise circuit control and monitoring have opened up ways for investigating the molecular, cellular, and systems-level mechanisms involved in converting sensory inputs into finely tuned behavioral responses (Sengupta et al.,, 2009). *C. elegans* is comprised of 302 neurons, whose connections have been well established with previous research. In addition, because *C. elegans* is transparent, has a short life cycle, and shares many of its fundamental biological processes and molecular pathways with humans, it is an ideal tool for understanding how the human brain functions (Sengupta et al., 2009).

Various simulations of neural circuits based on structural information of *C. elegans* neural circuits have been performed (Karbowski et. al, 2007, Dunn et al., 2004). On the other hand, although whole-brain activity data of *C. elegans* have been actively obtained in recent years (Schrodel et al, 2013, Kato et al, 2015, Nguyen et al, 2016), few studies have been made of integrating these whole-brain activity data with simulations of neural circuits.

The gKDR-GMM model has been developed to capture and reproduce whole-brain dynamics based on the structure of the neural circuit (Toyoshima 2022). By combining gradient Kernel Dimension Reduction (gKDR) and Gaussian Mixture Model (GMM), this method addresses the high-dimensional and stochastic nature of whole-brain activity dynamics. In fact, the gKDR-GMM method has been applied to whole-brain activity data of *C. elegans* to explore the communication patterns within the neural network. It is reported that the gKDR-GMM model successfully captured the overall relationships of neural activities and reproduced the stochastic yet coordinated dynamics observed in the data. However, the performance of the gKDR-GMM model has not yet been validated using whole-brain activity data produced by groups other than the authors.

In this paper, we test the performance of the gKDR-GMM model by using publicly available whole-brain activity of *C. elegans* (Kato et al, 2015). We compare the simulation (freerun) results of the gKDR-GMM model with the activities of real observed data in three aspects: heatmap of neural activities over time, correlation analysis, and Principal Component Analysis (PCA). The heatmap of neural activities over time enables a direct comparison between simulation and real data. In the correlation analysis, we test how the correlations of neural activities are reproduced by the gKDR-GMM model. Then PCA is applied to the high-dimensional data to verify whether higher-order relationships and temporal development characteristics of neural activities are captured by the gKDR-GMM. In addition, the stability of the performance of gKDR-GMM models is tested by linear regression of correlation coefficients of multiple samples. This multi-faceted analysis not only confirms the performance of the gKDR-GMM but also provides valuable insights into the dynamics and similarities between simulated and real neural activity. By conducting such validation studies, we can strengthen the confidence in the model's performance and enhance its potential as a valuable tool for the comprehensive analysis of brain activities and neural dynamics.

Methods–*Data*

Whole brain activity of *C. elegans* was obtained from Kato2015 whole brain imaging data on the database Center for Open Science (Kato et. al, 2015). The data with no stimulation (WT_NoStim.mat) was used. The data contain 5 animals and an average of 126 neurons were detected from each sample. Because the gKDR-GMM model uses information from neural circuits, particularly presynaptic neurons, only identified (named) neurons were used in this analysis. In addition, to extract neurons with high signal-to-noise ratios, neurons with autocorrelations greater than 0.3 at a lag of 20 time points were filtered (Toyoshima et al., 2022). Because the number of presynaptic neurons were too small, we used neurons connected

via two synapses as the presynaptic neurons in case the directly connected neuron was not observed. The average numbers of presynaptic neurons including the indirect connections for each sample are shown in the table below.

Table 1. This table shows the number of neurons detected, neurons identified, neurons filtered, and presynaptic neurons (including the indirect connections) for each sample.

Sample	Neurons detected	Neurons identified	Neurons filtered	Presynaptic neurons	Time points
1	109	38	30	13.7±7.2	3137
2	135	44	34	15.1±6.8	3134
3	131	32	23	12.4±5.2	3059
4	125	47	42	15.9±7.9	3311
5	129	31	23	12.3±5.8	3021

gKDR-GMM: Gradient kernel dimension reduction (gKDR)

The gKDR-GMM model utilizes gradient kernel dimension reduction (gKDR) to decompose basal dynamics of the network as well as a probabilistic model of Gaussian mixture modeling (Toyoshima et al., 2022). gKDR reduces the dimension of the dataset while maintaining important information and characteristics of the data through the use of the gradient kernel, which measures the statistical independence and correlation between variables (Fukumizu et al., 2013). Dimension reduction is important in data analysis to avoid complexity and problems caused by high dimensionality (Fukumizu et al., 2013). This helps identify which variables are most informative for predicting the output variable. gKDR makes use of the Reproducing Kernel Hilbert Space (RKHS), which is a specialized mathematical space where functions (in this case, variables) reside. RKHS provides a framework for analyzing the relationships and properties of functions (variables) using inner products and other mathematical operations. gKDR evaluates the statistical independence between the input variables and the output variable conditioned on the reduced subspace. It maximizes this independence to identify the most informative variables or combinations of variables for predicting the output variable (Wu et. al., 2019). The dimension reduction matrix is determined by solving a convex optimization problem and eigenvalue calculation. Because the optimization problem is convex, each solution will be unique and this can be solved by a matrix-based calculation (Fukumizu et. al., 2013).

gKDR-GMM: Gaussian Mixture Modeling (GMM)

Gaussian mixture modeling (GMM) is a probabilistic model that represents a dataset as a combination of multiple Gaussian distributions (Noh et al., 2022). It is a statistical method based on the weighted sum of probability density functions of multiple Gaussian distributions. Each Gaussian component captures a distinct cluster or pattern within the data. The parameters of GMM, which are means, covariances, and mixture coefficients, are estimated to describe the underlying statistical distribution of the dataset (Li et al., 2020). It allows for flexible modeling

of stochastic and complex data with multiple modes or clusters, providing a framework to understand the variability and structure within the data.

Overview of the gKDR-GMM model

A single gKDR-GMM model corresponds to a single neuron and is trained to predict the activity of that neuron 1 step ahead, based on the previous activities of itself and presynaptic neurons. This training is repeated for all identified neurons to obtain the trained gKDR-GMM model for each neuron. The set of these trained gKDR-GMM models can then represent the entire whole-brain activity data of the identified neurons. By combining these predictions, we have predicted the whole-brain activity 1 step ahead based on the previous time series of whole-brain activity data. Long-term prediction, which we refer to here as freerun simulation, was achieved by repeating the prediction of whole-brain activity at the next time point based on these predicted values.

Training of the gKDR-GMM model

The gKDR-GMM model was trained as described previously (Toyoshima et al., 2022). The gKDR-GMM was trained to predict neural activity 5-time points ahead, based on the past neural activities up to 300 time points ago with an interval of 10 time points. Namely, the activities of M+1 neurons for the past 30 time points were used as the predictor X to predict the activity of the target neuron 5 time points ahead (here we refer to as the objective variable Y), where M is the number of the presynaptic neurons including the indirect connections. For each target neuron, the pair of predictor X and objective Y were subsampled every 5 time points. Namely, the training data set contains N/5 pairs, where N is the time point included in the original data. The training data set was used to train the gKDR and GMM. The gKDR has weight matrix B that computes the intermediate variable U as U = XB, where the dimension of U was set as 3. The weight matrix B was trained so that the U was the most informative to predict the objective Y. Then the parameters of the GMM were optimized so that the GMM represented the distributions of the intermediate U and the objective Y.

The trained weight B of the gKDR and the parameters of GMM consist of the gKDR-GMM model for the target neuron. By repeating this training step for each target neuron, we finally obtained the gKDR-GMM models corresponding to a single sample of whole-brain activity. Because the numbers of identified neurons varied by sample, a different gKDR-GMM model was obtained for each sample. In the following analysis, the gKDR-GMM model for sample 3 was used, except for the linear regression analysis of correlation coefficients.

Simulation by the gKDR-GMM model

For predicting the activity of a target neuron 1 step ahead, the predictor X which contains activities of its own, and M presynaptic neurons for the past 30 time points was organized. Then the trained gKDR calculated the intermediate variable U by U=XB. Based on the values of U, the trained GMM generated a single realization of the probabilistic distribution of the neural activity.

This realization is used as the predicted activity of the target neuron 1 step ahead. By repeating this prediction for each target neuron, the 1 step ahead prediction of whole-brain activity was obtained. The predicted activity was used with the observed activity to predict the whole-brain activity 2 steps ahead. By repeating this step, long-term prediction (we refer to it as the freerun simulation) was achieved.

Results

Heatmap time series shows the gKDR-GMM model's success in the simulation of cyclic patterns

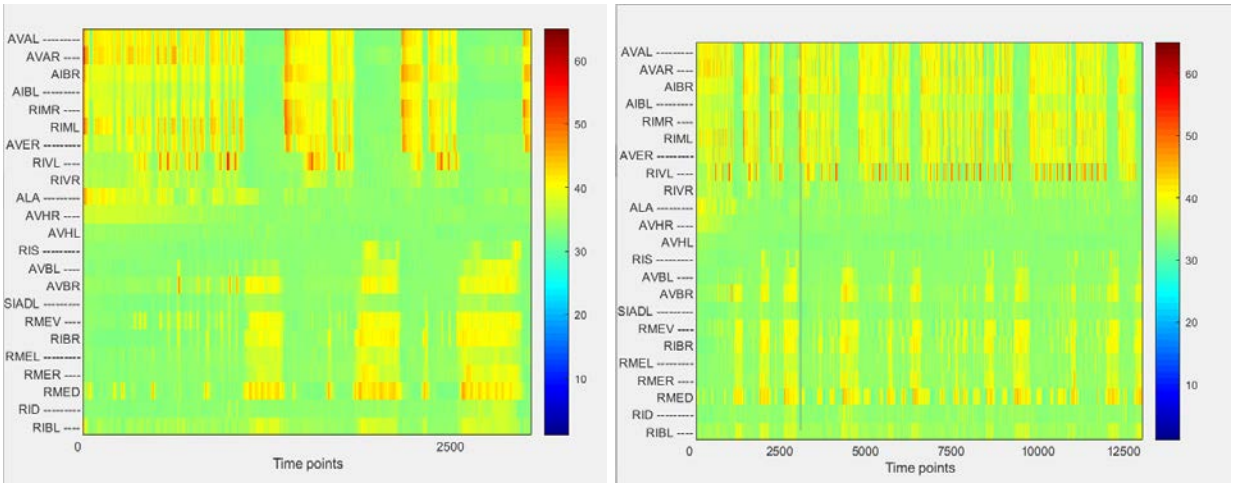


Figure 1: Time series of neuronal activity in actual observed data and simulation results. This figure presents a time series heatmap representing the dynamic activity of neurons over consecutive time points. The x-axis represents the time points, while the y-axis corresponds to individual neurons. The color within the heatmap indicates neuronal activity. The left panel shows actual observed data. The right panel shows simulation (freerun) results, where the first approximately 3000 time points are the observed data used for the simulation.

A comparison of heatmap time series was conducted to observe the creation of the "cycle" of active and inactive neurons. The simulation depicted distinct strips of yellow, indicating the alternating patterns of neuron activity; however, it should be noted that these patterns were not perfectly precise when comparing the results from the real and simulated results. This can be explained by the stochastic nature of the simulation process and the results will differ each trial. There was an inverse relationship between the activity of the top and bottom groups of neurons within the simulation, which indicates the successful simulation of the forward and backward movements of the worm (Kato et. al., 2015). Furthermore, the simulation managed to recreate the neuron activities in earlier times, which are characterized by frequent switching between levels of high and lower activities. Throughout the middle green section, consistent patterns of activity are present, suggesting stability in the overall neural dynamics during this period. In the bottom section of the figure, specifically involving the neurons RID and RIBL, relative consistency in activity levels was apparent throughout the simulation.

The gKDR-GMM model can capture neuron correlations as shown through correlation matrices

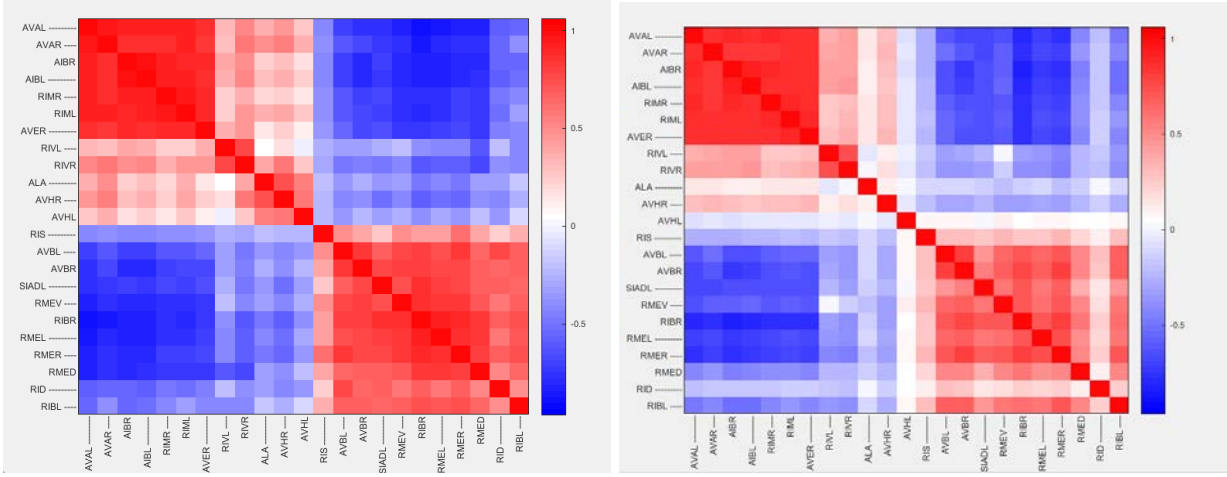


Figure 2: Correlation matrices of real data and simulation.
This figure presents two matrices of correlation coefficients, the left from real data and the right from simulation. Neuron names label both axes. Red indicates positive correlations, while blue indicates negative correlations. Darker colors indicate stronger correlations.

The performance of freerun in capturing correlations and activities within a simulated system was assessed. Several general correlations were successfully simulated, shown by the comparable size and locations of the red and blue regions in both matrices. Particularly, strong positive correlations were observed among AVAL, AVAR, AIBR, AIBL, RIMR, RIML, and AVER, indicated by the presence of dark red blocks in the top left region of the correlation matrices. These neurons are known to be associated with the backward movements of the worm. Similarly, strong positive correlations were also observed among AVBL, AVBR, RMEV, RIBR, RMEL, RMER, and RMED. These neurons are known to be associated with the forward movements of the worm. In addition, the strong negative correlations between the forward movement neurons and the backward movement neurons were accurately simulated. These negative correlations indicated by the blue regions were consistently observed in both matrices. Thus the gKDR-GMM model successfully captured the overall correlation structure within the whole-brain data.

Nonetheless, certain neural activities were not well-simulated in the freerun approach. The freerun simulation has visibly whiter areas, indicating weaker correlations. Notably, the activities of RID were poorly captured in the second from the bottom region of the heatmap. Similarly, the simulation inadequately represented the activity of AVHL in the 12th position from the top. These discrepancies may be due to missing data as the necessary neurons for prediction were not identified. The results suggest that the gKDR-GMM model captures some underlying relationships rather than blindly learning all apparent correlations and activities.

Correlation matrix linear regression shows clear correspondence between real and simulated correlation matrices

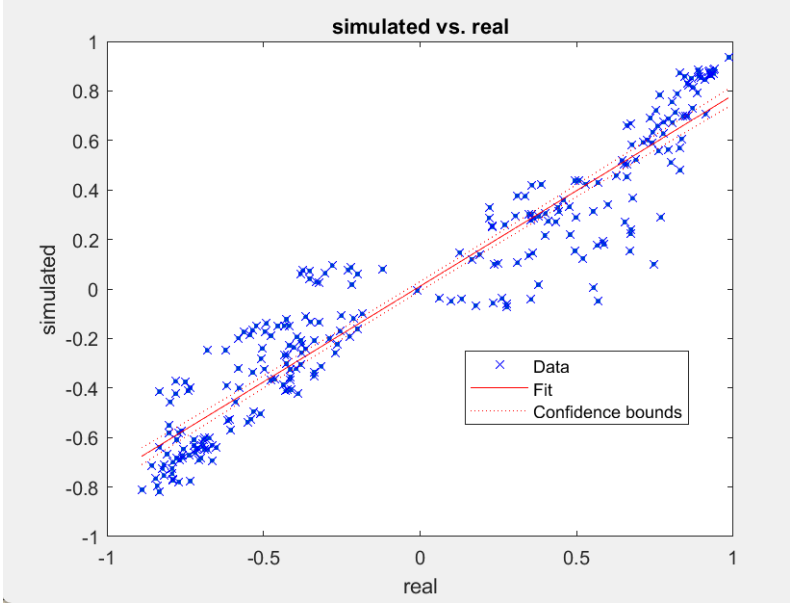


Figure 3: Linear regression of correlation coefficient based on real and correlation coefficient based on simulation

This figure shows linear regression comparing the correlation matrices obtained from real experimental and simulated data. The x-axis represents the correlations from the real dataset, while the y-axis corresponds to the correlations generated through simulation. The red line represents the best-fit line derived from the linear regression, indicating the relationship between the real and simulated correlation values. The blue 'x' symbols represent individual data points. The dotted lines surrounding the red fit line represent the confidence intervals.

A linear regression analysis was applied for comparing the correlation matrices obtained from real experimental data and the correlations generated through simulation. Figure 3 displays the results of the linear regression analysis. The red line in the plot represents the best-fit line obtained from the linear regression, revealing the relationship between the real and simulated correlation values. The linear regression analysis demonstrates a positive correlation between the two datasets, suggesting that the model successfully captured aspects of the underlying neuronal activity. The confidence intervals suggest that the relationship between the real and simulated correlation matrices is statistically significant.

R-squared of correlation matrices' linear regression

Table 2. This table shows the R-squared values for linear regression of correlation matrices comparison for 5 samples. All 5 samples are data from *C. elegans* without stimulation.

	R-squared Values			
Sample	Trial 1	Trial 2	Trial 3	Average
1	0.604	0.417	0.448	0.49
2	0.766	0.815	0.737	0.773
3	0.904	0.873	0.881	0.886
4	0.653	0.752	0.663	0.689
5	0.601	0.556	0.468	0.542

Here we test the stability of gKDR-GMM performance by using different time series. The publicly available dataset of whole-brain activity contains 5 samples. In the above analysis, we used sample 3. If the gKDR-GMM shows high R-squared values for each time series, it means that the gKDR-GMM can stably capture the correlation structure of the time series. Also, since the freerun simulation used to calculate the R-squared values is stochastic, the R-squared values themselves will vary from trial to trial. Therefore, we also examined the stability of the R-squared value from trial to trial.

Table 1 shows the R-squared values for the linear regression of the correlation matrix comparisons. The R-squared values are closer to 1, particularly in samples 2, 3, and 4. The variation of the R-squared values between trials is not large. These results indicate that the gKDR-GMM can stably capture the correlation structure of the time series, and its performance depends on the real time series.

gKDR-GMM simulates neuron activities and sequences as shown through Principal Component Analysis

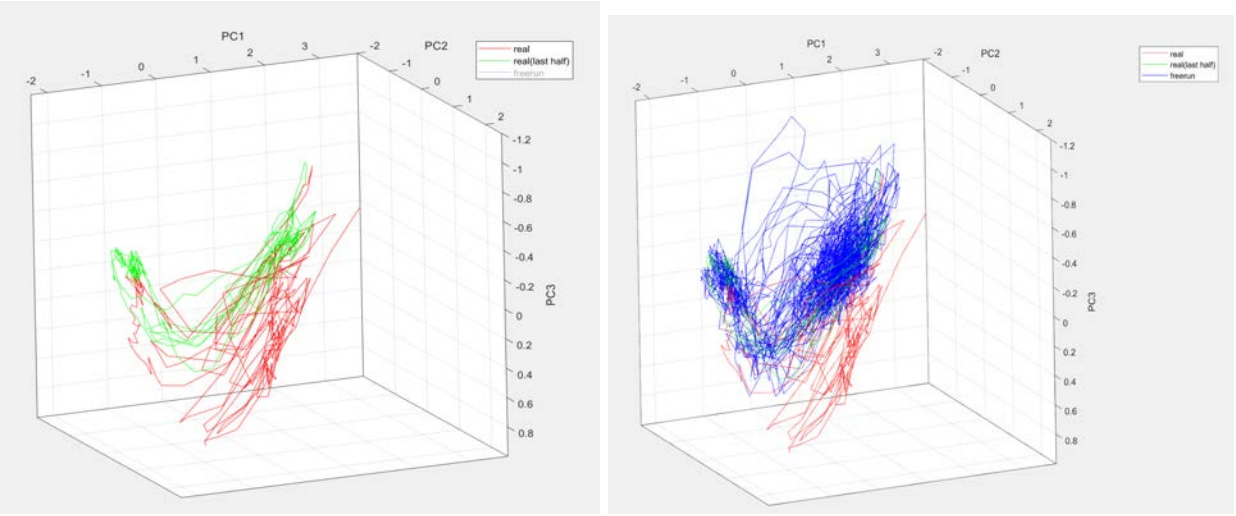


Figure 4: Phase diagram in PCA space - comparison of real and simulation freerun. This figure showcases a three-dimensional Principal Component Analysis (PCA) plot, with the x-axis representing the first principal component (PC1), the y-axis representing the second principal component (PC2), and the z-axis representing the third principal component (PC3). The red line in the plot corresponds to the PC scores of the first half of the real experimental data. The green line corresponds to the PC scores of the second half of the real experimental data. The blue line corresponds to the PC scores of the freerun simulation. The left panel shows the real data. The right panel shows freerun simulation data overlaid on the real data.

Principal Component Analysis (PCA) was applied for comparing the freerun and real data. PCA is valuable in that it is able to capture the relationships between more than 2 neurons compared to the correlation matrix, which only captures the relationships between 2 neurons. In addition, the agreement of trajectories of the real data and the simulation data in the principal component space means that the characteristics of the temporal development in both time series are consistent.

Real data was used to obtain the coefficients (PC axes). The coefficients are used to calculate the PC scores of real and simulation time series. The PC1-3 scores corresponding to the top 3 PCs determined the coordinates in the PC space. A time point in the time series was represented as a point in the PC space. Thus, the state and temporal development of the time series were represented as the position and trajectory of a point in the PC space. The top 3 PCs explained 87.6% of the variation in the real data and also explained 76.7% of the variation in the freerun simulation, indicating that the PCs successfully captured the overall dynamics of both the freerun simulation and the real data.

It is evident that the freerun data exhibited a similar manifold and trajectory as the real data in the PC space (Figure 4). This finding strongly suggests that the freerun simulation successfully captures similar activities and sequences as observed in the real dataset. The blue trajectory from the simulation shows a shape that is spiraling down, similar to the red and green trajectories from the real data. The alignment of the trajectories of real data and the simulation in the PC space indicates a high degree of correspondence between the simulated and actual neural dynamics, highlighting the efficacy of the freerun approach in accurately replicating key features of the real system's behavior.

Discussion

The gKDR-GMM model exhibits both strengths and limitations in reproducing activities and capturing neural dynamics within a simulated system. While it successfully captures general correlations and strong positive correlations among specific neural populations, it shows limitations in fully replicating all correlations and specific neuronal behaviors. The time series heatmap shows a successful simulation of inverse activities between different groups of neurons that are consistent with the real data. This clearly depicts the patterns of forward and backward

movements and stable dynamics during certain periods. The correlation matrices show that the model can successfully simulate positive or negative correlations between neurons. The linear regression for the correlation matrices also supports this idea, showing a clear linear relationship between the real and simulated correlation matrices. The alignment of the trajectories of freerun simulation with that of the real data in PCA space underscores the model's efficacy in capturing key features of the real system's behavior.

Certain neural activities, such as those of RID and AVHL, appear to be inadequately simulated. However, these discrepancies may be due to a lack of presynaptic data for those specific neurons. Nonetheless, the model provides valuable insights into relationships between certain neurons, as evidenced by strong negative correlations consistently observed in both simulations. The gKDR-GMM model successfully reproduced the reciprocal switching of the activity state of forward and backward movement neurons. This switching cycle seems to be unstable and longer than 300 time steps to which the gKDR-GMM model refers. This implies that the gKDR-GMM model determines the switching timing by some method instead of directly measuring the activity duration. The gKDR-GMM model might capture the determinant of the timing of the switching which is largely unknown. Further studies including eigenvalue analysis and event-triggered averaging may reveal the neural and computational implementation of this relatively long-period switching. Overall, the gKDR-GMM model shows promise but further refinement of the model and trials with other datasets may enhance its accuracy in simulating complex neural dynamics.

The success of the gKDR-GMM model in simulating brain activity from other data sources is of great significance for its application to other datasets in neuroscience. The fact that the gKDR-GMM model could accurately replicate brain activities using the data from the paper by Kato et al., which had a significantly smaller number of neurons compared to the in-house data (Toyoshima et. al., 2022), suggests the model's robustness and ability to handle datasets with varying sizes and complexities. For the in-house data, an average of 201.9 neurons were detected in the 24 samples, while for this study an average of 125.8 neurons were detected in the 5 samples. An average of 146.8 neurons were identified in the in-house data while an average of 29.4 neurons were identified in the Kato et. al., data. The samples in the dataset used in this study also had fewer presynaptic neurons, as opposed to 21.1 indirect links in the in-house data. The results highlight the gKDR-GMM model's ability to effectively leverage indirect connections and extract meaningful information from neural activity data, even when faced with limited neuronal information. There are 302 identified neurons in the *C. elegans* brain, but the model was able to capture neuron dynamics using only data from 20 neurons. Even with an unknown network structure and missing information on presynaptic neurons, this study proved that the model can still be used to simulate whole brain activity. The gKDR-GMM model shows its stability through accurate simulations even with different time scales and significantly different data.

The performance of the gKDR-GMM model on whole brain imaging data of *C. elegans* carries substantial implications for advancing our understanding of whole brain activity and

complex neural systems. By successfully reproducing the activities, the model provides insights into the intricate dynamics and emergent properties of neural circuits within the *C. elegans* brain. It reveals how individual neurons interact in response to stimuli for movement and contribute to the overall network dynamics, shedding light on the underlying mechanisms that govern neural activity. This knowledge not only enhances our comprehension of *C. elegans*' neural system but also has broader implications for neuroscience. The gKDR-GMM model is a foundation for understanding whole brain dynamics across different organisms, which can contribute to a greater understanding of the human brain as well. A significant advantage of gKDR-GMM is its ability to tackle the high-dimensional nature of brain data by reducing its complexity while preserving crucial information. gKDR-GMM can identify the most relevant variables and patterns underlying brain activity. Moreover, this approach facilitates the exploration of functional networks within the brain, helping to uncover distinct regions that collaborate during specific cognitive processes such as attention, memory, or language processing. By examining the dynamics of brain interactions, gKDR-GMM can shed light on the temporal relationships and functional connections between different brain regions. This, in turn, can contribute to our understanding of cognitive functions and the underlying mechanisms that drive them.

It is important to know that other methods to analyze neuron data exist. The Gaussian Mixture Model (GMM) alone offers a probabilistic approach to clustering and density estimation. GMM models data as a mixture of multiple Gaussian distributions, allowing it to capture complex data structures. However, GMM's performance may be limited when dealing with high-dimensional data, insufficient data, or complex distributions, which is where the gKDR-GMM model might offer advantages (Musab et al., 2021). Time series forecasting, such as the autoregressive integrated moving average (ARIMA) model, can help identify correlations through the analysis of relationships between variables, making it useful for predicting future trends and patterns (Kaur et. al., 2023). However, a disadvantage is that this method is not satisfactory when applied to data that is irregularly spaced. Causal inference is another method, and it is the process of determining the causal relationship between variables, which allows for better predictions (Pearl, 2010). However, establishing causality can be challenging due to the presence of confounding variables and biases that can distort the relationship between the cause and effect. In addition, causal inference often relies on high-quality data that accurately captures the relevant variables and their relationships, thus poor data quality can lead to unreliable conclusions of causality.

Additionally, the gKDR-GMM model can be used for other areas of study that involve stochastic processes and multi-elemental interactions. These include prediction of bacterial growth, disease spread, population growth in ecology, and prey and predator dynamics. For example, a previous study has been done on applying time series forecasting to data on measles, chickenpox, and marine phytoplankton populations (Sugihara et. al, 1990). The method here was able to make short-term predictions of the trajectories of chaotic dynamical systems.

The gKDR-GMM model's success represents a significant advancement in computational modeling that can be applied to neuroscience and other fields. Its ability to accurately capture and reproduce complex neural dynamics sets a precedent for future modeling approaches in studying whole brain activity.

Conclusion

This research paper tested the performance of the gKDR-GMM model by applying it to the whole brain activity dataset of *C. elegan*s. Based on analyses of correlation matrices, time heatmaps, and PCA, it was found that the gKDR-GMM model was successful in simulating the overall activities and patterns of neurons. The findings presented in this study have significant implications for neuroscience, as they highlight the relationship between neurons in movement and behavior as well as the inverse relationship between certain groups of neurons. It is hoped that the insights gained from this research will serve as a catalyst for further exploration and foster positive changes in computational neuroscience.

Acknowledgment

We thank Prof. Yuichi Iino for sharing the MATLAB code for gKDR-GMM.

Works Cited

Dabbura, I. (2018, September 17). K-means Clustering: Algorithm, Applications, Evaluation Methods, and Drawbacks. Towards Data Science. Retrieved July 31, 2023, from https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a

Dunn, N. A., Lockery, S. R., Pierce-Shimomura, J. T., & Conery, J. S. (2004). A neural network model of chemotaxis predicts functions of synaptic connections in the nematode Caenorhabditis elegans. Journal of computational neuroscience, 17(2), 137–147. https://doi.org/10.1023/B:JCNS.0000037679.42570.d5

Fukumizu, K., & Leng, C. (2011, September 2). *[1109.0455] Gradient-based kernel dimension reduction for supervised learning*. arXiv. Retrieved July 14, 2023, from https://arxiv.org/abs/1109.0455

Institute of Statistical Mathematics. (2004). Dimensionality Reduction for Supervised Learning with Reproducing Kernel Hilbert Spaces. Journal of Machine Learning Research. Retrieved July 14, 2023, from https://www.jmlr.org/papers/volume5/fukumizu04a/fukumizu04a.pdf

Jospin, M., Qi, Y. B., Stawicki, T. M., Boulin, T., Schuske, K. R., Horvitz, H. R., Bessereau, J. L., Jorgensen, E. M., & Jin, Y. (2009). A neuronal acetylcholine receptor regulates the balance of muscle excitation and inhibition in *Caenorhabditis elegans*. PLoS biology, 7(12), e1000265. https://doi.org/10.1371/journal.pbio.1000265

Karbowski, J., Schindelman, G., Cronin, C.J. et al. Systems level circuit model of *C. elegans* undulatory locomotion: mathematical modeling and molecular genetics. J Comput Neurosci 24, 253–276 (2008). https://doi.org/10.1007/s10827-007-0054-6

Kato, S., Kaplan, H. S., Schrodel, T., Yemini, E., & Lockery, S. (2015, October 15). Global Brain Dynamics Embed the Motor Command Sequence of *Caenorhabditis elegans*. Cell. https://www.cell.com/cell/fulltext/S0092-8674(15)01196-4?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0092867415011964%3Fshowall%3Dtrue

Kaur, J., Parmar, K. S., & Singh, S. (2023). Autoregressive models in environmental forecasting time series: a theoretical and application review. Environmental science and pollution research international, 30(8), 19617–19641. https://doi.org/10.1007/s11356-023-25148-9

k-Means Advantages and Disadvantages | Machine Learning. (2022, July 18). Google for Developers. Retrieved July 31, 2023, from https://developers.google.com/machine-learning/clustering/algorithm/advantages-disadvantages

Li, H., He, J.; Misra, S. (2020). Chapter 5 - Robust geomechanical characterization by analyzing the performance of shallow-learning regression methods using unsupervised clustering methods. In Machine Learning for Subsurface Characterization, Elsevier.

Musab T.S. Al-Kaltakchi, Mohammed A.M. Abdullah, Wai L. Woo, Satnam S. Dlay,

Chapter 7 - Closed-set speaker identification system based on MFCC and PNCC features combination with different fusion strategies, Editor(s): Nilanjan Dey, In Primers in Biomedical Imaging Devices and Systems, Applied Speech Processing, Academic Press, 2021, Pages 147-173, ISBN 9780128238981, https://doi.org/10.1016/B978-0-12-823898-1.00001-1.

Nguyen, J. P., Shipley, F. B., Leifer A. M. (2015, December 28). Whole-brain calcium imaging with cellular resolution in freely behaving *Caenorhabditis elegans*. https://www.pnas.org/doi/10.1073/pnas.1507110112

Noh, H. Y.,; Fagert, J. (2022, July 29). Big Data Analysis for Civil Infrastructure Sensing. Sensor Technologies for Civil Infrastructures (Second Edition). https://www.sciencedirect.com/science/article/abs/pii/B9780081027066000076

Pearl J. (2010). An introduction to causal inference. The international journal of biostatistics, 6(2), 7. https://doi.org/10.2202/1557-4679.1203

Sengupta, P., & Samuel, A. D. (2009). *Caenorhabditis elegans:* a model system for systems neuroscience. Current opinion in neurobiology, 19(6), 637–643. https://doi.org/10.1016/j.conb.2009.09.009

Schrödel, T., Prevedel, R., Aumayr, K. et al. Brain-wide 3D imaging of neuronal activity in *Caenorhabditis elegans* with sculpted light. Nat Methods 10, 1013–1020 (2013). https://doi.org/10.1038/nmeth.2637

Wu, C. (2019). Solving Interpretable Kernel Dimension Reduction. Retrieved July 22, 2023, from https://proceedings.neurips.cc/paper/2019/file/61f3a6dbc9120ea78ef75544826c814e-Paper.pdf

Yu Toyoshima, Hirofumi Sato, Daiki Nagata, Manami Kanamori, Moon Sun Jang, Koyo Kuze, Suzu Oe, Takayuki Teramoto, Yuishi Iwasaki, Ryo Yoshida, Takeshi Ishihara, Yuichi Iino bioRxiv 2022.11.18.517011; doi: https://doi.org/10.1101/2022.11.18.517011

Zambom, A. Z., & Dias, R. (2012, December). A Review of Kernel Density Estimation with Applications to Econometrics. https://arxiv.org/pdf/1212.2812 . Retrieved July 31, 2023, from Arxiv

Zhang, S., Jin, W., Huang, Y., Su, W., Yang, J., & Feng, Z. (2011). Profiling a *Caenorhabditis elegans* behavioral parametric dataset with a supervised K-means clustering algorithm identifies genetic networks regulating locomotion. Journal of neuroscience methods, 197(2), 315–323. https://doi.org/10.1016/j.jneumeth.2011.02.014

Leaders in Transition: A Comparative Study of Military and Political Figures During Epochal Transformations By Minggang Li

Introduction

"Empires wax and wane; states cleave asunder and coalesce." Such is the opening line of one of the greatest Chinese novels on war, The Romance of the Three Kingdoms. There existed two periods of the Roman government. The Roman Republic existed between 509 BCE to 30 BCE, and the Roman Empire between 30 BCE and 476 CE. Technically, Rome had an empire during the Roman Republic. However, the word "Roman Empire" defines the period when Augustus was the emperor, and during that period, Rome was led by emperors. Octavian became the sole leader of Rome when his army defeated Cleopatra and Antony during the Actium battle. The Roman government senate named him Augustus, which is referred to as the highest person. Octavian, known as Augustus, was made the Consul for life through the Senate. He was also made to be the princeps, which referred to the first citizen; imperator, which meant the victorious general; and pontifex Maximus, which referred to him as the chief religious' leader of the emperor. Pompey was a robust military figure during the fall of the Roman Republic, and Julius Caesar was known for power consolidation. In China, Sun Yat-sen was behind the modernization process and the decline of the Qing dynasty. Yuan Shikai attempted to centralize the Chinese government through the military, and Mao Zedong led the PRC in stabilizing China by winning over Chiang Kai-Shek. Through an examination of the military and political leaders during the pivotal transitions from the fall of the Roman Republic to the rise of the Roman Empire and from the fall of the Qing Dynasty to the height of the People of the Republic of China, this essay uncovers intriguing parallels and distinct divergences in their actions, motivations, and impacts. By analyzing the rise to power, leadership styles, consolidation of authority, military campaigns, and legacies of leaders in these historical contexts, we gain insight into the challenges and complexities leaders face navigating periods of profound societal change and the historical inevitability of war.

Contextual Background

The decline of the Republic of Rome and the subsequent rise of the Roman Empire was a transition period in ancient Rome history, accompanied by a series of military, social, and political developments that changed the Roman government. The Roman Republic existed between 509 BCE to 30 BCE, and the Roman Empire between 30 BCE and 476 CE. Technically, Rome had an empire during the Roman Republic (Burden-Strevens 12). However, the word "Roman Empire" defines the period when Augustus was the emperor, and during that period, Rome was led by emperors. Emperors Augustus and the emperors after him attempted to keep the façade that they were chosen as officials, not dictators.

The phrase "first among equals" was used to justify that the Roman Emperor was an important and prestigious member of the Senate. However, all the senators were regarded as significant. Roman emperors led the Empire with little contribution from other people. The Qing Dynasty, established during the 17th century, ruled the Chinese government for nearly three

centuries. The Manchu ethnic group established the Dynasty. The Qing empire was characterized by relative stability, territorial expansion, and prosperity during its early times. Despite the prosperity, it later experienced external threats and internal decay.

Decline of the Roman Republic

The economic reasons behind the fall of the Roman Republic were the disparity between the poor and the rich. Rome was divided according to social class, where there existed the haves and the have-nots (Watts 8). Also, there was a trade and manufacturing decline since a small number of people of lower socioeconomic status could buy goods, which decreased purchasing power. The large estates also became self-sufficient, and it also hindered trade and the manufacturing process. The improvised workers in Rome were tied to the acres of land as colonists. The military reasons that led to the fall of the Roman Republic included the over-reliance on the barbarian armies who were not entirely loyal to Rome; they were also not relied on since they were fighting their fellow barbarians and they were much interested in acquiring booty and not fighting for Rome.

Furthermore, the Roman Republic's military personnel were Roman servants, and the Roman Empire military unmade and made emperors (Taylor 76). Politically, there was a decline in patriotism since democracy was not there, the Roman citizens lost their rights to vote for the state, patriotism was based on loyalty to the ruling Emperor and not the Country, and most emperors did not command respect. The creation of two empires also created loyalty challenges. Politically, the disorderly successions characterized by civil wars, forced suicides, and murders led to the decline of the Republic. Socially, war, plagues, and hunger led to population decline. There was also a decreased intellectual culture and reduced public service activities. Instead, most citizens used leisure time to watch gladiatorial contests and chariot races. There were religious divisions where emerging religious domains such as Christianity challenged imperial divinity. Even though the Roman Republic served for many centuries, it was divided by the tensions within the government. Civil wars started between teams with distinct loyalties; they accelerated the change from a Republic to an Empire.

Rise of Rome as an Empire

The rise of the Roman Empire started during the Augustus period. Augustus took a leadership role after the death of Caesar. When Caesar died before Augustus, Rome experienced an increased rate of civil battles, making Marcus Lepidus, Octavian, and Mark Antony form a Second Triumvirate. This alliance objected to the return of order. Octavian defeated Cleopatra and Antony in 31 BCE in the Actium war and thus made Rome an Empire. Augustus initiated different reforms during his leadership to stabilize the economy and ensured Pax Romana, which was responsible for protecting the citizens against hunger and integrating peace.



The Roman Empire Map, history, facts and Timeline (Photo by BBC History Magazine)

Background of the fall of the Qing Dynasty and the emergence of the Republic of China

The Fall of the Qing Dynasty

The Qing Dynasty of 1644 to 1912 was the last Dynasty of China. Manchu ethnic group created it after overthrowing the Ming Dynasty. Qing Dynasty enjoyed stability and expansion over different ethnic groups and a vast territory (Tseng 27). Internal problems that faced the Dynasty included inefficiency and corruption due to expansion and population increase, which caused increased strain on the available resources, thus social unrest and famines. There were increased rebellions to the Dynasty, like the Taiping Rebellion of 1850 to 1864 and the Boxer Rebellion of 1899 to 1901 (Ramírez Ruiz and Niu 141). The external challenges that affected the Qing Dynasty were unequal treaties imposed on China's government by foreign countries that eroded China's territorial integrity and sovereignty.

Emergence of the Republic of China

Western power's imperialism and Japan's encroachment lead to wars such as the Opium Wars of 1839 to 1842 and 1856 to 1860. Chinese government officials adopted modernization strategies in response to imperialism, such as infrastructure and military modernization, but they were inefficient and limited to enhancing change. Reform movements emerged where reformers and scholars, including Liang Qichao and Kang Yuwei, were attributed for sweeping cultural, political, and social reforms to develop China. The Xinhai Revolution of 1911 led to Changes in the governance of China (Ferrer 1148). There was an emergence of a coalition of military officers, nationalists, and reformers led by political figures such as Sun Yat-sen, who aimed to overthrow the Dynasty. The coalition spread in China and abducted Puyi, the last Qing Emperor, in 1912 (Breslauer 132). The Republic of China was created in 1912, and Sun-Yat was made the initial provisional President. The

establishment of the Republic ended the Dynasty's reign in China, thus creating a new system of Republican governance.



Republic of China (1912-1949)

Overview of Military and Political Leaders–Leaders During the Roman Transition: **Julius Caesar**

Caesar was among the legendary leaders of the Roman Empire. Caeser played a crucial part in the transition of Rome from the Roman Republic to Rome as an Empire. While he rose to power, there was a combination of charisma, political maneuvering, and military prowess. His political career started after he was elected as a consul around 59 BCE (Bondioli et al.). Caesar created the First Triumvirate, which was an alliance that was informal together with Crassus and Pompey. Caesar carried out various military campaigns to expand the Roman Empire as a proconsul in Gaul. The movements in Modern-day France, such as the conquest of Helvetii and the war of Alesia, made him popular. Tensions between the Senate and Caesar in 49 BC, led by the conservative faction and Pompey, resulted in a civil war. Casear popularly crossed the river Rubicon, going against the laws of the land, and marched to the Roman Empire. He consolidated authority after emerging as the winner in the civil war. He was appointed dictator in perpetuity, thus closing the chapter of the Republican system. Caesar was assassinated in 44 BCE. His important economic transformation was grain purchase regulation, and he conducted the redistribution of land to the poor. Everyone would get adequate food to eat under his reign.

Pompey

Gnaeus Pompeius Magnus was a formidable political figure and military general at the time of the fall of the Roman Republic. Pompey's military victories, such as the victory against the Mediterranean pirates and the Eastern campaigns, gave him the name Magnus, which meant the

Great (Antonio et al. 710). Rivalry and political ambitions with Julius Caesar caused complications in the authority struggle in the Roman Republic. At first, he was with the conservative faction and the Senate, and they were against the rising popularity gained by Caesar (Barca). The desire to triumph caused tensions in the Senate. Pompey was defeated during the Pharsalus war and escaped to Egypt, where he was assassinated. He was known for using military power to expand the Roman territories in Italy



Pompey the Great: Stateman and General for Rome (Photo by Britannica)

Augustus (Octavian)

Gaius Octavius Thurinus was a crucial leader in the transition of the Roman political landscape. He was instrumental in the foundation of the creation of the Roman Empire. When Caesar was assassinated in 44 BCE, Augustus came up as a strong person in power (Bertrand 24). Together with Marcus Lepidus and Mark Antony, they created the Second Triumvirate, and the coalition enabled them to take revenge for Caesar's assassinations and take control of power. The alliance was later disbanded, and civil wars were experienced in the Empire (Patterson 94). After defeating Cleopatra and Mark Antony, he became the Roman Empire's sole ruler. Skillfully, he could keep Republican institutions appearance, but silently, he consolidated power and authority. Augustus was a personal proclaimed name when he took the throne as the first Emperor of Rome. The Principate was Augustus' Rule, which marked the transformation from a Republican to an Empire (Markov 297). Under Augustus's leadership, various reforms, such as creating a professional army,

were brought on board. Augustus introduced the inheritance and tax laws to improve morality. He was against adultery and awarded families with three children and above.

Leaders During Chinese Transition:**Yuan Shikai**

Yuan was famous during the last years of the Qing Dynasty. His military prowess was seen after he suppressed the Boxer Rebellion. His political ambitions were marked by his contributions to stabilize China and expand its territories. He was appointed the President of RoC, but his leadership style aimed to consolidate power and authoritarian tendencies (LI 283). The attempt he made to reinstate monarchy led to rebellion and inability. Salvation through political centralization to constitutional monarchy influenced his monarchy restoration decision.

Sun Yat-sen

Sun was a revolutionary ruler who overthrew the Qing Dynasty and formed China as a Republican. Sun was very determined to bring the Dynasty regime to an end. Sun was behind the formulation of the Three Principles of the People, which were livelihood, democracy, and nationalism, and these formed the basis of the revolutionary campaign (Jie 56). Sun's leadership style encompassed a visionary focus on bringing modern development to China. His inauguration as the Republic of China (RoC) provisional President was done after the decline of the Qing Dynasty in 1912 (Coddy 778). As the President of China, he encountered challenges such as regional divisions and warlords. Sun Yat Sen’s attempt at revolution had built the foundation on which the current PRC stands. Sun funded revolutions and advocated for Chinese revolutions. He organized campaigns to rebel against the Qing Dynasty.

Mao Zedong

Mao founded the People's Republic of China (PRC). During the Civil War, he was known for his guerrilla tactics, the Famous Long March, and strategic retreats that positioned him as a strong leader. His leadership at the PRC was marked by Marxism-Leninism ideological commitment and capability to mobilize followers. He was the leader of the Chinese Communist Party (CCP) in the protracted guerilla warfare against Chiang Kai-Shek's Nationalist government (ROC) (Kulacki 352). As the leader of the PRC, transformations were seen, such as the Grear Leap Forward. He was behind the launch of the reconstruction campaign of China to become an Industrialized country from an agrarian economy.

Parallels and Contrasts

A comprehensive analysis of these transition leaders and the courses of action they adopted during their leadership identifies intriguing contrasts and parallels. Some leaders attempted to reform and overturn the original political system. Julius Caesar overthrew the Roman Republic and tried and succeeded in establishing a new political system, the Roman Empire (Madsen 264). During the transition period, Casear experienced an increased rate of rebellion from the entrenched elites in the Empire.

Still, he used military force to curb the opposition, and he was able to create a Roman Empire. Sun Yat-sen of China, on the other side, overthrew the Qing Dynasty, and his attempts to create a new political system in China were successful. The Qing Dynasty had ruled China for over three decades, and Sun was interested and ambitious in changing the governance of China (Chen et al. 289). The Qing Dynasty, however, had conquered different territories, and it took a lot of work for Sun to accomplish his mission. In the process, he experienced rebellion from foreign countries and division of territories, but he successfully changed China's political system to the RoC. Under the leaders who preferred the old political system and their crafty use of the military to gain power, Pompey's support of optimates (a conservative faction of the Roman Senate) and his gain of power through military conquest (old power), defeated by Caesar is parallel to Yuan Shikai's rise of power through military, his attempt to reinstate monarchy in China.

By self-proclaiming himself as the Emperor, China eventually fell due to opposition. It is evident that both Yuan Shikai and Pompey supported the already-formed elites, and they were for the old political system. However, the main objective of Pompey was to keep the already existing republican order in Rome, but Yuan aimed at the process of restoration of the monarchical reign in China. Also, it was clear that Pompey had initial support from the Senate. However, Yuan's efforts were challenged by increased opposition in the territory, and thus it failed. Leaders came to the top after the tumultuous civil war and eventually stabilized the Country through the military.

Augustus (Octavian)'s success in winning the civil war against Mark Antony, his role in establishing and stabilizing the political system (Roman Empire) is similar to Mao Zedong's rise to power through his control of the military (Red Army), the defeat of Chiang Kai-shek (RoC), the establishment of PRC and stabilize the Country (Tong and Haijian 183). Both Mao Zedong and Augustus triumphed in the civil wars they were fighting and were vital in ensuring their countries were stable. They all relied on military loyalty to get their positions and control the territories.

However, while Augustus kept the façade of the institutions created by the Republicans while creating a centralized autocracy, Mao conducted open communist ideology and one-party Rule. While Mao's leadership witnessed increased ideological upheaval and turmoil, Augustus' reign saw increased prosperity and stability in the Roman Empire. Mao and Augustus experienced opposition during their Rule but defeated their rivals. Augustus was victorious over Mark Antony, and Mao defeated Chiang Kai-Sheks Nationalists. They all applied army forces to consolidate their Rule.

Legacies and Long-Term Impact

The legacies left behind by these transition leaders had lasting and profound impacts on the societies. The lasting influence of Julius Caesar and Sun Yat-sen: ending of the old political system (Roman Republic and Qing Dynasty). The overthrowing of the Roman Republic by Casear marked an essential change in Roman governance. Augustus initiated the Empire's development and civilization at large, and the civilization remained a landmark. Sun Yat-sen's legacy is evident just from the name he was given, "Father of Modern China" ("On the Reasons that Christians Supported the Revolution of 1911 Led by Dr. Sun Yat-sen" 93). His efforts aimed at modernizing China's social

and political systems made China remain one of the most industrialized countries in the world (Jan 149). His three principles of people, livelihood, democracy, and nationalism remained influential.

The two also left long-term impacts where Augustus made the Roman Empire rule the entire globe. Roman governance, law, and culture were spread throughout the Mediterranean and Europe. The administrative and legal systems of the Empire, together with their architectural achievements and cultural growth, remained influential to the next generations of civilizations. On the other hand, Sun's vision of democratic and modern China had a long-term impact on the Country's politics. It motivated the next generation of leaders, such as the Chinese Communist Party and Chinese Nationalist Party (KMT).

Pompey and Yuan Shikai's attempt to take over through the military and their failed attempt led to their respective downfall and hurt the Country, which remained legacies and had long-term impacts. Pompey failed to consolidate power and keep the Roman Republican system of leadership, thus leading to the fall of the Roman Republic. Rivalry and conflicts with Augustus led to political divisions in Rome. Yuan Shikai's plan to create a new Chinese monarchy by self-proclamation led to increased opposition, and the government's legitimacy was destroyed by his actions (Yang 996). The struggle for leadership between Caesar and Pompey led to the fall of the Republic, and the Roman Empire emerged. Moreover, the power struggle created a way to establish the autocratic Rule in the Empire. The actions of the Yuan had a long-term impact on China since the fall of the monarchy weakened the Presidency institution.

Augustus and Mao Zedong’s stabilization of the Country, ending the civil wars, and establishing new political regimes created legacies and long-term impacts on the Roman Empire and the Chinese government. The legacy left behind by Augustus remains to be the transition of Rome from a Republic to an Empire ("Plutarch and the Late Republican Civil Wars" 335). Through Pax Romana, he created stability, which led to the prosperity of the Empire. He laid the stone for the strong governance of the Roman Empire, marked with centralized bureaucracy. The legacy of Mao was unifying China and changing the Social-political Landscape of China. His reign created the foundation on which the People's Republic of China rested. The long-term impact left by Augustus on the Roman Empire was influencing civilization, which was transferred to the next generations. The Imperial Rule developed by Augustus shaped the governance rule for the next emperors. Mao impacted the political atmosphere of China by championing the formation of a single-party state.

Lessons and Reflections

Different reflections and lessons are learned from analyzing transitional leadership and the leaders responsible for the transformations. The first lesson is similarities in gaining power through military prowess: "Political power grows out of the barrel of a gun" by Mao. This statement aligns with the historical leaders' behaviors during the transition. From how some of them acquired power, using the armed forces can sometimes acquire political power. In cases where opposition is high, the military can be used to gain governance. Transitional leaders such as Yuan Shikai, Pompey, Caesar, and Mao used the military to attain their leadership ambitions.

Moreover, in the evitability of wars after a long period of stability: "Those long divided shall be

united; those long united shall be divided: such is the way of the universe" from Romance of the Three Kingdoms" statement implies that, transformations are achieved when there is long political stability, and some can be attained when parties disagree. China's transformation is a true lesson that for prosperity to be witnessed, there comes a series of struggles and battles between the ruling class and those seeking transformation. Sun had to fight the Dynasty to achieve China's modernization process (Liu 141). Through his efforts and struggles with the Dynasty, stability was conducted in China, and the Country embarked on a tremendous modernization process but, of course, accompanied by opposition and division of the Chinese territory.

Conclusion

The analysis of the political and military leadership during epochal transformations like the rise of the Roman Empire and the fall of the Roman Republic shows challenges in transition. The decline of the Qing Dynasty resulting in the Republic of China shows the complexities and problems leaders face during essential leadership transitions. Looking at the discussed leaders, such as Augustus of the Roman Empire, it is evident that during the transition period, they went through challenges that called for courage and leadership tactics to overcome them. Military skill is a core component of leadership since it can be utilized to end dominant regimes such as the Qing Dynasty of China. Contrasts and parallels among the historical transitions show the motivations and tactics of visionary leaders.

Visionary leaders who are after change, patience during turbulent periods for the leaders, and the consequences of their actions determine their success. Augustus enabled the transformation of Rome from a republic to an empire. Sun Yat-sen is considered the father of modernization in China just from his visionary skills and need for industrialization. History shows that leadership is essential in determining a country's destiny. Research about transitional leaders enriches our understanding of leadership dynamics during transition. The implications from the scenarios of these transitions are that change of governance can lead to the prosperity of a nation. China, through Sun, became industrialized until now due to efforts for modernization. Reforms by Augustus improved the living conditions of people in the Roman Empire. Future leaders must learn patient and consistent efforts to bring change to their nations.

Works Cited

Antonio, Margaret L., et al. "Ancient Rome: A genetic crossroads of Europe and the Mediterranean." *Science*, vol. 366, no. 6466, 2019, pp. 708-714, www.academia.edu/42665239/Ancient_Rome_A_genetic_crossroads_of_Europe_and_the_Mediterranean.

Barca, Natale. *Before Augustus: The Collapse of the Roman Republic*. Casemate, 2023, worldcat.org/en/title/1389808915.

Bertrand, Estelle. "Imperialism and the Crisis of the Roman Republic: Dio's View on Late Republican Conquests (Books 36–40)." *Cassius Dio and the Late Roman Republic*, 2019, pp. 19-35, bmcr.brynmawr.edu/2020/2020.05.29/.

Bondioli, Nelson, et al. "History, design and archaeology: The reception of Julius Caesar and the representation of gender and agency in Assassin's Creed Origins." *In die Skriflig / In Luce Verbi*, vol. 53, no. 2, 2019, indieskriflig.org.za/index.php/skriflig/article/view/2431/5904.

Breslauer, George W. "Consolidating Power and "Building Socialism" in China." *The Rise and Demise of World Communism*, 2021, pp. 126-134, academic.oup.com/book/40007/chapter/340344063.

Burden-Strevens, Christopher. "Cassius Dio's Speeches and the Collapse of the Roman Republic." 2020, brill.com/display/title/38725?language=en.

Chen, Kun, et al. "Song Brocade in the Ming and Qing Dynasties." *Clothing and Textiles Research Journal*, vol. 38, no. 4, 2020, pp. 285-297, www.researchgate.net/publication/357895082_The_Innovative_Practice_of_Artificial_Intelligence_in_the_Inheritance_of_Chinese_Xiangjin_Art.

Croddy, Eric. "Republic of China (Taiwan)." *China's Provinces and Populations*, 2022, pp. 777-811, . link.springer.com/chapter/10.1007/978-3-031-09165-0_35.

Ferrer, Alessandra. "A post‐imperial, pluralist nationalism? From the Five Nation Republic of China to multicultural Taiwan." *Nations and Nationalism*, vol. 29, no. 3, 2023, pp. 1142-1157, onlinelibrary.wiley.com/doi/abs/10.1111/nana.12946.

Jan, George P. "The Doctrine of Nationalism and the Chinese Revolution." *Sun Yat-sen's Doctrine in the Modern World*, 2019, pp. 138-159, www.taylorfrancis.com/books/edit/10.4324/9780429307935/sun-yat-sen-doctrine-modern-world-chu-yuan-cheng-yu-long-ling-hung-chao-tai-harold-schiffrin.

Jie, Wang. "Sun Yat-sen studies." *Contemporary Studies on Modern Chinese History III*, 2021, pp. 36-70, cir.nii.ac.jp/crid/1130287351415259269.

Kulacki, Gregory. "Nuclear Weapons in the Taiwan Strait Part II." *Journal for Peace and Nuclear Disarmament*, vol. 3, no. 2, 2020, pp. 342-365, www.tandfonline.com/doi/full/10.1080/25751654.2020.1834962.

LI, HUAIYIN. "Centralized Regionalism: The rise of regional fiscal-military states in China, 1916–28." *Modern Asian Studies*, vol. 55, no. 1, 2020, pp. 253-291, www.cambridge.org/core/journals/modern-asian-studies/article/centralized-regionalism-the-rise-of-regional-fiscalmilitary-states-in-china-191628/342F93D33C8532B04FB26496B524F

FE7.
Liu, Shih-Diing. "Book review: Illiberal China: The Ideological Challenge of the People's Republic of China." *China Information*, vol. 34, no. 1, 2020, pp. 140-141, www.academia.edu/42072733/China_Information_Article_Information_Book_review_Illiberal_China_The_Ideological_Challenge_of_the_Peoples_Republic_of_China.
Madsen, Jesper M. "Like Father Like Son: the Differences in How Dio Tells the Story of Julius Caesar and His More Successful Son." *Cassius Dio and the Late Roman Republic*, 2019, pp. 259-281, bmcr.brynmawr.edu/2020/2020.05.29/.
Markov, Konstantin V. "Towards the Conceptualization of Cassius Dio's Narration of the Early Career of Octavian." *Cassius Dio and the Late Roman Republic*, 2019, pp. 282-298, www.academia.edu/94061374/Studia_Dionea_novissima_Historical_Narrative_Interpretations_of_the_Past_and_Political_Contexts_of_Cassius_Dio_s_Roman_History_Part_II.
"On the Reasons that Christians Supported the Revolution of 1911 Led by Dr. Sun Yat-sen." *Yearbook of Chinese Theology*, 2020, pp. 89-102, bdcconline.net/en/stories/sun-yat-sen.
Patterson, John. "Military organization and social change in the later Roman Republic." *War and Society in the Roman World*, 2020, pp. 92-112, www.researchgate.net/publication/342791466_Military_organization_and_social_change_in_the_later_Roman_Republic.
"Plutarch and the Late Republican Civil Wars." *The Historiography of Late Republican Civil War*, 2019, pp. 320-350, . eprints.ncl.ac.uk/272892.
Ramírez Ruiz, Raúl, and Guanjie Niu. "Spain in China." *Examining Colonial Wars and Their Impact on Contemporary Military History*, 2023, pp. 132-147, www.igi-global.com/book/examining-colonial-wars-their-impact/305808.
Taylor, Michael J. "Tactical Reform in the Late Roman Republic: The View from Italy." *Historia*, vol. 68, no. 1, 2019, p. 76, www.academia.edu/22819197/Tactical_Reform_in_the_Late_Roman_Republic_The_View_from_Italy.
Tong, Liu, and Mao Haijian. "Military history." *Contemporary Studies on Modern Chinese History I*, 2020, pp. 173-189, www.taylorfrancis.com/chapters/edit/10.4324/9781351269407-6/military-history-liu-tong-mao-haijian.
Tseng, Hui-Yi K. "Sovereignty as the Axis of State-Making: the Re-sovereignisation of the Republic of China/Taiwan's South China Sea Claim." *International Journal of Taiwan Studies*, 2023, pp. 1-32, brill.com/view/journals/ijts/aop/article-10.1163-24688800-20231250/article-10.1163-24688800-20231250.xml.
Watts, Edward J. "The Eternal Decline and Fall of Rome." 2021, worldcat.org/title/1178641721.
Yang, Tianzhi. "The Inevitability of the Civil War: The Popularis and the Optimates." *Advances in Social Science, Education and Humanities Research*, 2022,

Machine Learning for Autonomous Decision-Making and Visual Computing for Robotics By Damla Köse

Abstract

This study explores the deployment of edge machine learning technology across various applications, including smart devices for robot control and unmanned aerial vehicles (UAVs). It emphasizes the use of machine learning at the edge to process data locally, ensuring uninterrupted decision-making. Edge machine learning technology and its implementation of various solutions have been presented in this study. It is used across various contexts and implemented across smart devices for robot handling, and UAV unmanned aerial vehicles where data is processed using ML locally for uninterrupted decision-making. Implementation of edge ML is presented. An IOT device integrated with a teleoperated robot is utilized for neural network inference which is trained in machine learning. Additionally, the study delves into the integration of UAVs with visual computing capabilities, achieved through the incorporation of machine learning in their development and implementation. This study presents the versatile use of edge machine learning in diverse contexts, showcasing its effectiveness in smart devices, robots, and UAVs. It underscores the importance of local data processing for seamless decision-making and highlights specific implementations in the field.

Introduction

Internet of Things devices software runs on small electronic equipment present in it that collects data from the environment and sensors shared on the cloud for analysis and processing. Protocols and standards are fragmented which could cause edge computing issues along with security. (Zhou et al.) combine Artificial intelligence and edge computing to give edge artificial intelligence where devices generated data is taken as input to algorithms and processing is performed on data locally. (Roshak) Using the Internet of Things has given rise to edge machine learning. Several smart devices have given rise to Edge ML as they relate to each other on IOT clouds and make solutions for equipping IoT smart devices with Edge ML. As the number of devices are operating each of them processes collected data transferring from the device to IoT clouds directly or through gateways. The IoT network may not always be available for huge IoT devices collecting data uploading demand. Connections may become confined due to huge amounts of data uploading and could cause challenges in processing IoT cloud computing and security purposes which are addressed by Edge ML where the processing of collected data is performed locally on edge devices by utilizing machine learning and artificial intelligence algorithms for making an informed decision so that there will no need to process data on cloud reducing latency and increase secure processing (Angel et al.). Processing data locally makes processing in real-time possible (Sheng et al.). Autonomous machine training requires artificial intelligence that includes algorithms techniques and machine learning sub-branches as AI domain (Langley). Machine learning focuses on algorithms that learn through accessed data autonomously. Training of models, support vector machines, and classification are different approaches included in machine learning.

New task learning in machine learning includes deep learning where neural networks process information as the human brain learns and processes new information. ML and DL algorithms are used locally in Edge ML for processing data. Fig 1. Depicts hierarchy followed for machine learning.

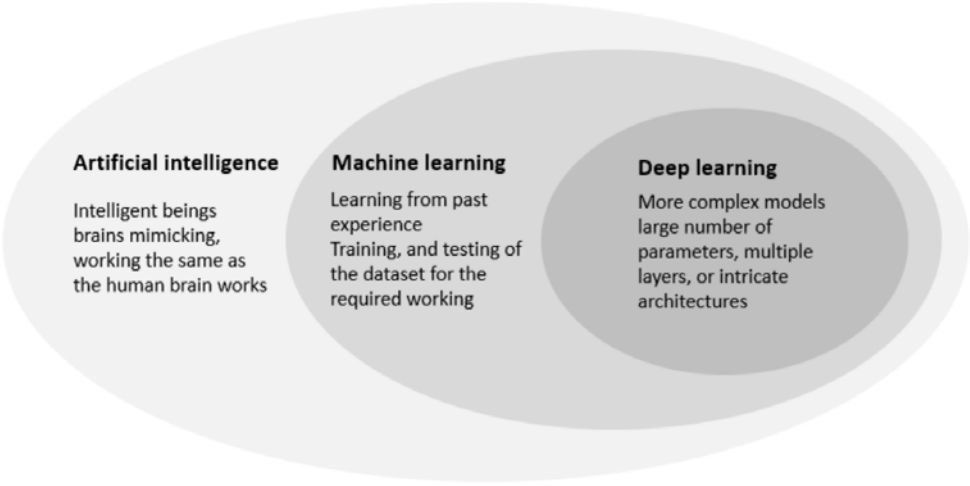


Figure 1. Hierarchy for Machine learning and artificial intelligence.

Big data is used for massive dataset processing collected from IoT systems. Predictive analysis of big data has improved over the years. Whereas in edge ML devices data requires powerful processing locally which removes the need for processing data in the cloud such as in Google home assistance software bot includes artificial intelligence and cloud computing in other words interaction of motionless robots. Similarly in Amazon echo speech recognition algorithms processes. Local storage contains data and upon asking processes info from there and gives results instead of connecting to the AWS cloud. In the case of local storage, it does not have information stored about any information or question asked; it will search in the cloud and provide correct information. They perform both edge and cloud computing algorithms. Robots are machines or pieces of machine programmed for specific task performance. They could perform a wide variety of tasks including simple to complex assignments. Robot processing units consist of compliable and readable code upon which they operate and execute (Wallén). Sensors in these devices perceive and respond to environmental variations. AI integration makes them smarter by allowing prediction more accurately and responding to queues. Teleoperated robots read environmental sensor data and return information for operation and are remotely controlled by users having a broad range of applications from taking samples to manipulating explosive materials and accessing dangerous environments for humans. It aims to reduce risks for humans such as rescue, finding leakage of gas, and detecting landmines after war. Many researchers are working to find solutions to perform jobs in safer conditions and extending the use of devices lacks rational thinking for progress. The study tried to detect UAV classification into categories through ML techniques that replace manual labor with automation in the industry (Bouguettaya et al.). The main question arises as applied research

that implements and provide a solution to edge machine learning for robots by implementing and optimizing experimental results in the following area:

- Can we effectively leverage edge machine learning to address limitations in embedded devices.
- Strategies can be employed to optimize internal memory and storage utilization in edge-embedded devices when implementing machine learning.
- Can we overcome network communication constraints to ensure seamless data transfer and real-time decision-making in edge machine learning applications.
- Do specific properties, such as security and mobility, impact the design and implementation of edge machine learning solutions within the underlying robotic systems.

These research questions address the core challenges and optimization opportunities identified in the study, providing a clear focus for further investigation and development in the field of edge machine learning for robots and drones.

Our Contribution

Proof of concept implemented in this paper using edge machine learning which is deployed successfully. Embedded devices integrated into IoT devices run AI neural network inference that handles teleoperated robots. Which has been trained in the machine learning cloud for accurate and improved control over teleoperated robots that have gas sensors and cameras. It finds areas where detection of gas leakage needs to be identified. Sends collected data about gas amount so it could be operated. ML processes on edge are the novelty of this approach which is improved by a small neural network applied to data on embedded device processing (Islam, Al Amin and Shin).

Literature Review

Edge machine learning previous studies are divided into training, inference, offloading, and caching (Zou et al.). Edge training focuses on architecture, optimization, and application along with security measures and privacy protocols application. It also involves edge training of typical applications and their cost for updating (McMahan et al.). Edge inference models design and compression in various fields i.e., face recognition (Duong et al.), human activity recognition HAR (Almaslukh, Al Muhtadi and Artoli), audio sensing (Georgiev et al.), and vehicle driving (Lane, Georgiev and Qendro), and pinpoints inference acceleration.

Edge caching replaces cache, deploys computation content, and implies data for practical approaches based on redundancy in computation that has been developed (Huitl et al.). Edge offloading deals with different types of offloading strategies i.e., device to device, device to edge, device to cloud, and hybrid offloading use cases such as smart industry and city (Tang et al.), intelligent transportation, and healthcare. IoT ecosystem uses communication protocols in the area of CoAP, MQTT, and IoT-NB 5G with applications in different fields are presented (El-Mottaleb et al.). In this study, our focus is on applications use cases where communication is lacking among modules and edge inference so that it must operate accurately from edge simulation to security perspective.

The classification of agricultural crops through imagery offers a means to enhance agricultural productivity by leveraging UAV technologies, resulting in reduced inspection durations and lowered crop management expenses technologies. Solutions typically come with limitations such as moderate area coverage, limited endurance, and susceptibility to weather conditions. When compared to satellites, they need more time to cover extensive areas and must adhere to aviation laws and regulations (Islam, Al Amin and Shin). The FBI employs a sophisticated system that integrates drones is a concern, and the middleware, which is implemented in Python, faces performance challenges when handling large volumes of data. Its efficiency depends on the sample rate of the connected IoT devices (Tang et al.).

The research focuses on the Microservices Framework within Edge Computing, exploring studies related to the orchestration of edge AI and microservices. At present, constraints associated with IoT edge devices limit the deployment of microservices containerization/virtualization on IoT nodes, excluding gateways (Liu et al.). The solution entails the utilization of unmanned aerial vehicles to capture multispectral images of a designated area within the upper Bogota River basin in Colombia. These images are then subjected to advanced processing techniques to estimate the levels of Total Suspended Solids (TSS) in the water. This application provides valuable insights into water quality and environmental conditions, facilitating more informed decision-making and resource management in the region.

The analysis involves the calculation of various statistical indexes, including regression, NDSSI (Normalized Difference Suspended Sediment Index), and NSMI (Normalized Suspended Material Index). Notably, machine learning (ML) and deep learning (DL) inference are not employed in this process. It's important to acknowledge that the accuracy of the results might be influenced by this approach. The research solution focuses on the Image Mosaic Method applied to UAV images in the context of earthquake emergency response. This method involves extracting scale-invariant feature points from remote sensing images using the Scale-Invariant Feature Transform (SIFT) technique. Regarding the system, it's important to note that it does not perform machine learning (ML) or deep learning (DL) inference at the edge. This omission may potentially impact the accuracy of the results obtained (Bi, Mao and Gong).

Recent studies have unveiled vulnerabilities and security threats within IoT architectures, primarily attributed to their heterogeneous nature. Consequently, network edge computing has gained prominence, increasingly relying on secure microservices to provide faster, more reliable, and more secure outcomes (Al-Doghman et al.).

Methodology– Edge Computing Machine Learning

Computer vision common applications are OCR, facial recognition, gesture recognition, and object recognition. Currently, it is used in supervised classification problems due to its reliability in detecting edges from images. Desired features in images are easily recognized by stacking detection even from video that contains thousands of frame images rapidly changing so it also works well for video. Typically, a computer vision task is structured in various steps which are as follows. Image acquisition is the first and important stage where the characteristics of the image followed are

illumination, quality, focus, and contrast (Davies). Then in the next step, pre-processing of images is performed for extraction of features from the input image which is an easy step as it removes unnecessary features from the image such as saturation in the image, hue effects, and image size reduction which leads to successful model implementation or to failure therefore these are important steps to perform before proceeding further.

Feature extraction is followed further after preprocessing of all image/video data which detects all important aspects of the image including edges, corners, and lines. Which becomes a component of the image processed and reduced to the background. After feature extraction data segmentation is performed to segment the data into region sets for optimal solution detection for the problem. These are important steps for reaching high-level processing validating input and estimation of size, and distance performed. Afterward, high-level processing is performed in which recognition tasks take place along with registration with final steps of making decisions based upon selected feature values i.e., input data matching. In a sensitive environment, these depend upon approval by humans (Davies).

Convolution machines resolve performance issues by both convolutional architectures and machines. Feature extraction is performed by using convolutional neural networks to extract features from images and directly from spatial data. It is a sequential process that associates overarching images utilizing multiple convolutional networks which removes ambiguity and allows training through backward propagation from spatial models dependent on input image (Toma et al.). Fig 2. Give overview representation of process followed for training for edge machine learning computation.

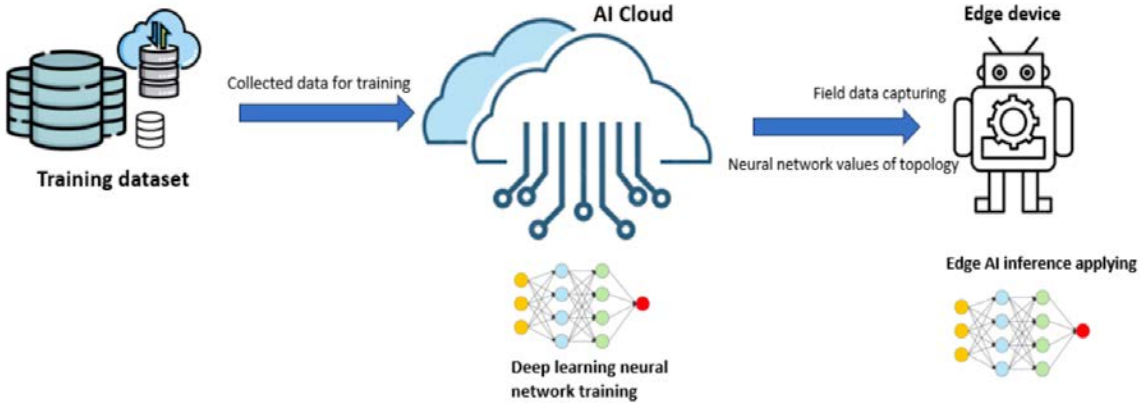


Figure 2. Machine learning and visual computing overview for network training.

Our method utilizes a feed-forward neural network to start mathematical formulation for letter ‘c’ inference. Feed-forward neural network along backpropagation is trained for recognizing canvas of java applet which have been trained on neural network for forward inference in ML cloud on IoT device. It contains a 30 elements array as input, the hidden layer having 10 nodes as an array, and the output layer with 4 neurons array. The input layer and hidden layers establish a connection with the weight of the matrix having 30 nodes as input elements and for the hidden layer 10 nodes or

columns. Whereas the connection between the hidden layer and output layer is established with the second weight matrix between the last two layers. It has 10 neurons in the hidden and 4 neurons in the output layer. The training is performed in a neural network based on proper epoch numbers. Training in the ML cloud is performed because of the specificity of data and the learning curve has been examined for convergence of model. Error loss and accuracy epochs are calculated to find network performance so that it may not be overfitted or under-fitted. Values for accuracy and loss have been stabilized after several training cycles epochs. There are two phases for this type of neural network feed forward layer with back propagation and hidden layer (Géron). The hidden layer calculated metrics and weights that are non deterministic having specific input as well as output. ML cloud consumes time. The phase involves the computation of random value weights at the start and is non-deterministic.

The inference phase involves a deterministic approach once values are established. Neural network classifies. It classifies letters written by hand the phase runs for edge ML. if CPU power has space for sorting neural arrays for float multiplication. If the float value of 4 bytes, then input 30 x 4 bytes + 10 x 4 + 4 x 4 + first matrix weight (30 x 10 x 4) + second matrix weight (10 x 4 x 4) = 1536 bytes (1.5 kbs). Hardware such as GPU and CPU are accessed by machine learning cloud providers. ML cloud access TensorFlow or deeplearning4j.org APIs in the context of software. Machine learning terms tasks perform data transformation, processing, training, inference, and deployment. Edge devices perform only inference tasks whereas embedded side ML hardware or software has no access for integrators and developers.

Backpropagation training shows feed-forward inference on smart devices. Normalization of pixels and the weights are chosen randomly between -1 and 1 within the matrix of weights as 32 bits.

Technically, a detailed explanation of the feed-forward process in a neural network is provided, with a focus on specific calculations and the limitations of floating-point precision in representing values and discusses the application of mathematical equations and the source code method used for the feed-forward process. Additionally, it is emphasized that the training of the neural network has already been completed in a machine learning (ML) cloud environment, involving back propagation techniques. The paper then transitions to presenting two proof-of-concept projects where edge machine learning has been successfully implemented in the fields of robotics and unmanned aerial vehicles (UAVs) or drones. These applications showcase the practical deployment of edge ML in real-world scenarios beyond the theoretical discussion of neural network calculations.

Machine Learning in the Teleoperated Robots

It gave robotic solutions and integration details for the detection of inflammable gasses inside the room. Robots have gas MQ-2 sensors along with cameras that help in searching for specific rooms. It creates graphs for identifying the root cause of leakage along with creating a dataset about the amount of gas present in the room to send to the operator for taking essential measures. A glove that is controlled remotely with an Arduino nano board and MPU6050 accelerometer that allows reading and translation into robot movements.

Gas leakage is life-threatening if not detected timely. Olfactory robots read leakage gas data efficiently and limit human exposure to such an environment. Methane gas does not have a specific odor and is used widely in industries for heating, combustion, and electricity generation. It is combined with pungent gas to make leakage detection easy. Lower environmental concentrations are not dangerous to humans, but higher concentrations could cause health issues along with deficiency of oxygen leading to asphyxiation. Methane gas of 1000 ppm (parts per million) is safe but at levels 50,000 to 150,000 ppm it could be combustible and dangerous. Beyond 150,000 ppm is not combustible as it replaces all oxygen that could contribute to ignition. Air quality depends upon the concentration of O2, CO, and O3. Device for smoke detection developed to reduce incidents that encounter limitations of low cost for mass production, its autonomic ability, less power consumption to influence the quality of reading data of gasses H2, CO2, CO. fire alarm also depends upon smoke detector and have chances of false alarms due to other conditions such as water vapor dust inside the device.

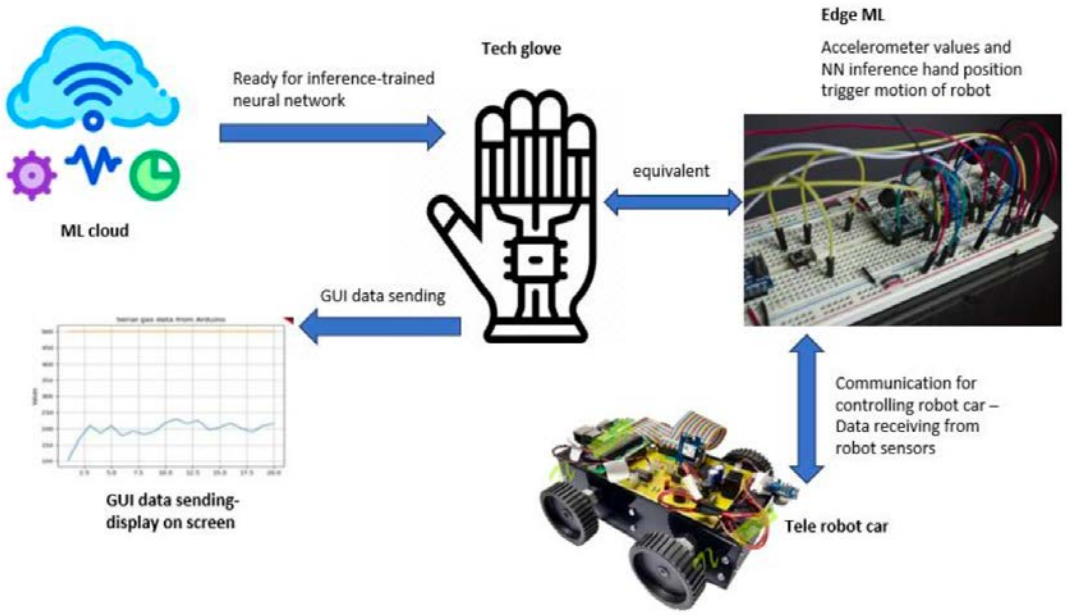


Figure 3. Machine learning architecture for Telerobot.

The olfactory telerobots have sensing ability along with obtaining information from environment datasets from speed, and wind along visual and auditory input streams. Fig 3. gives the overall architecture of telerobot. Sensors offer enhanced applications for sources of gas leakage by localization GSL of gas leakage in collapsed buildings, and industries. All along quite recently the smell-sensitive interface helps the operator in a great way to achieve maximum expectations that might not meet all expectations. Data is analyzed for accuracy determination of the system. Teleoperation robots allow input receiving and interaction with the environment important in terms of safety and demand reliability as compared to those offered by autonomous systems. They are useful due to sensing capability as visual, and audio are enough to tackle but combining both makes

it more advantageous as they smell from the surroundings and send information about dangerous chemicals in the environment. Chemometric data readings on the robot's e-nose are a direct approach. It has gas sensors that respond to the chemical environment. Information collected is processed by the ML algorithm for specific measures to hazardous environments. The advantage of e- nose is its low-cost versatility and integration. Robots require complementary sensor inputs i.e., anemometer, and gas mapping software.

The proposed architecture of the robotic unit detects a variety of gasses that are potentially dangerous for exceeding its limits having four main components. Machine learning cloud trains neural networks for movement-related decisions of robots by glove according to their position. Tech glove is an IoT smart device that has a sensor, Bluetooth module, and interface. It communicates with the robot through Bluetooth. And tell the robot to move in the direction and receive gas data. Gyroscope and Bluetooth data are enough for gloves but for more refined control use neural network inference. The robot is a car that moves and detects gas sources by reading atmospheric data. The attached sensor detects leakage of gas and sends data to a controller. The user interface generates a graph from the received data from the robot. It connects with Tech Glove and generates graphs for data along with live streaming to visualize the environment.

By using these modules robots collect data and help in the prevention of gas leakage on a large scale. Hardware utilizes Arduino UNO and NANO, two HC-05 Bluetooth, MPU0650 gyroscope, gas sensor, L293D motor shield driver, chassis, wheels, peripherals, and motor. Arduino is an open- source microcontroller board easily incorporated with LED sensors and boards and equipped with a power jack, rest button, ICSP header, and 20 digital input output pins, an AREF pin for voltage reference. HC-05 is used for Bluetooth communication and has a signal radius of 10m configured in slave pin value 0 and master pin value1 to set up modules. MPU6050 has a 3-axis gyroscope and 3D space to detect the motion and direction of objects. Its analog to digital pins give digital output and provide fast conversion of gestures. MQ-2 sensor for gas detects leakages and is adjusted through a potentiometer responsive to LPG, CH4, CO, C3H8, and C2H5OH. L293D motor driver shield amplifies low input current compatible with Arduino UNO. It is used to assemble a robotic unit. OLED display 128 x 64 is compatible and small displays data 128 x 64 points and consumes less energy. It communicates via a peripheral interface with microcontrollers. Used to display information on tech gloves and ensure robots are connected, the glove is powered on, and their battery level.

1. Robotic Parts: The project involves a robotic car with four DC motors. These motors can be controlled remotely via Bluetooth communication. The specific motor control instructions are sent to the car from the glove, allowing it to perform different movements based on the received commands.
2. Glove: The glove worn by the operator incorporates two Bluetooth modules. One module connects to the robotic car, enabling remote control, while the other connects to a PC's Bluetooth for data communication. The glove is also equipped with an MPU6050 gyroscope sensor to detect the tilt and position of the operator's hand.

3. MPU6050 Gyroscope Sensor: This sensor is used to analyze the tilt and position of the operator's hand. It provides data in the form of XYZ-axis values, where each axis corresponds to a different aspect of hand movement (X for up/down tilt, Y for left/right rotation, Z for hand rotation with a palm parallel to the ground).
4. Movement Detection: The XYZ values from the gyroscope sensor are used to detect specific hand movements. There are five distinct types of movements, including SPIN, STOP, START, BACK, and an undefined/faulty state.
5. Communication Protocols: Bluetooth serial communication is used for data exchange between the glove, robotic car, and PC. The glove sends movement codes to the robotic car, which interprets these codes to perform the corresponding actions. Additionally, gas sensor data is sent from the car to the glove via Bluetooth, and the glove relays this data to the PC.
6. Gas Sensor (MQ-2): The robotic car is equipped with an MQ-2 gas sensor, which measures gas levels and converts voltage values into meaningful data. This gas data is sent back to the glove, allowing the operator to monitor the gas levels in the environment.
7. Wi-Fi Streaming: The robotic car is also equipped with a mobile phone connected to the same Wi-Fi network as the PC. This phone streams live video and records the car's surroundings. This video streaming is crucial for providing visual feedback to the operator during teleoperation.
8. PC and GUI: The PC unit connects to the HC-05 Bluetooth module, establishing a connection with the glove. The GUI module on the PC continuously receives gas data values, which are stored in an array and plotted in real time. This allows the operator to monitor gas levels visually.
9. Security Testing: To test the system's functionality and safety, acetone gas is used as a test substance. The system is tested by exposing it to acetone gas, and the resulting gas level data is monitored and plotted. A threshold level is set to detect dangerous gas levels, and the system responds accordingly.

It combines robotics, sensor technology, Bluetooth communication, and a graphical user interface (GUI) to enable remote control of a robotic car while monitoring gas levels in the environment. It's designed with safety and real-time monitoring in mind, making it suitable for various applications, including hazardous environments where gas detection is critical.
The Edge Machine Learning (Edge ML) capability was considered for processing the video stream feed, but our preference was to leverage it specifically for enhancing the accelerometer data obtained from the technical glove. This strategic choice allowed for a more sophisticated refinement of the accelerometer values by employing a feed-forward neural network model within the glove module. This approach significantly improved the accuracy and quality of the accelerometer data, as detailed in the second section of this paper.

Results and Discussion

For proof of concept, we analyze performance from three points of view: efficiency, limitation, and improvements. This study successfully detected gas leakage and responded

spontaneously to the tech glove commands along with the creation of a map and graph of data collected without any delay. During testing gas levels were identified correctly and successfully for exceeding the threshold value of gas it generated a graph along with video generation that aids in identification and becomes easier in difficult scenarios. It faces limitations in terms of signal range which is 10m of the Bluetooth module. Therefore, it could operate successfully within this range which could be addressed using broader range signals i.e., Wi-Fi for future working. The sensor used for detection should be of good quality so it could limit false detection in case of extraordinary environmental conditions such as high humidity. Substituting the current MQ-2 gas sensor with a more robust alternative is advisable due to the MQ-2's susceptibility to various environmental factors, including long-term storage, physical shocks, and humidity, all of which can result in inaccurate data readings and interpretation. A superior gas sensor should be chosen for improved reliability and durability. Furthermore, an enhancement in the communication infrastructure can be achieved by replacing the Bluetooth modules with GSM IoT-NB/Wi-Fi modules. These upgraded modules would facilitate report generation through user-friendly software tools that don't require users to possess software development or programming expertise. To optimize this further, replacing the Bluetooth module connecting to the PC with an IoT-NB/Wi-Fi module running on a dedicated server would enable report accessibility from any computer through the plot's URL, ensuring greater convenience and accessibility.

The mean error and mean squared error calculations play a crucial role in monitoring the convergence of the neural network model across multiple epochs, particularly concerning the training, validation, and test datasets. These metrics serve as essential indicators for assessing how well the model is learning from the data. Fig 4. give a graph for performance comparison of different learning models.

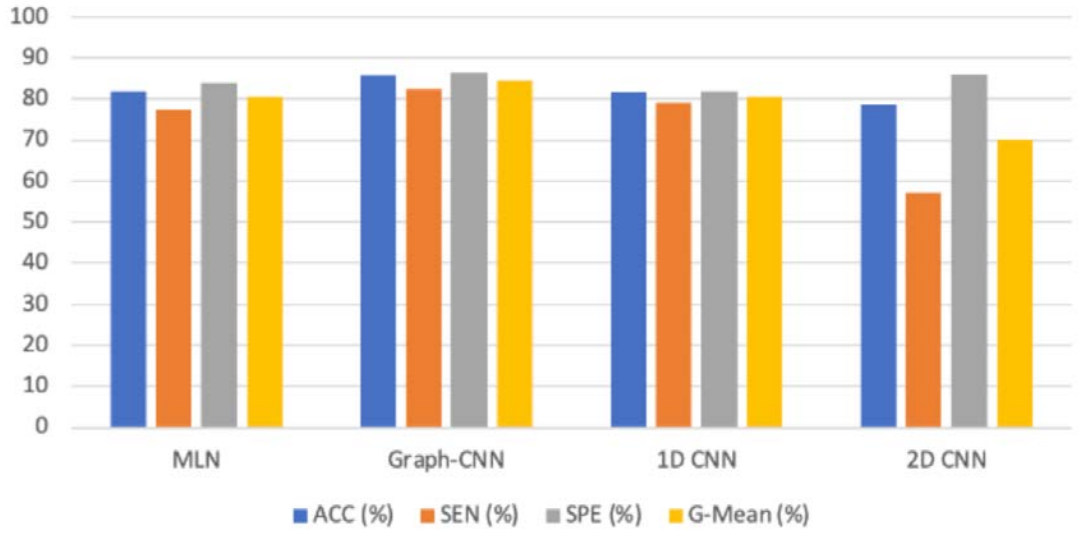


Figure 4. Deep learning models performance comparison

In terms of performance evaluation, when comparing various neural networks trained on the same task, such as image recognition, and utilizing the same number of training datasets, the

graphical representation typically showcases key performance metrics. These metrics include classification accuracy (ACC), which measures the overall correctness of the model's predictions, sensitivity (SEN), which gauges the model's ability to correctly identify positive instances, and specificity (SPE).

$$ACC = \frac{true\ positive + true\ negative}{true\ positive + true\ negative + false\ positive + false\ negative} \quad (1)$$

$$SPE = \frac{true\ negative}{true\ negative + false\ positive} \quad (2)$$

$$SEN = \frac{true\ positive}{true\ positive + false\ negative} \quad (3)$$

They assess the model's capacity to accurately identify negative instances. These performance metrics are fundamental for assessing and comparing the effectiveness of different neural network architectures on the given task.

Conclusion

The importance of AI and machine learning is evident in the integration of these technologies into various software and hardware products. One specific challenge in computer vision is training Convolutional Neural Networks (CNNs) leading to the adoption of Edge Machine Learning. This approach involves training neural networks in the cloud and conducting inference locally on devices, including IoT devices, and robots with a focus on the multi-layer feed-forward neural network (MLN) with backpropagation in some instances. This study explores the deployment of edge machine learning technology across various applications, including smart devices for robot control emphasizing the use of machine learning at the edge to process data locally for uninterrupted decision-making. The implementation of edge machine learning highlights the importance of local data processing. The study discusses the integration of gas sensors and cameras in teleoperated robots for detecting gas leakage, as well as the use of neural networks to improve control emphasizing the significance of edge machine learning in diverse contexts and its effectiveness in real-world applications. The study successfully demonstrated gas leakage detection and response, with limitations related to Bluetooth signal range and the quality of gas sensors. It recommends replacing the MQ-2 gas sensor with a more robust alternative and improving communication infrastructure by using GSM IoT-NB/Wi-Fi modules. The mean error and mean squared error calculations were used to monitor neural network convergence, and performance metrics like classification accuracy, sensitivity, and specificity were assessed for evaluating different neural network architectures. The paper underscores the growing importance of machine learning and AI in various applications and highlights the adoption of edge machine learning to address challenges in IoT devices, robots.

Works Cited

Al-Doghman, Firas, et al. "Ai-Enabled Secure Microservices in Edge Computing: Opportunities and Challenges." IEEE Transactions on Services Computing (2022). Print.

Almaslukh, Bandar, Jalal Al Muhtadi, and Abdel Monim Artoli. "A Robust Convolutional Neural Network for Online Smartphone-Based Human Activity Recognition." Journal of Intelligent & Fuzzy Systems 35.2 (2018): 1609-20. Print.

Angel, Nancy A, et al. "Recent Advances in Evolving Computing Paradigms: Cloud, Edge, and Fog Technologies." Sensors 22.1 (2021): 196. Print.

Research on Image Mosaic Method of Uav Image of Earthquake Emergency. 2014 The Third International Conference on Agro-Geoinformatics. 2014. IEEE. Print.

Bouguettaya, Abdelmalek, et al. "Deep Learning Techniques to Classify Agricultural Crops through Uav Imagery: A Review." Neural Computing and Applications 34.12 (2022): 9511- 36. Print.

Davies, E Roy. Computer and Machine Vision: Theory, Algorithms, Practicalities. Academic Press, 2012. Print.

Mobiface: A Lightweight Deep Learning Face Recognition on Mobile Devices. 2019 IEEE 10th international conference on biometrics theory, applications and systems (BTAS). 2019. IEEE. Print.

El-Mottaleb, Somia A Abd, et al. "A Qos Classifier Based on Machine Learning for Next-Generation Optical Communication." Electronics 11.16 (2022): 2619. Print.

Georgiev, Petko, et al. "Low-Resource Multi-Task Audio Sensing for Mobile and Embedded Devices Via Shared Deep Neural Network Representations." Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1.3 (2017): 1-19. Print.

Géron, Aurélien. Hands-on Machine Learning with Scikit-Learn, Keras, and Tensorflow. " O'Reilly Media, Inc.", 2022. Print.

Tumindoor: An Extensive Image and Point Cloud Dataset for Visual Indoor Localization and Mapping. 2012 19th IEEE International Conference on Image Processing. 2012. IEEE. Print.

Islam, Anik, Ahmed Al Amin, and Soo Young Shin. "Fbi: A Federated Learning-Based Blockchain-Embedded Data Accumulation Scheme Using Drones for Internet of Things." IEEE Wireless Communications Letters 11.5 (2022): 972-76. Print

Deepear: Robust Smartphone Audio Sensing in Unconstrained Acoustic Environments Using Deep Learning. Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing. 2015. Print.

Langley, Pat. "The Changing Science of Machine Learning." Machine learning 82.3 (2011): 275- 79. Print.

Liu, Chang, et al. "A New Deep Learning-Based Food Recognition System for Dietary Assessment on an Edge Computing Service Infrastructure." IEEE Transactions on Services Computing 11.2 (2017): 249-61. Print.

Communication-Efficient Learning of Deep Networks from Decentralized Data. Artificial intelligence and statistics. 2017. PMLR. Print.

Roshak, Michael. Artificial Intelligence for Iot Cookbook: Over 70 Recipes for Building Ai Solutions for Smart Homes, Industrial Iot, and Smart Cities. Packt Publishing Ltd, 2021. Print.

Sheng, Shuran, et al. "Deep Reinforcement Learning-Based Task Scheduling in Iot Edge Computing." Sensors 21.5 (2021): 1666. Print.

Tang, Bo, et al. "Incorporating Intelligence in Fog Computing for Big Data Analysis in Smart Cities." IEEE Transactions on Industrial informatics 13.5 (2017): 2140-50. Print.

Toma, Cristian, et al. "Edge Machine Learning for the Automated Decision and Visual Computing of the Robots, Iot Embedded Devices or Uav-Drones." Electronics 11.21 (2022): 3507. Print.

Wallén, Johanna. The History of the Industrial Robot. Linköping University Electronic Press, 2008. Print.

Zhou, Zhi, et al. "Edge Intelligence: Paving the Last Mile of Artificial Intelligence with Edge Computing." Proceedings of the IEEE 107.8 (2019): 1738-62. Print.

Edge and Fog Computing Enabled Ai for Iot-an Overview. 2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS). 2019. IEEE. Print

1995 Rugby World Cup; Illusory Racial Reconciliation in the Rainbow Nation By Ted Huneke

In the words of Nelson Mandela, "sport has the power to change the world" (Mandela).[1] Inspirational, motivational, and unifying, he saw rugby as the key to the coalescence of the so-called "rainbow nation:" a South Africa free from the racial segregation of the Apartheid system that persisted from 1948 to the early 1990s.[2] In the 1995 Rugby World Cup, despite its status as the host nation, the South African team, nicknamed the Springboks after the national animal, were still a heavy underdog. Nevertheless, they made a miraculous run to the final, where they faced the heavily favored All Blacks of New Zealand. After extra time, the Springboks upset the All Blacks, which delighted the nation, according to Mandela, the media, and the team. However, despite all the assumptions of how beneficial this victory would be to the entire nation, many Black South Africans remained indignant at the lack of progress following the abolition of Apartheid. The South African government leveraged their victory at the 1995 Rugby World Cup to unite their historically divided homeland in the eyes of the world, falsely overstating the importance of sports in nation-building, through a contrived, illusory sense of "imagined community" (Anderson 15).[3] In this invented narrative, the Springboks had catapulted the nation past its racially tumultuous history. In contrast, by acknowledging the persistent economic tribulations of black South Africans, this event is more aptly characterized as a moment of sportswashing; the idealistic representation, from the President, press, and players, overshadowed the continual racial inequality of their nation in an attempt to improve the world's perception of them.

Background Context

Following a century of discrimination, a young South African government, led by Nelson Mandela—an abolitionist of Apartheid and adopter of democratic governing—looked to the Springboks as a means of uniting the historically divided nation. By the time Apartheid rule was established in 1948, Black South Africans had been facing discrimination at the hands of European powers for decades, including the violence of the Anglo-Zulu War, imprisonment during the Boer War, and the British monopolization of the lucrative diamond trade.[4] Exploitations furthered when the Native People's Act was passed in 1913, which limited black South Africans to owning only 7% of the land.[5] The Nationalist Party's election in 1948 marked the beginning of Apartheid, the harshest example of legislated racism. Discriminatory laws dictated that the Black population were to pay higher taxes than their white counterparts, not allowed to teach fellow Black South Africans, or be taught to read and write by white employers

[1] Laureus, "Nelson Mandela's Iconic Speech - 'Sport Has the Power to Change the World' - Full Version," February 9, 2020: 0:00-0:04.
[2] Multiple sources in the bibliography include the idea of the "Rainbow Nation."
[3] Benedict Anderson, "Imagined Communities," as quoted in Jacqueline Maingard, "Imag(in)ing the South African Nation: Representations of Identity in the Rugby World Cup 1995," *Theatre Journal* 49, no. 1 (1997): 15.
[4] Class notes 2/17/23, 2/22/23.
[5] Class Notes 2/22/23.

(Wizarat 84-85).[6] In the realm of sports, from 1960 to 1972, South Africa was banned from all major sporting events; the country's lowly reputation persisted until the fall of Apartheid and the 1994 election of the first Black president, Nelson Mandela, who sought to unify his people through rugby (Catsam 8).[7]

Conditions of and Reactions to the 1995 World Cup

The new leader viewed the 1995 Rugby World Cup as a means of uniting the nation, and reversing the stigma that engulfed the nation. In his opening address of the tournament, during a global broadcast, Mandela stated that the international competition "[confirmed] the unity in diversity and the humanity in healthy contact that [their] young democracy [had come] to symbolize. On behalf of [their] rainbow nation [he] [welcomed] [them] all" (Dickow and Møller 198).[8] Mandela believed that the tournament presented the opportunity for South Africa to reaffirm their democracy on the international stage—an opportunity to reaffirm their unity, both in the eyes of the people and the world. In speaking on the behalf of his "rainbow nation," Mandela blindly assumed that his diverse populace, one of "unity," appreciated the presence of the Rugby World Cup in their country—a serious error in judgment that was clearly confirmed by the conflicted response to the Springbok's victory.

The opening ceremonies of the 1995 World Cup also contributed, through stereotyping, to the illusory narrative of "unity in diversity" and the invention of this "rainbow nation." The opening event, which preceded the first match, superficially depicted the complex facets of South African identity, hosting Zulu dance, with men wearing animal skins, and an interpretation of "Cape Carnivals," which were shows performed by Black South Africans that originally cropped up in protest of performances by blackface-wearing Afrikaners. The Zulu dance was hyper-stylized and the Cape Carnival interpretation was an "inauthentic copy" that seemed unrehearsed (Maingard 19).[9] These broad stereotypes attempted to enhance Mandela's narrative of "unity in diversity" through oversimplification, reducing the labyrinthine history of Black South Africans into the shallow stereotype of tribal life, more palatable to the less-knowledgeable outsiders watching the broadcast; the inaccuracy of the "Cape Carnival" only furthered the disrespectful nature of the ceremonies (Dickow and Møller 198).[10] Mandela's faulty assumption of his people's absolute appreciation of the Springboks, and the ceremony's generalization of the identities of Black South Africans were emblematic of larger societal insensitivities that reduced or ignored the cultural and racial nuances of the society.

---

[6] Talat A Wizarat, "APARTHEID AND RACIAL DISCRIMINATION IN SOUTH AFRICA— AN OVERVIEW OF THE CONTROL NETWORK." *Pakistan Horizon* 33, no. 4 (1980): 84-85.
[7] Derek Charles Catsam, "The Death of Doubt? Sport, Race, and Nationalism in the New South Africa," *Georgetown Journal of International Affairs* 11, no. 2 (2010): 8.
[8] Nelson Mandela as quoted in Helga Dickow and Valerie Møller, "South Africa's 'Rainbow People', National Pride and Optimism: A Trend Study," *Social Indicators Research* 59, no. 2 (2002): 198.
[9] Jacqueline Maingard, "Imag(in)ing the South African Nation: Representations of Identity in the Rugby World Cup 1995," 19.
[10] Dickow and Møller, "South Africa's 'Rainbow People', National Pride and Optimism: A Trend Study," 198.

The immediate reaction to the triumph from the players, and the following day from the press, further heightened the assumptive narrative that all South Africans supported (and benefitted from) the national team's success. After defeating the All Blacks in overtime, Francis Pienaar, the Afrikaaner captain of the Springboks, was interviewed about the victory, during which he stated, when referring to the team's support, that "[they] didn't have sixty thousand South Africans. [They] had forty three million South Africans" (World Rugby).[11] This sentiment was mirrored once again in the *Sunday Times* the morning after: "South Africa turned into one big street party last night as jubilation at the country's greatest sporting victory gripped the nation" (Van Der Riet 100).[12] Both Pienaar and the press assumed, like Mandela, that everyone in the nation was appreciative of the victory—that everyone supported the Springboks. Both furthered the idealistic "imagined community" that Mandela had claimed existed; the illusory sentiment that the nation-wide broadcast of the Springboks fostered unity (Maingard 15).[13]

Downside of the Celebration

The overly assumptive narrative of the President, players, and press—that all of South Africa supported the Springboks—was inherently inaccurate, neglecting the millions of South Africans that remained indignant toward the sport itself. To millions of Black South Africans, the Springboks represented "white supremacy at its apex;" there was only one player on the 1995 team who was black, especially troublesome considering that the vast majority of the population and youth rugby players were non-white (Catsam 9).[14] Mandela, his players, and the press knew this yet persisted in using the Springboks to bolster their narrative; rugby was, in reality, a symbol of racial disenfranchisement, not national unity. Many millions of Black South Africans felt that rugby was not emblematic of racial reconciliation but that, in fact, it was the exact opposite: a powerful symbol of Afrikaner domination. Thus, Mandela's attempts to unite the entirety of his "rainbow nation" through rugby were obviously unrealistic, indicating that Mandela was not acting solely for his people but rather using the event as a means of seeming more unified on the world stage.[15]

Not only was this narrative of broad support inaccurate, it also downplayed and overshadowed the economic tribulations of and the labor discrimination against Black South Africans. Despite all idealistic sentiments displayed to the world, during the 1995 Rugby World Cup and in the years following, racism was still pervasive in the nation. Despite the overly positive perception of the Springboks' victory from the team, government, and the press, Black South African were greatly disadvantaged in a society that still, economically, closely resembled Apartheid. From 1993 to 2008, racial economic inequality only increased in South Africa; a slew of unsuccessful economic plans failed to develop the post-Apartheid society and readjust racial

---

[11] "Rugby World Cup 1995 Final: South Africa v New Zealand," *YouTube*, World Rugby, 2020: 2:00:40-2:00:44
[12] *Sunday Times* as quoted in Jane Van Der Riet, "Triumph of the Rainbow Warriors: Gender, Nationalism and the Rugby World Cup," *Agenda: Empowering Women for Gender Equity*, no. 27 (1995): 100.
[13] Maingard, "Imag(in)ing the South African Nation: Representations of Identity in the Rugby World Cup 1995," 15.
[14] Catsam, "The Death of Doubt? Sport, Race, and Nationalism in the New South Africa," 9.
[15] Multiple sources in the bibliography include the idea of the "Rainbow Nation."

inequities (Leibbrandt et al Section 5).[16] In 1994, plans to redistribute 30% of all farmland were widely considered unsuccessful (Padayachee 271).[17] The educational system remained "inequitable and largely unable to provide relevant skills" (Padayachee 271).[18] In the late 1990s, research on racism in South African labor-markets revealed persistent discrimination against Black workers, even when taking into account the insufficient levels of schooling in Black communities compared to their white counterparts (Darity and Nembhard 310).[19] The presumptive, all-encompassing language of Mandela's opening speech, Pienaar's interview, and the Sunday Times attempted to sportswash: to distract the world from the economic and racial difficulties that was still pervasive—a concerted attempt to overshadow the economic realities of Black South Africans.

The reliance on rugby success that South African leadership exhibited in their nation-building efforts was imprudent; victory failed to provide any relief to the economic hardships of Black citizens. The explanation that Mandela and the corroborators of his narrative were simply ignorant to the ineffectuality of sports in nation-building is dismissible because the Truth and Reconciliation Commission, which was later set up to address the inequity of rights lingering from Apartheid, had not even met at the time of the Springboks' victory (Catsam 9).[20] Mandela, the players, and the press knew that inequality was persistent, and even growing, yet disregarded this when it came time to celebrate in front of the world.

This simplistic attempt to absolve the entire nation of its deeply troubled past further ostracized the black majority from the "imagined community" of rugby. Because South Africa did not honestly confront their economic struggles, attempting to convince the world of their progress, international pressure to address the issues never materialized. Indeed, it was the global anti-Apartheid focus, resulting in crippling economic sanctions, that dissolved the party and helped Mandela become president in the first place. In its myopic focus on the world's perception, a new South Africa abandoned its non-white populace, continuing the Apartheid legacy it supposedly sought to dismantle. The prioritization of global opinion, in turn, made black South Africans feel unwelcome in their own country, as seen in their general disgust at the marginalizing and superficial opening ceremony (Maingard 17).[21] While South Africa made large strides in terms of social rights, and democratic reform (i.e. the one-man, one-vote system), the Rugby World Cup represented progress in international standing, not the apex of racial reconciliation as portrayed by Mandela, the players, and the press. Rather than acknowledging a lack of economic advancement, the South African establishment chose to protect its reputation at the expense of its people.

---

[16] Murray Leibbrandt, Arden Finn, and Ingrid Woolard, "Describing and Decomposing Post-apartheid Income Inequality in South Africa," *Development of South Africa* 29, no. 1 (February 14, 2012): Section 5. https://www.tandfonline.com/doi/full/10.1080/0376835X.2012.645639.

[17] Vishnu Padayachee, "The South African Economy, 1994-2004," *Social Research* 72, no. 3 (2005): 271.

[18] Padayachee, "The South African Economy, 1994-2004," 271.

[19] William Darity and Jessica Gordon Nembhard, "Racial and Ethnic Economic Inequality: The International Record," *The American Economic Review* 90, no. 2 (2000): 310.

[20] Catsam, "The Death of Doubt? Sport, Race, and Nationalism in the New South Africa," 9.

[21] Maingard, "Imag(in)ing the South African Nation: Representations of Identity in the Rugby World Cup 1995," 17.

Conclusion & Legacy

Two decades following the fallacy of "unity in diversity," South Africa's racial inequality was still at "world beating levels" (Leibbrandt et al Section 6).[22] Despite the successful eradication of legislated Apartheid and the adoption of democracy, the economic realities of Black citizens on the ground largely remained unchanged, or, according to some statistics, even worse. The oversimplified nature of the "unity in diversity" narrative from the 1995 World Cup has persisted and even grown in prominence in the ensuing years: the Hollywood film *Invictus* (2009), starring Matt Damon and Morgan Freeman, exhibited the same romanticization of the immediate post-Apartheid era (Catsam 9). [23] By glamorizing the actions of Mandela and his Springboks while neglecting the persistent suffering of Black South Africans, viewers were once again not provided with a comprehensive narrative, echoing the misleading half-truths of the President, the players, and the press at the time of the tournament. The superficial storyline surrounding the Springboks in the mid-1990s, coupled with the more recent blockbuster that reiterated this mythology, suggests that audiences, when considering the content of their leaders and those that are held highly, should question their assumptions and seek to uncover the nuances that every story contains.

[22] Leibbrandt, Finn, and Woolard, "Describing and Decomposing Post-apartheid Income Inequality in South Africa," Section 6.
[23] Catsam, "The Death of Doubt? Sport, Race, and Nationalism in the New South Africa," 9.

Works Cited

Albert Fu, and Martin Murray. “Sentimentalizing Racial Reconciliation in the ‘New South Africa’: Cinematic Representation of the 1995 Rugby World Cup.” *Black Camera* 9, no. 1 (2017): 22–46. https://doi.org/10.2979/blackcamera.9.1.02.

Catsam, Derek Charles. “The Death of Doubt? Sport, Race, and Nationalism in the New South Africa.” *Georgetown Journal of International Affairs* 11, no. 2 (2010): 7–13. http://www.jstor.org/stable/43133837.

Darity, William, and Jessica Gordon Nembhard. “Racial and Ethnic Economic Inequality: The International Record.” *The American Economic Review* 90, no. 2 (2000): 308–11. http://www.jstor.org/stable/117241.

Dickow, Helga, and Valerie Møller. “South Africa’s ‘Rainbow People’, National Pride and Optimism: A Trend Study.” *Social Indicators Research* 59, no. 2 (2002): 175–202. http://www.jstor.org/stable/27527026.

Höglund, Kristine, and Ralph Sundberg. “Reconciliation through Sports? The Case of South Africa.” *Third World Quarterly* 29, no. 4 (2008): 805–18. http://www.jstor.org/stable/20455073.

Laureus. “Nelson Mandela’s Iconic Speech - ‘Sport Has the Power to Change the World’- Full Version,” February 9, 2020.https://www.youtube.com/watch?v=y1-7w-bJCtY.

Leibbrandt, Murray, Arden Finn, and Ingrid Woolard. "Describing and Decomposing Post-apartheid Income Inequality in South Africa." *Development of South Africa* 29, no. 1 (February 14, 2012): 19-34. Accessed May 17, 2023. https://www.tandfonline.com/doi/full/10.1080/0376835X.2012.645639.

Lukhele, Francis. “Post-Prison Nelson Mandela: A ‘Made-in-America Hero.’” *Canadian Journal of African Studies / Revue Canadienne Des Études Africaines* 46, no. 2 (2012): 289–301. http://www.jstor.org/stable/43860349.

Maingard, Jacqueline. “Imag(in)Ing the South African Nation: Representations of Identity in the Rugby World Cup 1995.” *Theatre Journal* 49, no. 1 (1997): 15–28. http://www.jstor.org/stable/3208422.

“1995 Rugby World Cup.” *Wikipedia*. Wikimedia Foundation, February 17, 2023. https://en.wikipedia.org/wiki/1995_Rugby_World_Cup.

Pandit, Puja. “One Team One Country: Sports for Development and Peace.” Vision of Humanity, November 23, 2022. https://www.visionofhumanity.org/one-team-one-country-the-importance-of-timing/.

Padayachee, Vishnu. “The South African Economy, 1994-2004.” *Social Research* 72, no. 3 (2005): 549–80. http://www.jstor.org/stable/40971781.

Poole, Eric. “The Greatest Day in Sports History: Mandela, Brutus and the 1995 Rugby World Cup.” *Atticus Review*, August 7, 2016. https://atticusreview.org/the-greatest-day2/.

“Rugby World Cup 1995 Final: South Africa v New Zealand.” *YouTube*, World Rugby, 2020, https://www.youtube.com/watch?v=RYCBAHhIjVY.

Van Der Riet, Jane. “Triumph of the Rainbow Warriors: Gender, Nationalism and the Rugby World Cup.” *Agenda: Empowering Women for Gender Equity*, no. 27 (1995): 98–110. https://doi.org/10.2307/4065979.

Wizarat, Talat A. “APARTHEID AND RACIAL DISCRIMINATION IN SOUTH AFRICA—AN OVERVIEW OF THE CONTROL NETWORK.” *Pakistan Horizon* 33, no. 4 (1980): 84–87. http://www.jstor.org/stable/41403900.

Vaping: Comprehensive Review and Health Effects on Youth By Kevin Cao

Abstract

Electronic cigarettes, or vaping, particularly ENDS products, have become the most commonly used tobacco product among youth. According to a recent report from the CDC, 14.1% of high school students (2.14 million) and 3.3% of middle school students (380,000) are currently using e-cigarettes. This article reviews the statistics and trend of e-cigarette use in youth. It covers the structure of vaping devices, nicotine, solvent, flavoring and other components in the e-liquid, and how they are delivered and transformed when vaped. The toxicity of major components of vaping liquid and corresponding transformation products are also discussed, as well as various adverse health effects on users. The article further discusses how e-cigarettes can negatively impact brain development, physical fitness, academic performance, and mental health in youth and young adults.

Introduction

Tobacco has been used by humans for thousands of years, but it was not until the 1960s that the general public recognized the harm it had on human health. The U.S. Surgeon General released the first report on the health effects of smoking with a conclusion that smoking causes lung cancer and bronchitis. Later research found that smoking is especially harmful to the adolescent population whose lung functions are not fully developed and are prone to damage by harmful chemicals. When lit, tobacco leaves in conventional cigarettes burn and produce smoke containing vaporized nicotine that the user inhales. The smoke also contains thousands of other substances, including at least 70 carcinogens, which are compounds that promote the development of cancer. Cigarette smoking is the leading preventable cause of death in the United States and according to the CDC, contributes to nearly 1 out of 5 deaths (Kimbrough).

With more and more people becoming aware of the adverse effects of smoking and wanting to quit, tobacco use has been declining globally (O'Connor et al.). In 2007, a new type of product, the electronic nicotine delivery system (ENDS, also called e-cigarettes or vaping devices), came to the market and became increasingly popular as a substitute to traditional cigarettes for those addicted to smoking. This device is promoted as a healthier alternative for those wanting to quit cigarette smoking and a way to reduce its appeal and addictive potential for youth (Bonner et al.). Electronic cigarettes are tobacco-free, and nicotine is delivered to users via a non-combustion method, i.e., aerosol mist. The device acts as more of a humidifier that delivers the mist to the lungs where nicotine and other components are absorbed.

Youth Vaping: Statistics and Trend

Since e-cigarettes entered the U.S. Market, especially since 2014, the use of ENDS devices has ballooned among young users. In fact, vaping had such a surge in popularity that the Oxford English Dictionary named "vape" as its 2014 word of the year. In 2018, the FDA and Surgeon General of the United States declared youth vaping a public health "epidemic"

("Surgeon General"; FDA Commissioner).

In late 2022, the FDA and CDC released the most recent survey result on youth tobacco use in middle and high school students (Park-Lee et al.). Approximately 11.3% (3.08 million) of middle and high school students reported current use of any tobacco product, including 16.5% (2.51 million) of high school and 4.5% (530,000) of middle school students. E-cigarette use among youth remains a top concern, as e-cigarettes were the most commonly used tobacco product among high school (14.1%, 2.14 million) and middle school (3.3%, 380,000) students. Overwhelmingly, 85% of current e-cigarettes users chose flavored products, with fruit flavors being the most popular (Cooper et al.).

There are many factors that contribute to the popularity of e-cigarettes in adolescents and young adults. Novelty and curiosity are among the top appeals of use. Flavors, in particular, attract both youth and adults to use ENDS. They decrease harm perceptions and increase willingness to try and initiate the use of ENDS. Other attractive factors include variety and choice of feature (design, nicotine levels, adjustable device settings, producing 'clouds'), packaging and price (especially relative to cigarettes), avoidance of smoking restrictions, positive sensory, and other physical experiences, such as social acceptability and perceived safety for bystanders. Disposable and pod-style devices have become popular, especially among adolescents, due to their attractive design, user-friendliness, less aversive vaping experiences, desirable flavors, and discreetness (O'Connor et al.).

With major tobacco companies moving toward harm reduction and away from combusted cigarettes with a focus on a "smoke-free" world, more and more ENDS and other delivery systems are promoted to smokers with better appeal and improved marketing. On the other hand, policy makers and regulatory agencies have taken a multitude of actions to keep ENDS out of the hands of youth, including policy making, compliance and enforcement, public education, and scientific research ("Congressional Testimony").

Vaping Devices

Compared to tobacco smoking, vaping devices apply other forms of nicotine delivery. These include electronic nicotine delivery systems (ENDS) and heated tobacco products (HTPs). Evolved over time, some of these products now deliver nicotine at levels comparable to cigarettes and may serve as effective substitutes for smokers (O'Connor et al.).

The electronic nicotine delivery system (ENDS) is the focus of this review since it is predominantly used in the US. ENDS represents a class of products that commonly contain four parts: a mouthpiece, a rechargeable battery that powers a heating element to aerosolize a liquid, a cartridge with a liquid (or pod) that typically contains nicotine, water, flavorings and humectants, and an atomizer. The atomizer heats the solution from the cartridge, producing an aerosol mist of tiny liquid droplets (technically not a vapor) that the user inhales. When users inhale the mist, nicotine and other compounds are absorbed into the lung and upper airway (Kimbrough).

ENDS products have changed substantially over the past 16 years, starting with disposable cigalikes that resembled a cigarette, then rechargeable devices with refillable tanks

and more modifiability (e.g., adjustable power), and, more recently, pod-style devices like JUUL. JUUL was launched in 2015 and quickly became the predominant ENDS brand (O'Connor et al.). As of 2021, ENDS products were the second most used tobacco product in the US, with 4.5% of adults reporting current use. This number was even higher (11.0%) for the 18-24 age group (Cornelius et al.).

Vaping Liquid Components and Transformation

The e-liquid in ENDS device has a number of components that work together. This article focuses on the solvents, nicotine, flavorings, transformation products, and metal contaminations.

Solvents

The largest components of e-liquids are solvents, or humectants, that comprise 80% - 94.9% of the volume of e-liquids. The purpose of the solvents is to keep other components in suspension, to enhance absorption of the wicking material, and to generate aerosolized plumes when e-liquid is heated (Clapp and Jaspers). The two most common solvents used in e-liquids are propylene glycol (PG) and vegetable glycerin (VG), or glycerol. The FDA classifies PG as "generally recognized as safe" (GRAS) for ingestion, but not for aerosolization (Code 21CFR184.1666). PG is reported to produce a "throat hit" similar to cigarette smoking, while glycerol is used to enhance the flavor and increase the amount of aerosol. When heated, they generate toxic thermal decomposition products (Wang et al.).

Nicotine

Most e-liquids contain nicotine. Since nicotine causes addiction, it is important to understand its exposure and absorption via inhalation and the potential addictiveness of e-cigarettes. Nicotine is a type of alkaloid that can be found naturally in plants. Alkaloids are natural organic molecules that contain at least one nitrogen, have important physiological effects on humans, and are widely used in modern medicines. They are either naturally occurring or synthetically modified. Some examples are caffeine, morphine, codeine, and quinine (Heinrich et al.).

Nicotine (Fig. 1) has two nitrogens present in the molecules, and each has a different pKa. The pyridine nitrogen has a pKa of 8.02 and the *N*-methylpyrrolidine nitrogen has a pKa of 3.12. Depending on the acidity or pH, nicotine can exit as a free base (Nic, unprotonated), or its conjugate acid forms: monoprotonated $NicH^+$ and diprotonated $NicH_2^{2+}$ (Fig. 1) (Joullié). At the physiological pH of 7, only one of the nitrogens is protonated, making monoprotonated $NicH^+$ dominant.

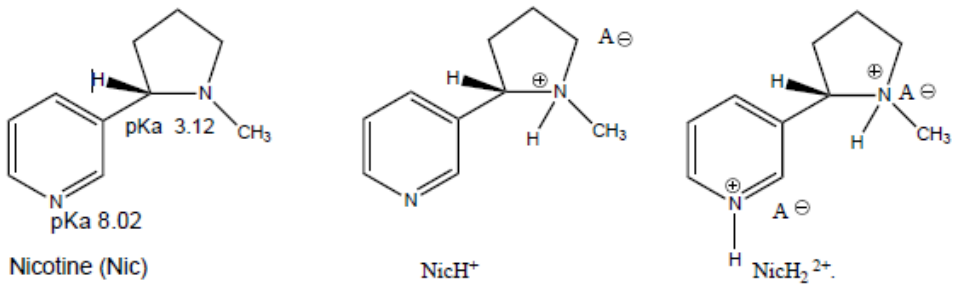


Fig. 1. Nicotine structure and its acid forms

Protonation plays an important role in nicotine absorption and sensation. It was discovered that the free base form of nicotine is more readily absorbed by the lungs, but users feel hard and scratchy at the back of the throat. On the other hand, the conjugate acid form of nicotine cannot be absorbed into the lungs as easily, but is reported to feel milder and smoother when inhaled.

The protonation can be changed by modifying the acid/base conditions of the medium. Traditional tobacco smoke aerosols produced during vaping are believed to contain a combination of Nic and $NicH^+$ because conditions in the aerosol are not sufficiently acidic. Free base of nicotine is the only form that can be volatilized upon heating.

The electronic nicotine delivery system (ENDS) is different. Many vaping liquids contain acid forms or nicotine salts (the main source of nicotine in JUUL pods). The introduction of salts in the chemical brew adds a new level of complexity. Here are some of the organic acids used to make nicotine salts and their pKa's (Joullié): formic acid: 3.75, acetic acid: 4.74, benzoic acid: 4.20, citric acid: 3.14 / 4.77, malic acid: 3.40 / 5.20. The addition of these acids resulted in a wide pH range in e-liquids and multiple chemical forms of nicotine, thus altering the nicotine/nicotine salt ratio in the aerosol, and changing nicotine delivery and absorption when inhaled. In a 2017 published study, e-liquid solutions with equal nicotine concentration and pH but different acid salts (formate, acetate and citrate) were prepared to assess the effect of the counter anion on nicotine partitioning. High performance liquid and gas chromatography methods were used to determine the counter-anions at the two protonated nicotine forms (mono- and di-), as well as to identify the free nicotine form in standard solutions of e-liquids and aerosols. Formate and acetate salts were found to evaporate intact while citrate salt decomposed, releasing the free-base of nicotine that was enriched in aerosols. The total nicotine yield in aerosols is a function of the battery power output, pH of the liquid, and the thermostability of the salt form (El-Hellani et al.).

Flavoring Chemicals

One of the features that attracts users is the wide choice of flavors. There have been a substantial amount of studies that investigated the chemical composition of hundreds of flavored e-liquids. The most common chemicals used to generate these flavors were menthone, p-anisaldehyde, menthol, cinnamaldehyde, vanillin, and ethyl maltol (Behar et al.). Some of

these flavorings had concentrations that exceeded those found in food or consumer products.

Transformation Products

When an e-liquid is heated to be converted to an aerosol, some of the liquid components are degraded or interact with each other to be transformed into new chemical substances. ENDS devices with variable power settings allow users to adjust wattage to experience different levels of vaping sensations. Changing the wattage alters the heating temperature, which in turn affects the type and amount of transformation products generated.

In a 2017 study, various transformation products of the carrier solvents were identified, including glycidol, prop-1-en-1-ol, dihydroxyacetone, acrolein, lactaldehyde, glycolaldehyde, glyceraldehyde, acetaldehyde, propanal, acetone, hydroxyacetone (acetol), and allyl alcohol (Jensen et al.). Glycidol and acrolein were primarily produced by glycerol degradation, while acetol and prop-1-pen-1-ol were produced mostly from propylene glycol (PG). On the other hand, thermal decomposition of flavoring compounds dominates the formation of aldehydes during vaping, which includes acetaldehyde and formaldehyde, producing levels that exceed occupational safety standards (Khlystov and Samburova).

Metal Contaminations

Metals are often found in e-liquids in ENDS devices and metal concentration varies from device to device. Increased voltages, resistance, temperature, and/or poor manufacturing practices can contribute to the degradation of heating coils, soldered joints, and other metal elements, releasing metals/metalloids into e-liquid.

In a 2018 study, 56 e-cigarette devices from daily e-cigarette users were sampled from the refilling dispenser, aerosol, and remaining e-liquid in the tank and analyzed in a 15-metal panel. Dramatic increases were observed in aerosol samples for Pb, Zn, Cr, Ni, and Sn concentrations. The result indicated that e-cigarettes are a potential source of chronic exposure to toxic metals (Cr, Ni, and Pb), and to metals that are toxic when inhaled (Mn and Zn) (Olmedo et al.).

Toxicity of E-Cigarette Components and Transformation Products–Solvent Toxicity

Propylene glycol (PG) and glycerol are both airway irritants. There are documented occupational exposures to aerosolized PG in theatrical fogs. Based on analysis of symptoms reported by 218 actors and performers, statistically significant associations between peak PG exposure and reported symptoms was established (Stratton et al.). These symptoms included mucous membrane, throat and nasal irritations, constriction of peripheral airways, etc.

Glycerol, on the other hand, is a natural product found in the human body. It is less irritating to upper respiratory airways than PG. When used as a drug, reported adverse effects following the oral administration of glycerol at unspecified doses included mild headache, dizziness, nausea, vomiting, thirst, and diarrhea. When combusted, glycerol can form thermal decomposition products which could cause adverse health effects (Stratton et al.).

Nicotine Toxicity

By stimulating the sympathetic nervous system, exposure to nicotine results in short-term increases in heart rate, blood pressure, and myocardial contractility. This mechanism has been involved in the short term effect of tobacco nicotine as a trigger for myocardial ischemia and myocardial infarction. On the other hand, there is insufficient evidence that e-cigarette use is associated with long-term changes in heart rate, blood pressure, and cardiac geometry and function. Based on known cardiovascular effects of nicotine, exposure to nicotine from e-cigarettes likely elevates the cardiovascular disease risk in people with preexisting cardiovascular disease(s) (Stratton et al.).

Nicotine-containing e-cigarette aerosols have the potential to adversely impact several host defense mechanisms in the lungs. Exposure to nicotine downregulates α7 nAChR activity, causing impaired mucociliary clearance (MCC), COPD, asthma hyperresponsiveness, and impaired cough. Research showed that nicotine triggers a signaling cascade and disrupts the endothelial cell barriers and cell proliferation, causing lung injury (Stratton et al.).

Flavoring Chemicals Toxicity

With thousands of e-liquid flavors available, flavoring is widely considered to be one of the most attractive characteristics of ENDS products. While the flavoring chemicals in e-liquids are generally recognized as safe (GRAS) for oral consumption, there is no evidence to indicate they are safe to inhale as aerosols. On the contrary, the limited evidence available would suggest that they pose a significant inhalation hazard.

Nearly all aldehydes are respiratory irritants when inhaled at significant concentrations. One study (Erythropel et al.) reported that flavor aldehydes, including benzaldehyde, cinnamaldehyde, citral, ethylvanillin, and vanillin, rapidly reacted with the e-liquid solvent PG, with > 40% of flavor aldehydes converted to flavor aldehyde PG acetals. These acetals showed a 50-80% carryover rate into aerosol, and remained stable with half-lives over 36 hours. Acetals activated aldehyde-sensitive TRPA1 irritant receptors and aldehyde-insensitive TRPV1 irritant receptors *in vitro*.

In buttery or caramel flavored e-liquids, diacetyl and its analogs are commonly found. They are associated with the severe respiratory disease bronchiolitis obliterans (popcorn lung) (Ind), and the inhaled exposure can easily exceed the NIOSH recommended limit.

Recent studies have shown that e-liquids may contain genotoxic/carcinogenic components such as safrole, estragole, 2-furyl methyl ketone, dimethylhydroxyfuranone, and pulegone (Barhdadi et al.). Pulegone, a mint oil extract, has been found in substantial concentration, although the FDA banned the synthetic pulegone as a food additive due to its association with several cancers in rodent models.

Transformation Products Toxicity

Solvents and flavoring chemicals are the main sources of transformation products,

generated from thermal decomposition when heated. Carbonyl compounds, such as formaldehyde, acetaldehyde, acrolein, and glyoxal, which have been found in e-cigarette aerosols, are potentially hazardous and may induce various health effects in users (Stratton et al.). Formaldehyde and benzene are classified as “carcinogens to humans” by the International Agency for Research on Cancer (IARC) and “known to be a human carcinogen” by the US National Toxicology Program (NTP). Exposure to formaldehyde can irritate the skin, throat, lungs, and eyes, while repeated exposure can lead to certain types of cancers ("Formaldehyde"). Exposure to benzene affects the eyes, skin, airway, nervous system, and lungs, and can result in blood cancers such as leukemia ("Benzene"). Exposure to acrolein can result in severe respiratory irritation and contributes to the morbidity and mortality associated with acute lung injury and COPD, and possibly asthma and lung cancer (Bein and Leikauf).

Metal Toxicity

Metal and metalloid concentrations have been measured in e-liquid, aerosols, and biosamples of ENDS users across device systems to evaluate metal/metalloid exposure levels and the potential implications on health outcomes. A systematic review indicated that most metal levels in the biosamples were similar or even higher in ENDS users compared to conventional cigarette users (Zhao et al.). The study supported the evidence that the presence of metals and metalloids (e.g., arsenic, chromium, lead, nickel) in ENDS aerosols is a major concern given their serious health effects, which includes cancer, cardiovascular disease, renal damage, and neurotoxicity. For example, Ni is an established inhalation carcinogen by IARC. Inhaled Ni is rapidly absorbed through the respiratory tract, and the lung is the most sensitive target of Ni toxicity. Ni exposure is related to decreased lung function, lung inflammation, induced rhinitis, sinusitis and allergic dermatitis (Zhao et al.).

Adverse Health Effect of e-Cigarettes

There is a perception that e-cigarettes are less harmful than traditional tobacco cigarettes due to fewer harmful or carcinogenic byproducts in the aerosol than the smoke from cigarette combustion, but the lower degree of harm is not easily quantified. On the other hand, the use of ENDs presents exposure to additional chemicals (solvents, flavorings, metals and their transformation products) and causes other adverse health effects, including respiratory symptoms and damage to cardiovascular, nervous, and immune systems. These adverse effects are also challenging to quantify due to various factors, such as ENDs device design and settings, frequency and duration of use, user lifestyles, and co-use with traditional tobacco and/or cannabis products. This article mainly discusses the adverse health effects on youth and young adults.

EVALI

EVALI (e-cigarette or vaping use-associated lung injury) emerged as a public outbreak in 2019. As of February 2020, the CDC had identified over 2800 cases of EVALI in the U.S., 15%

of which were under 18 years of age and 37% were 18-24 years old ("Outbreak of Lung Injury"). Symptoms of EVALI are characterized by: 1) respiratory symptoms including cough, chest pain, and shortness of breath; 2) gastrointestinal symptoms including abdominal pain, nausea, vomiting, stomach pain, and diarrhea; 3) non-constitutional symptoms including fever, chills, and weight loss ("Information for Healthcare Providers"). Many of these patients required intensive care. Although EVALI was seemingly linked to vitamin E acetate in THC-containing vaping products, e-cigarettes likely pose a host of other pulmonary toxicities. The outbreak highlights the urgent need to understand and regulate the hazards of vaping.

Bronchitic Symptoms, Acute Epiglottitis and Seizure

A 2018 study showed that 67 adolescent e-cigarette-only users had levels of five different volatile organic compounds (VOCs) detected in their urine in quantities up to three times greater than in matched controls (Rubinstein et al.). In addition to EVALI, high exposure to toxins associated with ENDS use had led to multiple adverse health outcomes, including chronic bronchitic symptoms (chronic cough, phlegm, and bronchitis) (McConnell et al.), acute epiglottitis (Bozzella et al.), and seizures (Benowitz). Seizures are a known potential side effect of nicotine toxicity and have been reported after vaping. Between 2010 and 2019, the FDA received reports of 114 cases of seizures in 123 ENDS users, the majority of which were youth or young adults (Faulcon et al.).

Nicotine on Adolescent Brain Development

As the major component of e-liquid, nicotine boosts the level of dopamine in the brain. Dopamine is a neurotransmitter, a type of chemical released by neurons to send signals to other nerve cells. Dopamine is involved in the activation of the brain's pleasure and reward centers. When nicotine is consumed, it artificially boosts dopamine above normal levels, causing an artificially positive feeling. When this effect wears off, users look to regain the same feeling and experience, and this pattern of use and reinforcement can lead to addiction.

During the period of transition from childhood to adulthood, adolescence is a sensitive period for the maturation of brain circuits that regulate cognition and emotion, with resulting vulnerability to the effects of nicotine and tobacco (Yuan et al.). At the adolescent stage, the brain has not completed its maturation and continues to develop until approximately age 25. The prefrontal cortex (PFC), the brain area responsible for executive functions and attention performance, is one of the last brain areas to mature. Smoking during adolescence increases the risk of developing psychiatric disorders and cognitive impairment in later life. In addition, adolescent smokers suffer from attention deficits, the severity of which aggregates with the years of smoking. Studies in rodents reveal the molecular changes induced by adolescent nicotine exposure that alter the functioning of synapses in the PFC and underlie the lasting effects on cognitive function (Goriounova and Mansvelder).

E-Cigarettes on Physical Fitness, Academic Performance and Mental Health of Youth

As demonstrated above, nicotine exposure can impact brain development and affect memory, attention, learning, and impulse control, making ENDS use among adolescents and young adults particularly troubling. This section reviews some of the relevant literature within physical fitness, academic performance and mental health categories.

Physical Fitness

Adverse effects of ENDS on physical performance were evaluated at the U.S. Army Public Health Center on data gathered from the Army Physical Fitness Test (APFT) (Dinkeloo et al.). The APFT is designed to test muscular strength, endurance, and cardiorespiratory fitness and it consists of three events: a timed 2-mile run, 2-minute maximal effort push-ups, and 2-minute maximal effort sit-ups.

A total of 2,854 young men were tested with an average age of 24.6 years among the four groups: current ENDS users, smokers, dual users and a reference group of non-users. The study identified the non-users as the most physically "fit", averaging the best on all 3 fitness tests. ENDs users were found to be slower (mean differences of +27 seconds) on the 2-mile run, performed fewer (-4.56) push-ups, and fewer (-2.01) sit-ups compared to non-users (Fig. 2). The study concluded that individuals who engaged in e-cigarette usage had a noticeably lower level of fitness compared to non-users.

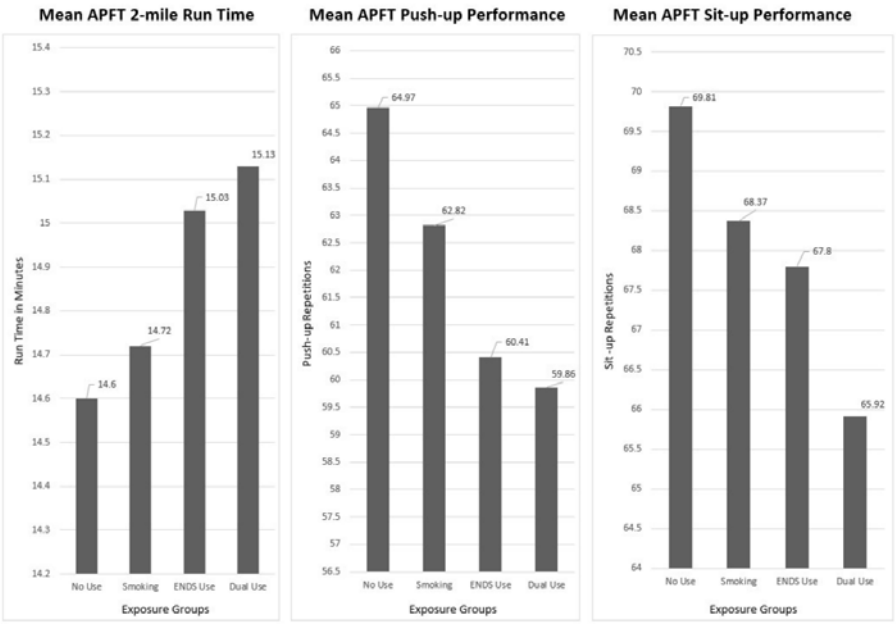


Fig. 2: Mean Army Physical Fitness Test (APFT) performance by exposure groups (n=2,854).

The respiratory impairment in individuals resulting from e-cigarette use was not just confined to this study. As of February 2020, there were 2,807 hospitalized e-cigarette, or vaping, product use-associated lung injury (EVALI) cases or deaths reported ("Outbreak of Lung

Injury"), stemming from the continued usage of vaping products. Regardless of high school demographics, vaping-addicted teenagers were recorded to have a high tendency to miss sport practices in order to fulfill a smoking break and subsequently became quickly fatigued in both practices and competition due to substantial lung damage.

Academic Performance

Cigarette smoking related to lower academic performance has been widely documented. Data from the 2019 national Youth Risk Behavior Survey (YRBS) show that students with higher grades are less likely to use tobacco products compared to students with lower grades ("Tobacco Product"). 26% of US high school students with mostly A's currently used an electronic cigarette or other electronic vapor products (including e-cigars, vape pens, e-hookahs, hookah pens, and mods [such as JUUL, Vuse, MarkTen, and blu], on at least 1 day during the 30 days before the survey), compared to 52% of students with mostly D/F's (Fig. 3). The survey result provided evidence of a significant association between tobacco product use and academic grades.

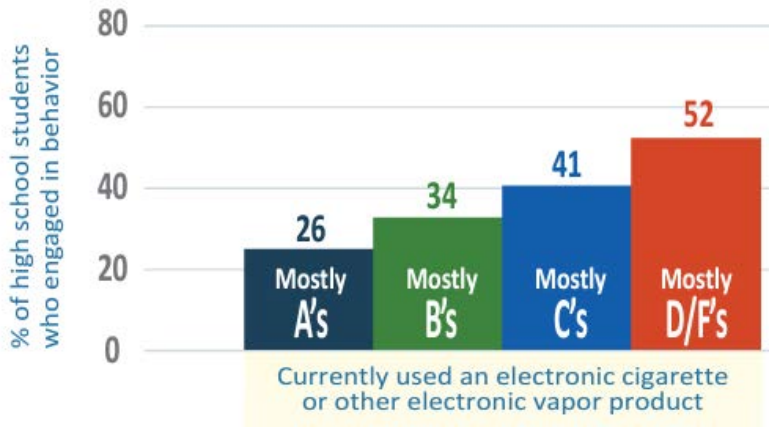


Fig. 3: The percentage of high school students who engaged in electronic cigarette or other electronic vapor products, by type of grades mostly earned in school.

A recent study (Dearfield et al.) examined the longitudinal association of e-cigarette initiation on subsequent academic performance. This study utilized prospective data from the Population Assessment of Tobacco and Health (PATH) study (2014-2018), which includes a national sample of youth, to examine how e-cigarette use initiation is associated with subsequent academic performance, which is an important factor for educational attainment, a well-documented social determinant of health. Youth (12–15 years) of non-users of tobacco products at Wave 2 were included in the analysis (n=4,960). Initiation of e-cigarettes and cigarettes were assessed at Wave 3. Association between e-cigarette and cigarette initiation at Wave 3 and academic performance at Wave 4 was evaluated. Results indicate that e-cigarette use initiation is associated with lower subsequent academic performance. The result reinforces the

importance of reducing all tobacco and ENDS use among youth, which may improve future educational attainment in addition to protecting their health.

A qualitative study conducted by Truth Initiative provided specific examples and anecdotes from teachers' perspective on how vaping impacts student behavior and academics ("A Classroom Crisis"). Truth Initiative convened an online discussion in October 2019 with 28 high school teachers from various regions of the United States gathered for a discussion on vaping's effects within the school environment. Each of the teachers had at least four years of teaching experience and taught at schools with at least 300 students. To minimize any potential bias, each of the teachers were representatives of different age groups that spanned from 25 to 71 years old. Teachers made various comments on their classroom experience in dealing with the presence of vaping across the student body. One of the biggest concerns that they stressed involved the drastic and apparent shift in behavior among many students, especially with those who were seemingly addicted. Half of the surveyed teachers noted that vaping had played a significant role in a student's declining attendance and increasing tardiness. Many were late to, or even neglected class because they had taken the time to vape. Some remarked that students had even resorted to vaping in the classroom. This in turn has led to many classroom disruptions, wasting class time and affecting the learning experiences of other students.

Almost one-third of teachers also warned of vaping's detrimental effects on grades. Teachers observed how nicotine dependency led to mood swings and altered students' academic priorities. Those who vaped identified themselves with other vapers and social groups that continued to have a negative influence on them, leading to widespread apathy for schoolwork.

Mental Health

E-cigarettes have commonly been advertised as an alternative to traditional tobacco products that can mitigate addiction. However, they still contain nicotine, which ultimately fuels addiction, rather than lower it. With nicotine addiction, there can be a variety of social implications that can result from potential withdrawal. Some of these symptoms include irritability, restlessness, general anxiety and depression, and problems in concentrating. Mental health problems, especially concerning depression, continue to persist with e-cigarette usage.

Studies found correlations between impulsivity, low self-esteem, high anxiety and e-cigarette use. Many adolescents turn to vaping as a means to lower stress. In a 2019 study (Jha and Kraguljac), aimed to assess the influence and mental health characteristics of adolescent vapers, 79 students between grades 10-12 in a co-ed high school in Toronto, Canada were surveyed, including both vapers and non-vapers. Survey questions were adapted from the 2019 Environics Research and Health Canada "Vapers Panel Survey", which measured family history of vaping and demographics, vaping habits, factors prompting individuals to vape, peer vaping patterns, and the self-esteem and stress levels of participants, respectively. In the section analyzing the various reasons for these high-schoolers to vape, it was found that the most common cause among participants was for stress relief (70%), followed by peer influence (60%).

The survey also measured stress level using the American Psychological Association

(APA) Perceived Stress Scale, and determined that vapers had higher stress levels than non-vapers. Ironically, those who had sought to lower stress by vaping found themselves more stressed than those who abstained from e-cigarettes. Thus, an important aspect of teen vaping prevention and cessation may be positive mental health initiatives, including support and guidance programs that can provide students with more healthy ways of coping with their stress than vaping.

Conclusion

Vaping is a serious public health issue, as evidenced by substantial research and studies. Although vaping appears to be less harmful than cigarette smoking, as the literatures overwhelmingly indicate, it can be profoundly hazardous and pose adverse effects on many aspects of health, especially for adolescents and young adults, whose bodies and brains are not fully matured to take the stress of nicotine and other toxins. Policy makers, education agencies, communities, and parents should work together to educate and create effective prevention programs and policies that move toward an environment free from vaping and a society in which those in need are helped with welcoming arms.

Acknowledgement

The author would like to thank Professor Madeleine Joullié at the University of Pennsylvania for her guidance and support in constructing this research review.

Works Cited

"A Classroom Crisis: How the Youth Vaping Epidemic Is Impacting Teachers." 2020, doi:truthinitiative.org/research-resources/emerging-tobacco-products/classroom-crisis-how-youth-vaping-epidemic-impacting.

"Benzene and Cancer Risk." 2023, doi:www.cancer.org/cancer/risk-prevention/chemicals/benzene.html.

"Congressional Testimony: An Epidemic Continues: Youth Vaping in America." 2021, doi:www.fda.gov/news-events/congressional-testimony/epidemic-continues-youth-vaping-america-06232021.

"Formaldehyde and Cancer Risk." 2022, doi:www.cancer.org/cancer/risk-prevention/chemicals/formaldehyde.html.

"Information for Healthcare Providers." 2020, doi:www.cdc.gov/tobacco/basic_information/e-cigarettes/severe-lung-disease/healthcare-providers/index.html.

"Outbreak of Lung Injury Associated with the Use of E-Cigarette, or Vaping, Products." 2020, doi:https://www.cdc.gov/tobacco/basic_information/e-cigarettes/severe-lung-disease.html .

"Surgeon General's Advisory on E-Cigarette Use among Youth." 2018, doi:e-cigarettes.surgeongeneral.gov.

"Tobacco Product Use Behaviors and Academic Grades." 2020, doi:www.cdc.gov/healthyschools/health_and_academics/tobacco_product_use.htm.

Barhdadi, Sophia et al. "Toxicity Assessment of Flavour Chemicals Used in E-Cigarettes: Current State and Future Challenges." *Arch Toxicol*, vol. 95, 2021, pp. 2879–81.

Behar, Rachel Z. et al. "Analytical and Toxicological Evaluation of Flavor Chemicals in Electronic Cigarette Refill Fluids." *Scientific Reports*, vol. 8, 2018, p. 8288.

Bein, Kiflai and George D. Leikauf. "Acrolein - a Pulmonary Hazard." *Molecular nutrition & food research*, vol. 55, no. 9, 2011, pp. 1342-60.

Benowitz, Neal L. "Seizures after Vaping Nicotine in Youth: A Canary or a Red Herring?" *Journal of Adolescent Health*, vol. 66, no. 1, 2020, pp. 1-2.

Bonner, Emily et al. "The Chemistry and Toxicology of Vaping." *Pharmacol Ther*, vol. 225, 2021, p. 107837.

Bozzella, Michael J. et al. "Epiglottitis Associated with Intermittent E-Cigarette Use: The Vagaries of Vaping Toxicity." *Pediatrics*, vol. 145, no. 3, 2020, p. e20192399.

Clapp, Phillip W. and Ilona Jaspers. "Electronic Cigarettes: Their Constituents and Potential Links to Asthma." *Curr Allergy Asthma Rep*, vol. 17, no. 11, 2017, p. 79.

Cooper, Maria et al. "E-Cigarette Use among Middle and High School Students - United States, 2022." *MMWR Morb Mortal Wkly Rep*, vol. 71, no. 40, 2022, pp. 1283–85, doi:http://dx.doi.org/10.15585/mmwr.mm7140a3.

Cornelius, Monica E. et al. "Tobacco Product Use among Adults – United States, 2021." *MMWR Morb Mortal Wkly Rep*, vol. 72, no. 18, 2023, pp. 475–83,

doi:http://dx.doi.org/10.15585/mmwr.mm7218a1.
Dearfield, Craig T. et al. "E-Cigarette Initiation Predicts Subsequent Academic Performance among Youth: Results from the Path Study." *Prev Med*, vol. 153, 2021, p. 106781.
Dinkeloo, Eero et al. "Vaping, Smoking, and the Physical Fitness of Active Young Men." *American journal of preventive medicine*, vol. 58, no. 1, 2020, pp. e31-e37.
El-Hellani, Ahmad et al. "Carboxylate Counteranions in Electronic Cigarette Liquids: Influence on Nicotine Emissions." *Chemical Research in Toxicology*, vol. 30, no. 8, 2017, pp. 1577-81.
Erythropel, Hanno C. et al. "Formation of Flavorant-Propylene Glycol Adducts with Novel Toxicological Properties in Chemically Unstable E-Cigarette Liquids." *Nicotine Tob Res*, vol. 21, no. 9, 2019, p. 1248.
Faulcon, Lisa M. et al. "Adverse Experience Reports of Seizures in Youth and Young Adult Electronic Nicotine Delivery Systems Users." *Journal of Adolescent Health*, vol. 66, no. 1, 2020, pp. 15-17.
FDA Commissioner. "Statement from FDA Commissioner Scott Gottlieb, M.D., on New Steps to Address Epidemic of Youth E-Cigarette Use." FDA, September 11 2018. doi:www.fda.gov/news-events/press-announcements/statement-fda-commissioner-scott-gottlieb-md-new-steps-address-epidemic-youth-e-cigarette-use.
Goriounova, Natalia A. and Huibert D. Mansvelder. "Short- and Long-Term Consequences of Nicotine Exposure During Adolescence for Prefrontal Cortex Neuronal Network Function." *Cold Spring Harb Perspect Med*, vol. 2, no. 12, 2012, p. a012120.
Heinrich, Michael et al. "Alkaloids Used as Medicines: Structural Phytochemistry Meets Biodiversity—an Update and Forward Look." *Molecules*, vol. 26, no. 7, 2021, p. 1836.
Ind, Philip W. "E-Cigarette or Vaping Product Use-Associated Lung Injury." *Br J Hosp Med*, vol. 81, no. 4, 2020, pp. 1-9.
Jensen, R. Paul et al. "Solvent Chemistry in the Electronic Cigarette Reaction Vessel." *Sci Rep*, vol. 7, 2017, p. 42549.
Jha, Vedika and Alan Kraguljac. "Assessing the Social Influences, Self-Esteem, and Stress of High School Students Who Vape." *Yale J Biol Med*, vol. 94, no. 1, 2021, pp. 95-106.
Joullié, Madeleine M. "Electronic Cigarettes: The Many Factors Influencing Vaping." unpublished manuscript, 2019.
Khlystov, Andrey and Vera Samburova. "Flavoring Compounds Dominate Toxic Aldehyde Production During E-Cigarette Vaping." *Environ Sci Technol*, vol. 50, no. 23, 2016, pp. 13080-85.
Kimbrough, Doris. "Vaping: What You Need to Know." *ChemMatters*, 2019, pp. 5-8.
McConnell, Rob et al. "Electronic Cigarette Use and Respiratory Symptoms in Adolescents." *Am J Respir Crit Care Med*, vol. 195, no. 8, 2017, pp. 1043-49.
Olmedo, Pablo et al. "Metal Concentrations in E-Cigarette Liquid and Aerosol Samples: The Contribution of Metallic Coils." *Environ Health Perspect*, vol. 126, no. 2, 2018, p. 027010.

O'Connor, Richard et al. "Evolution of Tobacco Products: Recent History and Future Directions." *Tob Control*, vol. 31, 2022, pp. 175-82.

Park-Lee, Eunice et al. "Tobacco Product Use among Middle and High School Students - United States, 2022." *MMWR Morb Mortal Wkly Rep*, vol. 71, no. 45, 2022, pp. 1429–35, doi:http://dx.doi.org/10.15585/mmwr.mm7145a1.

Rubinstein, Mark L. et al. "Adolescent Exposure to Toxic Volatile Organic Chemicals from E-Cigarettes." *Pediatrics*, vol. 141, no. 4, 2018, p. e20173557.

Stratton, Kathleen et al. *Public Health Consequences of E-Cigarettes.* The National Academies Press, 2018.

Wang, Ping et al. "A Device-Independent Evaluation of Carbonyl Emissions from Heated Electronic Cigarette Solvents." *PLoS ONE*, vol. 12, no. 1, 2017, p. e0169811.

Yuan, Menglu et al. "Nicotine and the Adolescent Brain." *J Physiol*, vol. 593, no. 16, 2015, pp. 3397–412.

Zhao, Di et al. "Metal/Metalloid Levels in Electronic Cigarette Liquids, Aerosols, and Human Biosamples: A Systematic Review." *Environ Health Perspect*, vol. 128, no. 3, 2020, p. 036001.

Long-term Health Consequences of Bulimia Nervosa By Catherine Jian Xi Teng and Cindy Xin Wen Zhang

Abstract

Overall, this paper provides a review of the long-term consequences of Bulimia Nervosa. The targeted study group is mostly among teenagers and mainly composed of primary papers or data. As the paper shows, the long term consequences of bulimia are related to gastrointestinal, cardiovascular, and mental illness. Most of the consequences may lead to extreme health problems and even will increase the mortality rate. This paper hopes to provide some effective information about bulimia and its effects for reference.

1. Introduction

Bulimia is one of three main ED (eating disorders) and it presents as binge eating, eating a large amount of food due to loss of control during eating episodes. Binge eating episodes are followed by self-induced vomiting referred to as purging. Bulimia differs from the other two main types of eating disorders, anorexia and binge eating. Anorexia is characterized by restrictive eating and the patients will have a below normal range of weight. Binge eating is characterised by eating excessive food. Foods individuals with bulimia" consume are often high in calories, fat and sugar content (Nitsch,A., et al, 2021). Patients' self-purging behaviour is caused by their guilt of consuming too much and the guilt from eating too many calories from overeating which will cause excessive weight gain (Engel S, et al., 2018). Episodes are followed by negative emotional states. Overall, a patient's persistent and excessive obsession with their body weight and appearance is the main trigger of bulimia (Harrington, B. C. et al, 2015). The diagnostic and statistical manual of mental disorders, edition 5 (DSM 5) classifies bulimia as binge eating plus purging behaviours occurring in combination at least twice a week for 3 months.The importance of this recognition of bulimia as a distinct ED is that individuals with bulimia can be at a significant risk for morbidity and mortality secondary to the compensatory behaviours they use to prevent weight gain, as well as from associated comorbid psychiatric conditions (Castillo, M., & Weiselberg, E. 2017). Patients with bulimia have average to above-average body weights, so they can often go unnoticed by their relatives, friends, and doctors. Individuals may be ashamed to share their symptoms due to secrecy and shame over the binge or purge behaviours, so that is why many patients are not diagnosed and treated in time. According to a survey with 52 women with eating disorders. They completed six questionnaires related to shame, anxiety, and depression,and the results from the survey is that a high amount of shame was observed. Additionally, many patients reported a lot of ambivalence emotions which makes them isolate themselves unwilling to seek assistance (Keith, L., Gillanders, D., & Simpson, S. 2009). According to community based studies, the prevalence of bulimia nervosa is between 0.5% and 1.0%, regardless of social class distribution (Hudson, J. I.,et al, 2007). The prevalence of eating disorders such as bulimia nervosa is lower in non-industrialised populations and varies across ethnic groups (Hay, P. J., & Bacaltchuk, J.,2001). African-American women

have a lower rate of restrictive dieting than white American women but have a similar rate of recurrent binge eating. Bulimia is a modern disease that arose after people had an abundance of food sources. However, in regions where there is food insecurity/ hunger and malnutrition, people prioritize eating above social standards of beauty. Therefore, bulimia is a disorder of modern times that emerged as a result of increased food availability. People will pay more attention to their bodies, appearances in locations with high per capita living standards, and relatively abundant material resources. In these locations, food availability is more abundant and considerably less crucial due to social media influence and other factors. Moreover, the recovery of bulimia nervosa is also an important process.The treatment is divided into two types - cognitive behavioural therapy (CBT) and interpersonal therapy (IPT).

The previous studies suggest that CBT was significantly more effective than IPT treatment in the percentage of participants recovered (29% vs 6%), the percentage remitted (48% vs 28%), and the percentage meeting community norms for eating attitudes and behaviours (41% vs 27%). For treatment completers, the percentage recovered was 45% for CBT and 8% for IPT. The finding is that CBT was substantially more efficient than IPT at gaining improvements in bulimia nervosa patients. As a result, CBT may be the referred as the best psychotherapeutic approach for treating bulimia nervosa (Agras, W. S., et al ,2000)

First, cognitive behavioural therapy (CBT), interpersonal therapy (IPT), dialectical behavioural therapy (DPT), and selective serotonin reuptake inhibitors (SSRIs) are among the treatments available to adults suffering from bulimia. In CBT, the therapist directs the procedure and gives patients advice on creating efficient coping mechanisms for a range of problems, including eating disorders, anxiety, and depression (Murphy, R.,et al, 2010). For instance, therapists may instruct patients to modify how they perceive their bodies. The main objective of IPT is to improve a client's social life and ideas in interpersonal relationships in order to help lessen overall suffering and anxiety. IPT offers bulimia patients advice on how to stop caring about what other people think of their body image (Karam, A. M., et al, 2019).
Cognitive-behavioral strategies are combined with mindfulness exercises in DPT. It is now the most complete and scientifically supported affect regulation therapy for bulimia. This treatment was created by Linehan to treat borderline personality disorder, and DPT was mainly designed to impart emotional control skills to lower binge eating and purging rates (Safer DL, et al, 2001).

Moreover, DBT assists people in controlling emotional dysregulation and improving their coping methods for handling stress and unpleasant emotions, which lessens the need for binge eating and purging. Last but not least, SSRIs, which are commonly prescribed antidepressants, may be given to bulimics to assist control co-occurring mood disorders such anxiety and depression. Serotonin– a brain chemical that influences how signals are sent from the brain to the body and also has an impact on an individual's emotions. Availability in the brain is increased by SSRIs, helping to stabilize mood and lessen binge eating episodes(Lochmann, D., & Richardson, T. ,2019).Even though all of these treatments are effective for bulimia, patients still suffer long term and permanent consequences even when bulimia is treated. As such, this literature review will analyze the long term health consequences of bulimia in major organ systems impacted by

this eating disorder. This paper hypothesizes that bulimia will affect gastrointestinal, cardiac, and neurobiology systems the most.

2. Methods

This paper provides an analysis and collection of current research in eating disorders adolescents and young adults up to 30 years old. Primary papers and data were collected from NIH–The National Library of Medicine database on PubMed and Google Scholars. Moreover, both in terms of bulimia treatment and research, the development of modern medicine has greatly advanced in order to provide us with more data and study. This article focuses on a group of studies that have been published in English only over the last ten years that examine the effects and variations of bulimia and other eating disorders.

3. Results

3.1 Gastrointestinal System

According to a study by Santonicola, A, et al, 2012, individuals with bulimia and EDNOS (eating disorders not otherwise specified) frequently experience the feeling of fullness, upper stomach pressure, and nausea. The enormous quantity of food eaten during a binge in Bulimia patients causes them to feel as though they have lost control and at times, excruciating pain in their stomachs. Self-induced vomiting results in the latter [Stacher G. 2003].

Gastroesophageal reflux disease (GERD) is a long term gastrointestinal consequence of bulimia. According to the Montreal definition, GERD is a condition of troublesome symptoms and complications that result from the reflux of stomach contents into the esophagus. Self inducing vomiting may increase the possibility of this disease. Heartburn and acid reflux in the upper throat region when vomiting are the most typical signs of GERD. People who exhibit specific symptoms are frequently treated without having any additional testing done. Long-term GERD can cause esophageal inflammation and over time, lower esophageal cells mutate due to consistent exposure to acid.

Similar to GERD, Barrett's esophagus is also caused by repeated and prolonged vomiting and the destruction of esophageal tissue. Barrett's esophagus is a condition whereby the lining of the lower esophagus is replaced with intestinal lining, and it results directly from the vomiting that occurs in bulimia and is premalignant, which means that it could develop into cancer later. (Gregson, E., et al ,2016). Both long-term GERD and Barrett's esophagus increase risk of esophageal cancer. These mutations accumulate over time to severely raise the likelihood of developing esophageal cancer which may be fatal for the patients [Maret-Ouda, J., Markar, S. R., & Lagergren, J., 2020]It can take several years for Barrett's esophagus to turn into cancer; however, it is the most common cause/form of esophageal cancer.

The stated examples have some impact with damage caused by vomiting and also the long term effects such as increased risk of gastrointestinal cancer, and digestive problems such as insufficient nutrient absorption (Saunders, J., & Smith, T., 2010). According to figures from 2020, there were 1,806,590 new instances of cancer and 606,520 cancer deaths in the United

States, with gastrointestinal cancer ranking as one of the nation's most prevalent causes of death. The most common and deadly gastrointestinal cancer is colorectal cancer (CRC). Endoscopy is an effective examination technique for treating gastrointestinal cancer by detecting and treating early precursor lesions and diagnosing malignancies quickly. Endoscopic white light imaging allows for the real-time observation of the mucosal morphology of the tissues lining the digestive system, and tissue samples from highly questionable lesions can be collected to assist with the diagnosis. (Yao K., et al, 2020)

3.2 Malnutrition

Through force-vomiting behaviours, patients purge out food which contain various sources of necessary nutrients. This behaviour prevents patients from getting enough nutrients, leading to malnutrition. Malnutrition can have many long-term consequences for the gastrointestinal system by decreasing its ability to function. For example, chronic malnutrition affects pancreatic exocrine function which causes decreased hormone secretions that help humans digest and absorb food.Chronic malnutrition results in changes in pancreatic exocrine function, intestinal blood flow, villous architecture and nutrient absorption. The colon cannot reabsorb water and electrolytes, and the small and large bowels secrete ions and fluids. Diarrhoea could arise from this, and patients who are severely malnourished have a rather high fatality risk when they experience it (Saunders, J., & Smith, T., 2010).

3.3 Dental Health

Another example of long term bulimia consequences related to persistent vomiting is dental problems. Bulimia patients who often binge-purge will experience dental erosion, reduced salivary flow rate, tooth sensitivity, and dry mouth (xerostomia) (Valena V, Young WG., 2000). Dental erosions often appear on the surface of the maxillary teeth. Dental erosion and sensitivity may lead to teeth decay which make patients' eating process and the use of their teeth extremely difficult. Moreover, the consequences of frequent dry mouth are continuous sore throat, difficulty eating or swallowing, throat hoarseness, and increased thirst, which will somewhat affect patients' daily life routine.

The duration of the self-inducing vomiting process, the types of food consumed, oral cleanliness, the frequency of vomiting, and the baseline quality of the tooth structure may all have an influence on the pace and severity of erosions. (Mehler P. S. ,2011).

3.4 Cardiovascular Disease

Malnutrition caused by self purging behaviour will result in an imbalance of fluid and electrolytes, leading to an irregular heartbeat. This may in turn lead to more serious heart conditions such as heart attacks or arrhythmias (Bierma, P. 2022). Electrolyte imbalances such as hypokalemia–the deficiency of potassium in the blood can spur cardiac arrhythmias, including ventricular arrhythmias such as ventricular tachycardia or ventricular fibrillation. The patients' lives could be at danger due to these heart rhythm anomalies (James. L. L. 2023).

In addition, Mitral valve prolapse is a condition where the heart's mitral valve doesn't close properly, leading to blood leakage back into the left atrium. Although the exact relationship between bulimia and mitral valve prolapse is not fully understood, there is evidence by(Tith RM, et al., 2020) suggesting a higher prevalence of mitral valve prolapse among individuals with bulimia. The majority of cardiovascular illnesses were linked to bulimia nervosa. When compared to women hospitalized for pregnancy-related events, Mitral valve may lead to high risk of women with bulimia with atherosclerosis–the buildup of fats, and other substances in and on the artery walls that was 6.94 (95%), 2.99 (95%) times the risk of conduction disorders, and 5.48 (95%) times the risk of myocardial- of or relating to the muscular tissue of the heart infarction which are dangerous for people especially women (Tith RM, et al., 2020).

Moreover, bulimia nervosa may cause endocrine abnormalities or imbalances that further alter metabolic profiles(Warren MP, 2011). Menstrual irregularities may be present in bulimic individuals, but the prevalence varies greatly (94–97), from 37 to 64%, with amenorrhea(absence of normal menstrual flow) occurring in 5–40%. Patients may have anovulatory syndrome and irregular bleeding while secreting an appropriate amount of estrogen, which might explain this variability. Low estrogen levels are frequent in bulimic women, and these levels may eventually have a deleterious impact on cardiovascular health. Cardiac muscle and arteries may become stiffer and less elastic as a result of low estrogen levels. blood pressure may rise as a result, damaging patients' blood vessels and raising their risk of heart attack, stroke, and heart failure which often happens in prolonged bulimia patients(Tith RM, et al., 2020).

3.5 Mental illness:

The binge-purge cycle seems to be exacerbated by the shame, resentment, helplessness, and dissatisfaction with one's appearance that many people with eating disorders experience [Susan Cowden, M. 2020]. A person may experience additional anxiety and stress due to the strain of keeping their illness a secret. Some common destructive consequences are depression, anxiety, and body-dysmorphia although an additional scenario exists whereby these mental illnesses may have already been present prior to developing bulimia.

The DSM 5 characterizes body dysmorphic disorder (BDD), sometimes referred to as dysmorphophobia, as an obsession with a perceived physical deficiency that is either not obvious to others or is just marginally perceptible to them (Foster, A. C., et al, 2015). In other words, a person’s internal view of their face and body does not correspond or match with what they look like. The obsession is intense enough to noticeably affect functioning in social, physical, professional, or other domains. In order to fulfill diagnostic criteria, a person must exhibit repeated behaviors at some time throughout the course of the illness, such as frequent mirror-checking, weight checking etc. The original cause of bulimia is self esteem and body image problems, indeed this may last longer than the bulimia itself leading to continual body dysmorphia as a disease even longer than bulimia. They often have an extreme attention about their body images, which might lead to the increasing possibility of diagnosing body dysmorphia among bulimia patients (Nicewicz, H. R., & Boutrouille, J. F. 2022). Many BDD sufferers

experience depression. According to studies conducted by Phillips K. A., between 24 and 28 percent of people with body dysmorphia disorder attempt suicide. Suicidal thoughts and attempts are linked to a considerable rise in psychosocial dysfunction and symptom severity. Comorbid BD, inability to control drug use, major depressive disorder, post-traumatic stress disorder, and personality disorders, particularly borderline personality disorder, are risk factors for suicidal thoughts and attempts in BDD (Phillips K. A., 2007). An observational study (n = 200) found 94.6% of adolescents with BDD have reported distress and severe impairment socially, academically, and psychologically and 44.4% had suicide attempts(Phillips, K. A, et al, 2006). The consequences of body dysmorphic disorder can be extremely severe in all perspectives and fatal especially for adolescents.

Female adolescents and young adults are particularly susceptible to developing anxiety disorders, depression, and eating disorders. Study shows that the global prevalence of major depressive disorder was 5.8% in women and 3.5% in men. The gender difference in depression – generally believed to be twice as many females experiencing major depression as males(Ferrari et al., 2013). The degree of anxiety of bulimia patients is caused by the degree of their thoughts about their body image or severity of their overweight ideas. According to a study by Sander, J., et al, 2021) there was evidence of increased anxiety or depression in 18.4% of people. According to the data table indicated in this study, anxiety was suggested by 16.6%. 69% of those who supported the comorbidity also stated that the anxiety problem occurred before the ED commenced. (Swinbourne J.m et al, 2012).Although anxiety was high in both groups, the overweight group reported much greater rates of anxiety than the healthy weight group did, according to the authors (Balantekin KN, et al, 2021). Negative consequences of anxiety can be suicidal possibilities. According to a previous study,regardless of the stage of the disease or whether they are receiving curative or palliative care, anxiety affects up to 20% and 10% of cancer patients (Mitchell AJ, et al, 2011).

Participants also expressed higher perfectionism and mood dysregulation as well as generally low self-esteem. Their depression is often caused by the sense of losing control of their eating or their isolation from assistance. Severe depression is a risk factor for suicide which can cause a negative consequences for the child and their whole family (Sander, J., et al, 2021). The development and maintenance of depression are influenced by emotion regulation. Theoretical models of eating disorders assume that unhealthy emotion regulation techniques, including disordered eating, are encouraged to downregulate emotions and raise the likelihood of developing threshold eating disorders. Overall, the connection between eating disorders and mental illness like depression is very close and mutual (Fairburn C.G., et al, 2003).

There are several other mental illnesses that are comorbid such as personality disorders, bipolar disorder, schizophrenia that are not as common.

4. Discussion

Based on the information from this paper, the results from the review indicate that there are serious, long-term health consequences of bulimia nervosa that persist long after bulimia has

been treated. All the long-term effects listed in the essay will indirectly affect bulimia patients' physical health issues or even increase their mortality rates. Based on the data or the resources that provided, bulimia will lead to a variety of diseases. And then those diseases will slowly evolve into a variety of complications, thus devastating the patient's mental health and physical health in a variety of aspects.

The major health problems that arise as a consequence of bulimia are related to nutrition as per the nature of the disease. With bulimia in particular, chronic vomiting is the main cause of latter disease development as it causes stomach and intestinal damage and also ion imbalance. The chronic vomiting may also increase the risk of dental erosion due to the stomach acid or the increasing risk of cardiovascular disease such as heart failure which will lead to other negative physical consequences. However,The psychiatric component is nuanced. The psychological contributors that may have caused bulimia may persist beyond the treatment of bulimia and further, mental illness may arise during the time the patient suffers from bulimia and continue as well.

This review was necessary because many studies focus only on the disorder as it is happening and immediately after. However, these consequences stay with the patients for a long time if not forever and it is important to investigate how patients may have a decreased quality of life following the disorder. A limitation of this review is that we did not consider those that are at high risk for bulimia but are not yet fully diagnosed. Another limitation is that this paper only surveyed forty numbers of papers which might not include the complete information, not every single paper in the literature. Papers past the year 2000 were analyzed and only papers written in english. Moreover, not every single organ system was investigated, only the most commonly affected systems were analyzed in this research paper.

Longitudinal studies; For future studies, a longitudinal cohort of individuals who are at high risk for developing bulimia would be interesting to study. The study would follow these patients and observe whether or not they develop bulimia and what characteristics they possess that put them at high risk. Further, the study would follow these patients beyond their bulimia treatment to evaluate how the disease impacted their health in the long term.

Remove the above paragraph and insert concluding statement about your findings: The long term health consequences endured by bulimia patients are a result of their binge eating and vomiting behaviour. The physiological consequences in most instances are a direct result of chronic vomiting. In this way, treating this symptom during bulimia more heavily will yield preventative benefits for long term health beyond bulimia treatment.

Works Cited

Santonicola, A., Gagliardi, M., Guarino, M. P. L., Siniscalchi, M., Ciacci, C., & Iovino, P. (2019). Eating Disorders and Gastrointestinal Diseases. *Nutrients*, *11*(12), 3038. https://doi.org/10.3390/nu11123038

Santonicola, A., Siniscalchi, M., Capone, P., Gallotta, S., Ciacci, C., & Iovino, P. (2012). Prevalence of functional dyspepsia and its subgroups in patients with eating disorders. *World journal of gastroenterology*, *18*(32), 4379–4385. https://doi.org/10.3748/wjg.v18.i32.4379

Stacher G. (2003). Gut function in anorexia nervosa and bulimia nervosa. *Scandinavian journal of gastroenterology*, *38*(6), 573–587. https://doi.org/10.1080/00365520310002120

Castillo, M., & Weiselberg, E. (2017). Bulimia Nervosa/Purging Disorder. *Current problems in pediatric and adolescent health care*, *47*(4), 85–94. https://doi.org/10.1016/j.cppeds.2017.02.004

Saunders, J., & Smith, T. (2010). Malnutrition: causes and consequences. *Clinical medicine (London, England)*, *10*(6), 624–627. https://doi.org/10.7861/clinmedicine.10-6-624

Valena, V., & Young, W. G. (2002). Dental erosion patterns from intrinsic acid regurgitation and vomiting. *Australian dental journal*, *47*(2), 106–115. https://doi.org/10.1111/j.1834-7819.2002.tb00313.x

Mehler P. S. (2011). Medical complications of bulimia nervosa and their treatments. *The International journal of eating disorders*, *44*(2), 95–104. https://doi.org/10.1002/eat.20825

Maret-Ouda, J., Markar, S. R., & Lagergren, J. (2020). Gastroesophageal Reflux Disease. *JAMA*, *324*(24), 2565. https://doi.org/10.1001/jama.2020.21573

Castillo, M., & Weiselberg, E. (2017). Bulimia Nervosa/Purging Disorder. *Current problems in pediatric and adolescent health care*, *47*(4), 85–94. https://doi.org/10.1016/j.cppeds.2017.02.004

Keith, L., Gillanders, D., & Simpson, S. (2009). An exploration of the main sources of shame in an eating-disordered population. *Clinical psychology & psychotherapy*, *16*(4), 317–327. https://doi.org/10.1002/cpp.629

Hay, P. J., & Bacaltchuk, J. (2001, July 7). *Bulimia nervosa*. The BMJ. https://www.bmj.com/content/323/7303/33

Nitsch, A., Dlugosz, H., Gibson, D., & Mehler, P. S. (2021). Medical complications of bulimia nervosa. *Cleveland Clinic journal of medicine*, *88*(6), 333–343. https://doi.org/10.3949/ccjm.88a.20168

Hudson, J. I., Hiripi, E., Pope, H. G., Jr, & Kessler, R. C. (2007). The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication. *Biological psychiatry*, *61*(3), 348–358. https://doi.org/10.1016/j.biopsych.2006.03.040

Bierma, P. (2022, May 9). *Bulimia nervosa*. HealthDay. https://consumer.healthday.com/encyclopedia/children-s-health-10/kids-ailments-health-news-434/bulimia-nervosa-644424.html

III, James. L. L. (2023, August 10). *Hypokalemia - endocrine and metabolic disorders*. Merck Manuals Professional Edition. https://www.merckmanuals.com/professional/endocrine-and-metabolic-disorders/electrolyte-disorders/hypokalemia#:~:text=Hypokalemia%20may%20cause%20premature%20ventricular%20beats%20and%20premature,increasingly%20severe%20hypokalemia%3B%20eventually%2C%20ventricular%20fibrillation%20may%20occur.

Sander, J., Moessner, M., & Bauer, S. (2021). Depression, Anxiety and Eating Disorder-Related Impairment: Moderators in Female Adolescents and Young Adults. *International journal of environmental research and public health*, *18*(5), 2779. https://doi.org/10.3390/ijerph18052779

Susan Cowden, M. (2020, May 19). Understanding the binge-purge cycle in bulimia. Verywell Mind. https://www.verywellmind.com/the-binge-purge-cycle-1138380

Nicewicz, H. R., & Boutrouille, J. F. (2022). Body Dysmorphic Disorder. In *StatPearls*. StatPearls Publishing.

Engel, S. (2023, March 28). *Bulimia nervosa in adults: Clinical features, course of illness, assessment, and diagnosis*. UpToDate. https://www.uptodate.com/contents/bulimia-nervosa-in-adults-clinical-features-course-of-illness-assessment-and-diagnosis?search=Bulimia+nervosa+in+adults%3A+Clinical+features%2C+course+of+illness%2C+assessment%2C+and+diagnosis.&source=search_result&selectedTitle=1~150&usage_type=default&display_rank=1#H191477011

Agras, W. S., Walsh, T., Fairburn, C. G., Wilson, G. T., & Kraemer, H. C. (2000). A multicenter comparison of cognitive-behavioral therapy and interpersonal psychotherapy for bulimia nervosa. *Archives of general psychiatry*, *57*(5), 459–466. https://doi.org/10.1001/archpsyc.57.5.459

Murphy, R., Straebler, S., Cooper, Z., & Fairburn, C. G. (2010). Cognitive behavioral therapy for eating disorders. *The Psychiatric clinics of North America*, *33*(3), 611–627. https://doi.org/10.1016/j.psc.2010.04.004

Karam, A. M., Fitzsimmons-Craft, E. E., Tanofsky-Kraff, M., & Wilfley, D. E. (2019). Interpersonal Psychotherapy and the Treatment of Eating Disorders. *The Psychiatric clinics of North America*, *42*(2), 205–218. https://doi.org/10.1016/j.psc.2019.01.003

Safer, D. L., Telch, C. F., & Agras, W. S. (2001). Dialectical behavior therapy for bulimia nervosa. *The American journal of psychiatry*, *158*(4), 632–634. https://doi.org/10.1176/appi.ajp.158.4.632

Vakil N, van Zanten SV, Kahrilas P, Dent J, Jones R Global Consensus Group. The Montreal definition and classification of gastroesophageal reflux disease: a global evidence-based consensus. *Am J Gastroenterol.* 2006;101:1900–1920. [PubMed] [Google Scholar]

Gregson, E., Bornschein, J. & Fitzgerald, R. Genetic progression of Barrett's oesophagus to oesophageal adenocarcinoma. *Br J Cancer* 115, 403–410 (2016). https://doi.org/10.1038/bjc.2016.219

Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin.* 2020;70:7–30.

Yao K, Uedo N, Kamada T, Hirasawa T, Nagahama T, Yoshinaga S, et al. Guidelines for endoscopic diagnosis of early gastric cancer. Dig Endosc. 2020;32(5):663–98.

Swinbourne J, Hunt C, Abbott M, Russell J, St Clare T, Touyz S. The comorbidity between eating disorders and anxiety disorders: Prevalence in an eating disorder sample and anxiety disorder sample. Aust N Z J Psychiatry. 2012;46(2):118–31.

Balantekin KN, Grammer AC, Fitzsimmons-Craft EE, Eichen DE, Graham AK, Monterubio GE, et al. Overweight and obesity are associated with increased eating disorder correlates and general psychopathology in university women with eating disorders. Eat Behav. 2021;41:101482.

Mitchell AJ, Chan M, Bhatti H, et al. Prevalence of depression, anxiety, and adjustment disorder in oncological, haematological, and palliative-care settings: a metaanalysis of 94 interview-based studies. Lancet Oncol 2011;12:160-74. PubMed doi:10.1016/S1470-2045(11)70002-X

Fairburn C.G., Cooper Z., Shafran R. Cognitive behaviour therapy for eating disorders: A "transdiagnostic" theory and treatment. *Behav. Res. Ther.* 2003;41:509–528. doi: 10.1016/S0005-7967(02)00088-8.

Phillips K. A. (2007). Suicidality in Body Dysmorphic Disorder. *Primary psychiatry*, *14*(12), 58–66.

Phillips, K. A., Didie, E. R., Menard, W., Pagano, M. E., Fay, C., & Weisberg, R. B. (2006). Clinical features of body dysmorphic disorder in adolescents and adults. *Psychiatry research*, *141*(3), 305–314. https://doi.org/10.1016/j.psychres.2005.09.014

Warren MP. Endocrine manifestations of eating disorders. *J Clin Endocrinol Metab*. 2011;96(2):333-343. doi:10.1210/jc.2009-2304

94 Pirke KM , Fichter MM , Chlond C , Schweiger U , Laessle RG , Schwingenschloegel M , Hoehl C 1987 Disturbances of the menstrual cycle in bulimia nervosa. *Clin Endocrinol (Oxf)* 27:245–251

95 Cantopher T , Evans C , Lacey JH , Pearce JM 1988 Menstrual and ovulatory disturbance in bulimia. *BMJ* 297:836–837

96 Gendall KA , Bulik CM , Joyce PR , McIntosh VV , Carter FA 2000 Menstrual cycle irregularity in bulimia nervosa. Associated factors and changes with treatment. *J Psychosom Res* 49:409–415

97 Vyver E , Steinegger C , Katzman DK 2008 Eating disorders and menstrual dysfunction in adolescents. *Ann NY Acad Sci* 1135:253–264

Tith RM, Paradis G, Potter BJ, et al. Association of Bulimia Nervosa With Long-term Risk of Cardiovascular Disease and Mortality Among Women. *JAMA Psychiatry.* 2020;77(1):44–51. doi:10.1001/jamapsychiatry.2019.2914

Lochmann, D., & Richardson, T. (2019). Selective Serotonin Reuptake Inhibitors. *Handbook of experimental pharmacology*, *250*, 135–144. https://doi.org/10.1007/164_2018_172

Harrington, B. C., Jimerson, M., Haxton, C., & Jimerson, D. C. (2015). Initial evaluation, diagnosis, and treatment of anorexia nervosa and bulimia nervosa. *American family physician*, *91*(1), 46–52.

Foster, A. C., Shorter, G. W., & Griffiths, M. D. (2015). Muscle dysmorphia: could it be classified as an addiction to body image?. *Journal of behavioral addictions*, *4*(1), 1–5. https://doi.org/10.1556/JBA.3.2014.001

Ferrari, A. J., Somerville, A. J., Baxter, A. J., Norman, R., Patten, S. B., Vos, T., & Whiteford, H. A. (2013). Global variation in the prevalence and incidence of major depressive disorder: a systematic review of the epidemiological literature. *Psychological medicine*, *43*(3), 471–481. https://doi.org/10.1017/S0033291712001511

The Impact of Data and Analytics on Strategic Decision-Making in the National Basketball Association By Sarp Kologlu

Abstract

Analytics have had a profound impact on strategy and decision-making in sports. Through advanced statistical analysis, teams gain valuable insights into player performance, enabling more informed choices in scouting, drafting, and player signings. Analytics aid in identifying player strengths, weaknesses, and potential, while also assessing their value for free agency and trade negotiations. In-game strategies benefit from real-time data, allowing coaches to optimize lineups, adjust defensive schemes, and exploit opponents' weaknesses. This paper seeks to uncover the uses and impact of data and analytics in the NBA. I will examine the impact of analytics on scouting, drafting, signings, and trades by using quantitative and qualitative data from the NBA. The findings of this paper can impact current decision making in the NBA as well as support future analytical advancements.

Introduction

With the development of data and technology, the world of sports has seen a spectacular revolution, and the National Basketball Association (NBA) is no exception. How the game is played, controlled, and understood has transformed in recent years due to the growing significance of data analytics. The days of merely relying on instinct and expertise to make basketball judgments are long gone. In order to acquire a competitive advantage and make strategic decisions, the NBA now largely relies on data-driven insights. From player evaluation and performance improvement to team tactics and fan engagement, analytics have a significant impact on basketball in the NBA. Advanced measurements and statistical analysis have changed the game, altering the professional basketball environment and taking the game to new heights.

Teams now evaluate players' performance and potential using analytics. Advanced analytics include: Player Efficiency Rating (PER), Box Plus-Minus (BPM), and True Shooting Percentage (TS%) to determine a player's influence on both ends of the court rather than just traditional figures like points per game and rebounds. These measures offer a more thorough view of a player's contributions and aid in the discovery of underappreciated skills. Data analytics has dramatically changed NBA player shot selection. Teams are now more concerned with creating good shot opportunities, prioritizing three-pointers and shots close to the rim over jumpers from a distance. NBA game data supports that possessions with a three-point attempt or shot in the restricted area result in higher predicted points than those with other shot selections, thus supporting this strategic transition.

In addition, real-time data during games is becoming increasingly important for coaches and players to make educated judgments. Teams can evaluate the habits and patterns of their opponents using tracking technologies and advanced analytics, and then modify their game plans accordingly. This data-driven strategy improves in-game adjustments and has the potential to make the difference between success and failure.

The fan experience has also been improved using analytics. Even casual sports fans now have access to a multitude of statistics, player insights, and interactive material thanks to the development of sports analytics platforms and data visualization tools, such as Basketball Reference. This more intense interaction with the data enhances fans' relationships with their favorite clubs and players and creates a greater comprehension of the game.

Every NBA team makes use of data analytics. In fact, every single NBA team has a dedicated analytics department, a drastic change from just ten years ago, demonstrating the growing importance of data-driven decision-making. Further evidence of analytics' influence may be seen in the recent rise in three-point attempts. Teams attempted 18.1 three-pointers per game on average in the 2010–2011 season. Fast-forward to the 2020–2021 season, and that figure has climbed to a record–breaking 34.8 three-point attempts per game,[24] demonstrating the influence of data–driven shot selection techniques. The growing importance of data and technology in the NBA has undeniably transformed the way basketball is played and managed. Analytics has become a crucial tool in evaluating players, devising strategies, and engaging fans. As the NBA continues to evolve, data-driven insights will remain at the forefront of the league, propelling it into a new era of precision and competitiveness.

This paper seeks to analyze the impact of data and analytics on strategic decision-making in the NBA by looking at drafting and scouting, roster construction and management, and in-game strategy. To analyze this impact, I will be evaluating player scouting reports, contract data, and in-game data from recent NBA seasons. The analysis will be primarily qualitative, with some quantitative analysis to support certain points.

Drafting and Scouting

Analyzing the Top Picks of the 2016 NBA Entry Draft

The players drafted in the 2016 NBA Entry Draft combined to make an impressive class, with several top-tier players making an impact in the league. I evaluated the top 10 picks of the 2016 NBA Draft and how analytics might have been used to evaluate their strengths and weaknesses. The full player analyses are in the appendix, with highlights and key findings explained below.[25]

The top ten picks had a really intriguing blend of raw prospects with high potential, such as Jaylen Brown, Dragan Bender, Thon Maker, and Ben Simmons to some extent, and established players who could make an immediate impact, such as Jamal Murray, Buddy Hield, and Jakob Poeltl. There was also a good mix of players who succeeded and flopped. With this, we can see the drawbacks and advantages of drafting people based on potential against drafting an already established player who may not be the most freakish athlete. Murray, Poeltl, and Hield are all above average players in the NBA but have never been all stars; however, Simmons,

[24] *Basketball-Reference*. (n.d.). Basketball-Reference. Retrieved September 3, 2023, from https://www.basketball-reference.com

[25] Givony, J. (n.d.). *DraftExpress*. Retrieved September 3, 2023, from https://www.draftexpress.com/

Brown, and Ingram have all been all-stars. The risk of drafting off of potential is that the player has a much greater chance to be a bust, like Bender, Maker, and Dunn.

Looking at some of the raw prospects,[26] although Ben Simmons was injured for his entire first year, he had a very strong bounce-back year averaging 16 points, 8 assists, and 8 rebounds (16-8-8) with almost 2 steals and a block each game and winning rookie of the year. However, his jump shooting concerns never went away as he did not make a single three pointer in his second year, and he is a 60% career free throw shooter, extremely bad for the point guard (PG) position. Another prospect who had an interesting start to his NBA career was Brandon Ingram. He had an abysmal start to his NBA career, averaging 9-4-2 shooting 40% from the field, 29% from three, and 62% from the line. However, throughout his next two seasons he was able to come back and average 17-5-4 on a much improved 48% from the field and 36% from three, although his free throw shooting still needed some work (68%). With this improvement Ingram was quickly becoming one of the best young players in the league; however, that offseason Ingram was a key asset in the blockbuster trade that sent Anthony Davis to the Lakers. With more freedom to run the offense, Ingram broke out, averaging 24-6-4 with a steal and 46% from the field, 39% from three, and 85% from the line, as well as winning the Most Improved Player award and becoming a first time all star. Jaylen Brown's start was also quite rocky, averaging 7-3-8, shooting 45% from the field and 34% from three, along with an awful 69% from the free throw line (45/34/69). Additionally, Brown was a horrendous decision maker with the ball, averaging more turnovers than assists in his first two years in the league. Although he has gotten better, dropping his turnover percentage, calculated by estimating the average amount of turnovers per 100 plays or 100 * TO/(FGA + 0.44 * FTA + TOV), from 13 down to 11, he is still slightly above league average in turnover percentage (11.5 compared to the 2022-23 season average of 11) despite not being the primary ball handler. He also averaged 3 turnovers to just 3.5 assists. However, he has improved in all other facets of his game, being a 2-time NBA All Star, and 1-time All-NBA Second Team, with his most recent All Star appearance and his All-NBA selection coming in the 2022-23 season, when he averaged 27-7-3.5 with a steal on 49/34/77 and above league average expected field goal percentage (.550 vs. .545).

2023 NBA Entry Draft: Cam Whitmore

Cam Whitmore is a one-and-done forward out of Villanova, meaning that he only played the one required year of college basketball before entering the NBA Draft, who stands at 6'7" with a 232 pound frame and a 6'9" wingspan. In his one and only season in college he averaged 12.5 points and 5.3 rebounds, as well as 1.4 steals, very impressive for a freshman. In fact, his scouting report stated, "Cam Whitmore is one of the most promising two-way players on the board. His remarkable athleticism and high-level competitive spirit should have teams as early as No. 3 considering the Villanova product. While it will be difficult for some to overlook the inconsistency and even the relative lack of success for Villanova as a whole, smart teams will see

[26] *Basketball-Reference*. (n.d.). Basketball-Reference. Retrieved September 3, 2023, from https://www.basketball-reference.com

in Whitmore the crystal-clear outline of a future star. His particular blend of size, strength, and quick-twitch power is not common. Very few players can move and elevate like Whitmore. Even fewer can truck defenders like a lineman."[27]

Throughout the NCAA season and even right before the draft, Cam Whitmore was regarded as anywhere from a lottery (i.e., first half of first round) to a top-five pick in the draft, yet he slid all the way to the end of the first round where the Houston Rockets picked him with the 20th overall pick. What caused him to slide that much?

One potential reason that Cam Whitmore slid so far in the draft was injury concerns and health reasons. Whitmore only appeared in 26 games for Villanova due to a thumb injury that made him miss the first seven games of the season. Additionally, since Whitmore only played 26 games in college, this is a very small sample size and there is a chance that he simply couldn't sustain that level of play.[28] As soon as a few teams began to have concerns about Whitmore's injury history and lack of experience at the collegiate level, more teams caught on, and he began to slip even further in the draft.

There were also some concerns about his work ethic and his attitude. According to ESPN, Whitmore had very disappointing workouts and interviews with several NBA teams, furthering these concerns. Also, an Eastern Conference executive stated to *The Athletic* that Whitmore did not impress at the Nike Hoop Summit, an international basketball prospect showcase, where he allegedly did not put in much effort in practice and did not perform up to his expectations. Whitmore himself even admitted that he had issues with both his shooting and the amount of intensity he displayed in the pre-draft workouts. Questions also arose about his game IQ and general feel for the game, as during his one season at Villanova he had 42 turnovers to just 19 assists.[29]

As displayed by the Cam Whitmore situation during the 2023 NBA Draft, pre-draft interviews and workouts play a crucial role in the draft process as they provide teams with valuable insights into the skills, personalities, and potentials of prospective players. One reason why these processes are important is that they allow teams to assess a player in person. This firsthand experience supports what can be learned from statistics, game footage, or scouting reports, helping teams make more informed decisions. Interviews give teams a chance to gauge a player's maturity, coachability, and how well they might fit within the team's culture. Not every player will fit well with every team's playing style and system. Workouts and interviews give teams a chance to evaluate whether a player's skills align with their strategies and needs, which helps ensure that drafted players are more likely to thrive within their new team environment.

---

[27] Levy, I. (2023, June 22). *Why did Cam Whitmore drop so far in the NBA Draft?* Fansided. Retrieved September 3, 2023, from https://fansided.com/2023/06/22/ cam-whitmore-nba-draft-slide/

[28] Kalbrosky, B. (2023, June 22). *Explaining why Villanova star Cam Whitmore fell so far in the 2023 NBA draft*. ForTheWin. Retrieved September 3, 2023, from https://ftw.usatoday.com/2023/06/cam-whitmore-nba-draft-why-is-he-falling

[29] Noh, S. (2023, June 22). *Cam Whitmore NBA draft slide, explained: Injury concerns not only reason projected top-10 pick fell to Rockets at No. 20*. The Sporting News. Retrieved September 3, 2023, from https://www.sportingnews.com/us/nba/news/cam-whitmore-nba-draft-slide-pick-rockets-20/j1hdshye21p0kmwb9bxn1voy

This information is crucial in avoiding potential locker room conflicts and building a cohesive team.

Workouts and interviews can additionally provide an opportunity for teams to assess a player's physical condition. Medical evaluations help teams identify any potential injury risks or ongoing health concerns that might affect a player's long-term performance. Interviews, workouts, and other in-person evaluations could also be indicative of how a player handles the pressure of an interview or workout can be indicative of how they might handle pressure situations on the court. Players who exhibit confidence, composure, and professionalism during these interactions could be better equipped to handle high-pressure game scenarios.

In-person evaluations can further help validate or challenge the data collected from various sources, such as analytics, game footage, and stats. A player might have impressive numbers but struggle in certain aspects of the game that become evident during workouts or interviews. This live data could be very useful when looking at not highly-touted draft picks, as some players may not have received much attention or exposure before the draft process. Workouts and interviews provide these players with an opportunity to showcase their skills and potential, allowing them to move up in the draft rankings. These sessions also often take place closer to the draft day, allowing teams to make more up-to-date decisions based on the most current information available.

Pre-draft interviews and workouts are essential because they give organizations a thorough picture of athletes beyond their on-court performance. The knowledge gained through these exchanges impacts a team's draft choices and helps both the players and the organization succeed in the long run.

Drafting for Talent vs. Fit

Drafting for talent versus fit is a crucial decision-making process in the NBA draft. Teams often find themselves balancing the potential of a player's individual skills and abilities (talent) with how well they align with the team's existing roster, playing style, and overall strategy (fit). Both approaches have their merits, and the best approach depends on the team's circumstances and goals.

A positive when drafting for talent is drafting players with very high upside and ceilings. When a player possesses exceptional athleticism, scoring ability, defensive prowess, or other standout skills, teams may prioritize them based on their potential to become game-changers. This approach is common for teams in rebuilding phases or lacking star players. Teams may choose the most talented player available even if they don’t precisely match the present squad. This is a long-term investment with the expectation that the player's growth and contributions will ultimately improve the team's performance as a whole. High-talent players also frequently have higher trade values if their teams do decide to use them as trade assets in the future to fill out other particular needs. Getting outstanding talent can give the team freedom in determining its future direction.

Conversely, a positive when drafting for fit is better team chemistry. Teams aiming for immediate success may prioritize players who fit seamlessly with their existing core players. This approach emphasizes how well a player's playing style, position, and skills complement the team's overall strategy. Another positive is that drafting for fit can often lead to better immediate success since teams may draft players who fill in immediate, specific positional needs, such as a point guard or center, to balance out the roster and address glaring gaps in their lineup. Drafting for fit is also encouraged for teams with distinct playing styles or strategies. Drafting players who can excel in that system might be more important than simply selecting the most talented players. For example, a team that emphasizes ball movement might prioritize a player with better playmaking and court vision, rather than maybe just another scorer. Ultimately, the approach a team takes can vary greatly based on their current situation, long-term goals, and evaluation of the available talent. Successful franchises often strike a balance between talent and fit, recognizing that both elements play a role in building a competitive and cohesive team. Additionally, teams can employ trades, player development programs, and coaching strategies to help align a player's skills and the team's needs over time.

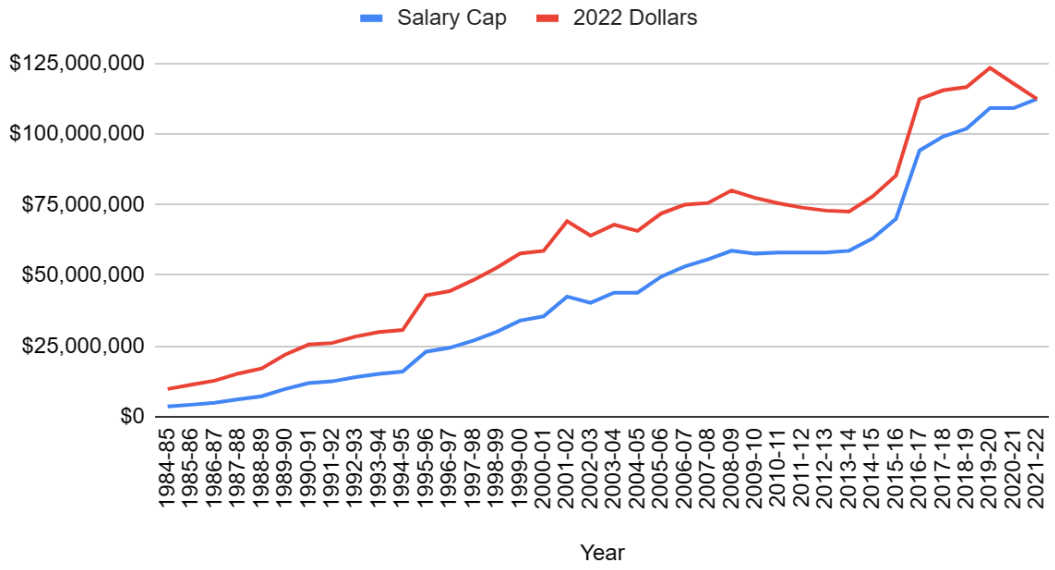
Roster Construction and Management

NBA Collective Bargaining Agreement Overview and History

The NBA first proposed a salary cap in order to maintain a healthy degree of competition and prevent any one team from outspending its rivals and taking over the league. The salary cap limits the total amount a team can pay its players during any given season (Figure 1). However, the salary cap and its intricacies have not stayed stagnant over the many years. The salary cap is one piece of a larger governing document called the Collective Bargaining Agreement (CBA). With each new CBA that the NBA and the National Basketball Players Association (NBPA) have negotiated, the salary cap has experienced substantial adjustments. The parameters of player contracts outlined below, including salary cap ceilings, luxury tax thresholds, and other financial components of the league are determined by CBAs, which are usually multi-year agreements.[30]

[30] Blum, B. (2023, June 15). *Exceptions to the Salary Cap in NBA Salaries: An Overview*. Somos NBA. Retrieved September 3, 2023, from https://www.somosnba.com/exceptions-to-the-salary-cap/

Figure 1: Salary Cap Growth



Salary Cap Introduction (1984–1985): During the 1984–1985 season, the NBA instituted its first salary cap. This cap was relatively straightforward compared to later iterations and intended to stop teams from paying their players too much money.

Bird Rights: The Larry Bird Exception, named after the Boston Celtics legend, was introduced in 1988 and permitted teams to pay more than the salary cap in order to re-sign their own free agents. Players who had played for the team for at least three seasons qualified for this exception. Later, this exception developed into what is now referred to as "Bird Rights."

Luxury Tax: The NBA instituted the luxury tax in 1999 as a means of discouraging teams from routinely overspending during the 1998–1999 season. A team has to pay a luxury tax if their payroll goes over a certain amount (i.e., the salary cap). The money gained from the luxury tax is then distributed to the teams that spent below the salary cap.

Maximum Player Contracts: In order to curb exorbitant player pay, the idea of maximum player contracts was established. The "supermax" contract was added in the 2017 CBA, enabling outstanding players to sign deals with their present team worth a higher proportion of the wage cap. Different types of max contracts include the rookie extension max (25% of the salary cap) and the typical max contract (30% of the salary cap). To qualify for the rookie extension max, players must have played for 4-6 years. One example of this contract is Anthony Edwards' rookie extension, which kicks in in his 5th year with the Minnesota Timberwolves. To qualify for the typical Max Contract, players must have 7-9 years of experience in the NBA. This contract is usually a player's third contract, following the rookie contract and extension. One example of the typical max is Zach Lavine, whose $37.1 million contract was 30% of the 2022-23 salary cap.

Supermax Contracts: The final max contract is the supermax, which was added in 2017 to enable outstanding players to sign extensions with their current team for a whopping 35% of

the salary cap. To qualify for this contract, players must have either played for 10+ years or be entering years 8 or 9 and had made an All-NBA or All-Defense team, or have won MVP/DPOY in either the season before or in two of the previous three seasons. One example of this is Jaylen Brown. Brown was given the largest contract in NBA history and in 2024-25 when the contract fully kicks in, he will be making $49.7 million, or 35% of the projected salary cap. However, the supermax can only be offered by the team that drafted the player or traded for them while they were still on their rookie contract. This is to the benefit of the team as it allows them to keep their stars as they can offer 35% while other teams can only offer 30%.

The Rose Rule: This rule, named after the youngest NBA MVP in league history, states that they can be eligible for the 30% max if they earn certain NBA awards in their fourth season or any of these twice in years two and three.

Changes to Cap Calculation and Exclusions: The NBA has modified the salary cap calculation over time and created a number of exclusions that let teams to sign players even if they are over the cap. These exemptions, which give teams opportunities to add new players without going over the cap, include the standard Mid-Level Exception (MLE), which maxes out around $12 million a year, is available once every season as long as you are entering that season below the cap. The Bi-Annual Exception is set to be around $4.5 million in 23-24. Only teams that are over the cap and under the tax apron are eligible for this, which can be used for contracts lasting up to two years and provides a 5% raise after year one. The Veteran Minimum exception, which allows teams to sign veteran players for a minimum salary even if they are over the cap. The contract cannot last for more than two seasons, and is a maximum of $1.6 million if they have played for over 10 years, and $1.5 million if they have played less.[31] However, the total values of these exceptions and contracts do not stay constant, as over the years, the pay cap has continually risen as a result of variables including rising league revenue, new TV agreements, and other monetary ones like inflation.

Using Advanced Stats to Evaluate Players and Build Rosters

PER (Player Efficiency Rating) calculates a player's performance rating per minute by adding up all of their positives, subtracts the negative accomplishments, and returns a per-minute rating of a player's performance.[32] Looking at the top 10 leaders in PER of the 2022-23, season we can see that they are Nikola Jokic (31.5), Joel Embiid (31.4), Giannis Antetokounmpo (29.0), Luka Doncic (28.7), Anthony Davis (27.8), Jimmy Butler (27.6), Shai Gilgeous-Alexander (27.2), Damian Lillard (26.7), Kevin Durant (25.9), and Steph Curry (24.1). It is important to note that the league average PER is always 15.00. Immediately, we can see that the top ten are all incredibly recognizable names and superstars. In fact, the average salary of these players was an astounding $38.74 million! To put this figure into perspective, the average salary in the NBA

[31] Adams, L. (2022, August 15). *Mid-Level, Bi-Annual projections for 2023/24*. Hoops Rumors. Retrieved September 3, 2023,

[32] Greenberg, N. (2017, April 13). *What is Player Efficiency Rating?* Washington Post. Retrieved September 3, 2023,

during the 2022-23 season was around $9.4 million.[33] Meaning that if you were top 10 in PER you were being paid just over four times more than the average player. This means that teams know and use PER to analyze players, as the entire top ten has salaries that are well over the average. In fact, looking at the entire top 20 for PER, we can see that only two of them were not All-Stars,[34] with one of them (Kawhi Leonard, 23.9 PER) not making it because he had missed too many games.

Types of Contracts on Championship Teams

(Player must have played 200+ minutes in the playoffs to qualify)[35] [36]

Figure 2: 2022-23 Denver Nuggets		
Player (minutes played)	**Contract Type**	**Value and Term**
Nikola Jokic (789)	Supermax contract extension	5yr $264 million
Jamal Murray (799)	Maximum rookie extension	5yr $170 million
Michael Porter Jr. (654)	Maximum rookie extension	5yr $179 million
Aaron Gordon (713)	Veteran extension	4yr $92 million
Kentavious Caldwell-Pope (671)	Veteran extension	2yr $30 million
Bruce Brown (530)	FA multiyear contract	2yr $13 million
Jeff Green (343)	FA multiyear contract	2yr $9 million
Christian Braun (247)	Rookie scale contract	4yr $14 million

[33] *Basketball-Reference*. (n.d.). Basketball-Reference. Retrieved September 3, 2023, from https://www.basketball-reference.com
[34] *Basketball-Reference*. (n.d.). Basketball-Reference. Retrieved September 3, 2023, from https://www.basketball-reference.com
[35] Ginnitti, M., & Allen, S. (n.d.). *Spotrac*. Retrieved September 3, 2023, from https://www.spotrac.com
[36] *Basketball-Reference*. (n.d.). Basketball-Reference. Retrieved September 3, 2023, from https://www.basketball-reference.com

Figure 3: 2021-22 Golden State Warriors		
Player *(minutes played)*	**Contract Type**	**Value and Term**
Steph Curry *(764)*	Max contract extension	4yr $215 million
Klay Thompson *(792)*	Veteran max contract	5yr $190 million
Andrew Wiggins *(767)*	Veteran extension	4yr $109 million
Draymond Green *(703)*	Veteran extension	4yr $100 million
Jordan Poole *(606)*	Rookie scale contract	4yr $10 million
Kevon Looney *(448)*	Free Agent/Bird contract	3yr $22.5 million
Otto Porter Jr. *(371)*	Veteran minimum	1yr $2.4 million
Gary Payton II *(203)*	Rest of season contract after 10 day contract	1yr $2 million

Figure 4: 2020-21 Milwaukee Bucks		
Player *(minutes played)*	**Contract Type**	**Value and Term**
Giannis Antetokounmpo *(800)*	Rookie extension	4yr $100 million
Khris Middleton *(922)*	Veteran contract extension	5yr $177 million
Jrue Holiday *(912)*	Veteran contract extension	4yr $135 million
PJ Tucker *(681)*	FA multiyear contract	4yr $32 million
Brook Lopez *(667)*	Veteran contract extension	4yr $52 million
Pat Connaughton *(546)*	Veteran contract	3yr $16 million
Bobby Portis *(366)*	FA multiyear contract	2yr $7 million
Bryn Forbes *(273)*	FA multiyear contract	2yr $5 million

A really interesting trend on all three of the most recent NBA championship teams (Figures 2, 3, and 4) is that there was a total of just one veteran minimum contract, Otto Porter's. This means that all three of these championship teams wisely invested in higher quality role players rather than simply signing whoever was left on a minimum contract. This was also due to the fact that none of the three attempted to build a superteam, as this would result in having more cap space available rather than spending a large majority of it on two or three players. With each team having at most 2 all stars that year and only having at most one true supermax contract per team.

In-Game Strategy

The Three-Point Revolution

The three-point revolution in the NBA, driven in large part by advanced analytics, has fundamentally transformed the way the game is played and strategized. The use of analytics has led to a shift in focus towards maximizing the efficiency of shot attempts, with an increased emphasis on three-pointers and shots at the rim.

Traditional basketball wisdom held that mid-range shots were valuable and that the closer a shot was taken to the basket, the better the chances of making it. However, analytics showed that three-pointers and shots near the rim are more efficient in terms of points per possession. This realization prompted teams to prioritize these high-efficiency shot attempts over long-range two-pointers. The way analytics impacted this is that advanced metrics such as Effective Field Goal Percentage (eFG%) and Points per Possession (PPP) provided a more accurate representation of a player's or team's offensive efficiency. These metrics accounted for the extra point earned from three-pointers, influencing teams to adapt strategies that optimized these metrics. As further proof, the eFG% in the 2022-23 season was .545, compared to the eFG% in 2012-13 which was .496. Despite this huge increase in eFG%, the changes in FG% and 3P% is not that sizable, with .22 and .02 differences respectively. Why is the eFG% so much higher now? Because the volume of 3PA is so much greater. In fact, the league-wide 3-point attempts per 100 possessions in 2022-23 (34.2) is over ten more than the 3-point attempts of just ten years ago (21.5).[37]

Analytics also changed how players are evaluated. Three-point shooting percentage and volume have become critical factors in assessing a player's value. "3-and-D" players (those who can shoot threes well and play strong defense) gained importance as teams looked for versatile players who fit the modern style. Teams began to adjust their defensive strategies as well. Traditional big men who couldn't defend on the perimeter became less effective, leading to a decline in their usage. Defenses had to adapt to guard against the increased three-point attempts, leading to more switching, zone defenses, and emphasis on defending the three-point line. As a result of this, coaching philosophies evolved as well. Coaches started to design offensive plays that generated open three-point looks, and defensive schemes were built around limiting

[37] *Basketball-Reference*. (n.d.). Basketball-Reference. Retrieved September 3, 2023, from https://www.basketball-reference.com

opponents' three-point attempts. Analytical insights became integral to game planning and in-game decision-making.

Young players entering the league increasingly focused on developing their three-point shooting skills, even big men. Centers and power forwards began practicing three-pointers to expand their offensive capabilities and fit into the modern style of play because teams started using analytics to inform their draft decisions. Players with high three-point percentages and efficient scoring tendencies gained an advantage in the draft process as teams saw their potential to contribute to the evolving offensive strategies.

Player Evolution

The three-point revolution also had a big impact on current players as well. Shooting was becoming an increasingly important factor in the NBA, and it wasn't just guards who were expected to have the ability to shoot. More and more coaches also desired their big man to have a three pointer. The reason for this is because when teams have spacing on the floor, the defense is not able to leave anyone open since all five players are, in theory, competent three point shooters. This strategy stretches opposing defenses, opening up driving lanes by not having the center clog the paint, and leads to drive and kick opportunities.

Several players, mainly centers, have had to develop a three pointer because of this, with the most notable ones being Al Horford, Kevin Love, and Brook Lopez. All players who just ten years earlier would never even think about shooting threes as much as they do now. First is Kevin Love, who during his first All-Star year with Minnesota in the 2010-11 season was attempting 14.1 shots per game, with 2.9 of those being threes. This means that only 20.6% of his shot attempts were beyond the arc. Fast forward to the 2022-23 season when Love, now playing for Miami and Cleveland, averaged 6.8 shot attempts with 4.8 of them being three's - a whopping 70.6%! An increase of 50% of his shots were now coming from three. In Brook Lopez's first All-Star year, 2012-13, he averaged 14.8 shots while taking one three-pointer the entire year; however, in the 2022-23 season he attempted 11.5 shots per game with 4.7 of them being threes, meaning 40.9% of his shots were threes, a drastic increase considering he attempted only seven three pointers in his first six years in the league. Another example of a player drastically shifting their style of play to change with the times is Al Horford. Just 10 years ago, he averaged 14.3 shots per game with just 0.1 being threes, a measly .7%. But just this previous year with the Celtics when Horford was attempting just 7.6 shots per game, he was attempting 5.2 threes. This meant that in ten years, Al Horford changed from shooting under 1% of his shots from three to shooting over 65% of his shots from deep.[38] All three of these big men realized that if they wanted to stay in the league, they had to adapt. As a result, all three of them drastically changed their playstyle, adapting from post-centric inside centers who rarely shot from beyond the arc to becoming some of the most consistent and valued stretch fives in the league.

It has not just been three point volume that has changed. The types of players shooting threes has also changed. If we look at the 2011-12 season, we can see that Ryan Anderson is the

[38] *Basketball-Reference*. (n.d.). Basketball-Reference. Retrieved September 3, 2023, from https://www.basketball-reference.com

clear-cut leader in threes made with 166 threes on the season. This past season, Klay Thompson led the league with 301 threes, almost double Anderson's figure in 2011-12. Their contracts are vastly different as well; Ryan Anderson was just the 10th highest paid player on his own team, making $2.2 million that year (just under $3 million accounting for inflation). Klay Thompson, on the other hand, was the 2nd highest paid player on his team, making around $41 million.[39] This isn't the only example of this pay disparity for 3-point shooters. If we look at the top 10 players for 3-pointers made in the 2011-12 season and compare their contracts to the top ten from 2022-23, we see the same trend (Figure 5).

Figure 5: 2011-12 vs. 2022-23 Comparison: 3-PT Made vs. Salary

	2011-12 NBA Season			***2022-23 NBA Season***		
Rank	**Player**	**3-PTs**	**Salary***	**Player**	**3-PTs**	**Salary***
1	Ryan Anderson	166	$2.9M	Klay Thompson	301	$40.6M
2	Jason Terry	138	$15.2M	Buddy Hield	288	$21.2M
3	Kevin Durant	133	$21.1M	Steph Curry	273	$48.1M
4	Steve Novak	133	$1.4M	Donovan Mitchell	245	$31.0M
5	Brandon Jennings	129	$3.4M	Damian Lillard	244	$42.5M
6	Wesley Matthews	129	$8.4M	Jayson Tatum	240	$30.4M
7	Randy Foye	127	$5.8M	Malik Beasley	235	$15.6M
8	Joe Johnson	125	$24.6M	Julius Randle	218	$23.8M
9	Danny Granger	123	$16.4M	Jordan Poole	214	$3.9M
10	Kyle Korver	118	$6.8M	Anthony Edwards	213	$10.7M

**Salary is expressed in 2022-23 dollars*

[39] *Basketball-Reference*. (n.d.). Basketball-Reference. Retrieved September 3, 2023, from https://www.basketball-reference.com

Just by looking at the salaries and amount of three pointers made, we can see that the player who made the tenth most threes in 2022-23 made almost 50 more threes than the player who made the most threes in 2011-12. Not only that, but we can see that three-point shooting in general has become much more valued, with the top ten of 2022-23 earning a substantially greater amount of money than the top ten of the 2011-12 season, even when accounting for inflation.[40]

Conclusion

In this essay, the influence of analytics on modern basketball is explained, spanning from the drafting and scouting process to roster construction and management, culminating in its effects on in-game strategies and player adaptations.

The first section of the paper illuminates the indispensable role of analytics in the drafting and scouting of basketball talent. Through data analysis, teams gain a profound understanding of player performance and potential. Concrete examples from the NBA landscape, such as Cam Whitmore falling in the 2023 NBA Entry Draft and the decision-making that led to the selections of the 2016 NBA Entry Draft emphasize the impact of analytics, showcasing instances where overlooked talents have thrived due to data-driven insights or where overhyped players fell short. These findings underscore the transformation of team decision-making processes, enabling them to make more informed choices that align with their long-term objectives.

Moving forward, we explore the relationship between analytics and the Collective Bargaining Agreement (CBA), the backbone of roster construction and management. By describing the pivotal components of the CBA, such as salary caps, contract types and exceptions, and luxury taxes, we lay the groundwork for understanding how analytics align with these regulations. Analytics-driven player valuation, such as using PER to determine a player's value, plays a central role in assembling competitive rosters while navigating financial constraints. To look at instances of successful roster building, we examined the rosters of past NBA championship teams to see the types of contracts that their players were on and how they utilized the salary cap to its fullest.

The final section delves into the dynamic landscape of in-game strategy, where analytics wield a transformative impact on gameplay decisions and player adaptations. As teams increasingly rely on data-driven insights, we observe shifts in offensive and defensive schemes, such as a much greater number of three-pointers taken. The three-point revolution, characterized by an emphasis on high-efficiency shots, stands as a quintessential example of how analytics have reshaped game strategies. Correspondingly, players such as Kevin Love, Brook Lopez, and Al Horford have evolved their skill sets to thrive in this new landscape, exemplifying their ability to notice the changing times and demands of the NBA and adapting to elevate their performance.

This paper navigates the intricate web of analytics in modern basketball, tracing its influence across drafting and scouting, roster construction and management, and in-game

[40] *Basketball-Reference*. (n.d.). Basketball-Reference. Retrieved September 3, 2023, from https://www.basketball-reference.com

strategies. The empirical evidence presented underscores the seismic shift in basketball operations and gameplay, highlighting the synergy between data-driven decision-making and tangible on-court outcomes. As the sport continues to evolve, it is evident that analytics will remain a cornerstone of success, reshaping the basketball landscape for teams, players, and fans alike.

Works Cited

Adams, L. (2022, August 15). *Mid-Level, Bi-Annual projections for 2023/24*. Hoops Rumors. Retrieved September 3, 2023, from https://www.hoopsrumors.com/2022/08/early-mid-level-bi-annual-projections-for-2023-24.html

Basketball-Reference. (n.d.). Basketball-Reference. Retrieved September 3, 2023, from https://www.basketball-reference.com

Blum, B. (2023, June 15). *Exceptions to the Salary Cap in NBA Salaries: An Overview*. Somos NBA. Retrieved September 3, 2023, from https://www.somosnba.com/exceptions-to-the-salary-cap/

Ginnitti, M., & Allen, S. (n.d.). *Spotrac*. Retrieved September 3, 2023, from https://www.spotrac.com

Givony, J. (n.d.). *DraftExpress*. Retrieved September 3, 2023, from https://www.draftexpress.com/

Greenberg, N. (2017, April 13). *What is Player Efficiency Rating?* Washington Post. Retrieved September 3, 2023, from https://www.washingtonpost.com/what-is-player-efficiency-rating/37939879-1c08-4cfa-aff3-51c2a2ae060e_note.html

Levy, I. (2023, June 22). *Why did Cam Whitmore drop so far in the NBA Draft?* Fansided. Retrieved September 3, 2023, from https://fansided.com/2023/06/22/cam-whitmore-nba-draft-slide/

Kalbrosky, B. (2023, June 22). *Explaining why Villanova star Cam Whitmore fell so far in the 2023 NBA draft*. ForTheWin. Retrieved September 3, 2023, from https://ftw.usatoday.com/2023/06/cam-whitmore-nba-draft-why-is-he-falling

Noh, S. (2023, June 22). *Cam Whitmore NBA draft slide, explained: Injury concerns not only reason projected top-10 pick fell to Rockets at No. 20*. The Sporting News. Retrieved September 3, 2023, from https://www.sportingnews.com/us/nba/news/cam-whitmore-nba-draft-slide-pick-rockets-20/j1hdshye21p0kmwb9bxn1voy

Application of CRISPR/Cas 9 by Biotechnology Companies to Treat Genetic Disease By Atiksh Vats

Abstract

Clustered regularly interspaced short palindromic repeats Cas9 (CRISPR-Cas9) is a recently developed gene editing technology that has the potential to cure many diseases by correcting DNA sequences. This possibility is being investigated by multiple companies, including Intellia Therapeutics, CRISPR Therapeutics, and Verve Therapeutics. These companies are currently in the process of conducting clinical trials using CRISPR technologies to test their therapies for safety and efficacy in a cohort of patients. If these treatments are successful, they can help treat some of society's most prevalent diseases. This paper will discuss the technologies being developed by companies listed above. For each therapy in clinical development, a mechanism is provided and the stage of the clinical trial is noted.

Introduction

The discovery of clustered regularly interspaced short palindromic repeats (CRISPR)-based gene editing, has provided new avenues for biomedical research. CRISPR is a technology that can alter genes by pinpointing precise spots in the genome to edit in a variety of ways, which can help correct genetic mutations that cause diseases.

The first component of this gene-editing technology is the Cas9 protein, which contains the endonuclease domains RuvC and HNH. RuvC cleaves the non-complementary DNA strand while HNH cleaves the complementary strand, which generates double-stranded breaks in the target DNA. The next component is a single guide RNA (sgRNA) that carries a scaffold sequence which binds to the Cas9 protein. The sgRNA then guides the CRISPR/Cas 9 protein complex to the intended location for editing. For the gene editing itself, one of two pathways are used: non-homologous end-joining (NHEJ), or homology directed repair (HDR). NHEJ is more susceptible to errors involving frameshift mutations and the random insertion and deletion of base pairs at the cut site. The HDR pathway, which is less error prone, involves using the homologous region of the unedited DNA as a template to create the damaged DNA. (Uddin et al.)

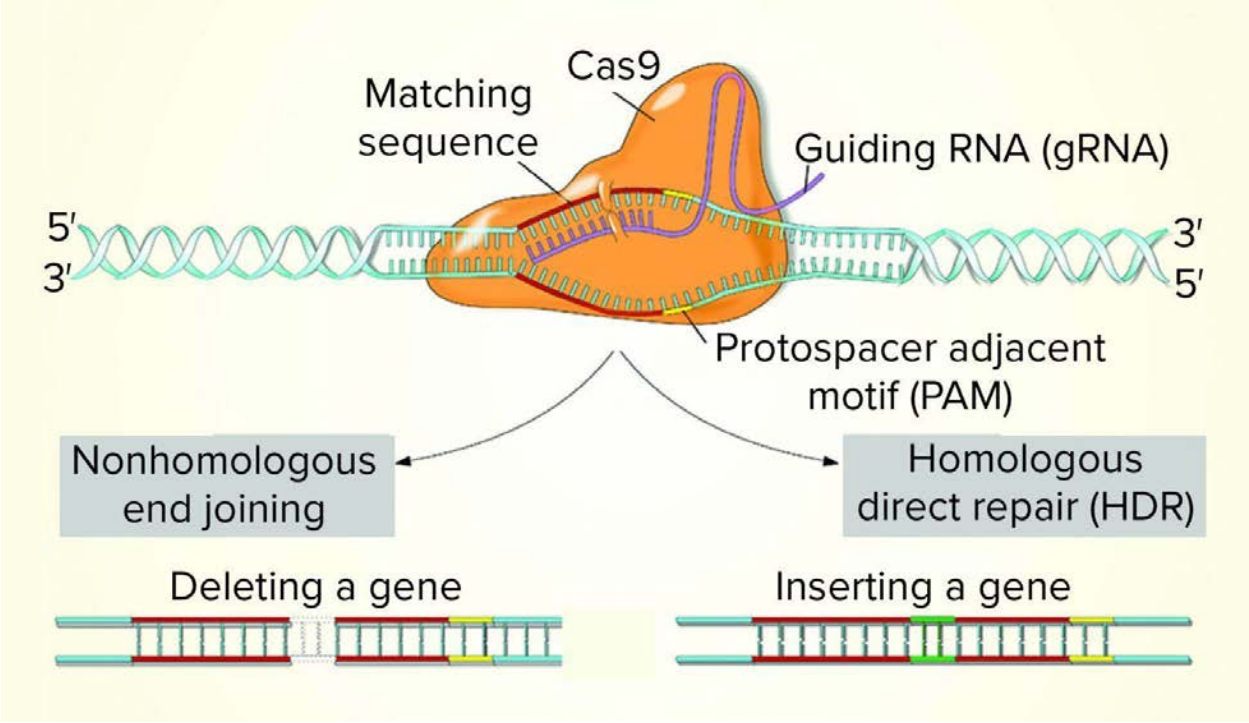


Figure 1 Mechanism of CRISPR gene editing technology (Piestun)

Researchers anticipate that CRISPR can correct disease-causing gene alterations, including some life threatening diseases such as genetic blood disorder β-thalassaemia and hereditary transthyretin amyloidosis (hATTR) (Chavez et al.). CRISPR could be especially effective when the genetic source of the disorder is well understood by scientists because they know the exact location of the genetic error.

For example, one major area where researchers are applying gene editing using CRISPR is in cancer. In a clinical trial in Sichuan University's West China Hospital, researchers extracted immune cells from patients and inserted them back into the patients after using CRISPR/Cas9 to modify TD1. PD-1 is a cell-surface protein on T cells that regulates T cell activation and prevents T cells from attacking healthy cells. However, CRISPR may block the expression of the PD-1 protein, which can then enable T cell response to the tumor. (Cyranoski)

CRISPR may also be applied to treat blood disorders, including sickle cell disease and transfusion-dependent thalassemia. These blood disorders occur when there are mutations in the β-globin (HBB) gene. These mutations affect the shape of the hemoglobin protein complex and adversely impact its function. CRISPR can be used to edit the gene mutation using two approaches. One approach is to correct the HBB mutations, which could be deletions or substitutions. Another approach is to activate γ-globin, which would compensate for the lack of HBB by replacing β-globin. It does this by inactivating the repressor of the γ-globin or inhibiting the repressor binding to the active site (Morshedzadeh et al.).

Currently, many companies are developing treatments using CRISPR technology, including CRISPR Therapeutics, Intellia, and Editas Medicine, which were selected for further

review because they have CRISPR technologies in clinical development. However, before the treatments can be marketed and sold to patients in the United States, they must first be approved by the United States Food and Drug Administration (US FDA).

The first step in drug development leading to FDA approval is preclinical research. Preclinical research is subject to FDA guidelines known as Good Laboratory Practices (GLP) which specify minimal standards for personnel, facilities, equipment, and operations. One task accomplished during preclinical research is toxicology testing, which involves testing the drug on animals to establish that the drug should be safe to test on humans.The next task is to test the drug in-vivo in animal models. The primary focus is on pharmacokinetics, or the study of how drugs move through living organisms. Researchers examine many factors such as absorption, excretion, and metabolism to make sure the drug has its intended effect and passes through the body safely. Then, an Investigational New Drug (IND) application is filed by the company with the FDA that includes all of the data collected in the preclinical phase involving animal testing and laboratory research in order to apply for the approval to conduct clinical trials on humans. If the IND is approved, clinical trials on humans may begin. There are three phases of clinical trials. A Phase 1 trial is primarily intended to test the safety and possible side effects of a therapy on a small group of patients (usually about 20 to 80) and to test the dosing of the drug necessary to achieve the desired result. After the safety is established in Phase 1, a Phase 2 trial is designed to test whether a particular drug can help treat a certain condition. This is a larger trial and often involves about 100 to 300 patients. A Phase 3 trial involves anywhere from a few hundred to a few thousand patients, which is designed to gather a broad set of information, such as the drugs' effect on different populations (i.e., gender, ethnicity), and to compare the drug with other drugs that are currently approved and marketed. Then, the manufacturer must file a New Drug Application (NDA), which is the formal application requesting that the drug be approved for marketing in the US. The NDA includes clinical trial data establishing the safety and efficacy of the proposed therapy. If the FDA approves the drug, then the company can manufacture and market the drug in the United States (Peterson). Generally, it takes about 10 to 15 years to get through all three phases of clinical trials.

Intellia Therapeutics

Intellia Therapeutics is a biotechnology company developing CRISPR technologies with a $3.43 billion market cap as of August 2023. Their primary goal is to "harness CRISPR-based genome editing technologies for human use" (Intellia Therapeutics Mission Statement). They are moving towards this goal by engaging in several partnerships with other companies for research and development, and conducting research on specific therapies that can treat diseases by gene editing with CRISPR/Cas-9.

One area of research includes Intellia's ATTR Amyloidosis program. ATTR Amyloidosis is a disorder that results from deposits of insoluble amyloid fibrils in organs and tissues leading to systemic organ failure (Cleveland Clinic). This is caused by a mutation in the TTR gene that results in defective TTR proteins. While the normal function of the TTR protein is to facilitate

transport of Vitamin A and the thyroid hormone thyroxine, the defective protein dissociates into an aggregate form that causes amyloid deposits in tissues, which is known as hereditary ATTR Amyloidosis (ATTRv). Deposits of TTR amyloid in the heart or nerves can cause ATTRv with polyneuropathy. Approximately 50,000 people suffer from this disease worldwide (Intellia Therapeutics). Some rare disease therapeutics can be considered for orphan drug designation. Orphan drug designation is given to pharmaceutical products that would otherwise not be developed due to a limited patient population, which the FDA defines as less than 200,000 persons. The company developing the drug can be incentivized with benefits such as seven years of market exclusivity and tax credits for clinical tests, which could potentially offset the costs of clinical trials and research (Jardine). The FDA awarded the company's treatment, NTLA-2001, orphan drug designation because of the relatively small patient population.

Intellia's treatment, NTLA-2001, is the first CRISPR therapy to be able to systematically edit genes inside the human body. NTLA-2001 is intended to treat the disease through a single dose, which will reduce the amount of TTR protein without chronic therapy. Studies have shown that there is a significant correlation between reducing TTR protein levels and therapeutic benefit. (Carroll et al.) The program is currently undergoing a Phase 1 clinical trial. Intellia reported interim data in 2022 from 12 patients who were treated for Class I-III heart failure because of ATTR. These patients were given single doses of 0.7 mg/kg and 1.0 mg/kg of NTLA-2001. The preliminary data showed reductions in the TTR protein of greater than 90% without any clinically significant abnormalities. These results indicate potential for NTLA-2001 to make further progress in the clinical trial process(Intellia Therapeutics).

Another major research program at Intellia is focused on Alpha-1 Antitrypsin Deficiency (AATD). AATD, a genetic disorder that increases the likelihood of incurring lung disease, affects approximately 250,000 individuals worldwide, out of a total of 1.25 million people with a defective genotype who are at greater risk of developing AATD. AATD results from a mutation in the SERPINA1 gene which causes a genetic variant of Alpha-1 antitrypsin (A1AT). A1AT is an inhibitor that is produced in the liver and has a variety of functions, such as inhibiting neutrophil elastase.

There are two primary effects incurred from the genetic variant of A1AT. The first is an accumulation of A1AT in the liver, increasing the risk of progressive liver disease. The second is the enhanced risk for emphysema due to the increase in neutrophil elastase.

Intellia's potential treatment for this disease is NTLA-3001, a CRISPR-mediated gene insertion therapy. It is designed to insert a healthy copy of the SERPINA1 gene, which has the potential to restore functional A1AT proteins. This therapy is promising; preclinical data has shown that insertion of the SERPINA1 gene in animals led to normal human A1AT levels. Although Intellia still plans to be working on IND-enabling activities (research which must be completed prior to clinical trials) until the end of 2023, because the current data shows promise for potential approval for clinical trials on humans. (Intellia Therapeutics)

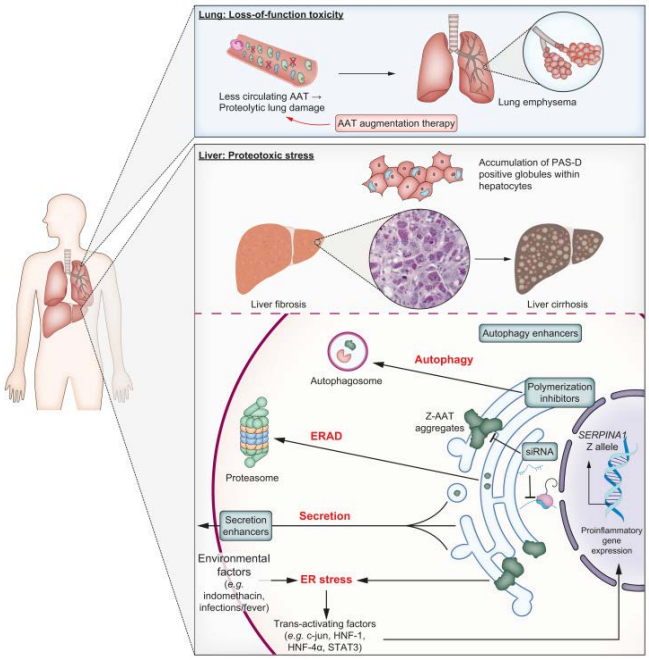


Figure 2 Pathophysiology, clinical manifestations, and treatment approaches for AAT deficiency. (Fromme et al.)

Intellia is also developing a therapy to treat AATD-associated lung disease. The treatment, NTLA-2003, is designed to inactivate the SERPINA1 gene that produces the abnormal A1AT protein. This helps stop liver disease from spreading further and provides an alternative to liver transplants. Preclinical data released by Intellia has shown that the elimination of the SERPINA1 gene has resulted in therapeutically relevant reductions of disease-related protein. This data indicates that this drug has great potential to succeed and effectively combat liver disease (Intellia Therapeutics).

CRISPR Therapeutics

CRISPR Therapeutics was one of the first biotechnology companies to use CRISPR gene therapies as a method for treating diseases, starting in 2013. One major area of active research for CRISPR Therapeutics is in the field of immuno-oncology. Immuno-oncology involves treatments such as engineered cell therapy that stimulate the immune system to attack cancer cells. Currently, most engineered cell therapies involve the creation of a unique, personalized therapy for each individual patient, which is costly and time consuming. CRISPR Therapeutics believes that CRISPR/Cas9 is vital for successful immuno-oncology products due to its suitability for multiplexed editing, the modification of multiple genes in a cell.(CRISPR Therapeutics 12) CRISPR/Cas9 can also make multiple edits with a single Cas9 protein and small gRNA molecules, unlike other gene editing platforms, which require different enzymes for each modification. (CRISPR Therapeutics 12)

CRISPR Therapeutics' solution to these issues is its flagship immuno-oncology product, CTX110, which is a donor-derived CAR T cell therapy that collects T cells from the patients, modifies them, and delivers them back into the patient. The main application of CTX110 is to treat many different patients from a single batch of CAR T cells, which would not require spending a large amount of money creating a unique, personalized cell line, improving accessibility for each patient with a tumor. CTX110 makes three modifications to the donor T cells. The first one eliminates the T-cell receptor to ensure that there is no risk of Graft versus Host Disease, or GvHD, which is where the recipients reject the donor cells immunologically, the second one involves insertion of a CD19-directed CAR into the TRAC gene, and the third one removes the class I major histocompatibility complex, MHC 1, in order to improve the persistence of CAR T cells. Currently, CTX110 is undergoing clinical trials in order to test its safety and efficacy. Although Phase 1 of the clinical trial is currently ongoing, due to the promising initial data from the clinical trial, CRISPR Therapeutics has decided to expand to include a Phase 2 clinical trial.

Phase 1 of the clinical trial was split into Part A and Part B. In Part A, patients were infused with one dose of CTX110 with dosage amounts from Dose Level 1 (30 million CAR + T cells) to Dose Level 4 (600 million CAR + T cells). In Part B, which was testing for a sustained response, patients received CTX110 at Dose Level 4 as well as a consolidated dose of CTX110 at the same level approximately four to eight weeks after the dosage for patients that performed well on the initial dosage. The clinical data from Phase 1 Part A reported by CRISPR Therapeutics indicates that there was a 67% overall response rate (ORR), which was the portion of patients whose tumors were substantially reduced, and a 41% complete response rate (CRR), which was the portion of patients whose tumor was completely eliminated, was for patients treated with at least one dose of CTX110. Also, three patients out of the 27 patient cohort have shown a consistent response without disease progression for two years, which speaks to the ability of CTX110 to produce sustained responses. Researchers observed clear evidence on the benefits of consolidation dosing, which increased the frequency of complete responses, and several patients had ongoing complete responses beyond six months (CRISPR Therapeutics 13-16). These early results show the great potential that CTX110 has for success in clinical trials and eventual approval by the FDA.

Another major program at CRISPR Therapeutics pertains to blood disorders such as beta thalassemia and sickle cell disease. Sickle cell disease (SCD) is an inherited blood disorder that occurs following a mutation in the beta globin gene that causes abnormal red blood cell function. This alters the shape of the red blood cells and causes them to become inflexible, leading to obstruction of blood vessels, which subsequently results in severe pain, infections, and possible death. Currently, approximately 300,000 individuals are born with sickle cell disease every year and approximately 20-25 million people worldwide suffer from the disease. (CRISPR Therapeutics)

The primary current treatment option available for both these diseases is chronic blood transfusions that are given every two to four weeks. However, this leads to high levels of iron in the body and subsequent organ damage. CRISPR Therapeutics' therapy, Exa-cel, is intended to increase the expression of the gamma globin protein, which creates a form of hemoglobin that is present in infants, fetal hemoglobin (HbF). Usually, the symptoms of beta thalassemia and SCD do not manifest until the amount of hemoglobin has declined considerably. Elevated levels of HbF result in a condition called hereditary persistence of fetal hemoglobin, which results in the patient either being asymptomatic or experiencing milder forms of the disease.

CRISPR Therapeutics's Exa-cel will do this by disrupting the erythroid-specific enhancer of the BC11A gene via CRISPR/Cas9, which encodes the BC11A protein that keeps the HbF level low. This will result in reduced expression of BC11A gene, which will increase HbF levels. One approach that CRISPR Therapeutics is using to achieve this is to isolate hematopoietic stem cells (HSC), which gives rise to red blood cells, treat those cells with CRISPR/Cas9 by modifying the aforementioned gene, and then transfuse these cells back into the patients.

Exa-cell has been given Fast Track, Orphan Drug, and Rare Pediatric Disease designation by the FDA. Currently, CRISPR Therapeutics is conducting clinical trials designed to assess the efficacy of a single dose of Exa-cel on patients ages 12 to 35 with SCD, in the CLIMB-111 trial and with TDT, in the CLIMB-121 trial. The initial data reported from the CLIMB-111 trial indicated that all 44 patients with TDT who were given Exa-cel showed clinically significant increases in total hemoglobin and HbF and reductions in the need for red blood cell transfusions. Also, 42 out of the 44 patients were transfusion free since the follow-up and the other two patients had reduced transfusion volumes by 75% and 89%, respectively (CRISPR Therapeutics 9-11).

CRISPR Therapeutics' CLIMB-121 trial involved 31 patients with severe SCD that were treated with Exa-cel. WIth a median follow-up of 10.2 months, all 31 patients avoided vaso-occlusive crises, the most painful complication of SCD, after being given Exa-cel. Also, all patients showed clinically significant increases in HbF levels. These clinical trial results show the great potential that Exa-cel has for future FDA approval and treating many patients with these kinds of blood disorders. This can also serve as a template for the manufacturing of similar CRISPR-based products to treat similar blood disorders. (CRISPR Therapeutics)

Verve Therapeutics

Verve therapeutics is another promising company developing CRISPR therapeutics, and is focused on solving the global health epidemic of cardiovascular disease. The market cap of Verve therapeutics is 1.04 billion USD as of August 2023 (Yahoo Finance).

According to the World Health Organization (WHO), cardiovascular disease (CVD) is responsible for approximately one in three deaths. It also costs the US healthcare system more than $350 billion per year. CVD refers to diseases of the heart and blood vessels, which are diagnosed as ASCVD (atherosclerotic cardiovascular disease). In ASCVD, cholesterol drives the development of atherosclerotic plaque, a mixture of cholesterol, cells, and cellular debris in the

wall of a blood vessel that results in hardening of the arteries, which increases the likelihood of heart disease and strokes.

Verve's novel approach to treat CVD is to target PCSK9 and ANGPTL3, genes to lower blood lipids, such as low-density lipoprotein cholesterol (LDL-C). This approach is justified through studies that have shown that reducing LDL-C by 39 mg/dL for five years in patients with ASCVD reduces the risk of a subsequent event by 21%, and a similar level given to people with no ASCVD reduces the risk of an event by 88%. (Verve Therapeutics 8-12)

Verve Therapeutics has designed a treatment, VERVE-101, to treat CVD. VERVE-101 is designed to use gene editing technology to permanently shut down the PCSK9 gene in the liver by making a single base change at a specific location in PCSK9. PCSK9 regulates the LDL receptor, which controls blood LDL-C levels.

Preclinical data validated this approach, indicating that after an infusion of a base editor targeting PCSK9 in non-human primates (NHP), there was an 89% decrease in PCSk9 protein levels and a corresponding decrease of blood LDL-C levels of 59%, which was maintained at an average of 71% after two years of treatment. The data also showed that VERVE-101 was well tolerated in non-human primates (NHPs) with only minor increases in liver function tests that were resolved in two weeks.

Another area where Verve is developing treatments is heterozygous familial hypercholesterolemia (HeFH). HeFH, a form of Familial Hypercholesterolemia (FH), is an inherited genetic disorder that can cause dangerously high cholesterol levels. HeFH compromises the liver's ability to remove particles that contain cholesterol from the blood. This results in LDL deposits in the blood vessels and arterial blockage, which can lead to heart attacks and strokes. Verve Therapeutics initially plans to use VERVE-101 to treat heterozygous familial hypercholesterolemia (HeFH). and then intends to expand it incrementally, culminating in usage on patients with ASCVD because they are both similar diseases with high cholesterol levels.

Currently, VERVE Therapeutics is planning the heart-1 clinical trial, which will administer VERVE-101 on 40 adult patients with ASCVD. The trial will evaluate the safety and efficacy of VERVE-101 along with supplemental analyses of reduction in PCSK9 and LDL-3. While the clinical trial has received approval in the United Kingdom and New Zealand, the FDA put the application on hold pending information including difference in potency between human and non-human cells, risks of germline editing, and off-target analysis on non-hepatocyte cell types. (Verve Therapeutics)

Another Verve Therapeutics product, VERVE-201, is designed to use gene editing technology to turn off the ANGPTL3 gene, a key regulator of cholesterol and triglyceride metabolism, which is expected to reduce LDL-C levels. The product was intended to treat patients with homozygous familial hypercholesterolemia (HoFH) and refractory hypercholesterolemia, patients with ASCVD that are not at LDL-C goal on oral therapy or a PCSK9 inhibitor. VERVE-201 will use GaINAc-LNP, Verve's proprietary technology, to deliver a base editor for ANGPTL3 to the liver.

Currently, VERVE-201 is in preclinical studies, where the data indicated a reduction of 94-97% in blood ANGPTL3 protein and a corresponding decrease in LDL-C levels of 100 mg/dL, which was a 35% reduction from baseline. The studies were conducted in an non-human primate (NHP) model of HoFH, which was created by editing the LDLR gene in wild-type NHPs . The data from both the preclinical studies indicates that the products show promise for success in clinical trials leading to FDA approval. Verve anticipates that it will begin clinical trials on this product in 2024. (Verve Therapeutics)

Conclusion

CRISPR gene-editing technologies have a very promising future in medicine, having the potential to treat disorders that were once thought to be incurable. One example of that is by treating cystic fibrosis (CF). CF is a lethal disease that affects the lung and digestive tract and is caused by mutation of the cystic fibrosis transmembrane conductance gene (CFTR) located on chromosome 7, which can cause chest infections and pneumonia and can ultimately lead to respiratory and heart failure. The CFTR gene regulates the ion channel and osmotic regulation in the body. Researchers have identified sheep as the most accurate model of CF in humans due to the similarities in lung development, size, and anatomy. CF symptoms of lambs also closely match those of humans, specifically, in liver and gallbladder diseases. This indicates that sheep will provide a useful template for researchers to develop treatment for humans with CF. Also, in another study, CRISPR was able to rectify 33% of the models. This indicates that CRISPR is a viable option to treat CF in the future. (Abdelnour et al)

Another promising application of CRISPR is in treating HIV. A Phase I clinical trial used gene editing technology to knock out the CCR5 co-receptor gene in T cells of HIV-positive patients, which made the T cells more resistant to the virus. The results indicated that this treatment is a viable option and further studies are planned. (Carroll)

CRISPR is proving itself to be a highly effective technology helping companies to make revolutionary strides in biotechnology, which may be applied to a number of diseases by treating the genetic basis of the disease. Intellia Therapeutics, CRISPR Therapeutics, and Verve Therapeutics are at the forefront of leveraging CRISPR/Cas9 gene-editing technology to develop revolutionary treatments; Intellia focuses on eliminating genetic diseases, CRISPR Therapeutics targets blood disorders, cancer, and immune system diseases, while Verve Therapeutics concentrates on cardiovascular diseases. Their contributions encompass developing pioneering therapeutics like Intellia's in vivo gene-editing therapies, CRISPR Therapeutics' CTX001 for sickle cell disease and beta-thalassemia, and Verve's gene-editing treatments designed to permanently reduce the risk of coronary artery disease.

Works Cited

Abdelnour, Sameh A., et al. *The Potential of CRISPR/Cas9 Gene Editing as a Treatment Strategy for Inherited Diseases*. 2021, https://dpl6hyzg28thp.cloudfront.net/media/Future4.pdf.

Carroll, Dana. "Genome Editing: Past, Present, and Future." 2017, https://dpl6hyzg28thp.cloudfront.net/media/Future2.pdf.

CRISPR Therapeutics. "Mission."

CRISPR Therapeutics. "2022 Annual Report." https://ir.crisprtx.com/static-files/6c9bf700-3261-447f-bd1a-33030f0a64e9.

Intellia Therapeutics. "2022 Annual Report." 2022, https://ir.intelliatx.com/static-files/48cb17cd-7245-48f4-b055-3b2100cc00dd.

Intellia Therapeutics.

Jardine, Jason J. "Orphan Drug Designation." 2017.

Morshedzadeh, Firouzeh, et al. "An Update on the Application of CRISPR Technology in Clinical Practice." 2022, https://dpl6hyzg28thp.cloudfront.net/media/12033_2023_Article_724.pdf.

Uddin, Fathema, et al. "CRISPR Gene Therapy: Applications, Limitations, and Implications for the Future." 2020. *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7427626/#:~:text=This%20review%20aims%20to%20%281%29%20provide%20a%20brief,that%20these%20and%20future%20trials%20may%20hinge%20upon.*

Verve Therapeutics. "2022 Annual Report." 2022, https://ir.vervetx.com/static-files/e0a1bcc3-6dff-4886-8bf9-bd46052b811c.

Carroll, Antonia, et al. "Novel approaches to diagnosis and management of hereditary transthyretin amyloidosis." June 2022, https://jnnp.bmj.com/content/93/6/668.

Chavez, Michael, et al. "Advances in CRISPR Therapeutics." 2023, https://www.nature.com/articles/s41581-022-00636-2.

Cleveland Clinic. *Transthyretin Amyloidosis (ATTR-CM)*.

Fromme, Malin, et al. "Alpha-1 antitrypsin deficiency: A re-surfacing adult liver disorder." 2022, https://www.sciencedirect.com/science/article/pii/S0168827821022273.

Piestun, Dan. "Use of CRISPR and Other Gene-Editing Tools in Cell Line Development and Engineering." September 2020, https://bioprocessintl.com/analytical/cell-line-development/use-of-crispr-cas9-and-other-gene-editing-tools-in-cell-line-development-and-engineering/.

Cyranoski, David. "Chinese scientists to pioneer first human CRISPR trial." 2016, https://www.nature.com/articles/nature.2016.20302.

Peterson, Leigh Ann. "FDA Drug Approval Process." 2022, https://www.drugs.com/fda-approval-process.html.

Parenting Styles and Academic Outcomes: Combatting Your Negative Learned Behaviors
By Cassandra Avallone

Abstract

Parenting styles are widely researched theories in the field of developmental psychology. Each style is categorized by their levels of demandingness and responsiveness. There are four main parenting styles recognized by psychologists: authoritative, authoritarian, permissive, and neglectful. The way in which children were parented influences them throughout their lives in the form of their academic motivation, social abilities, and performance. This review outlines the characteristics and academic outcomes associated with each parenting style and identifies strategies to help people combat negative learned behaviors that stem from the way they were parented.

Keywords: parenting styles, demandingness, responsiveness, academic outcomes

Introduction

The way you are raised stays with you forever, not only in memory, but also in the way that you think and behave. Even though it is subconscious, the way that you approach problems, socialize, and learn has been heavily influenced by your upbringing (Spera). Understanding who you are starts with learning how you were shaped by your caregivers. Certain parenting styles rear different child outcomes in traits like personality, patience, interpersonal abilities, and even self-confidence. Additionally, research shows that parenting styles impact childrens' academic performance, motivation, social life, and resilience (Spera). However, this does not mean that your academic abilities are fixed or that your academic fate is set in stone. While parenting does not entirely define your school outcomes, knowing the style in which you were parented can help you combat the negative effects it has had on you. After increasing self-awareness around your parents' childrearing style, you can use specific strategies to help you learn more effectively and achieve higher academic performance.

Parenting Styles

There are four types of parenting styles, categorized by levels of demandingness (the amount of authority parents exert on their children) and responsiveness (the extent to which parents are affectionate and present). These are authoritative, authoritarian, permissive, and neglectful (Zelster).

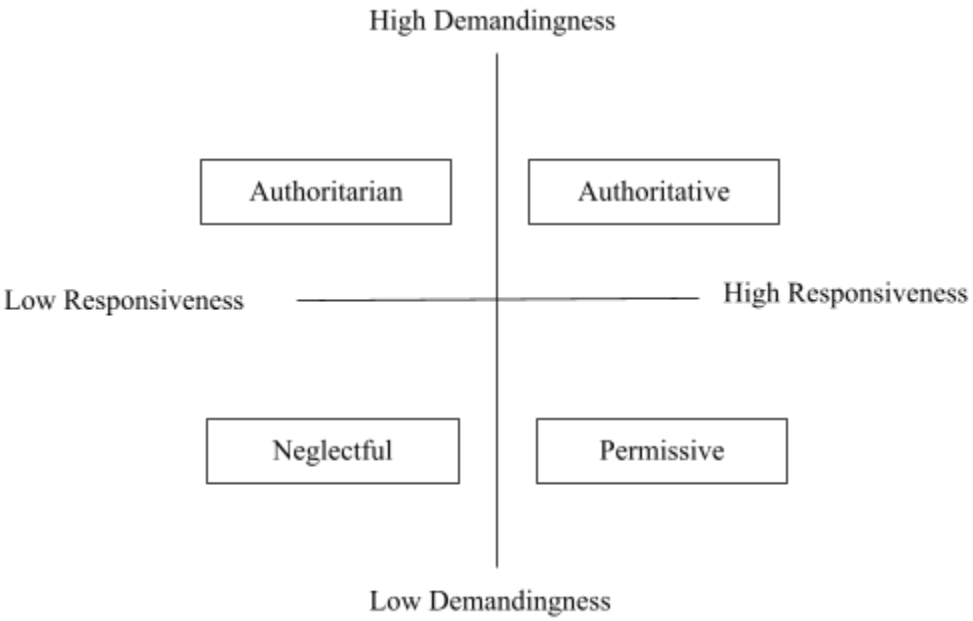


Figure 1: Parenting styles demandingness-responsiveness quadrant graph

Authoritative Parenting

Authoritative parents are in the high range of demandingness and responsiveness (Sanvictores). They expect maturity and effort from their children in school but also show lots of affection towards them. When parents use this style, communication goes two ways (between parent and child) and is reasonable (Spera). For example, when a parent or child raises a concern, they sort out the problem through a sensible conversation and genuine desire to help both parties. Hence, parent-child relationships are naturally closer when parents employ the authoritative style. Additionally, part of authoritative parents' communication with their children involves using a type of reasoning called induction (Spera). In the principle of induction, parents give explanations for their values, actions, and disciplinary behavior. When parents use induction, children better understand and more often adopt their parents' values, such as prioritizing academics. Along with authoritative parents having a much easier time getting their children to adopt their values around education, their children also tend to be more academically successful because of higher confidence, responsibility, and social skills than the kids of caregivers with different parenting styles (Sanvictores). Thus, authoritative style parenting yields the best academic outcomes because children are equipped with self-confidence and a better understanding of *why* school is important. Still, while authoritative parents' children tend to be the most academically successful, that does not mean that kids with other parental backgrounds cannot succeed too.

Authoritarian Parenting

Authoritarian parenting is characterized by expressing high levels of demandingness and low levels of responsiveness (Sanvictores). These parents offer very little nurturing and justify this stoic behavior as "tough love." Authoritarian parents respond to their children's mistakes by punishing them, so these children are very well behaved. Although, this constant fear of messing up often leads to their children having low self-confidence or struggling to make decisions (Sanvictores). Many children eventually rebel, which stems from their inability to reason with their parents and their one-way communication system (i.e. parents give demands and children obey them without discussion; Sanvictores). Because of the one-way communication system between authoritarian parents and their children, parents punish or ignore their kids if they attempt to give a different opinion.

In the classroom, children with authoritarian parents may be aggressive, socially inept, or shy (Sanvictores). Working in groups and making friends is harder for these children because relationships at home lack two-way communication and affection. Despite this, there are some advantages for children who have authoritarian parents. For instance, this parenting style creates a sense of structure for the child and increases work ethic (Rubin 11). These children tend to outwork and outperform their classmates in order to receive praise from parents, not necessarily because they truly value education. This is called extrinsic motivation, which often decreases when the child enters adulthood and no longer receives parental praise for doing well or punishment for failing. Therefore, their work ethic is often not high forever because their motivation only comes from wanting to please their parents.

Permissive Parenting

Permissive parenting entails parents exhibiting low levels of demandingness and high levels of responsiveness towards their children (Sanvictores). These are the types of parents that let their children do poorly in school but think they are helping them by giving unwavering praise and letting them make their own decisions. The children of these parents are rarely told, "no." So, rejection or failure in school is extremely distressing, and these kids are less likely to persevere through challenges (Spera). There are fewer positives to this parenting style than authoritative or authoritarian in the academic context. However, children often develop individual expression due to the amount of freedom they have been given by their parents (Rubin 17). This could manifest as freedom and exploration professionally, emotionally, or even sexually. Because of this, parents think that they are doing the right thing by being permissive with their child, no matter what. Regardless, kids need structure in their youth so that they can learn boundaries and develop the ability to use self-discipline, especially in academic contexts.

Neglectful Parenting

Neglectful parenting is characterized by parents showing low demandingness and low responsiveness towards their children (Sanvictores). As a result, parent-child relationships are distant, with little-to-no communication. Because of how scarcely these children are nurtured,

they have a harder time forming relationships with other people, particularly peers (Rubin 18). They also have trouble forming friendships because parents do not supervise or correct their childs' behavior towards other children. This impacts children academically because strong social skills are associated with higher academic performance (Moorhead). Additionally, neglectful parents do not monitor their children's academic progress, such as homework completion (Spera). Essentially, there are very few positives to this style. The only possible upside is that children often develop lots of independence (Sanvictores). One reason that some parents resort to this style is because they have mental health issues. Therefore, they do not have the ability to care for another person, let alone themselves (Kotchick 314). Still, due to the overwhelming amount of drawbacks for children, this parenting style should be avoided by caregivers. Overall, neglectful parenting creates an inconsistent, insecure environment, leading to confidence and mental health issues in children, as well as low academic performance.

External Circumstances That Impact Parenting Styles

Most studies show that raising children using an authoritative parenting style leads to having the most academically successful children (Spera). However, under specific circumstances, authoritarian parenting is necessary or academically beneficial for children. For instance, authoritarian parents use strict control and little warmth with their children, which has proven to be the most beneficial parenting style when living in a dangerous community (Kotchick 315). This is because if the neighborhood is unsafe, parents must prioritize their children's safety over academics by using greater control. In addition, parents in lower socioeconomic statuses tend to leverage authoritarian parenting to motivate their children academically because they rely on using strict rules to enforce the prioritization of education, rather than purchasing expensive educational toys, such as puzzles (Kotchick 313). Other studies show that being in a lower socioeconomic status group leads parents to adopt authoritarian styles because of the greater amount of stress that parents experience in order to survive and provide for their families (Kotchick 312). For example, almost all families suffered from low SES during the Great Depression (Kotchick 314). One study showed that authoritarian parenting was used dramatically more during this period due to financial strain. This strain led to irritability in parents, who then created stricter rules and more punishments.

Along with low socioeconomic status predicting the parenting style that caregivers adopt, parental mental health also shapes the way they behave when rearing their children (Moorhead). When parents suffer from poor mental health, they often become authoritarian, permissive, or neglectful. Anger issues and irritability, sparked by mental health issues (e.g., depression), often lead to strict authoritarian parenting. Similarly, if caregivers have serious mental issues and cannot support a child, they are likely to be neglectful parents (Kotchick 314). Finally, when parents have low self-confidence and trouble saying "no," even to their child, they are more likely to exhibit the permissive parenting style (Kotchick 315). Overall, research shows that external circumstances often lead parents to utilize negative parenting styles. This can create

child resentment towards parents when the children do not fully understand the circumstances (Sanvictores).

To take action to counteract the negative effects of poor parenting, it is important to first attempt to have compassion towards your parents' challenging life circumstances. Then, it is valuable to identify the areas in which you want to improve your life.

What to do if You Had Poor Parents to Improve Academic Performance

According to research, authoritative parenting often yields the best scholastic results (Spera). However, if children are raised with parents using one of the other three parenting styles and have socialization, motivation, or stress issues in the academic context, there are ways to counteract these effects.

Improving Social Skills

There is research showing that social skills and academic success are directly linked (Moorhead). If you are a student who has difficulty with socializing or working in groups, use these guidelines when teamwork feels unattainable, aggravating, or difficult: 1) Ask. Do not accuse. Speak respectfully and positively. 2) When clarifying ideas, take a deep breath. Use patience so you can fully express yourself, and others will listen. 3) If conflict arises in your group, work through it as a team. Teamwork requires compromise through respectful, mindful negotiation.Take a break when you cannot communicate peacefully or when you want to quit (The University of Queensland).

Self-discipline

If your parents were neglectful or permissive and did not create structure, it can be hard to learn those skills in school. Here are two things you can do to create a more structured environment for yourself: Keep an agenda, whether online or on paper. Being able to set aside specific time for each assignment will help you keep up with school's workload while still getting to schedule things for fun. Create a routine. With a routine, there is no room for excuses. Going to school and doing homework no longer become optional (Jessie).

Whether your parents enforce it or not, you can create a system for organization in your life, and practice discipline.

Managing Stress

Having poor parents takes a toll on your mental health and spikes stress levels. Stress drains your mental energy, making it hard to focus and perform academically (The Health News Team). Remember that mental energy requires physical energy. To fully optimize your performance in school, prioritize sleep. Getting eight hours of sleep alone has proven to improve focus, drive, and patience (The Health News Team). Additionally, mental breaks are vital to clearing your mind so that when it is time to work, you can focus more easily. Mental breaks can include a warm bath, a night out with friends, or a quiet walk in nature.

Conclusion

There is no instant fix for altering your maladaptive academic behaviors that you learned from your parents. However, you can identify the way in which you were parented ranging on two continuums: demandingness and responsiveness. Then, you can understand that your challenges academically may not be entirely your "fault." Therefore, with the right tools, you have the power to counteract your parents' negative effects on your motivation, social skills, and stress, and turn your academic life around as a result.

Works Cited

Jessie. "5 Ways to Create More Structure in Life (With Examples)." *Tracking Happiness,* 29 Jan. 2023, www.trackinghappiness.com/how-to-create-more-structure-in-your-life/.

Kotchick, Beth A., and Rex Forehand. "Putting parenting in perspective: A discussion of the contextual factors that shape parenting practices." *Journal of Child and Family Studies*, vol. 11, no. 3, Sept. 2002, pp. 255–269, https://doi.org/10.1023/a:1016863921662.

Moorhead, Joanna. "Horrible Parents: A Survival Guide." *The Guardian*, The Guardian, 11 Mar. 2017, www.theguardian.com/lifeandstyle/2017/mar/11/horrid-parents-how-survive-them-alyson-corner-angela-levin.

"Resolving Group Work Issues." *My.UQ*, University of Queensland Australia, my.uq.edu.au/information-and-services/student-support/study-skills/group-work/resolving-group-work-issues. Accessed 22 Sept. 2023.

Rubin, Stephanie N. "The Relationship Between Academic Motivation and Parenting Styles in Multiple Socioeconomic Status Areas." *Rowan University*, 2017.

Sanvictores, Terrence. and Magda D. Mendez. "Types of Parenting Styles and Effects On Children." *StatPearls*, StatPearls Publishing, 18 September 2022..

Spera, Christopher. "A Review of the Relationship Among Parenting Practices, Parenting Styles, and Adolescent School Achievement." *Educational Psychology Review*, vol. 17, no. 2, June 2005, pp. 125–146, https://doi.org/10.1007/s10648-005-3950-1.

The Health News Team. "Can Lack of Sleep Affect Your Mood?: Sharp Healthcare." *Sharp Health News*, Sharp Health News, 2 Mar. 2023, www.sharp.com/health-news/does-losing-sleep-make-you-cranky#:~:text=When%20someone%20is%20sleep%2Ddeprived.

Zeltser, Francyne. "A Psychologist Shares the 4 Styles of Parenting–and the Type That Researchers Say Is the Most Successful." *CNBC*, CNBC, 29 June 2021, www.cnbc.com/2021/06/29/child-psychologist-explains-4-types-of-parenting-and-how-to-tell-which-is-right-for-you.html.

NLP Sentiment Analysis of News Articles: Uncovering Opinion Dynamics in Contemporary Media by Akshay Shivkumar

Abstract

In the evolving landscape of digital news media, understanding the dynamics of opinion and sentiment within articles became crucial. This research delved into sentiment analysis of news articles to uncover the nuanced interplay of opinions, attitudes, and emotions that shaped public discourse. The investigation analyzed the extent of polarity and subjectivity across various domains, examined how sentiments evolved over time, explored differences among prolific authors, and identified the distinction between the sentiment of articles versus their headlines. Drawing on a dataset comprising 209,527 news articles from the Huffington Post across 40 diverse categories from 2012 to 2022, both VADER, a lexicon-based model, and BERT, a state-of-the-art transformer model, were employed for sentiment analysis. Our findings revealed distinctive polarity and subjectivity patterns across categories, with style, food, entertainment, and art exhibiting positive sentiment, while crime, politics, and education demonstrated a more negative sentiment. Shifts in sentiment over the years were identified, noting an overall decline with intermittent fluctuations attributed to external factors. Furthermore, our exploration of sentiment by author highlighted variations in subjectivity, with specific authors showing higher degrees of subjective expression. Finally, our analysis of headline sentiment relative to the article proves a disparity within news expression. By employing advanced sentiment analysis techniques and examining multiple facets, our study provides statistical conclusions about how news articles represent contemporary events. This research contributed to media discourse analysis by filling a gap in understanding opinion dynamics in the news more holistically.

Keywords: Sentiment, polarity, subjectivity, lexicon, machine learning, natural language processing

Introduction

Background

News has held a crucial role in societies throughout history, providing information and shaping public opinion. From handwritten bulletins to digital platforms, the evolution of news outlets has been driven by societal needs (The Evolution of Media). The accessibility of news articles in the digital age has facilitated a rapid exchange of information, leading to an abundance of news content available to a global audience.

While news articles have traditionally aimed to present objective and unbiased accounts of events, subjective perspectives and opinions often find their way into news reporting (Kavanagh et al.). Journalists, columnists, and commentators have utilized language as a means to express their viewpoints, influencing the narrative and shaping public perception, often through forms of bias, emotion, and self-interest (Chong). The presence of opinions within news

articles has become a subject of interest in studying media discourse and understanding the nuances of public sentiment.

Natural Language Processing (NLP)

The advancements in machine learning have revolutionized various domains by enabling computers to learn from data and extract meaningful insights (Jordan and Mitchell). Natural Language Processing (NLP), a subfield of machine learning, focuses on developing algorithms and models to process and understand human language (Khurana et al.). NLP techniques have been widely applied in areas such as text classification, information retrieval, and sentiment analysis.

Sentiment analysis, also known as opinion mining, is an important application of NLP that aims to automatically determine the sentiment expressed in a given text (Liu). It involves extracting subjective information, such as opinions, attitudes, and emotions, from textual data to gain insights into public sentiment. Sentiment analysis has gained significant attention due to its potential for understanding public opinion, brand perception, market trends, and more. There are two main methods to implement opinion mining: using either a lexicon-based or a machine learning-based approach.

Lexicon-based approaches, also called bag-of-words models, are standard methods employed in sentiment analysis. These approaches utilize sentiment lexicons, which are collections of words and phrases annotated with their corresponding sentiment orientation values (Taboada et al.). By comparing the terms in a text with the entries in the lexicon, sentiment scores can be assigned to individual words or entire sentences, enabling sentiment aggregation and analysis.

Machine learning-based approaches have gained popularity in sentiment analysis due to their ability to learn patterns from labeled datasets. These approaches involve training models, such as support vector machines, recurrent neural networks, and transformer-based models, on data where the sentiment polarity of texts is already known (Singh et al.). By learning from labeled examples, machine learning models can generalize patterns and make predictions about the sentiment of unseen texts.

Research Motivations and Contributions

The motivation behind this research stems from the increasing need to understand the dynamics of opinion and sentiment within news articles in the digital age. As the accessibility and volume of news content continue to expand, it becomes essential to analyze and quantify the presence of opinions in different domains, such as politics, entertainment, or sports. By examining the opinionated nature of news articles, we can gain insights into the shaping of public discourse, the influence of media on public opinion, and the potential impact on decision-making processes. This work addresses the evolving landscape of news media and its implications for society; policymakers, media organizations, and researchers are invested in

understanding the role of opinions in news reporting to make informed decisions and better understand the information ecosystem.

This research paper contributes to the field of sentiment analysis by conducting an in-depth analysis of a dataset comprising 210,000 news articles from the Huffington Post. The primary objective is to investigate the opinionated nature of news articles across various domains and draw conclusions by inductive reasoning. By leveraging state-of-the-art Natural Language Processing (NLP) techniques and machine learning algorithms, we apply sentiment analysis to quantify and categorize both the polarity and subjectivity expressed within the news articles. Additionally, we aim to identify patterns and trends in sentiment across different domains, shedding light on the degree of polarity within politics, entertainment, sports, and other news categories covered by the Huffington Post. Since the Huffington Post is a reputable news source, with millions of unique monthly visitors, it will provide credible and generalizable conclusions.

The rest of the research paper is structured into the following sections: Related Work, Methodology, Results, and Discussion. The Related Work section reviews the existing studies on applied sentiment analysis. The Methodology section describes the sentiment modeling techniques and data preprocessing steps. The Results section presents the findings obtained from the sentiment analysis of the news articles and the patterns and trends observed. The Discussion section provides a comprehensive interpretation and analysis of the results, highlighting the implications and potential applications of the findings. It also explains the limitations and suggests directions for future research in this domain.

Related Work

Several studies have explored sentiment analysis in the context of news media, uncovering valuable insights into public opinion and media discourse. For instance, Samuels and Mcgonical performed a polarity analysis on 2,225 BBC articles from 2004 to 2005 (S. Meher Taj et al.). They implemented the SentiWordNet 3.0.0 dictionary and found that business and sports topics were more positive (demonstrated enthusiasm, interest, optimism, etc.) whereas entertainment and tech were more negative domains (demonstrated hate, sadness, violence, discrimination, etc.). Similarly, Balahur et al. implemented the SentiWordNet dictionary on EMM applications NewsBrief and MedISys, using 1,592 quotes (Balahur et al.). They also created a new approach towards news subjectivity analysis specifically. Rozado et al. conducted a longitudinal analysis of emotion in 23 million headlines from a variety of US news outlets, using the SiEBERT transformer language model, and found an increase in sentiment negativity over time as well as among right-leaning media (Rozado et al.). Studies have investigated trends like agenda-setting in the news during the 2016 presidential campaign using sentiment analysis tools from Stanford CoreNLP, in order to study public perception and coverage (Albanese et al.). Reis et al. looked at 69,907 news headlines from The New York Times, BBC, Reuters, and Dailymail (Reis et al.). Using the Sentistrengh machine learning model which is optimized for short texts, they discovered that the sentiment of a news headline is strongly related to the popularity of that article.

While research, including the studies above, has explored sentiment in the news, no research has been conducted that analyzes several facets of news articles' polarity and subjectivity, including both article content and headlines, to evaluate opinion dynamics more holistically. This research gap is filled in this work as it provides empirical evidence regarding the opinion dynamics present in contemporary news media and offers insights into the role of opinions in shaping public discourse. Hence we pose and answer the following research questions:

1) What are the most polarized and subjective domains within news media? Do polarity and subjectivity trends align?
2) How has the sentiment of the news media changed over time?
3) How subjective are the most popular authors within the Huffington Post?
4) How does the sentiment of a news headline/title compare to that of the article itself, and what does this reveal about news representation?

Methodology

The original base dataset, titled the News Category Dataset, was curated by author and engineer Rishabh Misra and was last updated in October 2022 (Misra). The dataset consisted of 209,527 articles along with a corresponding link, headline, category, short description, authors, and dates. The articles spanned from 2012 to 2022, ensuring a diverse representation of news across 40 unique domains, such as U.S. News, Culture and Arts, and Sports.

To preprocess and adapt the dataset, we performed an intensive data-cleaning process. We first removed duplicate rows, and then screened for articles from sources other than the Huffington Post, with invalid links, or with no body content. Afterward, the content of each article was extracted by leveraging the Beautiful Soup web scraping library to retrieve the main body text from the respective URLs provided in the dataset. This process allowed us to gather the complete content of each article for subsequent analysis.

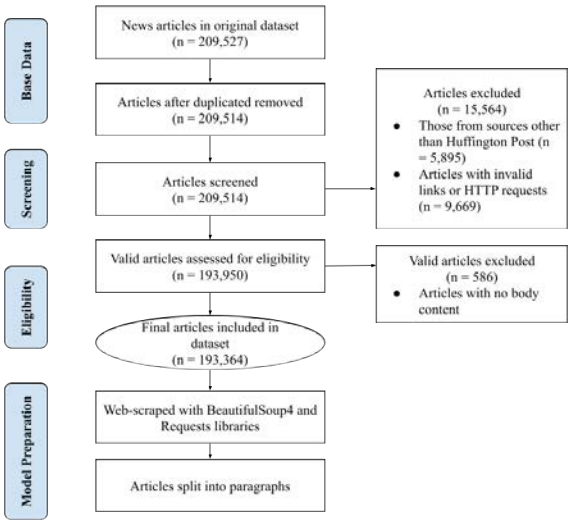


Fig 1: Flowchart displays steps taken to preprocess the dataset through screening data, excluding articles, and web scraping.

To perform sentiment analysis on the dataset, we broke down the concept of “sentiment” into polarity and subjectivity. Polarity is defined as simply how positive, negative, or neutral the semantic orientation of a particular text is (Liu). Polarity value typically ranges from -1 to 1, with -1 being the most negative, 1 being the most positive, and 0 neutral. Subjectivity is defined as whether a text is considered more objective or opinionated based on word choice (Liu). It typically ranges from 0 to 1, where 0 is the most objective and 1 is very subjective. By breaking down sentiment into both polarity and subjectivity, we capture a more holistic representation of opinion dynamics.

Lexicon VADER and TextBlob Models

Our sentiment analysis approaches employed VADER and TextBlob, lexicon-based approaches, and BERT models, which utilize a machine learning-based approach.

For our lexicon approaches to sentiment analysis, we utilized the VADER model to measure sentiment polarity and the TextBlob model to measure sentiment subjectivity. VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon-based sentiment analysis model developed by C.J. Hutto and Eric Gilbert in 2014 (Hutto and Gilbert). It was chosen for this research due to its ability to handle short, informal, and context-rich texts effectively, making it suitable for analyzing the polarity of article paragraphs. The model is built around a pre-established lexicon containing words and phrases, each associated with a polarity score ranging from -1 (extremely negative) to 1 (extremely positive).

To analyze the sentiment of an input text, VADER first tokenizes the text into individual words and punctuation marks. For each word, the model looks up the corresponding sentiment score in the lexicon, which serves as an indicator of the word's sentiment intensity. Additionally, VADER takes capitalization into account, increasing the sentiment score for words in all caps to reflect stronger emotional meaning. The model is also equipped to handle degree modifiers, such as intensifiers and degree adverbs, which can modify the sentiment of adjacent words, adjusting the sentiment scores of words based on these modifiers, and capturing their strengthened or weakened sentiment.

Furthermore, VADER is aware of valence shifting, where certain words may change their polarity based on the context in which they appear. For example, "not good" has a negative sentiment, while "not bad" has a positive sentiment. The model reverses the polarity scores of words following negations (e.g., "not," "never") to correctly account for valence shifts.

After processing all the words in the input text, VADER calculates a compound sentiment score for the entire text. The compound score represents the overall sentiment intensity of the text and is calculated as the sum of all the individual word polarity scores, normalized between -1 and 1. The model then applies valence-based rule heuristics and grammatical rules to further adjust the sentiment scores based on specific patterns and contexts in the text, such as handling contrasting sentiments with conjunctions like "but." This simplicity, speed, and effectiveness in handling social media data have made VADER an approach well-suited for this particular task.

In conjunction with the VADER model, we employed the TextBlob model to assess the subjectivity of the article content in the news dataset. TextBlob is a versatile natural language processing library that offers sentiment analysis capabilities. While VADER primarily focuses on polarity, TextBlob extends its functionality to capture the degree of personal opinion or factual information present in a given text.

Subjectivity, in the context of TextBlob, is quantified as a value ranging from 0 to 1. A lower subjectivity score suggests that the text contains more factual information and less personal opinion, while a higher subjectivity score indicates a greater presence of subjective language and personal viewpoints within the text. One distinctive feature of TextBlob's subjectivity calculation is its consideration of word intensity. In the English language, adverbs often act as modifiers to intensify or diminish the sentiment of the words that follow. For instance, the phrase "very good" intensifies the positive sentiment of "good." When an adverb modifies a word, it can affect the subjectivity score by indicating the strength of the sentiment expressed. TextBlob takes such intensity into account when determining subjectivity. This nuanced approach allows TextBlob to discern whether a word contributes to the personal opinion expressed in the text or merely presents factual information, and thus makes TextBlob a valuable lexicon-based subjectivity analysis approach.

BERT Model

BERT (Bidirectional Encoder Representations from Transformers) is a transformative language model introduced by Google for natural language processing tasks (Devlin et al.). It operates through pre-training on a vast corpus of text using two primary steps: Masked Language Model (MLM) and Next Sentence Prediction (NSP). In MLM, BERT learns to predict masked words in a sentence, allowing it to comprehend bidirectional dependencies and capture context effectively. Meanwhile, NSP helps the model understand relationships between sentences. BERT's architecture consists of multiple transformer encoder layers with self-attention mechanisms that process input tokens in parallel, enabling the model to weigh word importance and focus on relevant contextual information.

To perform sentiment analysis using the state-of-the-art BERT transformer, we employed two models: a polarity model and a subjectivity model.

The polarity detection model we used was TweetEval from Cardiff University, which used the RoBERTa structure (Barbieri et al.). To perform sentiment analysis with TweetEval, researchers employ RoBERTa, an enhanced version of BERT (Liu et al.). RoBERTa follows a similar pre-training process to BERT but refines the methodology, leading to better contextual representations. RoBERTa is fine-tuned on the training set and optimized using the validation set with techniques such as hyperparameter tuning and early stopping. TweetEval was fine-tuned on labeled Twitter sentiment datasets to learn the relationships between the textual features of tweets and sentiment labels. The model predicts the sentiment polarity of an article by outputting probabilities for each sentiment class (positive, negative, or neutral). The polarity score is calculated based on these probabilities, typically by assigning the class with the highest

probability as the predicted sentiment and using the corresponding probability as the polarity score. TweetEval was chosen because it was trained on an extremely large dataset of informative and content-rich tweets, which applies well to detecting traces of sentiment in news article paragraphs.

The subjectivity model, developed by Bilendi Technology, was based on the BiBert architecture and was trained on subjectivity-labeled datasets. It predicts the degree of subjectivity expressed in an article, providing a subjectivity score ranging from 0 (extremely objective) to 1 (extremely subjective).

Facets Analyzed

We specifically analyzed four main facets: sentiment versus category, sentiment versus time, sentiment by author, and the Title-Article Difference Score (TADS) versus category. Furthermore, for all metrics, we compared both VADER and BERT results amongst each other. The Title-Article Difference Score is a metric that takes into account the polarity of the article relative to its title. It is calculated by subtracting the absolute value of the polarity score of the title from that of the article; values closer to -1 indicate that the title is more polarized than the article, and values closer to 1 indicate that the article is more polarized than the title, and values around 0 display no difference between article and title polarity. This metric reveals the extent to which titles display an accurate representation of an article's polarity.

Results

Sentiment by Category

We examined the sentiment polarity scores for different categories of news articles. The results (presented in Figure 3) indicate patterns in polarity across different domains; categories related to style, food, entertainment, and art all produced highly positive polarity across both models. Additionally, categories related to crime, politics, environment, or education consistently displayed moderately negative sentiment across both models.

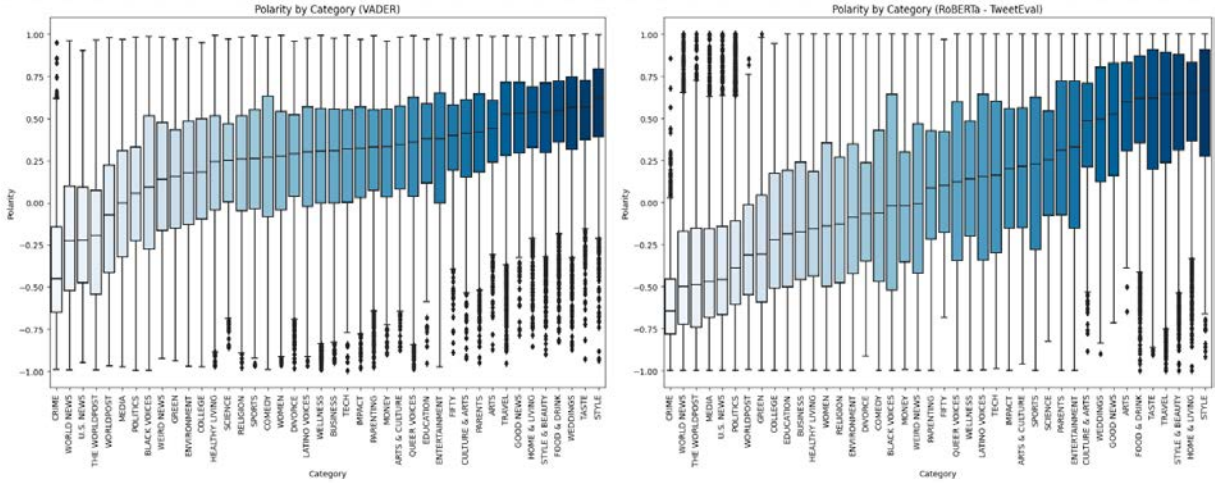


Fig 2: Box and whisker plot of the sentiment polarity distribution across various categories using the VADER and BERT models.

As for the subjectivity analysis (presented in Figure 4), we found that RoBERTa and TextBlob models provided a similar order as to which categories were more and less subjective. The RoBERTa model predicted almost a similar pattern to its sentiment prediction, with style, food, entertainment, and art being the most subjective and crime, politics, environment, or education the most objective. The TextBlob model showed that crime, news, and politics were the most objective, while taste, food and drink, and style were the most subjective.

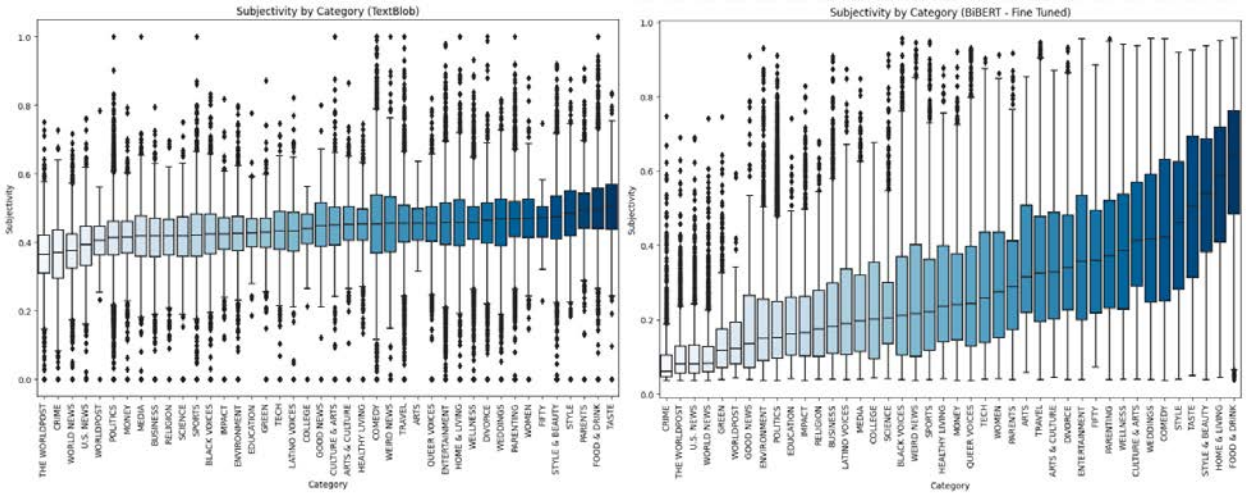


Fig 3: Box and whisker plot of the sentiment subjectivity distribution across various categories using the TextBlob and BERT models.

Sentiment Over Time

To observe how sentiment patterns have changed over time, we plotted the average sentiment polarity scores of the articles dating from 2012 to 2022. Figure 5 showcases the temporal trends in polarity using the VADER and BERT models. The plot reveals an overall decline in polarity from 2012 to 2022 (average change in polarity (ACiP) across VADER and RoBERTa equal to ~0.32). Specifically, there was an increase in polarity from 2012 to 2013, followed by a slow decrease from 2013 to 2018. From 2018 to 2019, there was a sharp downward spike, which was soon reverted by 2020. After 2020, the average polarity declined until 2022. Although the increases and decreases in values are minor, it is important to note that the fact that the average sentiment of almost 200,000 articles shifted is significant and indicates the influence of external factors on sentiment expression.

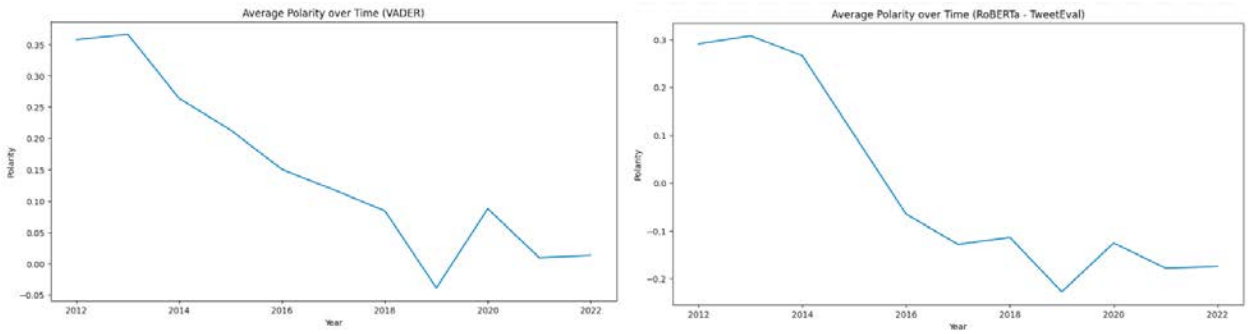


Fig 4: Line plot of the sentiment polarity over the years 2012 to 2022 from both VADER and RoBERTa models.

As for the subjectivity analysis (presented in Figure 6), we found that RoBERTa and TextBlob models provided similar temporal trends to each other and to the polarity graphs. Average subjectivity scores were stagnant until 2014 when they experienced a strong decline until 2019. From 2019 to 2020, there was an increase, and then after 2020 a decline until 2022.

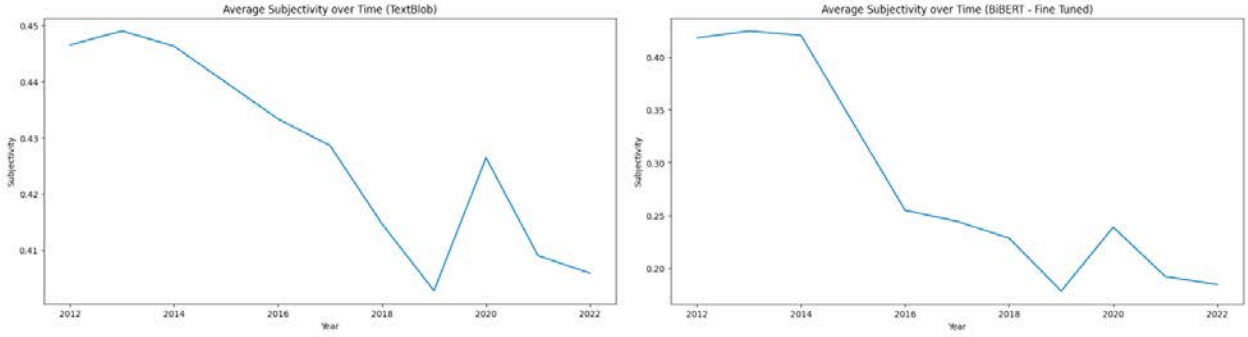


Fig 5: Line plot of the sentiment subjectivity over the years 2012 to 2022 from both TextBlob and RoBERTa models.

Sentiment by Author

We investigated the sentiment expressed by the 20 authors who have published the most articles in the Huffington Post that were in our dataset. Figure 7 illustrates the sentiment subjectivity scores associated with each author, allowing us to discern variations in sentiment based on authorship. Notably, authors Dana Oliver, Rebecca Adams, Andy McDonald, and Carly Ledbetter displayed the most subjectivity on average in their work.

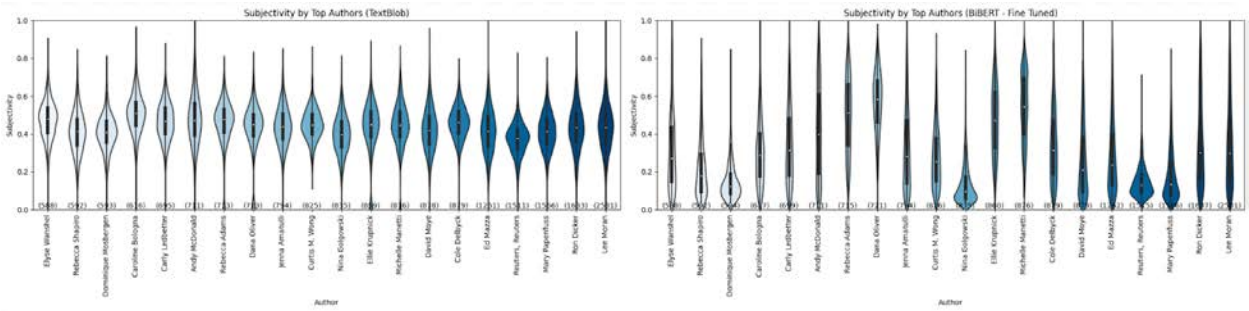


Fig 6: Violin plot of the sentiment subjectivity across the 20 authors with the highest mean subjectivity, between VADER and RoBERTa models.

Title-Article Difference Score

In order to analyze how the sentiment of the title differs from that of the article, we ran the VADER and RoBERTa polarity models on the headlines of each article. The TADS was calculated by the following formula: $|\text{Polarity}_{\text{Article}} - \text{Polarity}_{\text{Title}}|$. Figure 8 illustrates the TADS across categories for both VADER and BERT models, thus displaying which domains have a more stark contrast between article and headline sentiment. The results indicate that categories related to crime, news, and entertainment all produced higher TADS across both models. Additionally, categories that had moderate polarities as proven by part 1 of the results consistently displayed lower TADS across both models.

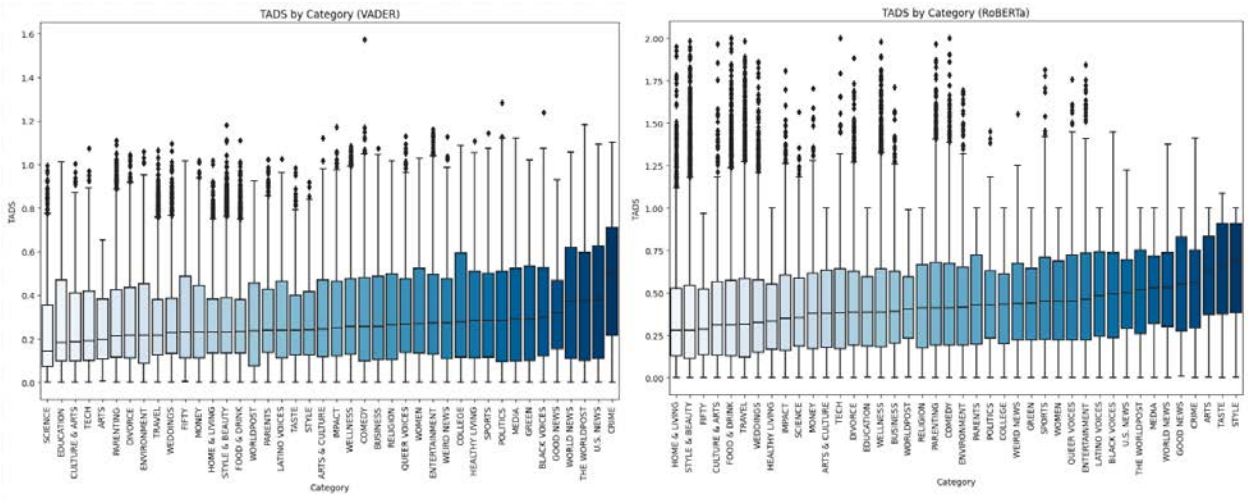


Fig 7: Box and whisker plot of the Title Article Difference Score (TADS) distribution across various categories using the VADER and BERT models.

Discussion

Sentiment by Category

In the analysis of sentiment by category, our study revealed noteworthy findings that shed light on the underlying dynamics of sentiment expression within news articles. The observed patterns in polarity across different domains provide valuable insights into the emotional tone and subjective nature of news content. To better understand the findings, it's essential to delve into the possible explanations and the existing literature that may explain these patterns.

Positive polarity and subjectivity in categories like style, food, entertainment, and art can be attributed to the inherently positive and enjoyable nature of these subjects. Conversely, the consistently negative sentiment in categories like crime and politics aligns with prior research on the "negativity bias." Studies have shown that negative news stories, especially about politics, tend to garner more attention and audience engagement (Trussler). This preference for negative news can be attributed to evolutionary factors where humans have evolved to prioritize

information with potential threats (Liu et al.). In the context of politics, the polarization of public opinion and the adversarial nature of political discourse can contribute to negative sentiment (Geiger).

These findings directly address our first research question: "What are the most polarized and subjective domains within news media? Do polarity and subjectivity trends align?" The analysis shows that certain domains exhibit distinct patterns of sentiment, with some consistently skewing positive or negative. Results suggest that media outlets may inherently approach certain topics with preconceived sentiment orientations, shaping the public's perception of these subjects.

This has implications for media organizations, journalists, and policymakers who aim to engage their audiences effectively. Understanding the emotional resonance of different news categories can guide editorial decisions and content strategies. For instance, recognizing that lifestyle and entertainment topics generally evoke positive sentiment can inform content creation strategies to foster reader engagement and satisfaction.

Moreover, the broader significance of these findings lies in their potential impact on public perception and societal discourse. The consistent negativity observed in certain categories, such as politics and environmental issues, underscores the need for responsible journalism that strives to provide balanced and constructive coverage of these topics. Acknowledging the emotional context within which news articles are received can help media outlets make informed decisions about framing and tone, thereby influencing public opinion and discourse.

Sentiment Over Time

In order to understand the aforementioned temporal patterns, we must consider the various factors that influence public sentiment and media representation.

- The gradual decline in sentiment from 2013 to 2018 may be related to increasing political polarization and the proliferation of negative news. Research has shown that political polarization can lead to more negative media coverage as media outlets cater to their audience's preferences (Tucker et al.).
- The sharp decline in polarity in 2018-2019 could be attributed to significant global events such as political controversies, international conflicts, and natural disasters. Research suggests that major events can lead to increased negative sentiment in news coverage (Sufi).
- The partial rebound in 2020 may be linked to the COVID-19 pandemic and related developments. During crises, the media often plays a critical role in disseminating information and fostering community support, potentially leading to more positive sentiment (Xie et al.).
- The continued decline in sentiment after 2020 may reflect the ongoing challenges and uncertainties associated with the pandemic and other global issues. Prolonged crises can lead to increased negative sentiment as the novelty of the situation wears off, and public fatigue sets in (Barrett and Chen).

These temporal trends directly address our second research question: "How has the sentiment of the news media changed over time?" The analysis demonstrates that sentiment within news media has undergone noticeable fluctuations and suggests that external events and societal changes exert a substantial influence on media sentiment.

The findings regarding sentiment over time carry several implications for media professionals, policymakers, and researchers. Media organizations should be aware of the impact of external events on news sentiment and strive for balanced reporting, particularly during times of crisis. Policymakers can leverage these insights to anticipate shifts in public sentiment and tailor communication strategies accordingly.

Moreover, information science researchers can use this longitudinal analysis as a foundation for further exploration of the intricate dynamics between media representation, public sentiment, and external factors. Understanding these dynamics is crucial for accurately interpreting societal attitudes and the role of media in shaping public discourse.

Sentiment by Author

Our analysis identified authors whose work consistently exhibited higher subjectivity, indicating that they tend to express their opinions more openly in their articles. Notably, authors Dana Oliver, Rebecca Adams, Andy McDonald, and Carly Ledbetter stood out in terms of their subjectivity levels.

To understand these findings, it is important to consider the factors that may influence the sentiment expressed by authors in their articles. One significant potential factor is that different journalists have varying writing styles and areas of expertise. Some journalists may specialize in opinion pieces or analysis, where subjectivity is expected and encouraged. Dana Oliver, one of the highly subjective and popular authors, writes many articles about beauty and style; in these fields, objective news reporting is less expected or required. Similarly, Andy McDonald, a popular yet subjective author, is a comedy writer who engages with political and entertainment content; thus it makes sense that his writing would report highly subjective due to the nature of his work. In contrast, other journalists may focus on objective reporting, which requires a more neutral tone.

Additionally, news organizations often have editorial guidelines and standards that influence how journalists frame their stories. Journalists may align their writing with the editorial stance and tone of their news outlet. Furthermore, journalists, like all individuals, hold personal beliefs and values that can influence their writing. These beliefs may manifest as subjectivity in their articles, particularly in topics they feel passionate about. Finally, journalists may consider their target audience when crafting their articles. They may adjust their writing style and level of subjectivity to cater to the preferences of their readers. Carly Ledbetter, another popular and subjective author, posts many articles related to pop culture; therefore, her audience would likely be searching for the most intriguing stories motivating her to publish only significant and polarizing events.

These findings directly address our third research question: "How subjective are the most popular authors within the Huffington Post?" By identifying authors who exhibit higher subjectivity levels in their work, we provide insights into the diversity of voices and perspectives within news media.

The implications of understanding author-specific sentiment extend beyond the realm of academic research. While many argue that objectivity in the news has long been the gold standard of American journalism and must be maintained (Ward), others claim that certain kinds of subjective reporting allow authors to show a personal voice, tell a story, and represent their diverse perspectives to allow readers to understand other points of view (Birgitte Kjos Fonn et al.). Media organizations can leverage these insights to recognize and celebrate the diversity of voices within their ranks. Embracing a range of writing styles and perspectives can enhance the richness of news reporting and cater to a broader readership.

Furthermore, readers benefit from being aware of the writing style and subjectivity levels of specific authors. This knowledge allows them to approach news articles with a critical eye, considering the author's perspective and potential biases.

Title-Article Difference Score

To understand these findings, it is essential to consider the factors that may contribute to the differences between headlines and article content. Firstly, in an effort to capture readers' attention and generate clicks, news outlets may employ sensational or provocative headlines that do not necessarily reflect the full nuance of the article content (Jang et al.). This practice can lead to higher TADS as headlines prioritize engagement over accuracy. Especially within domains of crime (Conklin) and politics (Andrew), events are often exaggerated to draw attention. Additionally, news headlines often have limited space to convey information. As a result, they may oversimplify or emphasize certain aspects of a story, potentially leading to disparities with the article's comprehensive sentiment. Furthermore, editorial teams may influence headline creation, emphasizing certain angles or perspectives to align with the editorial stance or the outlet's brand. These decisions can impact the sentiment conveyed in headlines. Finally, the preferences of readers can also influence headline choices. News outlets may tailor headlines to cater to reader expectations or biases, potentially leading to disparities in the article's actual polarity.

These findings directly address our fourth research question: "How does the sentiment of a news headline/title compare to that of the article itself, and what does this reveal about news representation?" By quantifying the difference between headline and article sentiment, we reveal the extent to which headlines accurately represent the content, thereby offering insights into the potential impact of headline sensationalism or simplification on reader perception.

Our study contributes to filling a research gap by quantifying the extent of headline-article disparity across diverse news categories. While prior research has explored headline sensationalism or clickbait tendencies, our analysis offers a numerical measure of this

disparity and its variation across different domains, contributing to a deeper understanding of media practices.

The implications of understanding headline-article disparity extend to media practitioners and news consumers. For media organizations, recognizing the potential impact of headline sensationalism on reader perception is vital. Striking a balance between engaging headlines and accurate representation of article content is crucial for maintaining trust and credibility.

News consumers can benefit from being aware of the potential disconnect between headlines and article sentiment. Critical thinking and a willingness to delve into the full article can help readers better assess the accuracy and nuance of news reporting.

Conclusion

This research has uncovered the intricate dynamics of sentiment within news articles, shedding light on the emotional tone and subjectivity across diverse categories, temporal shifts, author-specific styles, and headline disparities. The findings not only address the research questions posed but also offer valuable insights with far-reaching implications. Understanding the emotional resonance of news categories informs editorial decisions and promotes responsible journalism. The study underscores the role of external factors in shaping media sentiment, emphasizing the importance of balanced reporting during crises. Identifying highly subjective authors acknowledges diversity in news media voices but encourages critical reading by news consumers. Moreover, quantifying headline-article disparities provides a numerical measure of media practices, aiding both media practitioners and discerning readers. In an era of evolving news media, this research contributes to a deeper understanding of sentiment dynamics and their impact on public perception and discourse.

Acknowledgments

I extend my sincere gratitude to Dr. Jerry Ng (Ph.D., MIT) for his invaluable support and mentorship throughout the course of this research project. His expertise, guidance, and insight have played a crucial role in shaping the direction and quality of this study. I am deeply appreciative of his dedication and patient encouragement, without which this research would not have been possible.

Works Cited

Lule, Jack. “Understanding Media and Culture: An Introduction to Mass Communication.” BiblioBoard Library Catalog (Open Research Library), 2016, https://doi.org/10.24926/8668.2601. Accessed 25 Sept. 2023.

Kavanagh, et al. News in a Digital Age. RAND, 2019, www.rand.org/pubs/research_reports/RR2960.html. Accessed 25 Sept. 2023.

Chong, Phillipa. “Valuing Subjectivity in Journalism: Bias, Emotions, and Self-Interest as Tools in Arts Reporting - Phillipa Chong, 2019.” Journalism, 2019, journals.sagepub.com/doi/10.1177/1464884917722453. Accessed 25 Sept. 2023.

“Machine Learning: Trends, Perspectives, and Prospects.” Science, 2015, www.science.org/doi/10.1126/science.aaa8415. Accessed 25 Sept. 2023.

Khurana, Diksha, et al. “Natural Language Processing: State of the Art, Current Trends and Challenges.” Multimedia Tools and Applications, vol. 82, no. 3, Springer Science+Business Media, July 2022, pp. 3713–44, https://doi.org/10.1007/s11042-022-13428-4. Accessed 25 Sept. 2023.

Liu, Bing. Sentiment Analysis and Opinion Mining. 2012, www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf.

Taboada, Maïté, et al. “Lexicon-Based Methods for Sentiment Analysis.” Computational Linguistics, vol. 37, no. 2, Association for Computational Linguistics, June 2011, pp. 267–307, https://doi.org/10.1162/coli_a_00049. Accessed 25 Sept. 2023.

Singh, Jaspreet, et al. “Optimization of Sentiment Analysis Using Machine Learning Classifiers.” Human-Centric Computing and Information Sciences, vol. 7, no. 1, Springer Nature, Dec. 2017, https://doi.org/10.1186/s13673-017-0116-3. Accessed 25 Sept. 2023.

S. Meher Taj, et al. “Sentiment Analysis of News Articles: A Lexicon Based Approach.” 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (ICoMET), Jan. 2019, https://doi.org/10.1109/icomet.2019.8673428. Accessed 25 Sept. 2023.

Balahur, Alexandra, et al. “Sentiment Analysis in the News.” ArXiv.org, 2013, arxiv.org/abs/1309.6202. Accessed 25 Sept. 2023.

Rozado, David, et al. “Longitudinal Analysis of Sentiment and Emotion in News Media Headlines Using Automated Labelling with Transformer Language Models.” PLOS ONE, vol. 17, no. 10, Public Library of Science, Oct. 2022, pp. e0276367–67, https://doi.org/10.1371/journal.pone.0276367. Accessed 25 Sept. 2023.

Albanese, Federico, et al. “Analyzing Mass Media Influence Using Natural Language Processing and Time Series Analysis.” Journal of Physics, vol. 1, no. 2, IOP Publishing, July 2020, pp. 025005–5, https://doi.org/10.1088/2632-072x/ab8784. Accessed 25 Sept. 2023.

Reis, Julio, et al. “Breaking the News: First Impressions Matter on Online News.” ArXiv.org, 2015, arxiv.org/abs/1503.07921. Accessed 25 Sept. 2023.

Misra, Rishabh. “News Category Dataset.” ArXiv.org, 2022, arxiv.org/abs/2209.11429. Accessed 25 Sept. 2023.

Liu, Bing. “Sentiment Analysis and Subjectivity.” ResearchGate, unknown, 2010, www.researchgate.net/publication/228667268_Sentiment_analysis_and_subjectivity. Accessed 25 Sept. 2023.

Hutto, Clayton J., and Éric Gilbert. “VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text.” Proceedings of the International AAAI Conference on Web and Social Media, vol. 8, no. 1, May 2014, pp. 216–25, https://doi.org/10.1609/icwsm.v8i1.14550. Accessed 25 Sept. 2023.

Devlin, Jacob, et al. “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding.” ArXiv.org, 2018, arxiv.org/abs/1810.04805. Accessed 25 Sept. 2023.

Barbieri, Francesco, et al. “TweetEval: Unified Benchmark and Comparative Evaluation for Tweet Classification.” ArXiv.org, 2020, arxiv.org/abs/2010.12421. Accessed 25 Sept. 2023.

Liu, Yinhan, et al. “RoBERTa: A Robustly Optimized BERT Pretraining Approach.” ArXiv.org, 2019, arxiv.org/abs/1907.11692. Accessed 25 Sept. 2023.

Trussler, Marc. “Consumer Demand for Cynical and Negative News Frames - Marc Trussler, Stuart Soroka, 2014.” The International Journal of Press/Politics, 2014, journals.sagepub.com/doi/10.1177/1940161214524832. Accessed 25 Sept. 2023.

Liu, Yuting, et al. “The Priority of Goal-Relevant Information and Evolutionarily Threatening Information in Early Attention Processing:Evidence from Behavioral and ERP Study.” Scientific Reports, vol. 10, no. 1, Nature Portfolio, May 2020, https://doi.org/10.1038/s41598-020-65062-5. Accessed 25 Sept. 2023.

Geiger, Abigail. “Political Polarization in the American Public.” Pew Research Center - U.S. Politics & Policy, 12 June 2014, www.pewresearch.org/politics/2014/06/12/political-polarization-in-the-american-public/. Accessed 25 Sept. 2023.

Tucker, Joshua A., et al. “Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature.” Social Science Research Network, RELX Group (Netherlands), Jan. 2018, https://doi.org/10.2139/ssrn.3144139. Accessed 25 Sept. 2023.

Sufi, Fahim. “Identifying the Drivers of Negative News with Sentiment, Entity and Regression Analysis.” International Journal of Information Management Data Insights, vol. 2, no. 1, Elsevier BV, Apr. 2022, pp. 100074–74, https://doi.org/10.1016/j.jjimei.2022.100074. Accessed 25 Sept. 2023.

Xie, Lola, et al. “Building Community Resilience on Social Media to Help Recover from the COVID-19 Pandemic.” Computers in Human Behavior, vol. 134, Elsevier BV, Sept. 2022, pp. 107294–94, https://doi.org/10.1016/j.chb.2022.107294. Accessed 25 Sept. 2023.

“Social Repercussions of Pandemics.” IMF, 29 Jan. 2021, www.imf.org/en/Publications/WP/Issues/2021/01/29/Social-Repercussions-of-Pandemics-50041. Accessed 25 Sept. 2023.

Ward, Stephen. “Objectivity and Bias in Journalism.” Oxford Research Encyclopedia of

Communication, Mar. 2019, https://doi.org/10.1093/acrefore/9780190228613.013.853. Accessed 25 Sept. 2023.
Birgitte Kjos Fonn, et al. “Putting a Face on It.” Directory of Open Access Books (OAPEN Foundation), 2017, https://doi.org/10.23865/noasp.28. Accessed 25 Sept. 2023.
Jang, Joonwon, et al. “Detecting Incongruent News Headlines with Auxiliary Textual Information.” Expert Systems with Applications, vol. 199, Elsevier BV, Aug. 2022, pp. 116866–66, https://doi.org/10.1016/j.eswa.2022.116866. Accessed 25 Sept. 2023.
Conklin, Michael. “The Truth Can Be Deceiving: How Criminal Justice Headlines Are Misinterpreted.” Ssrn.com, 26 Apr. 2020, papers.ssrn.com/sol3/papers.cfm?abstract_id=3585743. Accessed 25 Sept. 2023.
“Political Journalism Represented by Headline News: Canadian Public and Commercial Media Compared on JSTOR.” Jstor.org, 2013, www.jstor.org/stable/43298263. Accessed 25 Sept. 2023.

Comparing Prominent Generative Language Models for Classifying Political Alignment Of Limited Context Bigrams By Sankalp Singh

Abstract

Generative Language Models (GLMs) have transformed artificial intelligence by enabling human-like text generation across diverse applications. This study delves into GLM-generated content, focusing on the ability of GLMs to classify politically charged bigrams from congressional speeches with minimal context by creating a Python[41] script for each GLM to prompt the models en masse. The investigation studies three major GLMs: Google's Bard, OpenAI's GPT-3.5 Turbo, and OpenAI's GPT-4. Using prompts encompassing target bigrams, congress details, and polarity values, the study assesses the models' proficiency in aligning bigrams with left-leaning or right-leaning ideologies. The dataset originates from Stanford University[1], comprising of parsed political bigrams from congressional speeches and corresponding political polarity values for each bigram. Despite expected deviations from the exact Stanford benchmark polarity values, the GLMs show varying degrees of accuracy in political classification, with GPT-4 exhibiting the highest proficiency. The findings underline GLMs' capacity to consider context and infer political associations based on their training data. They also emphasize the complexities of language, ideology, and context. This research contributes to understanding GLMs' strengths, limitations, and implications in political discourse analysis.

Key Terms*: Generative Language Model (GLM), Bard, GPT-3.5 Turbo, GPT-4, Few-Shot Learning*

Introduction

Generative language models represent a groundbreaking advancement in artificial intelligence (AI), revolutionizing the way machines comprehend and produce human-like text. Their significance spans across a spectrum of applications, encompassing creative content generation, language translation, text summarization, and even human-like conversation. By harnessing the power of deep learning techniques, GLMs have unlocked unprecedented capabilities, contributing to the evolution of various industries and sectors.

As versatile and impressive as GLMs are, they are not without their shortcomings. One of the glaring concerns associated with these models is their susceptibility to generating misinformation. This vulnerability arises from the nature of their training data, which, despite being extensive, can inadvertently encompass erroneous, biased, or unverified information. Consequently, GLMs may inadvertently propagate false narratives, as they may lack the discernment to differentiate between accurate and inaccurate content.

An illustrative study by Bolukbasi et al. (2016)[2] brought to light the issue of bias in GLMs. In their work, they exposed how word embeddings trained on datasets of news articles

[41] https://github.com/cubicmight/GLMPoliticalTest

could inadvertently perpetuate biased or prejudiced language, thereby fostering negativity toward certain societal groups. The implications of this discovery are profound, as it sheds light on the unintended consequences of deploying AI models in contexts that demand neutrality and accuracy.

Against this backdrop, this research elucidates the models' potential capacity to replicate realistic nuanced political discourse. We make this comparison because certain standout terms and bigrams in prompts to language models can greatly affect the information they output. The investigation focuses on evaluating the proficiency of three prominent GLMs in accurately classifying politically charged bigrams with limited contextual information. The contextual cues provided to the models include the specific congress during which the bigram was articulated, alongside the corresponding start and end dates of that legislative session. These four contextual elements, along with directions, collectively serve as the input prompt for the GLMs to generate a polarity value, indicating the model's assessment of the bigram's alignment with left-leaning or right-leaning political ideologies.

To benchmark the generated polarity values against a credible and well-established source, the research employs the Congressional Record for the 43rd-114th Congresses: Parsed Speeches and Phrase Counts dataset provided by Stanford University. Even though the primary nature of this task involves regression, wherein the model generates a continuous numerical value on a scale, we are opting to treat it as a classification task. In doing so, we aim to ascertain whether the model arrives at the same conclusion as the original polarity values in terms of the political standing conveyed. Here, political standing is specifically defined by the ideological association attributed to a given bigram, signified by the polarity value's sign. In this context, negative values correspond to left-leaning affiliations, whereas positive values indicate a right-leaning inclination. We did this because of two key factors: First, the realization that the polarity values provided by Stanford might not comprehensively capture the entire spectrum of global political intricacies. Second, the constrained information environment within our study prevents the model from replicating the exact numerical values of the original polarity assessments.

Data

The data utilized for this research is obtained from Stanford University's dataset named "Congressional Record for the 43rd-114th Congresses: Parsed Speeches and Phrase Counts dataset." This extensive dataset is organized to encompass the proceedings of diverse congressional sessions, offering polarity values linked to specific bigrams. The core emphasis of this study centers on Congresses 95 through 114.

The calculation of partisanship bigrams in the Stanford database is achieved by computing the expected posterior probabilities for a given bigram being associated with either the Republican or Democratic party. This involves incorporating parameters like probabilities of party affiliation (qR and qD) and a polarity function (ρt(xit)) that captures session-specific context. The partisanship of a particular bigram in a session is quantified as an average of these

values across active speakers in that session. This measurement offers both the direction and magnitude of partisanship, with positive values representing Republican-leaning bigrams, negative values indicating Democratic-leaning bigrams, and the absolute value signifying the extent of partisanship. The model evaluates the political leaning of individual bigrams within specific legislative sessions.

To ensure the data's quality and suitability for analysis in this study, a series of preprocessing steps were applied. The initial phase involved addressing issues with misspelled words within bigrams that were present in Stanford's dataset. These misspellings likely occurred during the creation of the Stanford dataset when stemming and lemmatization techniques were applied, as most of the misspelled bigrams were bigrams with words that were simply roots and were not real words. This step was executed using the Python programming language in combination with the Natural Language Toolkit (NLTK) library. Additionally, the NLTK corpus was used to identify and exclude non-English bigrams and bigrams that contained non-English words. Consequently, the dataset was refined to encompass exclusively English-language bigrams.

Moreover, we implemented a systematic removal of bigrams that comprised words with minimal occurrence. This elimination process was driven by the criterion that all words forming a bigram must exhibit a minimum frequency of five instances across the entire corpus. This measure mitigated inaccuracies arising from infrequent, obscure, or misspelled words. Furthermore, to eliminate instances of tautological phrasing, bigrams comprised of identical words were expunged from the dataset.

By enforcing this criterion, we aimed to enhance the reliability of the bigrams used in the analysis. We aimed to ensure that the models' interpretation of these bigrams remains robust and accurate. Removing bigrams with unfavorable characteristics was intended to prevent the models from encountering mistakes or misclassifications caused by factors such as misspellings.

Following the preparatory stages, we compiled a refined list of bigrams that exhibited the highest degrees of polarity across all congresses. From the expansive dataset spanning congresses 95 through 114, we chose the forty bigrams with the largest magnitudes in polarity scores.

In total, there were fifteen left-leaning terms and twenty-five right-leaning terms. In terms of their polarity value ranges, left-leaning bigrams ranged from a minimum of eleven to a maximum of negative one hundred and four, while right-leaning bigrams spanned from a minimum of seven to a maximum of eighty-five. Finally, we associated each bigram with its corresponding congress and assigned start and end dates based on the congress number to finish preprocessing the data.

Methodology and Results–Methodology

A standardized prompt (Figure 1) was then formulated for the GLMs, which included the target bigram, the relevant congress, and its associated start and end dates. The prompt also provided a clear framework for interpreting polarity values, where positive, negative, and neutral

numbers represented right-leaning Republican, left-leaning Democratic, and neutral bigrams, respectively. The prompt also directed the model to limit the polarity values to a range of -110 to 110, aligning with the original polarity values of the forty within that specific range. To facilitate comprehension, example outputs were included, transforming the task into a few-shot learning scenario. In few-shot learning, the model is presented with a limited number of examples or data points, as was the case here with the provided instances of target bigrams, congressional contexts, start and end dates, and their associated polarity values.

In this context, few-shot learning allowed the model to acquire the ability to generalize from these limited examples and apply its learned knowledge to make accurate predictions for similar cases that it had not explicitly encountered during training. This capability is particularly valuable when dealing with nuanced and context-dependent tasks like our classification of political bigrams, where each bigram may have unique characteristics that aren't easily captured with the limited amount of data points we provide the model to conclude from.

Model testing involved the selection of three prominent GLMs: Google's Bard, OpenAI's GPT-3.5 Turbo, and OpenAI's GPT-4. A consistent prompt was presented to each model to ensure uniform input conditions. To account for potential response variance, each bigram was subjected to three trials per model. This approach aimed to capture a comprehensive range of potential model-generated outcomes and establish an average result for subsequent analysis.

The analysis phase entailed the averaging of the three trial results for each bigram and model. These averaged predicted polarity values were then graphed against the corresponding polarity values from the Stanford dataset. This graphical representation provided a visual means of comparing the predictions of the selected GLMs with established reference values.

Given the below phrase, the congress the phrase was said in, and the start and end date of that congress, and the fact that a positive number means that the phrase is a right-leaning Republican phrase, a negative number means that the phrase is a left-leaning Democratic phrase, and 0 means that it is a neutral phrase, output the polarity value for the given phrase. The polarity value should be on a scale of -110 to 110.

Example:
Phrase: red tape
Start Date: January 4, 1977
End Date: January 3, 1979
Congress: 95
Output: 30
Example:
Phrase: interest rate
Start Date: January 3, 2013
End Date: January 3, 2015
Congress: 113
Output: -100

Example:
Phrase: repeal afford
Start Date: January 3, 2013
End Date: January 3, 2015
Congress: 113
Output: -39
Phrase:
Start Date:
End Date:
Congress:

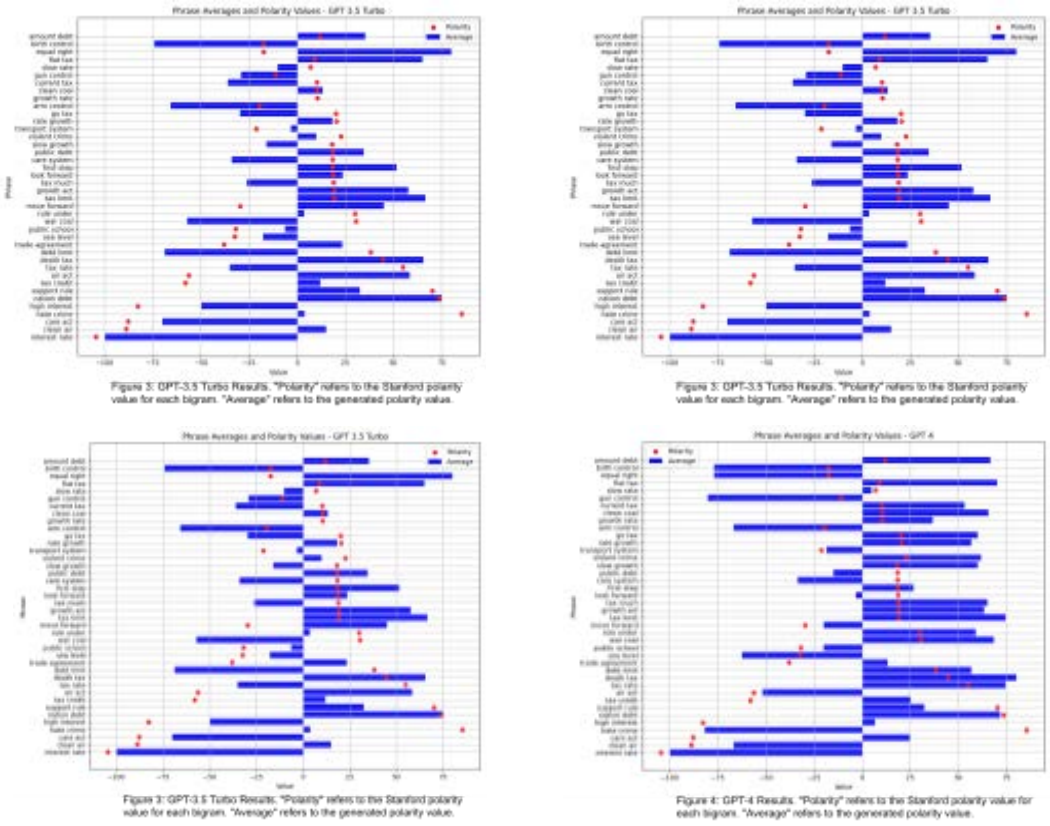


Figure 1: The prompt used for each of the GLMs

Results

The methodology employed in this research yields valuable insights into the performance and capabilities of the selected GLMs in accurately classifying politically charged bigrams with limited contextual information. The findings provide a comprehensive overview of the models' strengths, limitations, and potential real-world implications.

Firstly, it's noteworthy that the three GLMs chosen for this study – Google's Bard GLM, OpenAI's GPT-3.5 Turbo, and OpenAI's GPT-4 – demonstrate varying degrees of accuracy in predicting the political standing of the selected bigrams.

The analysis unveils a clear pattern in which Bard, demonstrating the lowest proficiency among the three, as depicted in Figure 2, incorrectly categorizes the political stance of 19 bigrams. GPT-3.5 Turbo displays a marginal performance improvement with 16 misclassifications, as depicted in Figure 3, whereas GPT-4 excels with the highest accuracy, misjudging only 8 bigrams, as depicted in Figure 4. This performance hierarchy indicates that more advanced iterations of GLMs are better equipped to infer the political leaning of the provided bigrams based on limited contextual cues

Additionally, it's noteworthy to examine the distribution of errors when considering left-leaning and right-leaning bigrams. Figure 2 shows that Bard incorrectly classifies 9 out of 15 left-leaning terms, yielding an error rate of 60%, and 10 out of 25 right-leaning terms, resulting in an error rate of 40%. As for GPT-3.5 Turbo, Figure 3 shows that the error rate for left-leaning terms stands at 6 out of 15, approximately 40%, while its error rate for right-leaning terms is 10 out of 25, approximately 40%. Finally, as seen in Figure 4, GPT-4's error rate for left-leaning terms is 4 out of 15, approximately 27%, while its error rate for right-leaning terms is 4 out of 25, approximately 16%.

From these error rates, we can again see that GPT 4 performed marginally better than the other two GLMs in both right and left-leaning contexts. However, we must take into account the fact that the amount of left and right-leaning bigrams is not equal. Because there are more right-leaning terms, the models had more chances to correctly classify terms, which means that the results are somewhat skewed.

However, by calculating error rates as percentages, we are standardizing the measure of errors across different sample sizes. This normalization allows us to directly compare the models' accuracy in both left-leaning and right-leaning contexts. The key advantage of this approach is that it accounts for the fact that there are more right-leaning terms than left-leaning terms, ensuring that the models are not unfairly penalized for having to classify a larger number of right-leaning terms. Therefore, in low-context environments, it can be concluded that Bard demonstrates a relatively better performance in classifying right-leaning terms, although both of its error rates are relatively high. GPT-3.5 Turbo, on the other hand, exhibits roughly similar accuracy levels for both left-leaning and right-leaning terms, albeit with somewhat elevated error rates. Finally, GPT-4 excels in categorizing right-leaning terms with a notably lower error rate, with the left-leaning error rate not far behind.

The fact that these models exhibit any degree of accuracy in classifying the political standing of the bigrams is a testament to their ability to understand and interpret context (Zhang et al., 2023)[3]. The fundamental principle of context-based learning is evident here, where the models recognize patterns and associations between certain bigrams and political ideologies based on the training data. This phenomenon underscores the models' capacity to leverage extensive datasets to draw meaningful conclusions, even when the provided contextual cues are

minimal during inference, as highlighted in Devlin et al.'s work on BERT (Devlin et al., 2018)[4]. BERT, a deep bidirectional transformer model for language understanding, demonstrated the significance of pre-training models on massive text corpora to develop a rich contextual understanding of language. Such contextual understanding enables models to capture nuanced relationships between words and bigrams, allowing them to make informed predictions in diverse contexts. The success of the models in predicting the general political standings of these bigrams suggests that they have effectively learned from the real-world usage of these bigrams in context, both in congressional discourse and online discussions.

These results highlight each of the models' capability to capture the nuanced ways in which these bigrams are employed in political communication across different contexts. It is worth noting that these bigrams in and of themselves are not politically charged. Rather, it is the contexts in which these bigrams are used which give them underlying connotations. By depriving the models of those politically charged contexts and simply providing them with a bigram, we constrain the models' understanding of the broader historical, socio-political, and cultural contexts that may influence the political connotations of a bigram and make them draw conclusions based on their training data. Due to the vast amount of training data, these models were able to use the context that these bigrams generally appeared in within their training data and draw conclusions based on those contexts, resulting in all three of them showing satisfactory levels of proficiency at this task.

Model performance in natural language processing tasks has shown correlation with the number of model parameters. Generally, larger models with more parameters tend to exhibit improved performance due to their enhanced capacity to capture complex patterns and nuances in language (Brown et al., 2020)[5]. In this study, the same correlation holds true. Bard, with the least satisfactory performance, also has the least parameters, at 137 billion. GPT 3.5-turbo has the second most, at 175 billion, and GPT 4 has by far the most, at 1.7 trillion. Bard and GPT 3.5-turbo most likely performed similarly because they had fewer contexts to rely on when classifying the terms, whereas GPT 4 had a much larger sample size to conclude from.

With regard to the actual polarity values generated by the models, it's important to note that the polarity values provided by Stanford serve as benchmarks and do not encapsulate the full complexity of global political nuances. These values are, in fact, the output of a model themselves, and as such, they reflect the predictions made by that specific model. This distinction is important because it underscores that we are comparing one model's output to another's in this study. Therefore, we are not directly measuring each of the GLM’s ability to match the "real world" political landscape. Despite this, the Stanford dataset is the largest, most comprehensive, and realistic polarity value database and, thus, was the best dataset for this study.

As a result, the models' generated polarity values reflect their interpretations based on the given prompt, and they are shaped by the patterns and associations they have learned from their training data. The deviations from the Stanford values can stem from a variety of factors, including the models' inherent biases, the nuances of political discourse not fully captured in the

prompt, and the complex interplay of political ideologies that cannot be entirely distilled into a single polarity value.

Despite the deviations from the benchmark polarity values, these values remain significant for several reasons. First, they serve as a foundation for evaluating the GLMs' ability to generalize political associations from limited context. The fact that the models can even approximate the political leaning of bigrams based on these prompts underscores their capacity to understand the patterns of language use across different contexts in their training data.

Also, while the polarity values may not capture the entirety of global political intricacies, they represent an attempt to quantify and categorize political associations within a constrained framework. These values, though simplified, offer a measure of political alignment that can be useful for comparative analysis and for understanding the broad ideological connotations of bigrams.

Furthermore, the divergence between the models' predictions and the benchmark values highlights the complexities of language, ideology, and context. It emphasizes that the accurate interpretation of political sentiment requires a rich understanding of historical, cultural, and socio-political factors that extend beyond the immediate textual cues. This recognition underscores the need for models to be continually refined and developed to better grasp the intricate layers of meaning embedded in human communication.

One of the significant takeaways from this analysis is the real-world application potential of GLMs. Their ability to interpret and generate contextually appropriate content has wide-ranging implications. Beyond the scope of this research, these models could be harnessed for tasks like fact-checking, content generation, and political analysis. However, the inherent limitations of context-driven learning must also be acknowledged, highlighting the importance of ongoing research and fine-tuning to enhance the models' discernment capabilities.

Conclusion

In conclusion, this research delves into the intricate landscape of generative language models (GLMs) and their ability to interpret and classify politically charged language with limited contextual information. The study's methodology involved a rigorous analysis of three prominent GLMs—Google's Bard GLM, OpenAI's GPT-3.5 Turbo, and OpenAI's GPT-4—utilizing a standardized prompt framework to assess their proficiency in predicting the political standing of selected bigrams. The results underscore a hierarchy of accuracy among the models, with GPT-4 exhibiting the highest classification accuracy, followed by GPT-3.5 Turbo and Bard. This performance hierarchy aligns with the models' respective parameter sizes and highlights their capacity to draw on extensive training data to understand nuanced political nuances, even in scenarios with limited context.

The findings underscore the potential of GLMs in comprehending and generating contextually appropriate content. However, the research also highlights the models' susceptibility to biases and inaccuracies inherent in their training data, underscoring the need for ongoing refinement and vigilance to ensure responsible deployment. The study not only sheds light on the

capabilities and limitations of GLMs but also emphasizes the intricate layers of meaning embedded in human language, political ideology, and context. As these models continue to shape the landscape of AI and information dissemination, further research is imperative to mitigate the risks associated with misinformation and to maximize the potential for responsible and effective use.

Works Cited

Gentzkow, Matthew, Jesse M. Shapiro, and Matt Taddy. "Congressional Record for the 43rd-114th Congresses: Parsed Speeches and Phrase Counts." Palo Alto, CA: Stanford Libraries [distributor], 2018-01-16. https://data.stanford.edu/congress_text

Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., & Kalai, A. T. "Man is to computer programmer as woman is to homemaker? Debiasing word embeddings." In Advances in neural information processing systems, pp. 4349-4357 (2016).

Zhang, Y., Wang, X., Li, Y., & Liu, Z. "How do generative language models learn from the internet? A survey of methods and challenges." arXiv preprint arXiv:2303.18223 (2023).

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. "Language models are few-shot learners." arXiv preprint arXiv:2005.14165 (2020).

Gentzkow, Matthew, Jesse M. Shapiro, and Matt Taddy. "Measuring Group Differences in High‐Dimensional Choices: Method and Application to Congressional Speech." Econometrica, Econometric Society, vol. 87, no. 4, July 2019, pp. 1307-1340.

Environmental and Genetic Risk Factors of Breast Cancer and Disparities By Jia Yue (Joyce) Deng

Abstract

Breast cancer is among the most common cancers in women in the United States. It occurs due to the uncontrolled growth of tissue in the breast, creating a tumor. A significant aspect of breast cancer is heritable, which accounts for 5% to 10% of breast cancers. Germline (inherited) or somatic (acquired) mutations contribute to breast cancer. Among these mutations, the two most commonly studied genes associated with breast cancer are *BRCA1* (Breast Cancer gene 1) and *BRCA2* (Breast Cancer gene 2). While the primary *BRCA* genes have been extensively studied for their association with breast cancer risk, emerging evidence suggests the involvement of additional genes in this multifactorial condition. Mutations in the genes *TP53* (Tumor protein p53) and *CDH1* (Cadherin-1) are also associated with breast cancer but influence the development of the disease less commonly compared to BRCA genes. Human genetic diversity is important to consider when evaluating the effect of genetic mutations, yet people of color are not equitably represented in breast cancer research, particularly in genome-wide association studies. This paper explores the impact of mutations in *BRCA1* and *BRCA2* and other variants at breast cancer genes, in addition to examining issues related to breast cancer research and treatment.

1. Introduction

According to the World Health Organization, the number of incidences of breast cancer was 2.3 million, with 685,000 cases of death in 2020 (World Health Organization, 2023). In 2023, approximately 300,000 new cases of breast cancer are estimated to be diagnosed, along with an estimated 43,700 deaths in the United States (American Cancer Society, 2023). Breast cancer is the most common cancer diagnosed among women and is the second most common cause of death in women from cancer in the world (Alkabban & Ferguson, 2022). Approximately 1 in 8 women are diagnosed with invasive breast cancer in the United States in their lifetime, meaning that the cancer has infiltrated surrounding breast tissues (Feng et al., 2018). On the other hand, non-invasive breast cancer refers to localized cancer cells, where the cancer remains in the lobules or ducts where they originated. The prevalence of this disease highlights the importance of breast cancer awareness and early detection. Among the many factors that cause breast cancer development, mutations in the *BRCA* genes (*BRCA1* and *BRCA2*) that are inherited from a parent are the most commonly known mutation and have been extensively researched. However, other high-penetrance inherited mutations, including *BARD1*, *BRIP1*, *PALB2*, and *TP53*, have been discovered to be associated with an increased risk of breast cancer (Hu et al., 2020). Penetrance refers to the probability of an individual with a genetic mutation causing the associated or expected genetic trait or disorder (Koellner, 2020). A highly penetrant mutation will have a high chance of causing traits or illnesses. On the other hand, low-penetrance mutations have a smaller effect on increasing risk of the associated disorder. Additionally, it is

important to understand that mutations can be either benign or pathogenic, where the first causes no impact on an individual's health although the structure of a gene may have changed, and the latter demonstrates a change that leads to the development of a disease (Cleveland Clinic, 2020). Most of the mutations addressed in this paper will be referring to pathogenic mutations, which can lead to the development of breast cancer.

Furthermore, studies show that race and ethnicity play a distinctive role in the risk of breast cancer (Henderson et al., 2012). According to the National Cancer Institute's (NCI) Surveillance, Epidemiology, and End Results Program (SEER), non-Hispanic White women have consistently placed highest in terms of the lifetime risk of being diagnosed with cancer in the United States, at 13.8% based on data from 2017-2019 compared to the second-highest lifetime risk of 12.0% for non-Hispanic Black women (SEER, 2020). However, non-Hispanic Black women have a much higher mortality rate from breast cancer compared to other races or ethnicities in the United States, with a mortality rate of 2.9% based on data from 2018-2020, compared to White women (second-highest) with a mortality rate of 2.5%. The risk of breast cancer diagnosis and mortality for all other races and ethnicities (Hispanic, non-Hispanic Asian/Pacific Islander, and non-Hispanic American Indian/Alaska Native) consistently remains the lowest, indicatingthat race is associated with breast cancer outcomes. Such disparities may be caused by genetic and environmental differences across populations, including factors such as menarche (first menstrual cycle), age of first birth, and frequency of alcohol consumption.

Moreover, this disparity calls for the examination of how mutations in *BRCA1*, *BRCA2*, and other germline breast cancer genes influence the risk of breast cancer diagnosis and mortality across women of diverse ancestral backgrounds in the United States. By understanding the relationships between ancestry and environmental and genetic risk factors, researchers can mitigate the disparity in populations of different ancestry in efforts put toward breast cancer research.

2. Breast Cancer Classification and Factors Influencing Susceptibility

Breast cancer is characterized by the malignant and uncontrolled growth of breast tissues. It can be categorized into two primary types: sarcomas and carcinomas (Feng et al., 2018). Sarcomas are much rarer forms of breast cancer, constituting <1% of breast cancer cases. Sarcomas originate from the connective tissues of the breast, affecting the stromal components including blood vessels and myofibroblasts. Conversely, carcinomas are the most common forms of breast cancer, and they are primarily associated with abnormal growth in the epithelial breast cells, leading to the formation of tumors in the milk ducts or lobules (Feng et al., 2018). The most common types of breast cancer are invasive ductal carcinoma (IDC) (70-80% of breast cancer cases), invasive lobular carcinoma (ILC) (10-15% of breast cancer cases), and ductal carcinoma in situ (DCIS, intraductal carcinoma, making up 20-25% of breast cancer cases) (American Cancer Society, 2021; Feng et al., 2018; Cleveland Clinic, 2022; van Seijen, 2019).

Invasive or infiltrating breast cancer means that cancer cells have spread outside the breast lobules and ducts and have invaded the surrounding stromal tissue. Furthermore, these carcinomas can invade other organs in the body, bringing the patient into the metastatic stage of breast cancer, also known as stage IV breast cancer, a late stage of the disease (Feng et al., 2018). The spread of cancer to other areas of the body initiates the growth of metastases, and secondary malignant growths in areas away from the primary tumor, and poses an unpredictable threat to both near and distant sites of the body relative to the breast, such as the armpit, lung, and bone. The danger of this stage lies in the fact that microscopic cancer cells may remain in the body even after the primary tumor has been surgically removed, which leads to cancer relapse. The risk of relapse varies from person to person, differing based on the patient's condition, the distinctive molecular structure of the tumor, and the stage of cancer, complicating predictions of whether the disease will return (Feng et al., 2018).

DCIS constitutes 20% of breast cancer cases and is often a noninvasive condition in which the cells of the milk ducts become cancerous, but do not spread (non-invasive, or in situ) (Sun, n.d.). Therefore, the disease can be relatively easily treated using surgery, radiation, and/or medicine. The risk of local recurrence after surgical removal of the breast (i.e., mastectomy) is less than 2%, which can be further lowered with hormonal therapy (Sun, n.d.).

Breast cancer results from a combination of genetic and environmental factors. According to the American Cancer Society (ACS), about 5-10% of breast cancers are hereditary (American Cancer Society, 2021). Hereditary breast cancer refers to the presence of the disease in the extended family for several generations (van der Groep et al., 2011). Moreover, women with an inherited *BRCA1* gene mutation will have a 55-65% risk of developing breast cancer in their lifetime (Feng et al., 2018). Furthermore, women who have inherited a *BRCA2* mutation have a 45% risk. Research has also shown that particular germline gene mutations may occur in ancestral populations at different rates, creating disparities in the severity of breast cancer; therefore, certain ancestry groups may be at a greater risk of breast cancer than others (Henderson et al., 2012). Furthermore, Black women have a higher likelihood of dying from breast cancer at all ages (Yedjou et al., 2020). In contrast, Asian/Pacific Islander women exhibit the lowest incidence and death rates for breast cancer. Consequently, the underrepresentation of groups of racial minorities limits the available data for further understanding of breast cancer and its potential treatment methods (Henderson et al., 2012). The disparity in the amount of breast cancer genomic research across ancestry groups could limit our ability to reduce the variation in breast cancer prevalence in ancestry groups.

The various environments to which diverse ancestry groups are exposed play a significant role in their susceptibility to breast cancer as well (Henderson et al., 2012). For instance, systemic racism leads to socioeconomic, employment, educational, legal, and many other inequalities. Unequal access to opportunities and resources (including healthcare) continues to impede the detection, diagnosis, and treatment of breast cancer, causing Black, Indigenous, and People of Color (BIPOC) individuals to be more vulnerable to the disease. Poverty, for example, is often closely associated with healthcare quality and accessibility, and BIPOC individuals are

more likely to be in poverty compared to Whites (Yedjou et al., 2020). According to the United States Census, Blacks have the greatest rate of poverty (18.8%) compared to Whites which had the least rate of poverty at (7.3%) (Creamer, 2021). Low income or poverty leads to lower rates of cancer detection through mammography screening, which will often lead to late-stage diagnosis, making the cancer much more difficult to treat (Yedjou et al., 2020).

3. The Role of *BRCA* Genes in Breast Cancer

BRCA genes, discovered in the mid-1990s, are the most well-known genes for familial breast and ovarian cancer and are classified as high-penetrance breast cancer genes (Somasundaram, 2011; Winters et al., 2017). *BRCA1* and *BRCA2* are tumor suppressor genes responsible for maintaining the normal growth of tissues in the breasts, limiting the proliferation of cancerous cells (Winters et al., 2017). However, *BRCA* gene mutations may be caused by genetic deletions or insertions, which could directly impede its functions and cause uncontrolled cell proliferation (Scaglione et al., 2018; Casaubon et al., 2022). Additionally, the mutated *BRCA* genes are autosomal-dominant genes (Roy et al., 2011). As a result, inherited *BRCA* mutations can directly cause hereditary breast and ovarian (HBOC) syndrome, and are correlated with higher susceptibility to various cancer types including pancreatic, stomach, and prostate cancer. More specifically, individuals possessing HBOC syndrome have an increased lifetime risk of breast cancer diagnoses at rates of 50-80% (Roy et al., 2011).

The process of classifying *BRCA* genes began in 1990 as the Human Genome Project (Hurst, 2014). Dr. Mary-Claire King's research on genetic associations and influences with early-onset familial breast began with the recruitment of 23 extended families who contained growing numbers of male members with breast cancer, members diagnosed at a young age, or other similar traits demonstrating signs of hereditary breast cancer (Hurst, 2014). Through chromosome analyses, tests analyzing the chromosomes to identify any changes or abnormalities, Dr. King and her team discovered a link between early-onset breast cancer and chromosome 17q21. Amidst competitive efforts to clone and sequence the gene responsible for breast cancer, the Myriad Genetics team led by Mark Skolnick finally sequenced *BRCA1* in February 1994. Dr. Skolnick and his team sequenced the genotypes of 214 families worldwide. Before the discovery of *BRCA1* in 1991 by Dr. King (Hurst, 2014), Dr. Skolnick's team found and analyzed the blood samples of a family (named K2082), with members of the family in different generations possessing breast or ovarian cancer, through the Utah Population Database (UPDB) (Goldgar et al., 1994). The researchers conducted genetic testing by extracting DNA from blood samples, and they acquired additional lifestyle information on living and deceased family members. They genotyped four short tandem repeat (STR) loci on chromosome 17q to assess possible linkage to *BRCA1*. Linkage analysis, used to detect the chromosomal location of a mutated gene, revealed a significant connection between the breast and ovarian cancer phenotype on chromosome 17q, indicating a potential association with *BRCA1*, leading to the conclusion that *BRCA1* mutations increase breast cancer risk (Goldgar et al., 1994). Later in 1994, the *BRCA2* gene and its mutations were discovered by Michael Stratton and Richard

Wooster through a genomic linkage analysis on chromosome 13 to be a second high penetrance gene associated with elevated risks of breast cancer (Hurst, 2014).

BRCA1 mutations were later identified as the cause of approximately 35% of familial cancer cases and were found in 50% of incidences (Casaubon et al., 2022; Deng & Scott, 2000). *BRCA1* is composed of 24 exons, which are pieces of DNA that encode proteins, and in humans, the protein consists of 1863 amino acids (Deng & Scott, 2000). This protein has various functional regions, including a zinc finger domain that is associated with DNA directly or indirectly through protein interactions. At the carboxyl terminus, there is a beta-transducin repeat containing the E3 ubiquitin-protein ligase (BRCT) domain, commonly seen in proteins involved in DNA damage response (DDR) checkpoints (Deng & Scott, 2000). Additionally, exon 11 of the *BRCA1* gene plays a significant role, constituting over 60% of the protein. It encodes two potential nuclear localization signals and a domain that interacts with gene *RAD51* (RAD51 Recombinase), which is responsible for providing information on DNA repair.

BRCA2 was found to account for about 25% of hereditary breast cancer cases, slightly lower than *BRCA1* (Casaubon et al., 2022). *BRCA2* is most notably known for its role in DNA repair, playing a vital role in repairing DNA (Shahid et al., 2014). Mutated *BRCA2* genes impede the repair of double-strand breaks (DSB). Furthermore, *BRCA2* mutations are more commonly associated with breast cancers of the luminal subtype, often associated with estrogen receptor (ER) positivity (Roy et al., 2011). Another important aspect of *BRCA2*'s role is its mediation of *RAD51* nucleoprotein filament formation (Shahid et al., 2014). These filaments are essential for guiding *RAD51* to the sites where DNA repair is needed. Due to its large size, however, challenges have been posed towards gaining a complete understanding of the precise mechanisms of *BRCA2* in promoting homologous recombination (HR), particularly in association with *RAD51* and single-stranded DNA. To gain deeper insights into these interactions and functions, researchers are utilizing biochemical and structural approaches such as X-ray crystallography (Scott, 2021; Shahid et al., 2014).

The prevalence of *BRCA* gene mutations differ across ancestry groups, suggesting a link between ancestry and breast cancer risk (Yedjou et al., 2020). The highest mutation rates occur in Ashkenazi Jewish women (8.3%), followed by Hispanic women (3.5%), non-Hispanic White (2.2%), Black (1.3%), and Asian (3.5%) women (Yedjou et al., 2020). While genetics play a significant role, environmental factors such as smoking and age of menstruation also contribute to disparities (Yedjou et al., 2020). Different populations have varying environmental exposures that could influence breast cancer incidence. In conclusion, *BRCA1* and *BRCA2* mutations might contribute to disparities in breast cancer risk and outcomes among ethnic groups.

4. An Overview of Additional High-Penetrance Genes

In familial cancer cases where *BRCA* gene mutations are not proven to be present, various additional inherited gene mutations result in the development of hereditary breast cancers (Roy et al., 2011). Based on the breast cancer gene, the effect of the pathogenic variants (PVs) differs. Although *BRCA1* and *BRCA2* mutations are most commonly associated with breast

cancer development, numerous additional genes that increase breast cancer risk to 20%, including *ATM*, *BARD1*, *BRIP1*, *CDH1*, *CHEK2*, *MSH6*, *NBN*, *NF1*, *PALB2*, *PTEN*, *RAD51C*, *RAD51D*, and *TP53* (Hu et al., 2020). In particular, *TP53* (Tumor Protein 53) and *CDH1* (Cadherin-1 or Epithelial Cadherin) are also high-penetrance genes commonly associated with increased risk of breast cancer, often identified as closely correlated with invasive breast cancers (invasive ductal carcinoma and invasive lobular carcinoma), in addition to *BRCA1* and *BRCA2* (Dossus & Benusiglio, 2015).

4.1. Tumor Protein P53 and Li-Fraumeni Syndrome

Tumor protein p53 (*TP53*) is a gene that provides the code for the production of a tumor suppressor protein, p53 (NCBI, 2023). The protein is responsible for maintaining cell stability and can cause cells to self-destruct, repair DNA, apoptosis, and more (NCBI, 2023). Consequently, mutations in *TP53* could impede the protein's ability to perform the regulating mechanisms. Furthermore, *TP53* mutations occur differently in that they are most commonly missense, meaning they are characterized by alterations in a single amino acid in a protein; this differs from other tumor suppressor genes (such as *BRCA1*) where mutations cause the suppressing mechanism to halt (Rivlin et al., 2011). Mutations in *TP53* can be inherited or acquired later in life; hereditary TP53 mutations cause Li-Fraumeni syndrome (LFS) (Rivlin et al., 2011). LFS dramatically increases an offspring's risk of cancer development, bringing male lifetime risks to 70% or greater, and female lifetime risks to 90% or greater (Schneider, 2019). The most common LFS-caused cancers include adrenocortical carcinomas, breast cancer, central nervous system tumors, osteosarcomas, and soft-tissue sarcomas. Due to the hereditary nature of LFS, early onset of cancer is prevalent in individuals, with the majority of breast cancer cases caused by LFS occurring before menopause around the age of 33 (Schneider, 2019). Furthermore, breast cancers associated with LFS are generally invasive ductal carcinoma (IDC – breast cancer originating from the breast ducts and have breached the ductal barriers to spread to other areas of the body) (in addition to *BRCA1* and *BRCA2* mutations) and being human epidermal growth factor receptors 2 (HER-2), estrogen and progesterone receptor (ER and PR) positive (Dossus & Benusiglio, 2015). A Utah study revealed the high hereditary risk, as a result of mutated genes, of invasive lobular carcinoma (ILC – breast cancer originating from the lobules and having spread to lymph nodes or other parts of the body), raising interest in additional (high-penetrance) genes (Dossus & Benusiglio, 2015).

4.2. Cadherin-1

Cadherin-1 (*CDH1*), is a gene located on chromosome 16 and serves as another tumor suppressor, coding for the E-cadherin protein (Dossus & Benusiglio, 2015). The E-cadherin protein functions as a cell-cell adhesion for the formation of organized tissues in the epithelial cells (Dossus & Benusiglio, 2015). Mutations in *CDH1* promote cancer growth, spread, and invasion; gastric, breast, colorectal, thyroid, and ovarian cancers are most commonly associated with *CDH1* mutations. Originally known for its connection to hereditary diffuse gastric cancer

(HDGC), research has identified E-cadherin protein inactivation as the early cause of HDGC development (Dossus & Benusiglio, 2015). While families carrying the *CDH1* mutation have been found to have high cases and risks of diffuse gastric cancer (DGC), high cases of ILC have also been identified within such families (Dossus & Benusiglio, 2015). Furthermore, research has linked individuals with a history of early-onset ILC without cases of DGC within the family to mutations in *CHD1*, allowing researchers to identify this correlation between *CDH1* mutations and ILC (Corso et al., 2020).

5. Methods for screening and finding breast cancer genetic variants

Several methods of identifying mutations can be implemented to inform patients of possible breast cancer risks, including DNA sequencing and hybridization, cytogenetic testing methods such as karyotyping, and comparative genomic hybridization (CGH) (Mahdieh & Rabbani, 2013).

5.1. Cytogenetic Testing Methods

Cytogenetic testing methods such as karyotyping, fluorescence in situ hybridization (FISH), and comparative genomic hybridization (CGH) involve the study of chromosomes and their structure to identify irregularities (Mahdieh & Rabbani, 2013). Conventional karyotyping involves analyzing the overall arrangement of chromosomes in a cell (Mahdieh & Rabbani, 2013). It is used to detect chromosome abnormalities and is especially valuable in cases of suspected chromosomal disorders, sexual disorders, congenital anomalies, developmental issues, infertility, miscarriages, and malignancies (Mahdieh & Rabbani, 2013). Therefore, conventional karyotyping can be applied in cancer cases when there is suspicion of genetic abnormalities that could induce cancer growth. For instance, karyotyping may be used to detect greater chromosomal abnormalities in breast cancer tumor cells as it is placed in comparison with normal cells (Morris et al., 1997; Kytola et al., 2000). The process of conventional karyotyping begins with cells arrested during the metaphase stage of cell division, spread on a slide, and stained to create a visual karyotype that helps classify chromosomes based on their centromere position and banding pattern (Mahdieh & Rabbani, 2013).

Next, FISH is used to localize specific genes on chromosomes. It employs probes that bind to the gene of interest and are visualized using fluorescence under UV light. FISH can rapidly diagnose conditions such as trisomies and microdeletions by observing the location of the gene on the chromosomes (Mahdieh & Rabbani, 2013). As a result, FISH is a quicker method to detect genetic abnormalities and aid in deciding treatment methods. For example, Herceptin-targeted therapy is used for HER2 over-expressed breast cancer. FISH is one of the methods used to assess the HER2 protein level, RNA expression, and DNA copy numbers, and specific FISH scoring criteria and interpretive guidelines have been established for determining the ratio between the number of copies of the HER2 gene and chromosome 17, helping clinicians determine the suitability of Herceptin therapy for breast cancer patients with HER2 overexpression (Cui et al., 2016).

Finally, CGH is a variation of FISH and is used to detect changes in the number of copies of DNA sequences in an individual (Mahdieh & Rabbani, 2013). CGH has been used to identify new genes involved in different subgroups of breast cancer as it involves comparing the total genomic DNA of a sample (e.g., tumor DNA) with that of normal cells (van Beers & Nederlof, 2006). The sample DNA and unmutated DNA are labeled with fluorescent dyes and then hybridized. The analysis compares the fluorescent signals to determine copy number alterations in the genome. CGH is particularly useful for identifying copy number variations in cancer and complex karyotypes. Array-CGH increases the resolution by replacing metaphase slides with DNA arrays on glass slides.

5.2. Genetic Variation Analysis Methods

Molecular testing methods such as sequencing, DNA hybridization, and restriction enzyme digestion methods identify single nucleotide polymorphisms (SNPs), a single nucleotide base change , of a specific gene to identify variations responsible for the disease (Gunter, 2023; Mahdieh & Rabbani, 2013). The use and development of gene mutation detection methods are vital in various genetic diagnostic applications, including preimplantation genetic diagnosis, prenatal diagnosis, presymptomatic testing, conformational diagnosis, and forensic/identity testing (Mahdieh & Rabbani, 2013). The choice of method depends on the specific genetic condition and the level of detail required for accurate diagnosis. Furthermore, with knowledge of their genetic mutations and potential risks, patients have the choice to implement desired precautionary measures of breast cancer such as preventive mastectomy, which decreases breast cancer risk by surgically removing one or both breasts (Mahdieh & Rabbani, 2013).

A method to identify mutations is genome-wide association studies (GWAS), which analyze genetic variations across the genome to identify the ones that show association with particular traits or diseases (Uffelmann et al., 2021). This technique has uncovered numerous strong links between various traits and diseases, and this list of connected genetic variations is anticipated to expand as GWAS study populations become larger (Uffelmann et al., 2021). The findings from GWAS have diverse applications, including deepening our understanding of the biological basis of a trait, assessing its hereditary nature, determining genetic connections, providing insights for clinical risk prediction, contributing to the development of new drugs, and suggesting possible cause-and-effect relationships between risk factors and health results (Uffelmann et al., 2021).

6. GWAS Representation in Breast Cancer

The UK Biobank released a GWAS on breast cancer cases across various ancestry groups, with populations of African, Asian, and European descent (UK Biobank, 2018). The GWAS data were analyzed to reveal insufficient data recorded from people of color. As shown in Figure 1, the database consisted of a European sample size of n = 13,491, compared to an African sample size at n = 138, a Central South-Asian sample size at n = 170, and an East-Asian sample size at n = 78 (UK Biobank, 2018). Moreover, it was found that only the European

population exhibited significant associations with a total of 2086 SNPs (UK Biobank, 2018). Polygenic risk scores (PRS) are values used to estimate an individual's genetic predisposition to traits or diseases by using findings from GWAS to calculate an individual's genetic predisposition to the particular traits or diseases (Choi et al., 2020). However, since GWAS data (as shown above) yields greater sample sizes of individuals of European descent, PGS will be more accurate in individuals of European descent than those of non-European descent.

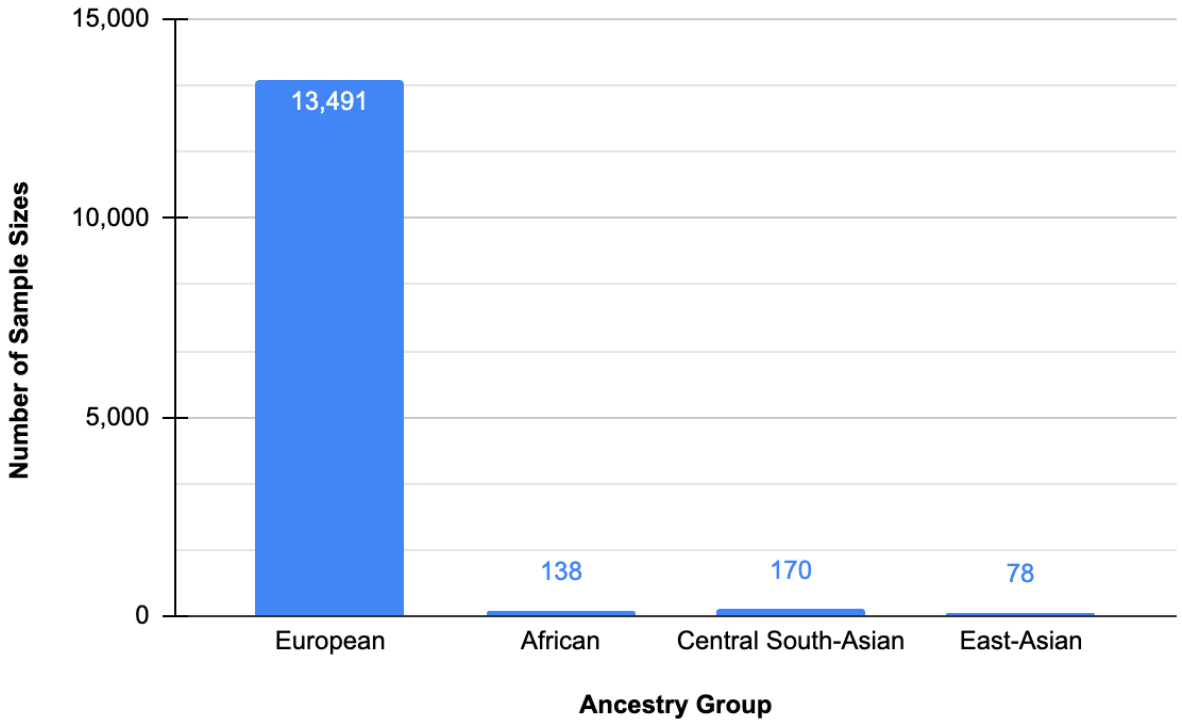


Figure 1. Number of Breast Cancer Sample Sizes Across Ancestry Groups. The graph shows a comparison between the UK Biobank GWAS sample sizes for breast cancer across European, African, Central South-Asian, and East-Asian ancestry groups.

As a result of unequal case representation, racial minority groups can be at a disadvantage for breast cancer as limited research will have been conducted on these groups as a result of a limited data pool, leaving healthcare providers with a confined idea of cancer subtypes that could affect certain ancestry groups more than others. Underrepresentation in GWAS is a general issue in the field of genetic studies. Popejoy & Fullerton reported that 96% of GWAS participants were of European descent in 2009 (2016). Although the number of individuals of non-European descent that participated in GWAS rose by 20% by 2016, 78% of the increase was a result of increased GWAS studies conducted in Asian countries such as China, Korea, India, and Japan (Popejoy & Fullerton, 2016). Moreover, the analysis reveals other ethnicities such as those of African, Latin American, and Indigenous descent are highly underrepresented;

furthermore, these are also the ethnic groups that are often considered to be under-resourced and underserved in the world. However, a majority of the largest biobanks are based in North American and European countries; therefore, the results of GWAS will likely reveal a much more diverse and holistic dataset for the European population (Ju et al., 2022), allowing for a greater understanding of the different rates of risks as a result of genetic mutations in breast cancer-associated genes. As a result, non-European ancestry groups could be at a disadvantage when breast cancer subtypes more closely associated with them are less studied due to underrepresentation in large human genomic studies such as GWAS. In conclusion, the unequal representation in GWAS greatly limits our ability to gain a full understanding of genetic risk factors of breast cancer across all populations; therefore, more research must be conducted in underrepresented populations.

7. Understanding Breast Cancer's Complexity and Potential Solutions

5-10% of breast cancers are hereditary, underscoring the importance of early cancer screening (American Cancer Society, 2021). Genetic testing provides a valuable tool for identifying both *BRCA* and non-*BRCA* mutations in individuals to signal more attention to how such mutations could initiate or accelerate breast cancer development (Valencia et al., 2017). The information from genetic testing aids physicians in tailoring treatment approaches to the patient's genetic profile, leading to more personalized and effective care (Valencia et al., 2017).

7.1. Environmental Risk Factors of Breast Cancer

Breast cancer's multifaceted nature, with genetic and environmental factors, means that simply targeting genetics alone is not sufficient. Environmental risk factors accumulated throughout life also play a substantial role in breast cancer cases, from factors such as breastfeeding and menopause to everyday habits including diets, physical activity, and contraceptive use (Kashyap et al., 2022). Late onsets of menopause or early menarche cause extended periods of exposure to estrogen, which has been shown to exacerbate the proliferation of tumor cells and induce cell mutations, increasing breast cancer risks (Kashyap et al., 2022; Yue et al., 2016). For instance, for every year younger than menstruation onset, the breast cancer risk is increased by a factor of 1.050 (Beral et al., 2012). Conversely, pregnancy helps maintain estrogen balance by increasing breast tissue differentiation and decreasing exposure to the detrimental non-estrogenic mutagens and genotoxic estrogens, natural hormones that proliferate breast tissue growth, ultimately lowering breast cancer risks (Dey et al., 2009). Such factors of biological processes differ in onset ages between various ancestry groups (Dey et al., 2009). As a result, certain groups with, for instance, early onset of menstruation would have prolonged exposure to estrogen in comparison to ancestry groups with generally later onset of menstruation. Therefore, the differences in the period of time exposed to hormonal imbalances that could trigger breast cancer development could contribute to differences in incidence rates of various ancestry groups. Another example is that statistically, Black women are more likely to have many children at a young age while not breastfeeding them as much as White women, which

contributes to an increase in risk for basal-like breast cancer (Yedjou et al., 2020; Breast Cancer Research Foundation, 2023). According to a 2013 study, 66.3% of black infants were breastfed compared to 84.3% of white infants (Anstey, 2018). The lactation process of breastfeeding extends the delay of menstruation periods, which exposes women to more estrogen (Chao, 1987; Yue et al., 2016). Therefore, decreasing the exposure period to estrogen through lactation from breastfeeding lowers breast cancer risks. In fact, the process of breastfeeding has been shown to decrease the risk of breast cancer by 4.3% for every 12 months the mother breastfeeds (Anstey, 2018). However, the disparities in breastfeeding rates vary based on additional factors such as a lack of adequate resources such as lactation care providers, who guide families through issues with breastfeeding such as plugged ducts, sore nipples, and so forth (Anstey, 2018). Additionally, breastfeeding has not yet been generally accepted by the Black community, lowering the rates of breastfeeding in black women due to embarrassment and inconvenience (Anstey, 2018).

Comorbidities such as obesity, physical inactivity, alcohol consumption, smoking, and more can all contribute to elevated breast cancer risks as well (Kashyap et al., 2022). Women who consume two to three alcoholic drinks each day have an approximately 20% elevated risk of breast cancer compared to those who don't consume alcohol (Feng et al., 2018). For instance, studies have shown that overconsumption of alcohol intake is also associated with increased estrogen and other breast cancer-related hormone circulation (Kashyap et al., 2022). Hence, being aware of the risks and causes of breast cancer to minimize tumor development through controllable factors would contribute greatly to alleviating breast cancer risks (decreasing smoking and alcohol intake, increasing physical activity, becoming pregnant within the recommended time frame, and more). Considering the diverse array of risk factors associated with breast cancer and understanding these influences is crucial for developing effective treatments and prevention methods.

7.2. Advancements in Breast Cancer Treatment With Car-T Cell Therapy

A promising new method of cancer treatment has emerged: Chimeric antigen receptor (CAR)-T cell therapy which utilizes artificially engineered T cell receptors designed to redirect immune cells, mainly T cells, to identify and eliminate cells that display a specific antigen (Sterner & Sterner, 2021). Unlike the traditional major histocompatibility complex (MHC) receptor, CARs interact directly with surface antigens, triggering T-cell activation and potent anti-tumor responses. Notably, the success of CAR-T cell therapy against B cell cancers led to its approval by the US FDA in 2017. Despite these achievements, significant challenges remain for CAR-T cell therapy, such as severe toxicities associated with CAR-T cells, limited effectiveness against solid tumors, resistance in B cell cancers, antigen escape, short persistence, inadequate infiltration into tumors, and the suppressive tumor environment. Various strategies have been proposed to address these issues, including combining CAR-T cell therapy with other cancer treatments, using creative CAR engineering techniques to enhance antitumor effects, broaden clinical applications, and minimize risks. Furthermore, recent advances in CAR-T cell therapy for breast cancer, particularly in treating triple-negative breast cancers (TNBCs), which is a

breast cancer subtype that is particularly aggressive and difficult to treat, are showing promising results in clinical trials (Dey et al., 2023). Innovative approaches, like the combination therapy of EGFR CAR-T cells and CDK7 inhibition, are being explored. Such a combination has shown success in suppressing immune resistance, reducing tumor growth, and inhibiting TNBC metastasis in various models. These advancements underscore the potential of CAR-T cell therapy as a transformative approach in the treatment of breast cancer, especially for challenging subtypes like TNBC.

8. Conclusion

In conclusion, breast cancer remains a significant global health concern, with 2.3 million women diagnosed in 2020, making breast cancer the most common cancer in the world (World Health Organization, 2023). Genetic factors play a pivotal role in breast cancer susceptibility, with 5-10% of breast cancer cases being hereditary with mutations of the genes *BRCA1*, *BRCA2*, *TP53*, and *CDH1* most strongly associated with increased risk. These gene mutations vary in prevalence among different ancestry groups, contributing to disparities in breast cancer incidence and mortality rates across these various groups, signifying the importance of awareness in unevenly distributed cases of the disease. Additionally, the data analysis conducted on the UK Biobank underscores concerns about the limited diversity in contemporary research and data collection. This is reflected in the insufficient availability of statistical information for several ancestry groups, potentially leading to a decreased level of quality of care for these groups. Additionally, the advancement of genetic testing methods, including molecular and cytogenetic techniques, has proven to be valuable tools for identifying genetic mutations and for physicians to tailor treatment approaches accordingly based on genetic information on the patient. However, the complex nature of breast cancer involves not only genetics but also environmental factors. Moreover, emerging therapies such as CAR-T cell therapy present promising avenues for improving treatment outcomes, especially for challenging aggressive subtypes such as TNBC. The combined efforts of genetics research, comprehensive risk assessment, and innovative therapies hold the potential to revolutionize breast cancer detection, prevention, and treatment.

Works Cited

Alkabban, Fadi M, and Troy Ferguson. "Breast Cancer - Statpearls - NCBI Bookshelf." *StatPearls*, StatPearls Publishing, 26 Sept. 2022, www.ncbi.nlm.nih.gov/books/NBK482286/.

American Cancer Society. "American Cancer Society: Cancer Facts & Statistics." *American Cancer Society | Cancer Facts & Statistics*, 2023, cancerstatisticscenter.cancer.org.

The American Cancer Society. "Types of Breast Cancer: About Breast Cancer." *About Breast Cancer | American Cancer Society*, American Cancer Society, 19 Nov. 2021, www.cancer.org/cancer/types/breast-cancer/about/types-of-breast-cancer.html.

Anstey, Erica H., et al. "Breastfeeding and breast cancer risk reduction: Implications for black mothers." *American Journal of Preventive Medicine*, vol. 53, no. 3, 1 Sept. 2017, https://doi.org/10.1016/j.amepre.2017.04.024.

Beral, V, et al. "Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies." *The Lancet Oncology*, vol. 13, no. 11, 2012, pp. 1141–1151, https://doi.org/10.1016/s1470-2045(12)70425-4.

"Breast Cancer." *World Health Organization*, World Health Organization, 12 July 2023, www.who.int/news-room/fact-sheets/detail/breast-cancer.

Casaubon, Jesse T, et al. "BRCA1 and BRCA2 Mutations." *StatPearls*, StatPearls Publishing, 23 July 2023, www.ncbi.nlm.nih.gov/books/NBK470239/.

Chao, S. "The effect of lactation on ovulation and fertility." *Clinics in Perinatology*, vol. 14, no. 1, Mar. 1987, pp. 39–50, https://doi.org/https://pubmed.ncbi.nlm.nih.gov/3549114/.

Choi, Shing Wan, et al. "Tutorial: A guide to performing polygenic risk score analyses." *Nature Protocols*, vol. 15, no. 9, 2020, pp. 2759–2772, https://doi.org/10.1038/s41596-020-0353-1.

Cleveland Clinic. "Lobular Breast Cancer: Symptoms & Treatment." *Cleveland Clinic*, 20 Sept. 2022, my.clevelandclinic.org/health/diseases/21180-lobular-breast-cancer.

Corso, Giovanni, et al. "Hereditary gastric and breast cancer syndromes related to Cdh1 germline mutation: A multidisciplinary clinical review." *Cancers*, vol. 12, no. 6, 17 June 2020, p. 1598, https://doi.org/10.3390/cancers12061598.

Creamer, John. "Inequalities Persist despite Decline in Poverty for All Major Race and Hispanic Origin Groups." *Census.Gov*, United States Census Bureau, 9 Dec. 2021, www.census.gov/library/stories/2020/09/poverty-rates-for-blacks-and-hispanics-reached-historic-lows-in-2019.html.

Cui, Chenghua, et al. "Fluorescence in situ hybridization: Cell-based Genetic Diagnostic and Research Applications." *Frontiers in Cell and Developmental Biology*, vol. 4, 5 Sept. 2016, https://doi.org/10.3389/fcell.2016.00089.

Deng, Chu-Xia, and Frank Scott. "Role of the tumor suppressor gene BRCA1 in genetic stability and mammary gland tumor formation." *Oncogene*, vol. 19, 7 Mar. 2000, pp. 1059–1064, https://doi.org/10.1038/sj.onc.1203269.

Dey, Anuvab, et al. “Recent advancement in breast cancer treatment using car T cell therapy:- A Review.” *Advances in Cancer Biology - Metastasis*, vol. 7, 1 Mar. 2023, https://doi.org/10.1016/j.adcanc.2023.100090.

Dey, Subhojit, et al. “Risk factors according to estrogen receptor status of breast cancer patients in Trivandrum, South India.” *International Journal of Cancer*, vol. 125, no. 7, 2 Apr. 2009, pp. 1663–1670, https://doi.org/10.1002/ijc.24460.

Dossus, Laure, and Patrick R Benusiglio. “Lobular breast cancer: Incidence and genetic and non-genetic risk factors.” *Breast Cancer Research*, vol. 17, 13 Mar. 2015, https://doi.org/10.1186/s13058-015-0546-7.

Feng, Yixiao, et al. “Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and Molecular Pathogenesis.” *Genes & Diseases*, vol. 5, no. 2, 12 May 2018, pp. 77–106, https://doi.org/10.1016/j.gendis.2018.05.001.

Goldgar, D. E., et al. “A large kindred with 17Q-linked breast and ovarian cancer: Genetic, phenotypic, and genealogical analysis.” *JNCI Journal of the National Cancer Institute*, vol. 86, no. 3, 2 Feb. 1994, pp. 200–209, https://doi.org/10.1093/jnci/86.3.200.

Gunter, Chris. “Single Nucleotide Polymorphisms (Snps).” *National Human Genome Research Institute*, 2023, www.genome.gov/genetics-glossary/Single-Nucleotide-Polymorphisms.

Henderson, Brian E., et al. “The influence of race and ethnicity on the biology of cancer.” *Nature Reviews Cancer*, vol. 12, no. 9, 2 Aug. 2012, pp. 648–653, https://doi.org/10.1038/nrc3341.

Hu, Chunling, et al. “The contribution of germline predisposition gene mutations to clinical subtypes of invasive breast cancer from a clinical genetic testing cohort.” *JNCI: Journal of the National Cancer Institute*, vol. 112, no. 12, 24 Apr. 2020, pp. 1231–1241, https://doi.org/10.1093/jnci/djaa023.

Hurst, Jillian H. “Pioneering geneticist Mary-Claire King receives the 2014 lasker~koshland special achievement award in medical science.” *Journal of Clinical Investigation*, vol. 124, no. 10, 8 Sept. 2014, pp. 4148–4151, https://doi.org/10.1172/jci78507.

Ju, Dan, et al. “Importance of including Non-European populations in large human genetic studies to enhance precision medicine.” *Annual Review of Biomedical Data Science*, vol. 5, no. 1, 10 Aug. 2022, pp. 321–339, https://doi.org/10.1146/annurev-biodatasci-122220-112550.

Kytola S, Rummukainen J, Nordgren A, Karhu R, Farnebo F, et al. Chromosomal alterations in 15 breast cancer cell lines by comparative genomic hybridization and spectral karyotyping. Genes Chromosomes Cancer. 2000;28:308–317.

Kashyap, Dharambir, et al. “Global increase in breast cancer incidence: Risk factors and preventive measures.” Edited by Yuvaraja Teekaraman. *BioMed Research International*, vol. 2022, 18 Apr. 2022, pp. 1–16, https://doi.org/10.1155/2022/9605439.

Koellner, Christine M., et al. “Basic concepts in human molecular genetics.” *Essential Concepts in Molecular Pathology*, 2020, pp. 81–99, https://doi.org/10.1016/b978-0-12-813257-9.00005-x.

Mahdieh, Nejat, and Bahareh Rabbani. “An Overview of Mutation Detection Methods in Genetic Disorders.” *Iranian Journal of Pediatrics*, vol. 23, no. 4, Aug. 2013, pp. 375–388, https://doi.org/https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3883366/.

Morris, Joanna S, et al. Genes, Chromosomes and Cancer, vol. 20, no. 2, 7 Dec. 1998, pp. 120–139, https://doi.org/https://doi.org/10.1002/(SICI)1098-2264(199710)20:2<120::AID-GCC3>3.0.CO;2-5.

National Center for Biotechnology Information . “CDH1 Cadherin 1 [Homo Sapiens (Human)].” *National Center for Biotechnology Information*, National Library of Medicine, 30 Aug. 2023, www.ncbi.nlm.nih.gov/gene/999.

National Center for Biotechnology Information. “TP53 Tumor Protein P53 [Homo Sapiens (Human)].” *National Center for Biotechnology Information*, National Library of Medicine, 30 Aug. 2023, www.ncbi.nlm.nih.gov/gene/7157.

“Not All Mutations Are Bad: Researchers Identify Differences between Benign and Pathogenic Variants.” *Lerner Research Institute*, Cleveland Clinic, 26 Oct. 2020, www.lerner.ccf.org/news/article/?id=94566bf013a834affc533e0c48ef73a361685c79.

Popejoy, Alice B., and Stephanie M. Fullerton. “Genomics is failing on diversity.” *Nature*, vol. 538, no. 7624, 12 Oct. 2016, pp. 161–164, https://doi.org/10.1038/538161a.

Rivlin, Noa, et al. “Mutations in the p53 Tumor Suppressor Gene: Important milestones at the various steps of tumorigenesis.” *Genes & Cancer*, vol. 2, no. 4, 18 May 2011, pp. 466–474, https://doi.org/10.1177/1947601911408889.

Roy, Rohini, et al. “BRCA1 and BRCA2: Different roles in a common pathway of genome protection.” *Nature Reviews Cancer*, vol. 12, no. 1, 23 Dec. 2011, pp. 68–78, https://doi.org/10.1038/nrc3181.

Scaglione, Giovanni, et al. “A whole germline BRCA2 gene deletion: How to learn from CNV in silico analysis.” *International Journal of Molecular Sciences*, vol. 19, no. 4, 2018, p. 961, https://doi.org/10.3390/ijms19040961.

Schneider, Katherine, et al. “Li-Fraumeni Syndrome - GeneReviews® - NCBI Bookshelf.” *GeneReviews*, GeneReviews, 21 Nov. 2019, www.ncbi.nlm.nih.gov/books/NBK1311/.

Scott, Duncan E., et al. “A small-molecule inhibitor of the BRCA2-RAD51 interaction modulates RAD51 assembly and potentiates DNA damage-induced cell death.” *Cell Chemical Biology*, vol. 28, no. 6, June 2021, pp. 835–847, https://doi.org/10.1016/j.chembiol.2021.02.006.

SEER Incidence Data. “Seer*Explorer.” *SEER*Explorer Application*, Nov. 2022, seer.cancer.gov/statistics-network/explorer/application.html?site=55&data_type=1&graph_type=2&compareBy=race&chk_race_6=6&chk_race_5=5&chk_race_4=4&chk_race_9=9&chk_race_8=8&rate_type=2&sex=3&age_range=1&stage=101&advopt_precision=1&advopt_show_ci=on&hdn_view=0&advopt_show_apc=on&advopt_display=2#resultsRegion0.

Shahid, Taha, et al. “Structure and mechanism of action of the BRCA2 Breast Cancer Tumor Suppressor.” *Nature Structural & Molecular Biology*, vol. 21, no. 11, 5 Oct. 2014, pp. 962–968, https://doi.org/10.1038/nsmb.2899.

Somasundaram, Kumaravel. “BRCA1 and BRCA1 genes and inherited breast and/or ovarian cancer: Benefits of genetic testing.” *Indian Journal of Surgical Oncology*, vol. 1, no. 3, 11 Mar. 2010, pp. 245–249, https://doi.org/10.1007/s13193-011-0049-7.

Sterner, Robert C., and Rosalie M. Sterner. “Car-T cell therapy: Current limitations and potential strategies.” *Blood Cancer Journal*, vol. 11, 6 Apr. 2021, https://doi.org/10.1038/s41408-021-00459-7.

Sun, Bonnie. “Ductal Carcinoma in Situ (DCIS).” *Johns Hopkins Medicine*, Johns Hopkins Medicine, www.hopkinsmedicine.org/health/conditions-and-diseases/breast-cancer/ductal-carcinoma-in-situ. Accessed 31 Aug. 2023.

Uffelmann, Emil, et al. “Genome-wide association studies.” *Nature Reviews Methods Primers*, vol. 1, 26 Aug. 2021, https://doi.org/10.1038/s43586-021-00056-9.

UK Biobank. *UK Biobank*, Neale Lab, 1 Aug. 2018, www.nealelab.is/uk-biobank/.

Valencia, Olivia M., et al. “The role of genetic testing in patients with breast cancer.” *JAMA Surgery*, vol. 152, no. 6, June 2017, pp. 589–594, https://doi.org/10.1001/jamasurg.2017.0552.

van Beers, EH, Nederlof, PM. Array-CGH and breast cancer. *Breast Cancer Res*. 2006, 8(3):210. doi:10.1186/bcr1510

Van der Groep, Petra, et al. “Pathology of hereditary breast cancer.” *Cellular Oncology*, vol. 34, no. 2, 19 Feb. 2011, pp. 71–88, https://doi.org/10.1007/s13402-011-0010-3.

Van Seijen, Maartje, et al. “Ductal carcinoma in situ: To treat or not to treat, that is the question.” *British Journal of Cancer*, vol. 121, no. 4, 9 Aug. 2019, pp. 285–292, https://doi.org/10.1038/s41416-019-0478-6.

Winters, Stella, et al. “Breast Cancer Epidemiology, prevention, and screening.” *Progress in Molecular Biology and Translational Science*, vol. 151, 2017, pp. 1–32, https://doi.org/10.1016/bs.pmbts.2017.07.002.

Yedjou, Clement G., et al. “Health and racial disparity in breast cancer.” *Advances in Experimental Medicine and Biology*, vol. 1152, 2019, pp. 31–49, https://doi.org/10.1007/978-3-030-20301-6_3.

Yedjou, Clement G., et al. “Health and racial disparity in breast cancer.” *Advances in Experimental Medicine and Biology*, vol. 1152, 28 Aug. 2019, pp. 31–49, https://doi.org/10.1007/978-3-030-20301-6_3.

Yue, Wei, et al. “Effects of estrogen on breast cancer development: Role of Estrogen Receptor Independent Mechanisms.” *International Journal of Cancer*, vol. 127, no. 8, 20 Aug. 2010, pp. 1748–1757, https://doi.org/10.1002/ijc.25207.

Music and Bridging Divide on National and International Scales By Anjali Schoetz

Abstract

All throughout history we have observed and researched how music has been successful in bridging national and international divides. The Singspiel opera has historically been used as a case study for music bridging divides through history. In a world that is more politically, socially, and mentally divided than we have ever seen before due to media, polarization, and more it is important to study how and why music is used as a successful means to bridge divides in the modern and historical worlds. Perhaps by studying what makes music so successful in bridging divides we can implement the core aspects of music in other areas of life to bridge divides. Throughout the research we discovered that the core aspect of music that allows it to be so successful in bridging divides is its multifaceted nature. The fact that music is created as a collective, that it is an international language, that it is created with unifying purposes, and is accessible to be created and enjoyed by all people.

Keywords: Germany, India, Music, Singspiel Opera, Bridge Divides

Introduction

To a layperson music is simply sound or a piece of string with wood, but in this paper a pause is taken to reimagine music as a multidimensional force that bridges cultural and national divides within people and countries all over the world. When simply observing the definition of music, there is no concrete definition. It is within the beautiful ambiguity of music that comes its power. In the book, “Rethinking Music” ethnomusicologist Philip Bohlman describes 13 different definitions of music. He describes music as “Our Music/Their Music” because of its ability to create social harmony (Cook, 1999). Music is described as the footsteps heard in the city and as a “flexible, permeable, and blurred” force that is a language with thousands of other languages within it (Ward, May 2021, 19). Other scholars like Shakespeare have described music as a form of expression used when words cease to convey. Some even believe that music is a measure of time. While coming to a specific definition of music is questionable and deeply argued, the power of music in society today is undeniable. When attempting to understand why music is so powerful in bridging divides we must take these multidimensional factors of music and study how they have interacted in the world throughout history and today.

This paper is an exploration of the specific facets of Music that makes it such an effective tool for bridging national, cultural, and political divides between people. Through two case studies of German and indian history and modern day examples, five major reasons why Music is an effective bridge between people are explored:

1. The way that music is created (it's a collective art)
2. The purpose that music is created for

3. That it is a form of human communication that can both express and be understood without language
4. Its multidimensional nature
5. Its simultaneous subjectivity

Music has been used as a tool to break down social barriers in Germany all throughout history and today. During the 18th century the German Opera genre of Singspiel helped break down class divides by reinventing the old Italian conventions of the Opera. The genre of Singspiel Opera is characterized by the spoken dialogue sounds fused into the music (recitative) ("*Johann Adam Hiller", n.d.*). The credited creator of Singspiel is a renowned musician, Johann Adam Hiller. Johann Adam Hiller was a school teacher, flutist, bass player, conductor, and composer. He used a scholarship to go to gymnasium, study music, and was the definition of a middle class German citizen in the day. Some other popular Singspiel composers include Jiří Antonín Benda and Karl Ditters von Dittersdorf. Karl Dittersdorf was a classical violinist born in Austria Vienna to a father who was a military Taylor, and Jiří Antonín Benda was born into a prominent yet humble family of musicians. The main commonality between the composers and creators of the Singspiel Opera was that they were middle class, common men, during their time. The genre of Singspiel eventually became known as "The Middle Class Opera" (*The Editors of Encyclopedia Britannica*, n.d.). Aristocracy and extremely wealthy in Italy. Additionally, the middle class did not just have access to the opera but they were composing the opera, they resonated with the stories in the opera, and were encouraged to see the opera. The composers of Singspiel created an opera that would travel from small town to town in direct rebellion against the exclusive practices of the Italian Opera (*The Editors of Encyclopedia Britannica*, n.d.) . This way all people could go to the opera because they did not have to travel to huge metropolitan areas to see it. While, having access to something like the Opera might not seem like a gargantuan deal in the 18th and 19th century when Europe was ravaged by the Napoleonic wars and social class battles between the declining Aristocracy and rising merchant class was at an all time high the Singspiel Opera was used as an escape (The Editors of Encyclopaedia Britannica, n.d.). People could laugh at the opera, forget about the social tension and groupings, and all come together to appreciate music. Unlike the complex and dramatic plots of the Italian Opera, Singspiel represented more simple characters in a magical world. It dove into the genre of magical realism ("*Johann Adam Hiller", n.d.*). The simple characters and plots not only made the middle class feel heard, special, and understood but also taught the aristocracy about their lives and put the aristocracy in the shoes of the middle class in a light hearted way. The only way to bridge divides is through empathy and knowledge. The Singspiel opera gave the gift of knowledge and empathy to the upper class Germans. This is reflected in the fact that Germany's "First National Operates", *Die Liebe auf dem Lande* (1768; "Love of the Land") and *Die Jagd* (1770; "The Hunt"), were in the genre of Singspiel ("*Johann Adam Hiller", n.d.*). This not only gave Germany musical prominence on a national scale, but also represented the German middle class at the national scale, and used the stories of the middle class to represent Germany as a

nation (Stark, 2021). A person who had been oppressed for many years by industrialism could be heard, felt, and seen through the opera.

The success of the Singspiel opera had rippling effects into music and the future conventions of opera all over Europe. Renowned composers such as Motzart started to fuse the traditional grand scale of the Italian “Opera Seria " with the style of Singspiel (*The Editors of Encyclopedia Britannica,* n.d.). . The infamous opera, “The Magic Flute” is a good example of the fusion of Italian and German opera. The high Coloratura Soprano that is the Queen of The Night represents the Italian opera while the German Folk Music like elements and the magical realism within the story is a strong representation of Singspiel. This is the definition of an opera that can appeal to both the aristocracy, upper class, middle class, and lower class.
The Singspiel genre of German opera not only started to include the middle class in the music but in the process the genre helped redefine social barriers between social classes, and break some of those barriers.

While throughout history the opera has broken down divides between born and raised Germans within the country today it even bridges divides between the German people and the immense number of refugees present in modern day Germany (Barone, 2018).
Syrian refugees who find themselves in a linguistically and culturally unfamiliar setting, miles away from home, have been using music as a means to integrate into German society and bridge the divides between their wartorn and peaceful worlds.

As of today 6.7 million refugees have been displaced from Syria and through Angela Merkel's immigration policies Germany has been the host of 532,100 of those Syrian refugees (Ward, May 2021, 9). In order to manage so many refugees Germany has systems in place that force refugees to leave their camps and take German integration classes, where the refugees have to practice German and be prepared to join German society (Ward, May 2021, 10). In order to create community and a small semblance of home for the Syrian community music groups all over Germany started to step up and create joint Syrian German collaborative music projects (Ward, May 2021, 10). The Bavarian Youth State Opera wrote an opera called “Moses" which allows for commonality between Bible, Quran, and Torah users (Barone, 2018). The work focuses on immigrant life in Germany after fleeing home. Primarily the work follows the story of Moses through a fun, energy filled, and fear filled libretto. Critics describe the Libretto as not having a heavy immigrant story feeling but rather a meaningful, light hearted, and beautiful story (Barone, 2018). The opera is written in the famous German Brechtian Style while also going through a multitude of refugee experiences. The actual cast of the opera is both Germans and many Syrian refugees. The opera was even used as a way to teach Syrians more of the German language (Barone, 2018). The fact that music is a universal language with symbols that both Germany and Syrians are able to read facilitated both Germany and Syrians to create a mixed community where they share their own languages with each other and can be a little less alone. Additionally, the reason the music was created to facilitate bridges and bring awareness to the Syrian refugee community is beautiful and has truly achieved its purpose. All performances of Moses at the Munich Opera have been sold out, receiving large audiences and viewers from all

over Germany and Europe (Barone, 2018). The issue of immigration is threatening Angela Merkel's power and leading to the creation of far right groups but this heartwarming story is helping to create less hostility around the issue of immigration (Barone, 2018). Furthermore the fact that the Opera is a children's opera is incredible because the stigma around refugees is indoctrinated into youth all over Germany and the world. Having children view this opera at such a young age is combating the stigma surrounding refugees at its core, because when operas such as this are shown to youth at such a young age the stigma is not created in the first place. The Bavarian youth state opera did not just create “Moses” but also created “Noah” the opera which was inspired by actual refugee and immigrant stories, like Ali Madad Qorbani who had to flee from Afghanistan to Iran to Europe (Barone, 2018). The efforts to create unity between the Syrian and German communities has been occurring through music groups all around Germany. The Frankfurt Based Orchestra For All International Musicians runs a program called Bridges which facilitates international musicians from all over the world and hundreds of refugees each year. The music chosen by the Frankfurt orchestra is picked by the orchestra members and allows for individuals to reflect on their experiences and express emotions through music (Le Blond, 2018). Amir Vafa sang the opening solo of the orchestra's song called “Das Heimweh" which translates to “homesickness”. He came to Germany because he was going to be persecuted in Iran and said that, “The song is special because all of us feel it” (Le Blond, 2018). The beauty of music is that it is able to connect all the international musicians not just physically but emotionally. The idea that everyone is following a single conductor, and all the woodwinds are taking every breath at the same time creates a literal and emotional connectedness like no other. The orchestra works to incorporate a mixture of music into their programs and also creates mixed music. The collaboration between all the musicians is truly key to the success of the organization. Waseem Hariri, an orchestra member who fled from Damascus Syria to Germany says, “ If everyone chose the way of music, there would be no more war.” (Le Blond, 2018). Music has truly been used to bridge divides in modern day Germany between the refugees and the country of Germany and the people of Germany. The key is that this kind of music is not just allowing for an assimilation of Syrians into German society, but it is truly allowing for integration between German and Syrian people (Ward, May 2021, 9). Instead of forcing the Syrian people to simply practice the way of traditional German society and strip away parts of their identity, this collaboration of music allows the Syrian refugees to bring pieces of home with them and practice and share them in Germany, while also learning about German culture, traditions, and ways of life.

Music being used as a means to bridge divides is a universal concept that is not solely prevalent in the country of Germany. Music is also being used to bridge divides in modern day India.

India

Music has been used as a tool to break down social barriers between modern India and Pakistan. India and Pakistan are nuclear rivals, countries battling over the territory of Kashmir,

and have been enraptured in outbreaks of violence that date all the way back to partition (*Time of India*, 2012). Despite all the tension, music is bringing members of each country together. An article by India Today reflects on the programs and success of a reality TV show based in India called Sur Kshetra that focuses on uniting musicians from India and Pakistan (*India Today*, 2012). Three renowned South Asian singers, Asha Bhosle (from India), Abida Parveen (from Pakistan), and Runa Laila (from Bangladesh) all worked together to find musicians from both Pakistan and India to create music together (*India Today*, 2012). People from India and Pakistan are becoming inspired by the collaboration. Infamous Indian music producer Himesh Reshmmiya has even said that the Sur Kshetra show has inspired him to use both Indian and Pakistani singers in his films and projects to facilitate collaboration on set (*India Today*, 2012). The fact that music is made in collaboration with so many individuals from all over the world makes the divide get overlooked so that music gets created. On August 27th at the Line of Control Pakistani soldiers began waving their hands and dancing to the song Bambiha Bole with the Indian soldiers on the other side of the border (*Times of India*, 2022). Soldiers that are very literally trained to fight one another, and stand on the physical border seeing the divide every day could come together and dance. Though the soldiers visually see and stand on the divide, music allows them to mentally break the divide. Music is universal and able to be enjoyed with everyone. By nature music is made to be shared and heard by every person in the same vicinity. The recent film, “Song of Lahore” has been even more transformative focusing on Lahore's vibrant musician community helping to merge East-West cultural divides (*Times of India*, 2022). Two time academy award winner Sharmeen Obaid attributes the success of the film to the inherent nature of art and music. He says, "I think that art, culture, music, film - all of these do not have any language. They are all intrinsically linked to humanity.” (*Business Standard*, 2016). Ultimately, in the case of India and Pakistan the countries that are extremely politically and physically divided however, music has had the power to stop the people of the countries from being extremely divided. IPA (The International Phonetic Alphabet)

In addition to bridging divides inside of nations, music bridges divides all over the world.

Through the use of IPA (the International Phonetic Alphabet) singers have been easily able to sing songs in other languages and create mixed music. The creation of the International Phonetic Alphabet commenced in the 19th century with a man named Otto Jesperson, who was inspired by the Russian linguist Rasmus Rask. The language was finally published in 1888 forever changing the world of singing (*International Phonetic Alphabet (IPA) | Definition, Uses, & Chart*, 2023). IPA essentially uses Roman Characters, Diacritical Marks, and letters from numerous different scripts to create a list of sounds. Any song from anywhere in the world can be translated into IPA so that singers can properly pronounce and convey the song. According to voice professor Duane Karna IPA is extremely useful in helping choirs and students efficiently and correctly pronounce foreign texts while also giving students further insight on the languages they are singing (Shearer, 2013). The importance of IPA is further demonstrated in the fact that choir directors are not just found using IPA to help pronounce Italian, German, Spanish, Latin,

and French but also languages like Swahili, Hindi, and Arabic (Shearer, 2013). IPA is in the Diction for Singers curriculum at almost every major music conservatory. The concept of IPA is incredible because it allows singers and choirs in all parts of the world to better learn languages and music from any other part of the world without needing to have a native speaker of the language in the room. The amount of worldly connectedness IPA facilitates in terms of music and language is enormous. Like the notes and rests of music is universal, so is IPA. Music stops language from being a cultural barrier. The rising amount of Mixed music is also reflective of a barrier free world within music. In Lambarena's: Bach To Africa written by Albert Schweizer music influenced by Johan Sebastian Bach becomes combined with African beat, drums, and dance. Albert Schweizer says that the music is inspired by much of his life spent in Gabon Africa (Post, 2014). Ravi Shankar, famous Indian Sitar player has been combining classical Indian music and Western Music since the 1960s. He even collaborated with the Beatles (Post, 2014). In a world where countries have sanctioned each other, wars are brewing, and there is so much hate, people are making music and the divide can slightly fade away. The emotional connection facilitated by music is stronger than the world's emotions of hate. People are able to collaborate, create mixed music, use IPA and notation so that language is no longer a barrier, and break the divide.

Conclusion

Ultimately, by studying how music has bridged social and cultural divides through a deep dive on the German Singspiel opera, Music programs for Syrian refugees, Music as a force regarding the India-Pakistan divide, and the International Phonetic Alphabet it has been discovered that music is extremely powerful as a force to bridge divides due to it is multifaceted nature. The fact that music is a collective art, is created with human purpose, is understood without language, and has great subjectivity. In a world with extreme political polarization, numerous nuclear rivals, and unending amounts of political, cultural, and social divide, music helps human beings remember the similarities and harmony amongst one another. After discovering the main reasons music is an effective force to bridge divides these facets can now be implemented into other areas of the world. Humans can create more universal languages that can be understood, humans can facilitate more political international collective projects, as a humanity we realize the importance of purpose and intention within projects, political endeavors, and more. The next steps of research are to create case studies on how to take these 5 main aspects of music and apply those to other aspects of the world and policy making to ultimately make a more united world.

Acknowledgements

Thank you to my friends, family, and especially mentors who supported me while writing this paper and while going on an exploration of music.

Works Cited

Barone, J. (2018, August 12). *A German Opera Spotlights the Refugee Crisis, With Refugees (Published 2018)*. The New York Times. Retrieved July 31, 2023, from https://www.nytimes.com/2018/08/12/arts/music/germany-refugee-crisis-opera-munich.html

Benjamin, A. (2023, January 16). *refugee-musicians-hit-right-note-german-orchestra*. UNHRC. Retrieved July 29, 2023, from http://www.unhcr.org/news/stories/refugee-musicians-hit-right-note-german-orchestra.

India and Pakistan soldiers jam to Sidhu Moosewala's songs at LOC; fans say late singer is 'bridging the divide' - WATCH | Punjabi Movie News - Times of India. (2022, August 27). The Times of India. Retrieved July 31, 2023, from https://timesofindia.indiatimes.com/entertainment/punjabi/movies/news/india-and-pakistan-soldiers-jam-to-sidhu-moosewalas-songs-at-loc-fans-say-late-singer-is-bridging-the-divide-watch/articleshow/93814240.cms

International Phonetic Alphabet (IPA) | Definition, Uses, & Chart. (2023, July 3). Britannica. Retrieved July 31, 2023, from https://www.britannica.com/topic/International-Phonetic-Alphabet

Johann Adam Hiller Biography - German composer and conductor (1728–1804). (n.d.). Pantheon. Retrieved July 31, 2023, from https://pantheon.world/profile/person/Johann_Adam_Hiller/

Le Blond, J. (2018, June 26). *Refugee musicians hit right note with German orchestra*. UNHCR. Retrieved July 31, 2023, from https://www.unhcr.org/news/stories/refugee-musicians-hit-right-note-german-orchestra

Music can bridge India-Pakistan divide: Pakistani filmmaker (IANS Interview). (2016, May 20). Business Standard. Retrieved July 31, 2023, from https://www.business-standard.com/article/news-ians/music-can-bridge-india-pakistan-divide-pakistani-filmmaker-ians-interview-116052000439_1.html

Music to bridge India-Pakistan divide: Atif Aslam. (2012, September 10). India Today. Retrieved July 31, 2023, from https://www.indiatoday.in/movies/celebrities/story/music-to-bridge-india-pakistan-divide-atif-aslam-115639-2012-09-09

Post, J. C. (2014, June 20). *Hybridity and Postcolonial Music – Postcolonial Studies*. ScholarBlogs. Retrieved July 31, 2023, from https://scholarblogs.emory.edu/postcolonialstudies/2014/06/20/hybridity-postcolonial-music/

Shearer, D. (2013, October 15). *The Use of the International Phonetic Alphabet in the Choral Rehearsal*. The IFCM Magazine. Retrieved July 31, 2023, from http://icb.ifcm.net/use-international-phonetic-alphabet-choral-rehearsal-edited-duane-richard-karna/

Singspiel | German, Musical Comedy & Comic Opera. (n.d.). Britannica. Retrieved July 31, 2023, from https://www.britannica.com/art/singspiel

Stark, J. (2021, March 18). *Singspiel – what is that?* Jonathan Stark – Conductor. Retrieved August 26, 2023, from https://starkconductor.com/singspiel-what-is-that/

Ward, C. (May 2021). Musical Textures of Migration: Music's Role in the Syrian Refugee Experience in Germany. 59. https://cdr.lib.unc.edu/concern/file_sets/v405sk344

Fostering Academic Advancement in São Luís: Enhancing Research Opportunities and Laboratory Infrastructure for High School Students By Camila W. Maalouf

1. Introduction

High school constitutes the final stage of Brazilian basic education. Over the course of its three-year duration, educational institutions are expected to empower students by solidifying the knowledge acquired during elementary education and awakening skills that will be beneficial in the academic lives of the learners, qualifying them for entry into higher education.

However, these objectives are challenged by an outdated educational methodology that is limited to content exposition and subjects students to a passive stance, leading to disinterest in the taught disciplines. The areas of Natural Sciences, specifically, due to their content nature of well-defined theories conceived within a context of technical rationality reminiscent of positivism, encourage the memorization of concepts, thereby inhibiting the implementation of more active methodological approaches (Mota, 2019). Resources such as encouragement of high school research and practical classes in laboratories, however, present themselves as alternatives to address this issue.

This article, a bibliographic review with qualitatively analyzed data, aims to examine the advantages of promoting research for high school students, as well as the benefits of implementing science laboratories during this stage of basic education, presenting them as pathways to educational advancement. Furthermore, it seeks to reveal the current landscape of these two aspects in the city of São Luís, the capital of the state of Maranhão, Brazil.

The relevance of this work is justified by considering that it addresses topics that concern the entire student body of high schools in São Luís, which, in 2021, accounted for 44,713 enrollments, with 37,597 in public schools and 7,116 in private schools (QEdu, 2021). The study also demonstrates importance by advocating for situations such as the promotion of science, supported by national legislative documents (Brasil, 1988, 1996, 2013).

The paper consists of this introduction, the objectives (subdivided into general and specific), the theoretical framework (subdivided into aspects related to high school research and science laboratories in high schools), the methodology, the conclusion, and the works cited. The chapters and subchapters are arranged sequentially.

2. Objectives

2. 1. General objective

- Examine the advantages of promoting research opportunities during high school, as well as the benefits of implementing science laboratories during this stage of basic education, presenting them as pathways to educational advancement.

2. 2. Specific objectives

- Reveal the current landscape of research engagement by students in high schools in the city of São Luís.
- Present data corresponding to the number of high schools in São Luís that have science laboratories.

3. Theoretical framework

3.1. Research in high school

Research is an experience that allows students to engage in the process of scientific production through the development of academic material. Typically, it takes place at the undergraduate level and serves as the students' first exposure to research practices. However, certain initiatives enable high school students to enjoy the benefits of research production, often referred to, in Portuguese, as "Iniciação Científica Júnior" or "Pré-Iniciação Científica" in these cases (Oliveira, 2020). A free translation to "Iniciação Científica Júnior" and "Pré-Iniciação Científica" could be “high school research”.

Exposure to this type of experience during basic education allows for an early awakening to “Scientific Literacy”, which, as described by Costa (2015, p.13), corresponds to "a contemporary concept that reflects people's critical thinking regarding the basic understanding of Sciences and their utilization"[142], bringing the young student closer to the educational atmosphere of higher education. According to the author, the implementation of research as a pedagogical tool in high school goes beyond academic benefits, as it also promotes the exercise of autonomy, reflective thinking, and the development of critical discernment about the reality in which the individual lives. This underscores the relevance of this activity in the formation of personal and collective values (Costa, 2015).

The study conducted by Costa and Zompero (2017) analyzes official Brazilian educational documents, highlighting passages that support the obligation to promote research and provide indications of the need for the curricular integration of research production in high school. The studied documents included the “Constituição da República Federativa do Brasil de 1988” (Constitution of the Federative Republic of Brazil of 1988), the “Lei de Diretrizes e Bases de 1996” (Law of Guidelines and Bases of 1996), and the “Diretrizes Curriculares Nacionais Gerais para a Educação Básica de 2013” (National General Curricular Guidelines for Basic Education of 2013).

Regarding the “Constituição da República Federativa do Brasil de 1988”, the highest law in Brazil, Costa and Zompero (2017) highlight Article 218, which deals with state incentives for scientific and technological development and research practices. This article does not exclude the involvement of basic education in these competencies, nor does it restrict their exercise to higher education. This suggests that the constitutional encouragement for scientific initiation is not dependent on the level of education:

[42] Translated from Portuguese

> § 1º - Basic scientific research shall receive priority treatment from the State, with a view to the public good and the progress of science, technology, and innovation.
> § 2º - Technological research shall be predominantly focused on solving Brazilian problems and on the development of the national and regional production system.
> § 3º - The State shall support the training of human resources in the fields of science, research, technology, and innovation, including through support for technological extension activities, and shall grant those engaged in these activities special means and conditions of work. (Brasil, 1988)[43]

Regarding the “Lei de Diretrizes e Bases de 1996” (LDB 9394/96), which regulates national education, Costa and Zompero (2017) draw attention to Article 35, item IV, which states that one of the objectives of high school education is the "understanding of the scientific-technological foundations of the production processes, relating theory to practice in the teaching of each discipline" (Brasil, 1996, p.14)[44]. The item I of § 1º of Article 36 of LDB 9394/96 is also highlighted by Costa and Zompero (2017). Concerning this passage:

> § 1º - The content, methodologies, and evaluation methods will be organized in such a way that by the end of high school, the student demonstrates:
> I - mastery of the scientific and technological principles that govern modern production. (Brasil, 1996, p.15)[45]

Based on the excerpts from the LDB 9394/96 highlighted above, it becomes evident that high school education should contribute to the scientific education of the student. Therefore, the promotion of research practices during this stage of basic education qualifies as a means to achieve this goal. Early exposure to research exposes the student to the scientific method and provides the opportunity for practical application of theoretical knowledge acquired in the classroom. This promotes educational advancement by ensuring the realization of the intentions presented in the fragments of LDB 9394/96.

Regarding the “Diretrizes Curriculares Nacionais Gerais para a Educação Básica de 2013”, which is the document that guides the processes related to the pedagogical proposals of Brazilian educational networks, Costa and Zompero (2017, p. 20) highlight several passages that explicitly support the implementation of research practices as an educational tool (Brasil, 2013). Therefore, the guidelines reaffirm the legitimacy of this activity as an educational catalyst.

[43] Translated from Portuguese
[44] Translated from Portuguese
[45] Translated from Portuguese

Access to research practices for high school students in Brazil was facilitated after 2003 with the establishment of the program "Iniciação Científica Júnior (ICJ)" (Junior Scientific Initiation), introduced by the "Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)" (National Council for Scientific and Technological Development), an entity associated with the Brazilian Ministry of Science, Technology, and Innovations that promotes Brazilian research (Costa and Zompero, 2017). According to Oliveira and Bianchetti (2018), the ICJ consists of three initiatives: the "Programa Institucional de Bolsas de Iniciação Científica para o Ensino Médio (PIBIC-EM)" (Institutional Program for Scholarships in Scientific Initiation for High School), the "Programa de Iniciação Científica Júnior (IC-Jr/FAPS)" (Junior Scientific Initiation Program), and the "Programa de Iniciação Científica da Olimpíada Brasileira de Matemática das Escolas Públicas (PIC-OBMEP)" (Scientific Initiation Program of the Brazilian Olympiad of Mathematics for Public Schools).

The PIBIC-EM is operated by educational and research institutions that have applied for a quota of scholarships from the CNPq. The scholarships are reserved for students from public schools or private application schools (Brasil, 2022). In Maranhão, institutions such as the "Instituto Federal de Educação, Ciência e Tecnologia" (Federal Institute of Education, Science and Technology of Maranhão) (IFMA, 2021) and the "Universidade Federal do Maranhão" (Federal University of Maranhão) (UFMA, c2022) implement the PIBIC-EM. In the case of IFMA, the positions are exclusively for students enrolled in the Technical Education Professional Course offered by the institution (IFMA, 2021). The IC-Jr/FAPS, on the other hand, is implemented through "Fundações de Amparo à Pesquisa" (State Foundations for Research Support) and is intended for students from public schools (Brasil, [n. d.]). In Maranhão, the scholarships are granted by the "Fundação de Amparo à Pesquisa do Maranhão (FAPEMA)" (Maranhão Research Support Foundation) (Maranhão, c2021). As for the PIC-OBMEP, it is provided to award-winning students of the "Olimpíada Brasileira de Matemática das Escolas Públicas (OBMEP)" (Brazilian Mathematical Olympiad for Public Schools), who are enrolled in public schools (OBMEP, 2022).

It is also possible to consider that schools may have their own research programs for high school students. However, data on this topic was not gathered, as collecting such information would involve individually searching for these programs on the websites of each high school in the municipality. Based on the data compiled on the QEdu portal from the 2021 School Census, there are a total of 161 high school institutions (QEdu, 2021). Additionally, it's worth noting the existence of a state-level research program intended for high school students, established through a partnership between FAPEMA and the "Secretaria Extraordinária de Juventude do Maranhão (SSEJUV)" (Extraordinary Secretariat for Youth of Maranhão) (Maranhão, 2019).

Knowing how many students enrolled in high schools in the city of São Luís are involved in research practices is an important factor for measuring the impact of public policies aimed at promoting research among this demographic at the municipal level. It also helps verify the fulfillment of the objectives mentioned in the legislative documents quoted on this article. Due to the challenge of tracking how many scholarships from the described high school research

initiatives were granted to students in São Luís, as the allocation of scholarships to students in most programs is determined by teachers who qualify as mentors within the respective programs, the analysis chose to examine the number of students from São Luís who were finalists in the "Feira Brasileira de Ciências e Engenharia (FEBRACE)" (Brazilian Science and Engineering Fair) over the last five years. This choice takes into consideration the wide reach of FEBRACE, considered the largest science fair in the country by the "Universidade de São Paulo (USP)" (University of São Paulo) (2020), the ease of accessing the desired information, and the scientific quality of the finalist projects, as indicated by their "finalist" status. It is noteworthy that research projects presented in exhibitions or fairs like FEBRACE also contribute to science communication, which, according to Rocha (2012), is regarded as a means of social inclusion that fosters public understanding of science and contributes to upholding democratic values.

The lists of finalists for the years 2018, 2019, 2020, 2021, and 2022 of FEBRACE were consulted. In 2018, Maranhão had six qualified projects: five from the city of Imperatriz and only one from São Luís. In the subsequent years, there were no records of finalists from São Luís. In the years 2019, 2020, 2021, and 2022, Maranhão had seven projects, seven projects, two projects, and three projects qualified, respectively. All the qualified projects from Maranhão were from the city of Imperatriz, except for two projects in 2019, which individually represented the municipalities of Açailândia and São João do Paraíso (FEBRACE, 2018, 2019, 2020, 2021, 2022).

Although this sample is not ideal for quantifying the proportion of high school students in São Luís engaged in research practices, as it is limited to projects submitted to FEBRACE, it is still an indicator of the low expressiveness of high school research in São Luís, especially when compared to the city of Imperatriz. Possible reasons for this scenario could be: the structure of the scholarship allocation system for PIBIC-EM and IC-Jr/FAPEMA, which instead of allowing students to directly apply for scholarships, assigns this task to teachers who serve as mentors, thus making it difficult for students without prior connections; the limited accessibility and promotion of the hosting websites for calls for proposals, which are rarely disseminated to the general public; the exclusivity of the PIC-OBMEP scholarships, which restrict scientific activities to those who have won awards, rather than providing an opportunity to expose students to Exact Sciences more broadly in basic education, as a means to foster education in these areas; the limited availability of research programs that include students from private schools as well.

The limited participation in research programs in the city of São Luís consequently leads to a decrease in student engagement in fairs and exhibitions, which are conducive environments for individuals to be exposed to science communication and its benefits. It also results in lesser integration of theoretical knowledge into practice, distancing individuals from the goals of higher education, and a lack of familiarity with the scientific method.

3. 2. High school science laboratories

High school science laboratories, as they provide an environment conducive to conducting experiments and practical lessons, promote the breaking of passive teaching molds

described by Mota (2019), and they allow for a more active interaction of students with their subjects of study. Furthermore, they serve as infrastructural resources that facilitate the execution of research practices, particularly in the field of Natural Sciences. The purposes of such spaces are also in line with the aspirations of educational and science-promoting themes taken from legislative documents referenced earlier on this article, demonstrating the significance of implementing science laboratories for educational advancement (Brasil, 1988, 1996, 2013).

In the municipal context, data from the 2021 school census, collected by the "Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep)" (National Institute for Educational Studies and Research Anísio Teixeira) and compiled by the QEdu portal, reveal that the city of São Luís, the capital of the state of Maranhão, had 161 Regular High Schools, with 100 belonging to the public sector and 61 to the private sector. Out of these educational institutions, only 50% of public Regular High Schools and 34% of private Regular High Schools had science laboratories (QEdu, 2021).

The potential causes for this unsatisfactory situation could be, in the case of public educational institutions, the high expenses associated with establishing and maintaining a laboratory, as well as the lack of physical space to accommodate such infrastructure (Mota, 2019). Furthermore, in line with Mota's study (2019), the low prevalence of laboratories in private schools corresponds to the lack of prioritization of practical lessons. This is due to the underestimation of the benefits of this methodology compared to predominantly theoretical classes, combined with the heavy workload of students and teachers. The latter often leaves no time for conducting laboratory activities, making the implementation of such infrastructure seemingly unnecessary.

Mota (2019, p. 60), however, disagrees regarding the dismissal of laboratory installation, despite the obstacles faced in establishing these spaces in high schools: "the investment in time and energy, the cost of providing space for specialized laboratories, equipment, and consumable materials are fully justified when we consider the importance of scientific work and the good results that are achieved and often presented at Science Fairs"[46].

Therefore, it can be inferred that the low number of science laboratories in high schools leads to a reduction in the number of practical classes, less integration of theory and practice, hindrance in implementing scientific research proposals, especially those related to the Natural Sciences field, and the promotion of the maintenance of passive teaching methodologies (Mota, 2019).

4. Methodology

The present study is a bibliographic review with qualitatively analyzed data on the topic "Fostering Academic Advancement in São Luís: Enhancing Research Opportunities and Laboratory Infrastructure for High School Students". Its general objective is to examine the advantages of promoting research opportunities during high school and the benefits of

[46] Translated from Portuguese

implementing science laboratories in high schools at this stage of basic education, presenting them as pathways to educational advancement.

For the collection of materials, searches were conducted on the scientific databases Google Scholar and Scientific Electronic Library Online (SciELO) using the following search descriptors in Portuguese: "iniciação científica," "laboratório de ciências," "ensino médio." The descriptors were combined using the Boolean operator "and" to broaden the scope of material. However, during the selection of materials, this article also includes studies referenced by the authors of the works collected in the databases, which were evaluated as relevant for the preparation of this work.

The inclusion criteria used in this research encompass materials such as articles and theses, published in full, within the time frame from 2012 to 2022, in the Portuguese language, written in their entirety. Additionally, reports from scientific fairs, high school research program announcements, and legislative documents were also considered.

As exclusion criteria involved discarding articles and theses published prior to 2012, those not written in the Portuguese language, and those that were not available in their entirety. However, exceptions to this rule were made for legislative documents, which, for the most part, date back to a period before 2012.

A total of 23 materials were selected. The initial screening of the material, concerning theses and articles, was carried out by reading the abstracts to identify works that directly addressed the topic of this research.

5. Conclusion

Given the presented information, the relevance of encouraging research in high school, as well as the establishment of new science laboratories, is confirmed. Both situations facilitate the student's engagement with the scientific method and contribute to the learning process, as the connection between theoretical knowledge and practical applications will be enhanced.

Regarding the study's objectives, it can be concluded that the general objective and the second specific objective were achieved. The first specific objective was partially attained, as the inability to track how many scholarships from the described high school research programs were awarded to students in the municipality of São Luís led the study to adopt a makeshift approach. However, this approach still demonstrated the limited prominence of high school research in São Luís, especially when compared to the municipality of Imperatriz. This point, therefore, could be a target for future research aiming for greater accuracy in these data.

In terms of suggestions to address the challenges encountered during the research, it is recommended to increase the reach of scientific initiation in high school among students in São Luís by redesigning the websites hosting the notices for the “PIBIC-EM” and “IC-Jr/FAPEMA” programs. These websites should be made more user-friendly, with better-organized information. Additionally, calls for program applications should be more effectively promoted to the general public, which can be accomplished through the social media channels of the organizations involved in running the programs. Furthermore, the creation of new science fairs and exhibitions

through the Secretaria de Estado da Ciência, Tecnologia e Inovação is encouraged, as it enhances the visibility of the topic.

Finally, regarding the establishment of new science laboratories in São Luís, the most plausible approach is to allocate state funds to public high schools that do not have access to such resources through programs from the state of Maranhão such as "Programa Escola Digna", which aims to improve and transform the teaching and learning process in basic education. If there are still infrastructure limitations for creating these spaces, turning to digital tools such as virtual laboratories can be an alternative to promote practical classes. These virtual resources can be funded, if necessary, through the same program.

Works Cited

Brasil. Constituição (1988). Constituição da República Federativa do Brasil. Brasília, 1988, http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Accessed June 16, 2022.

Brasil. Lei nº 9.394, de 20 de dezembro de 1996. Estabelece as diretrizes e bases da educação nacional. Brasília, 1996, http://www.planalto.gov.br/ccivil_03/leis/L9394.htm. Accessed June 16, 2022.

Brasil. Ministério da Ciência, Tecnologia e Inovações. ICJ - Programa de Iniciação Científica Júnior. Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq. Brasília, n.d., https://memoria2.cnpq.br/ic-jr/faps/. Accessed June 18, 2022.

Brasil. Ministério da Ciência, Tecnologia e Inovações. PIBIC-EM. Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq. Brasília, 2022, https://www.gov.br/cnpq/pt-br/acesso-a-informacao/acoes-e-programas/programas/programas-ict/pibic-em. Accessed June 18, 2022.

Brasil. Ministério da Educação. Diretrizes Curriculares Nacionais Gerais da Educação Básica. Brasília: MEC, 2013, http://portal.mec.gov.br/docman/julho-2013-pdf/13677-diretrizes-educacao-basica-2013-pdf/file. Accessed June 17, 2022.

Costa, W. L. da. "A CTS (Ciência, Tecnologia e Sociedade) na compreensão dos alunos que participam da iniciação científica no Instituto Federal do Paraná." Londrina, 2015, https://repositorio.pgsskroton.com/bitstream/123456789/832/1/A%20CTS%20(Ciência,%20Tecnologia%20e%20Sociedade)%20na%20compreensão%20dos%20alunos%20que%20participam%20da%20iniciação%20cient%C3%ADfica%20no%20Instituto%20Federal%20do%20Paraná.pdf. Accessed June 14, 2022.

Costa, W. L. da., & Zompero, A. de F. "A iniciação científica no Brasil e sua propagação no ensino médio." Londrina, 2017, https://revistapos.cruzeirodosul.edu.br/index.php/rencima/article/view/988/865. Accessed June 16, 2022.

Feira Brasileira de Ciências e Engenharia - FEBRACE. "Lista de finalistas Região Nordeste 2018." 2018, https://febrace.org.br/wp-content/uploads/2021/06/lista_finalistas_2018_NE.pdf. Accessed June 11, 2022.

Feira Brasileira de Ciências e Engenharia - FEBRACE. "Lista de finalistas Região Nordeste 2019." 2019, https://febrace.org.br/wp-content/uploads/2021/06/lista_finalistas_2019_NE.pdf. Accessed June 11, 2022.

Feira Brasileira de Ciências e Engenharia - FEBRACE. "Lista de finalistas Região Nordeste 2020." 2020,

https://febrace.org.br/wp-content/uploads/2021/06/lista_finalistas_2020_NE.pdf. Accessed June 11, 2022.

Feira Brasileira de Ciências e Engenharia - FEBRACE. "Lista de finalistas Região Nordeste 2021." 2021, https://febrace.org.br/wp-content/uploads/2021/06/lista_finalistas_2021_NE.pdf. Accessed June 11, 2022.

Feira Brasileira de Ciências e Engenharia - FEBRACE. "Lista de finalistas Região Nordeste 2022." 2022, https://febrace.org.br/wp-content/uploads/2021/12/lista_finalistas_2022_NE.pdf. Accessed June 11, 2022.

Instituto Federal de Educação, Ciência e Tecnologia do Maranhão - IFMA. PIBIC Ensino Médio Complementar - Vigência 2021/2022. São Luís, 2021, https://portal.ifma.edu.br/concursos-e-seletivos/?d=KyMzdWRdMEtRIkMmUENcRX5oc0B6RHxGZFdEQUNHVXNTRVBBUkFET1JASUZNQTE0Y2E3ZmY2NzU4YWQ1MGI4ODMyNjc3ZjJhYWQ3Zlt8XTAwMV9Qcm9ncmFtYV9JbnN0aXR1Y2lvbmFsX1JFSVRfRURJVEFMX1BSUEdJX04lQzIlQkFfMTM1MjAyMS5wZGY=. Accessed June 18, 2022.

Maranhão. Secretaria de Estado da Ciência, Tecnologia e Inovação. Edital FAPEMA N° 010/2018 – Geração Ciência. Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão - FAPEMA. São Luís, 2019, https://www.fapema.br/wp-content/uploads/2019/04/Edital-010_2019-GERAÇÃO-CIÊNCIA_retificado.pdf. Accessed June 19, 2022.

Maranhão. Secretaria de Estado da Ciência, Tecnologia e Inovação. Modalidades e valores de bolsa. Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão - FAPEMA. São Luís, c2021, https://www.fapema.br/modalidades-de-bolsa/. Accessed June 19, 2022.

Mota, M. D. A. Laboratórios de Ciência/Biologia nas escolas públicas do estado do Ceará (1997 - 2017): realizações e desafios. Fortaleza, 2019, https://repositorio.ufc.br/bitstream/riufc/45994/1/2019_tese_mdamota.pdf. Accessed June 13, 2022.

Olimpíada Brasileira de Matemática das Escolas Públicas - OBMEP. Programa de Iniciação Científica Jr. (PIC). 2022, http://www.obmep.org.br/pic.htm. Accessed June 19, 2022.

Oliveira, A. de, & Bianchetti, L. "Iniciação Científica Júnior: desafios à materialização de um círculo virtuoso." Florianópolis, 2018, https://www.scielo.br/j/ensaio/a/hTpGjQgGW4Mv6S5NsFMk6YL/?format=pdf&lang=pt. Accessed June 17, 2022.

Oliveira, S. F. do A. Iniciação Científica: Guia básico para iniciantes. Morrinhos, 2020, https://educapes.capes.gov.br/bitstream/capes/574724/2/Iniciação%20Cient%C3%ADfica-Guia%20básico%20para%20Iniciantes.docx.pdf. Accessed June 9, 2022.

QEdu. Censo Escolar 2021. QEdu: Use dados. Transforme a educação. 2021, https://novo.qedu.org.br/municipio/2111300-sao-luis/censo-escolar. Accessed June 8, 2022.

Rocha, M. B. O potencial didático dos textos de divulgação científica segundo professores de ciências. Curitiba, 2012, https://revistas.utfpr.edu.br/rbect/article/download/1263/847. Accessed June 12, 2022.

Universidade de São Paulo - USP. Jornal da USP. "Maior feira de ciências do País, Febrace tem inscrições abertas para projetos." São Paulo, 2020, https://jornal.usp.br/universidade/maior-feira-de-ciencias-do-pais-febrace-tem-inscricoes-abertas-para-projetos/. Accessed June 11, 2022.

Universidade Federal do Maranhão - UFMA. Agência de Inovação, Empreendedorismo, Pesquisa, Pós-Graduação e Internacionalização - AGEUFMA. PIBIC-EM. São Luís, c2022, https://portais.ufma.br/PortalProReitoria/ageufma/paginas/pagina_estatica.jsf?id=346. Accessed June 18, 2022.

Using Machine Learning to Improve the Accuracy of Sentiment Analysis on Amazon Consumer Reviews By Mahatru Guddamsetty

Abstract

There are many prominent e-commerce websites (e.g. Amazon, eBay, etc.) which have a variety of products available with a colossal volume of respective reviews per product. In the face of an immense influx of consumer reviews, it is vital for companies to utilize efficient methods to gain valuable insights and improve their products. By utilizing machine learning to enhance sentiment analysis accuracy on consumer review datasets, companies can appropriately make data-driven decisions and enhance their products based on customer feedback. This research paper aims to compare a machine learning model (RoBERTa) to a lexicon-based model (VADER) and show how the former more accurately performs sentiment analysis on a large consumer review dataset.

1. Dataset Description

The steps taken to preprocess the consumer review data for the machine learning and lexicon-based model are described in this section. Both models utilize a dataset consisting of a diverse and large collection of approximately 3.6 million consumer reviews from Amazon, a prominent e-commerce platform. Each review is originally accompanied by either of the following two sentiment labels: 0 indicating a negative review and 1 indicating a positive review.

1.1: Preprocessing

For every line of the dataset, a split function is used to separate each review from its sentiment label; each sentiment label is mapped to an integer with 0 denoting negative sentiment and 1 denoting positive sentiment. Any trailing white spaces or null entries are erased from the reviews to ensure consistency and cleanliness with the data. The data was split into three partitions: training, validation, and test. The train partition retains a small portion of the original dataset for the purpose of training. The validation partition aids in hyperparameter tuning. The test partition serves as an aid for evaluating generalization to unseen review data.

2. Methodology–2.1: Diagrams

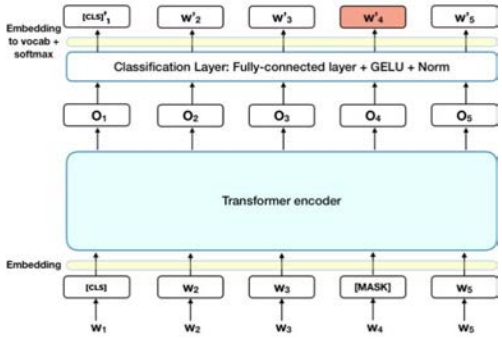


Fig 1: A breakdown of RoBERTa's architecture and framework

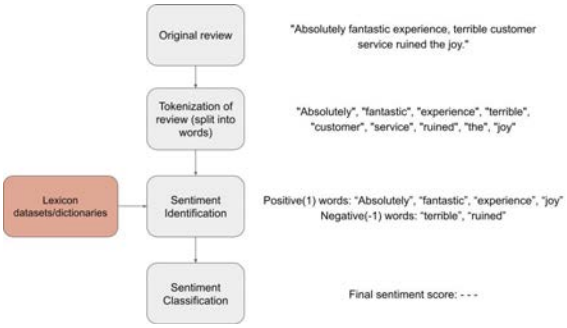


Fig 2: A breakdown of VADER's architecture and framework

2.2: The architecture behind RoBERTa

RoBERTa, an advanced natural language processing model based on BERT (which stands for Robustly Optimized BERT-Pretraining Approach), builds upon its predecessor in several ways. RoBERTa trains on a massive 160GB of uncompressed text from various domains, significantly beating the 10GB corpus used to train BERT. RoBERTa abandons the previously-used next-sentence prediction task in favor of only bidirectional training, which uses left and right contexts, and an increase in training steps. RoBERTa also employs larger mini-batches and higher learning rates for more efficient learning, and adopts the byte-level BPE tokenizer from BERT. Lastly, whereas BERT was trained using masked language modeling using a fixed mask for each sentence, RoBERTa expands upon the approach by dynamically changing the masking pattern during pre-training, allowing the mask to vary for each training sentence of a dataset.

2.3: How does RoBERTa work?

RoBERTa operates through a series of intricate steps that enable it to perform natural language processing tasks, such as sentiment analysis in the case of this project. Refer to Figure 1 for a detailed breakdown of RoBERTa's architecture; at its core, RoBERTa leverages a bidirectional transformer-based architecture. The model tokenizes text through its input layers, passing it through multiple transformer encoder layers, generating intermediate representations at different stages of processing. These intermediate representations hold essential contextual information crucial for understanding the nuances in text. To make predictions, a classification layer is added over the encoder's first output vector, corresponding to the [CLS] token in the input. This classification layer projects the resulting vector into the vocabulary dimension and produces a probability distribution over the classes using soft-max. The resulting probabilities guide the model in generating meaningful predictions based on the input text.

In the sentiment analysis program, RoBERTa requires various steps to determine the sentiment for each review from the Amazon dataset: model initialization and definition, data tokenization and encoding, training on the cross-entropy loss with Adam, and performance evaluation through the validation dataset. After preprocessing the consumer review data, a pre-trained RoBERTa model for sentiment classification tasks is loaded. The preprocessed consumer review data is converted into a pandas dataset to interface with PyTorch's DataLoader

for handling during RoBERTa's training, validation, and testing phases. While training, the cross-entropy loss and Adam optimizer are used to aid RoBERTa in adapting its parameters, minimizing its loss, and efficiently optimizing its sentiment predictions. Validation ensures that there is not excessive over-fitting with the dataset, which would lead to RoBERTa performing well with the train dataset but failing to generalize any new and unseen data. After choosing the hyperparameters on the validation dataset, the final model is evaluated on the test set. The model was fine-tuned for 1 epoch with a learning rate of 5e-5 and batch size of 16.

2.4: How does the lexicon-based VADER model work?

VADER (short for Valence Aware Dictionary and Sentiment Reasoner) is a lexicon-based model designed to analyze the sentiment of text data. Refer to Figure 2 for a diagram which breaks down VADER's architecture and sentiment analysis process. VADER utilizes a dictionary that contains an extensive variety of words, each associated with their respective sentiment scores. Such scores are essential as they allow VADER to determine between positive and negative words, while also considering factors such as intensity modifiers (i.e. "very", "extremely", etc.) and punctuation (i.e. exclamation marks, question marks, etc). VADER also uniquely accounts for emojis, slang phrases, and acronyms. VADER's sentiment analysis process involves parsing through the text review data, identifying and assigning sentiment scores within it, and then calculating a final valence sentiment score. By summing up the amount of positive and negative terms detected in a sentence, VADER provides a quantitative measure, being the final valence sentiment score, allowing for the polarity of the text data to be gauged.

2.5: Method of Comparison

To assess the accuracy and performance of RoBERTa and VADER on the Amazon consumer review dataset, a Receiver Operating Characteristic (ROC) curve was constructed on the test partition to visualize the models' ability to distinguish between positive and negative sentiment. There were additional metrics used for RoBERTa for further evaluation (i.e. train and validation datasets' loss and accuracy).

The ROC curves for RoBERTa and VADER were created by systematically adjusting the classification threshold to compute their true positive rates (TPR) and false positive rates (FPR). Refer to \textbf{Figure 3} for a visual representation of the ROC curves showing a trade-off between correct classification of positive sentiment as TPR and incorrect classification of negative sentiment as FPR.

To quantify the classification power of both models, the area under the ROC curve (AUC-ROC) was calculated; a higher AUC-ROC signifies a more accurate sentiment analysis model. If the model exhibits a curve closer to the top-left corner, there is indication of strong performance through the achievement of higher TPR while maintaining lower TPR.

3. Results–3.1 AU-ROC Curve Graph

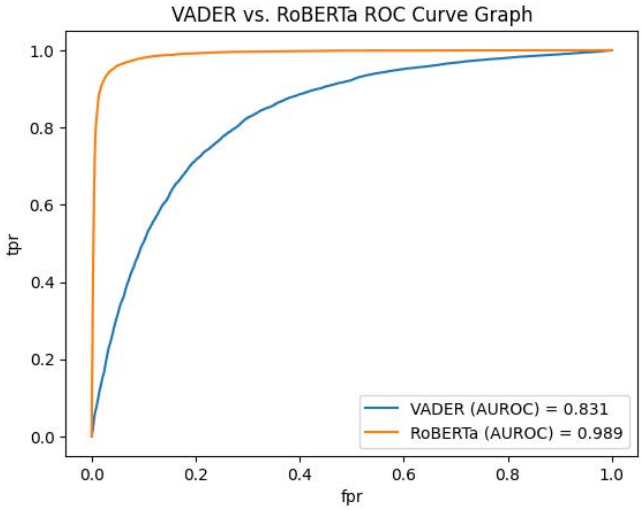


Fig 3: VADER vs. RoBERTa AUROC Comparison Graph

3.2 Weights and Biases Graphs

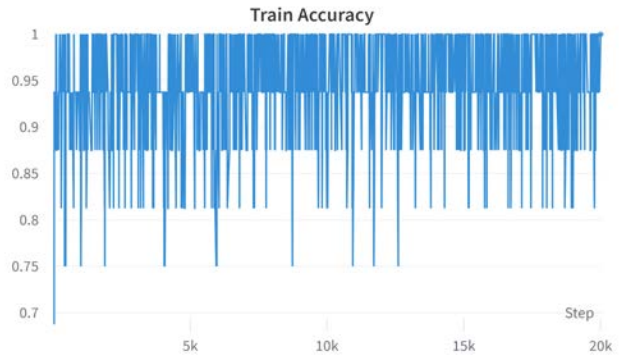


Fig 4: Train Dataset Accuracy Graph

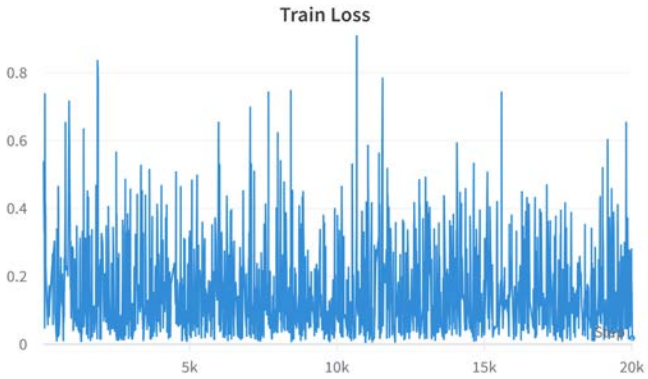


Fig 5: Train Dataset Loss Graph

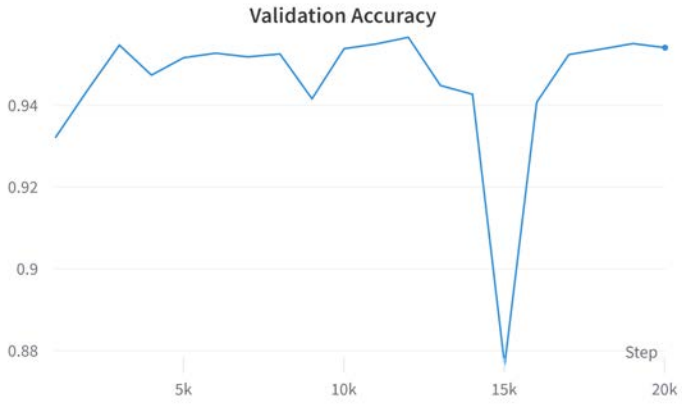


Fig 4: Validation Dataset Accuracy Graph



Fig 4: Validation Dataset Loss Graph

3.3: Results

VADER achieved an AUC-ROC value of 0.831, while RoBERTa achieved an AUC-ROC value of 0.989. During its training phase, RoBERTa achieved a perfect accuracy value of 1, while maintaining a low loss value of 0.03064. During the validation phase, a high accuracy of 0.9542 was maintained by RoBERTa with a minimal loss of 0.003879. In further validation of RoBERTa's performance, a low test loss value of 5.3e-6 was achieved across all test datasets.

3.4: Discussion

Significant performance differences were observed between RoBERTa and VADER. The ROC curve analysis, as reflected by the AUC-ROC values, shows RoBERTa's superiority over VADER. RoBERTa's AUC-ROC score of 0.989 significantly outperforms VADER's score of 0.831, indicating that RoBERTa excels in accurately distinguishing between positive and negative sentiments.

RoBERTa's accuracy metrics further validate its effectiveness. Achieving a perfect accuracy on the train dataset while maintaining a low loss value emphasizes its capacity to fit the training data closely and minimize errors. The high accuracy of 0.9542 on the validation dataset, coupled with a low validation loss value, illustrates RoBERTa's ability to generalize effectively

to previously unseen data. As indicated by the small test loss value of 5.3e-6, it further reaffirms RoBERTa's predictive accuracy and robustness with the test datasets.

3.5: Conclusion

In this report, two models for sentiment analysis were compared: RoBERTa, a deep learning model, and VADER, a lexicon-based model. The goal was to assess their performances in distinguishing positive from negative sentiments within a vast dataset of Amazon consumer reviews. The results indicate a clear distinction between these two methods. RoBERTa achieved an impressive AUC-ROC score of 0.989, demonstrating its proficiency in sentiment analysis. It maintained high accuracy, scoring perfectly on its training dataset and maintaining a high accuracy of 0.9542 on the validation dataset. ON the other hand, VADER, while a valuable tool, showed limitations in accurately discerning nuanced sentiments within the dataset. Its AUC-ROC score was lower, and its ROC curve appeared to be more parabolic and less sharp. These findings highlight the advantages of utilizing machine learning models, and the struggles with contextual nuances and diverse language styles that non-machine learning models may not address.

3.6: Future Refinement

For future research purposes, other machine learning models apart from RoBERTa can be considered for their metrics in sentiment analysis (i.e. BERT, GPT-2, XGBoost, LightGBM, and Naive Bayes). Similarly, other non-machine learning models can be used in place of VADER for comparison in terms of their classification power (i.e. SenticNet, AFINN, and N-gram models). While this study focuses on an Amazon consumer review dataset, other types of datasets may be used such as posts from social media platforms (such as X), news articles, and other specialized domains.

3.7: GitHub Repository with Code + Weights and Biases Run Information

https://github.com/mgudd001/Sentiment-Analysis-Research-Paper

3.7: Acknowledgement

I would like to thank my mentor for giving me brief pointers where needed. This research project took an immense amount of time, so whenever I was stuck on debugging my code, my mentor was really helpful in pointing me in the right direction toward figuring out what areas of my code were wrong.

Works Cited

Liu, Yinhan, et al. "RoBERTa: A Robustly Optimized BERT Pretraining Approach." arXiv, 26 July 2019, doi:10.48550/arXiv.1907.11692.

Zhu, Yukun, et al. "Aligning Books and Movies: Towards Story-like Visual Explanations by Watching Movies and Reading Books." arXiv, 22 June 2015, doi:10.48550/arXiv.1506.06724.

Devlin, Jacob, et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." arXiv, 11 Oct. 2018, doi:10.48550/arXiv.1810.04805.

Hutto, C. and Eric Gilbert. "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text." ICWSM, vol. 8, no. 1, 16 May. 2014, pp. 216-25, doi:10.1609/icwsm.v8i1.14550.

Korea's Independence Movements: Influence of International Context in the East Asia-Pacific Region By Yejun Cheon

Abstract

The relationship between Korea and Japan is ancient, and its quality has varied over time. There have been periods when diplomatic exchanges or connections between the two countries were beneficial, but there were also several political incidents that left deep wounds. Japanese colonial rule from 1910 to 1945 remains a major impediment to repairing relations and boosting collaboration between Korea and Japan. Understanding precisely how Korea was freed from Japanese colonial rule is a crucial part of creating a peaceful concord in East Asia-Pacific region; Korean independence was not easily attained. Numerous large and small domestic efforts, including the March First Movement, served as the cornerstone, but the international context of the East Asia-Pacific region, which includes Russia, Japan, and even the United States, also had a significant influence. Both local and global circumstances shed light on how Korea progressed toward independence. Rather than just a conflict between Korea and Japan, Korea's independence was a complicated matter linked to greater East Asian and Western forces. It is largely helped by virtue of Korea's relative position in geopolitics that it arrived at its liberation. A close examination of a series of Korean independence movements that took place between 1910 and 1945, and the consequence of the Pacific War, provides a clearer understanding of the liberation of Korea from Japan.

Japanese Colonialism in Korea

The Japanese colonization of Korea was pervasive, touching almost every area of public and private life. In the very late 19th century, as Japan emerged from its victory in the first Sino-Japanese War, it began to exercise tighter and tighter control over the peninsula. Some prominent Koreans resisted Japan's growing influence, calling for closer ties with Russia to counter the trend. In response, Japan's resident minister sent assassins into the imperial palace to murder Korea's leading opponent of Japanese rule, Empress Min. When the Russo-Japanese War ended in 1905, Japan further strengthened its grip by establishing a protectorate and implementing reforms that included gutting Korea's national army. Five years later, Japan fully annexed the peninsula.

Formally annexed to Japan in 1910, Korea's once-vigorous public life all but disappeared through police surveillance and censorship. In the 1920s, following the independence demonstrations of March 1919, Koreans of various ideological stripes began to pry open a limited space for political action as the fetters of state control loosened; the public, led by the vernacular press, burst forth through a web of cultural activities and voluntary associations. However, contrary to the assumptions of Koreans, colonial Japan had no intention of capitulating to Korea's desire for independence. Although nominally offering freedom of expression and assembly to Koreans was one of colonial Japan's key foreign policies in the 1920s, Japan never intended on granting these rights to Koreans. As colonial Japan ostensibly allowed free

publication, *Chosun Ilbo* ran an editorial critical of the Japanese suppression of the Korean press. However, this criticism led to the indefinite suspension of the paper by the Japanese Government-General. Even from the 1930s through World War II, Japan formulated policies to assimilate Korea economically and socio-culturally into its empire, thereby eliminating the cultural identity of the Korean people. They prohibited the use of the Korean language at home and in schools, and they 'Japanized' Korean names until Korea was liberated in August 1945, when Japan surrendered to the Western Allies at the end of World War II– 35 years after Japan forcibly annexed the country in 1910. Forced assimilation festered in the minds of oppressed Koreans. The colonial period (1910–45) was bitter. Japan's military ruled directly and brutally, replacing Korean officials, censoring the media, and violently suppressing dissent. Thousands of Korean nationalists were jailed, and many were tortured during detention; others were summarily executed.

Cultural and Economic Oppression

Japan came to view the peninsula as an important part of its empire, and it pushed to 'Japanize' the economy and the culture. Japanese merchants moved into Korean cities, establishing shops and factories, while Japanese farmers took over Korean agricultural land. Especially in the 1910s, the early period of Japanese colonialism was characterized by policies that explicitly sought to discourage the development of Korean business. In the 1910s, to impede Korea's business, colonial Japan introduced a unique coercive policy known as the Japanese Government-General's Ordinance for Company Establishment. This decree was in effect from 1911 to 1920, during which time the Japanese seized control of a substantial portion of Korean industry. The core of the ordinance was the requirement to obtain permission from the Japanese Government-General of Korea to establish a company; it would not have been easy for Koreans to obtain such permission. The Japanese Government-General of Korea granted permission to establish Korean firms at a rate of more than 80 percent, but most Koreans who gained permission were part of pro-Japanese groups that had the financial resources to manage a business. In the 1920s, this system changed, but the influx of Japanese capital continued. 95 percent of the capital in the larger companies on the peninsula was under the control of Japanese businessmen by 1940.

Japanese was the language of instruction in all schools, and Japan's history was covered rather than Korea's. People were encouraged, if not necessarily obliged, to take on Japanese names. According to Uchida, Korean memory is fraught with resentment, alienation, and humiliation, at a minimum, often fermenting from as early as school age; these reminiscences pose an awkward match to Japanese claims to innocence and silence on colonial oppression. As a result of Japan's policy of assimilation, Koreans had a rough life and were referred to as 'another Japanese.' Koreans were under constant pressure to adopt a loyal Japanese identity in order to be called up for war, but in practice they were classified as an 'inferior Japanese' and were never truly allowed to fit in.

American Exceptionalism

The idea of American exceptionalism, though partly problematic, extended an olive branch to oppressed Koreans. During their long struggle against Japanese colonialism, Korean independence activists in the United States embraced the rhetoric of American exceptionalism and did what they could to encourage the American belief that the United States had a special mission to intervene in global affairs. Koreans argued that the United States' commitment to the spread of Christianity and the promotion of international law included a responsibility to help Korea. After American missionaries had introduced Protestantism to Korea during the 1880s; by 1919, Korea was one of the fastest growing Christian mission fields in the world. Koreans upheld the pride of the American Christians in their accomplishments in Korea, but also warned that Japanese policies threatened to reverse Christianity's gains in Korea if the United States did not take action. This line of thought was very enticing to the American public, which regarded itself as the defender of all just causes. Korean political activists in both America and China were struck by the relevance to their situation of President Woodrow Wilson's Fourteen Points speech of January 8, 1918. Article Five of Wilson's address stressed the principle of self-determination, especially the more specific idea that colonial claims should be adjusted to reflect the interests of the colonized population, not those of the colonial governments. Wilson called for a free, open-minded, and impartial adjustment of all colonial claims, based upon a strict observance of the principle that, in determining such questions of sovereignty, the interests of the populations concerned must have equal weight with the equitable claims of the government whose title is to be determined. Considering the war in Europe, Wilson's reference was technically limited to the colonies of the defeated powers, but his speech sparked hope nonetheless. Hence, many Koreans believed in Wilson's America as the ally, patron, and supporter of Korean independence.

Korean Nationalism and The March First Movement

Korean nationalism was a prerequisite for the liberation of Korea. Korean nationalism spawned from a complex interaction between 1880 and 1900 in the process of reforming the traditional Korean social system and resisting external imperialist forces. The failure of the traditional state to maintain its autonomy against the heavy pressure of Japan in the 1890s led to an increasingly iconoclastic tone in the writings of Korean reform thinkers. Questions of how societies develop, and what political and social forms they should take, gained in popularity. The combination led to a new, more sophisticated, reform movement that gave birth to modern Korean nationalism. After 1910, anti-Japanese sentiment provided the most powerful symbol for Korean national consciousness. There was no strong consensus as to the positive aspects of the nation. There was no doubt, however, about the sincere desire of Koreans for independence and their hatred of the Japanese colonial masters.

Occurred based on this spirit, The March First Movement was a catalyst for a nationwide call for Korea's independence. Koreans responded in unison to the Japanese mandate of assimilation on March 1, 1919 by hitting the streets to call for independence. Originally the

priority of a small group, the mandate generated a powerful emotional energy, which produced strong feelings of solidarity, confidence, and righteousness among those not in favor. In the ensuing days and weeks, spontaneous demonstrations spread across the country, with marching crowds waving Korean flags and chanting, "*Manse!*" ("May Korea live ten thousand years!"). The Japanese were ruthless in suppressing the peaceful protests. The demonstrations, which came to be known as the March First Movement, hastened the Japanese resolve to eradicate all vestiges of Korean nationalism. Thousands of Koreans were killed, arrested, imprisoned, and tortured. The movement included over two million people; around 1,500 demonstrations were held, 7,509 people were killed, 15,961 were wounded, and 700 houses and 47 churches were burned or otherwise destroyed.

Although the March First Movement failed in its ultimate objective, independence, it did cause the Japanese to revise their colonial policy. Following the March First Movement, the Japanese Governor-General of Korea shifted its policy from military to cultural policy (*bunka seiji*). The cultural policy was not simply a euphemism for colonial exploitation, but a set of policies and discourses based on colonial culture. At the same time as the cultural policy was instituted in Korea, culturalism became a dominant intellectual orientation in Japan and in the colonies. Ideas of culture were gradually linked to political domination. Within strict limits, Koreans got the right to publish, organize, and participate nominally in colonial governance. Japan, through close surveillance, attempted to control the organized resistance of Koreans who had gained this freedom; in many respects, Japan was able to achieve its goal. A nationwide struggle as large as the March First Movement did not take place after 1919 until Korea became independent in 1945. Nevertheless, there were also variables that Japan didn't predict, many spurred by its cultural policy. Expansion of educational opportunity, especially in the growing urban sector, and a concomitant rise in literacy and political consciousness expanded the base of the independence movement.

The March First Movement highlighted the Korean identity shift from the local to the national. In the nineteenth and early twentieth centuries, Korea was a fragmented society. However, after this movement, the citizenry adopted a nationalist tone and launched a movement for self-determination. Nym Wales, writing in 1942, explained that Japan would not permit thoughts of independence among the colonies, but the intertwined nature of the Korea-Japan relationship meant that a revolution would be dangerous for Japan's systems as well.

The March First Movement was the culmination of a worldwide effort on the part of Koreans to preserve their culture and national identity. During the March First Movement, Korean nationalists also worked in the United States. Korean nationalists in America hoped to strike a chord with a superpower born from a colony. However, the Korean lobby was unique in its reliance on moral, humanitarian, and Christian values, rather than politics. The delegates at the April 1919 Korean Congress in Philadelphia drafted "An Appeal to America," which requested support and sympathy due to the American love for justice, the country's own fight for liberty and democracy, and its status as a bulwark of "Christianity and humanity." This lobby abroad had as much effect on the fight toward independence as the events happening on Korean

soil. Korean immigrants in the United States and Hawaii played a pivotal role in the Korean independence movement, though Korean immigration to America was rarely political. Koreans in America, most of whom had lived in America for years and reached some degree of assimilation, felt a great surge in Korean nationalism. The period from 1919 to 1921 saw a rise in Korean activism in the United States and Hawaii, but activism that Koreans consciously blended with a display of loyalty towards America. As America aligned against Japan, Koreans saw the possibility of a very favorable partnership. The March First Movement put Korea on the world stage, and it was there that Korea began to build its unofficial alliances. Anti-colonialism was an ideology held by nearly every Korean, though each was affected and disturbed by different elements of the occupation. The desire to end colonial culture roused the Korean people to the fullest, albeit scant, extent feasible under Japan's thumb.

Religious groups contributed to Korean independence

Religious groups played an indispensable role in the liberation of Korea, too. That the 33 representatives who proclaimed the Declaration of Korean Independence on March 1, 1919 were all religious figures (fifteen Ch'ŏndoists, sixteen Christians, and two Buddhists) reflected the exceptional reality that these religious organizations voluntarily assumed the role of representing the people when the Koreans could not be represented by their own government under the colonial rule, although it is also true that this grouping was inevitable after other prominent figures refused to step forward.

In the meanwhile, missionaries faced a dilemma. If Japanese colonial rule in Korea was legitimate, then the uprising was a political matter, and therefore, none of their business. However, they saw the behavior of the Japanese in suppressing the movement as an affront to human rights. When the protests began, Western missionaries were careful not to become involved and discouraged their Korean members from participating in the movement. However, the torture and wanton massacre of Koreans by the Japanese military police drove numerous missionaries to publicly denounce the deadly suppression process. The missionaries helped bring international attention to the atrocities committed during the suppression of the March First Movement, even raising the issue before the U.S. Senate. Although efforts to appeal to the Christian churches in America to act on behalf of Korean independence failed to gain traction, the missionaries succeeded in bringing negative attention to the Japanese on the international stage. Korean Protestants—by virtue of their large participation and dedication during the March First Movement—demonstrated beyond all doubts their commitment to the Korean nation. The missionaries to Korea—after vacillating between their empathy for the Koreans and willingness to display courtesy to the Japanese colonizers—decidedly demonstrated their sympathies for the Koreans. Religion had often brought on wars in the past, but in this case, peace was its goal. Korean American Christians gave the independence movement the moral high ground that it depended on.

The United States Versus Japan: The Pacific War

In seeking to maintain a high degree of patriotic fervor and international cooperation in the fight against the Axis, President Franklin D. Roosevelt and Prime Minister Winston S. Churchill relied, among other things, on abstract war slogans and the goal of unconditional surrender. Their policy of no negotiated termination of the war was never aimed at just smashing the fascist states. Its true objective was the military occupation and postwar reform—always the two together—of those states, so that the philosophies of fascism and militarism could be uprooted, and their societies democratized. The United States demanded Japan's 'unconditional surrender,' appearing prepared to immediately strike the mainland by air force.

When a broadcast went out on August 7, 1945, announcing that the United States had dropped an atomic bomb on Hiroshima, the States claimed it had "added a new and revolutionary crease in destruction"; it vowed to bomb Japan into extinction if it did not surrender. The bombing of Hiroshima and Nagasaki represented one of the final blows to Japanese colonialism, constituting a big step toward the liberation of Korea. Because Japan's top scientists deemed the production of atomic weapons impossible, the shock of Hiroshima reverberated throughout the surrender decision process. However, although the atomic bomb may be the direct cause of surrender, Japan had its own reasons for surrender, both domestic and international in nature.

Rural and urban populations under Japanese rule were aggrieved by the regularity of war and wartime quotas on food rations, which led to starvation and anti-war (and anti-establishment) sentiments. Moreover, starting in October 1943, the U.S. Navy surged past that of the Japanese. By March 1944, Japanese forces were almost all being either "pressed back or bypassed and isolated." Additionally, Germany, one of Japan's few allies, surrendered and it shook Japanese leaders. Even the most hardcore of the war hawks realized that the war was irretrievably lost. The surrender of Germany forced Japan to rethink its own investment in further conflict. The situation surrounding World War II became even more desperate for Japan in May 1945. On May 1, Hitler took his own life, and on May 7, Germany surrendered. Colonel Takayama Shinobu, a member of the Operation Section of the Army Headquarters, recalls the reaction of his colleagues in his office: "The surrender of Germany and the urgency of the war situation shocked all members of the Operation Section thoroughly. Although no one said anything, what all of us felt equally was that any hope of winning the war was completely extinguished."

Especially after the United States took over Iwo Jima and Okinawa, Japan's military position was so strategically poor that its leaders would likely have surrendered even if the United States had not employed the atomic bomb. All parts of Japan's main islands were exposed to virtually unlimited bombings by the United States. By July 1945, it became clear that Japan, cut off from all supplies and outside resources, had lost the war with no hope of regaining momentum. Today, 75 percent of the Japanese still believe that the country was headed for the surrender of its own accord and neither bombing was necessary. The United States exposed Japan to this threatening possibility, hoping to achieve its true objective of military occupation and post-war reform under the pretext of demilitarization. The Korean call for independence in the Pacific War theatre was a symbol of the Allied forces' principles. Under the uniting force of

democracy, America and South Korea stood side by side, if only for the mutual benefit of stopping Japan in its tracks.

The Russia-Korea Alliance

A Russian-Korean alliance, developed because Russia, which wanted to gain a privilege in the East Asia-Pacific region, stood on the opposite side of Japan. Though reluctant at an early stage of the war, Stalin was eager to attack Japan at the time of his choice, with or without American encouragement. Stalin played his hand carefully to exact maximum advantages for Russia before committing himself to fight Japan. Based on the Yalta Conference, the Soviet Union had to enter the war against Japan to acquire the vast interests it was promised. If Russia failed to enter the war before the Japanese capitulation, the terms of the agreement of the Yalta Conference would not be applied, and the Soviet Union would not be entitled to the Manchurian interests. With the Yalta concessions secure in its hand, the Soviet Union had no reason to provoke the wrath of the United States by negotiating with Japan behind its back. Besides, the concessions the United States had given were far more generous than what Japan was proposing. More importantly, the prospect of the Soviets securing privileges in Asia would be far more certain by taking sides with the United States, rather than with Japan, which was on the verge of imminent collapse. With the Hirota-Malik talks, Japan wasted the entire month of June; all it did was to expose Japan's desperation and vulnerability to Stalin, who in turn accelerated his preparation for war against Japan before its surrender, as the Soviet Union could secure the Yalta concessions only as a belligerent power, not as a neutral country. Japan needed an ally after Germany exited the war, but it was unable to make a friend out of an enemy. Russian interests left Japan alone at sea and ultimately secured their surrender, though it was late.

The Surrender of Japan

Avoiding more deaths was important to both the Japanese and the Americans. Almost all of the nearly 20,000 Japanese forces on Iwo Jima had died in March defending the island. American deaths numbered almost 6,800 killed and over 19,000 wounded. The American invasion of Okinawa resulted in an even bloodier encounter. From April to June, the U.S. Army, Marine Corps, and Navy units slowly defeated Japanese forces on Japan's southernmost island. More than 12,000 American military personnel were killed and over 36,000 wounded. Japanese military and civilian casualties soared to over 110,000 known deaths. War weariness among the American public became a major concern. Reduced casualties and a return to normalcy were demands made by thousands of anxious civilians and service personnel. The pretense of war required popular support if both countries were to avoid domestic turbulence, and Japan had courted discontent among Koreans for the entire colonial period. This is not to say that America believed in the true liberation of the Korean people. While the United States welcomed a force strong enough to assist in a U.S. occupation in governing the peninsula, the strength of this force would have to remain under American. This was done because they believed that they had to control the peninsula's government from the possibilities of forming alliances with indigenous

political groups that could potentially challenge the post-liberation plans. Therefore, America had an interest in lifting up Korea, but only so high. Still, the American desire to end the war went hand-in-hand with the Korean desire for independence.

The walls continued to close around Japan, and its defeat became inevitable. Even Japan's leaders knew this. But because its governmental machinery was, to a large extent, controlled by the military and hampered by a cumbersome system that required unanimity of views for any decision, Japanese leaders failed to translate defeat into surrender quickly. In the meantime, the Japanese mainland was struck by atomic bombs, closely followed by the Soviet Union's entry into the war, which compelled Japan to surrender. Hiroshima and Nagasaki made clear what some of Hirohito's cabinet refused to see. Still, even before the second bomb unleashed more death, influencers in the Japanese government, propelled by Emperor Hirohito and Foreign Minister Shigenori Togo, decided to promptly surrender on American terms, with one condition: retention of the emperor. Emperor Hirohito, specifically, considered war with the Allies to be unwinnable at that point, and persuaded political insiders to fully appreciate the realities of Japanese diplomacy. The Japanese were stuck between war with the United States and political submission, which paired severe sanctions with the abandonment of Chinese colonies. In either instance, the Korean occupation terminated, spelling the end to the Japanese control of the peninsula. The war was over and with it, Japan's colonization of Korea was also over.

Conclusion

During Japan's 35-year rule from 1910 to 1945, numerous independence movements took place in Korea against Japan. Each chipped away at the colonial structure with its own methods and brought Koreans closer to the freedom they enjoy today. Korean independence was hard-fought and hard-won, and the repercussions of the battle echoed across the world, inspiring and uplifting colonies in the East and West as major powers stood by Korea's side even if their support was hypocritical and conditioned by the needs of war. When examined, the series of Korean independence movements that took place between 1910 and 1945, and the consequence of the Pacific War, paint a more nuanced picture of the liberation of Korea from Japan.

The desperation for Korean independence fermented during the Japanese occupation. Japanese economic and cultural oppression turned Korea into a home for the Japanese while effectively disinheriting the whole of the Korean population. Their land was taken away and given to colonists, and their mother tongue was removed from the education system, transforming the cultural landscape of East Asia. Japanized names dragged Koreans further from their history, and the inability of Koreans to own businesses for many cases in the colonial period, paired with an influx of Japanese capital, set back their attempts at creating a future.

Korea's colonial struggle tugged at the heartstrings of the world, with no country more affected by their plea than America, which had, in many ways, become the only colony to make it to world power status. Many Americans supported Korea, and Korean Americans were moved not only by the plight of their home country but by the swell of support in the country to which

they had migrated. Religious leaders and missionaries preached America's sermons back to it, trapping it in its own principles and identifying the Korean fight as a replica of the American fight for independence from Britain. The parallel was motivating, and even the war-weary American public considered it a moral duty to uplift Koreans while serving the dual purpose of bringing Japan to its knees.

During World War II, despite the fall of Hitler and Germany, Japan was so desperate to hold on to power that it gave up the lives of thousands of Japanese and Korean soldiers on the front lines. However, with two atomic bombs dropped, Japanese colonialism of Korea came to a swift end.

Koreans, though certain of the injustice of their plight and willing to fight to death for its rectification, did not win their independence on their own. Korean independence became the joint enterprise of Koreans, Korean Americans, and all of those with a stake in the crumbling of the colonial mindset. The Pacific War was the board beneath belligerent pawns, and when the war games ended, the Korean condition was checkmate.

Works Cited

Asada, Sadao. "The Shock of the Atomic Bomb and Japan's Decision to Surrender: A Reconsideration." *Pacific Historical Review*, vol. 67, no. 4, 1998.

Baik, Youngseo. "The 1919 Independence Movement in Korea and Interconnected East Asia: The Incremental Unfolding of a Revolution." *The Asia-Pacific Journal*, no. 21, Nov. 1, 2021.

"Battle of Iwo Jima, 1945," *Naval History and Heritage Command*, 2020, https://www.history.navy.mil/research/library/online-reading-room/title-list-alphabetically/b/battle-for-iwo-jima.html Accessed Apr. 3, 2023.

Bernstein, Barton J. "Understanding the Atomic Bomb and the Japanese Surrender: Missed Opportunities, Little-Known near Disasters, and Modern Memory." *Diplomatic History*, vol. 19, no. 2, 1995.

Bernstein, Barton J. "The Perils and Politics of Surrender: Ending the War with Japan and Avoiding the Third Atomic Bomb." *Pacific Historical Review*, vol. 46, no. 1, 1977.

Bix, Herbert P. "Japan's Delayed Surrender: A Reinterpretation." *Diplomatic History*, vol. 19, no. 2, 1995.

Caprio, Mark E. "Marketing Assimilation: The Press and the Formation of the Japanese-Korean Colonial Relationship." *Journal of Korean Studies*, vol. 16, no. 1, 2011.

Caprio, Mark E. "The Eagle Has Landed: Groping for a Korean Role in the Pacific War." *The Journal of American-East Asian Relations*, vol. 21, no. 1, 2014.

Ch'oe, Yong-ho. "The Japanese Surrender and the Division of Korea in 1945: Why Japan Delayed Its Surrender Leading to the Division of Korea." *International Journal of Korean Unification Studies*, vol. 29, no. 2, 2020.

Chun, Clayton K. S. Rep. "US Army War College Guide to National Security Issues: Painting Yourself into A Corner: Conflict Termination, Unconditional Surrender, and the Case of Japan." *Strategic Studies Institute*, US Army War College, 2012.

Clark, Donald N. "Surely God Will Work out Their Salvation: Protestant Missionaries in the March First Movement." *Korean Studies*, vol. 13, no. 1, 1989.

Deede, Sara Elizabeth. "Activism and Identity: How Korea's Independence Movement Shaped the Korean Immigrant Experience in America, 1905-1945." 2010.

Fields, David P. "The Rabbi, the Lawyer, and the Prophet: American Exceptionalism and the Question of Korean Independence, 1919-1922." *The Journal of American-East Asian Relations*, vol. 22, no. 4, 2015.

Frank, Richard B. "Downfall: The End of the Imperial Japanese Empire." 1999.

Grunow, Tristan. "A Reexamination of the Shock of Hiroshima: The Japanese Bomb Projects and the Surrender Decision." *Journal of American-East Asian Relations*, vol. 12, no. 3-4, 2003.

Haggard, Stephan, David Kang, and Chung-In Moon. "Japanese Colonialism and Korean Development: A Critique." *World Development*, vol. 25, no. 6, 1997.

Hasegawa, Tsuyoshi. “The Soviet Factor in Ending the Pacific War: From the Hirota-Malik Negotiations to Soviet Entry into the War.” *University of California- Santa Barbara*, 2003.

Ha, Ki-Rak. “A History of the Korean Anarchist Movement.” *The Anarchist Library*, 1986.

Hatch, Walter. “Bloody History in Two Regions.” Essay in *Ghosts in the Neighborhood: Why Japan Is Haunted by Its Past and Germany Is Not.* Ann Arbor: The University of Michigan Press, 2023.

Hattori, Ryūji. “From the Second Sino-Japanese War to the Pacific War.” Essay in *Japan at War and Peace: Shidehara Kijūrō and the Making of Modern Diplomacy.* Canberra, ACT, Australia: Australian National University Press, 2021.

Houchins, Lee and Chang-su Houchins. “The Korean Experience in America, 1903-1924,” *Pacific Historical Review*, vol. 43, no. 4, 1974.

Judge, Sean M. Rep. “The Pacific War, 1941–45.” 2009.

Kane, Danielle and Jung Mee Park, “The Puzzle of Korean Christianity: Geopolitical Networks and Religious Conversion in Early Twentieth Century East Asia.” *American Journal of Sociology*, vol. 115, no. 2, 2009.

Kang, Jin. “Historical Changes in North Korean Nationalism.” *North Korean Review*, vol. 3, no. 1, 2007.

Kang, K. Connie. “The Resistance.” *Manoa*, Century of the Tiger: One Hundred Years of Korean Culture in America, 1903-2003, vol. 14, no. 2, 2002.

Kawamura, Noriko. “Emperor Hirohito and Japan’s Decision to Go to War with the United States: Reexamined.” *Diplomatic History*, vol. 31, no. 1, 2007.

Kim, Jimin. “Empire Versus Empire.” *Journal of American-East Asian Relations*, vol. 22, no. 4, 2015, pp.315-342.

Kim, Michael. “The Politics of Officially Recognizing Religions and the Expansion of Urban ‘Social Work’ in Colonial Korea.” *Journal of Korean Religions*, vol. 7, no. 2, 2016.

Kim, Richard S. “Inaugurating the American Century: The 1919 Philadelphia Korean Congress, Korean Diasporic Nationalism, and American Protestant Missionaries.” *Journal of American Ethnic History*, vol. 26, no. 1, 2006.

Kim, Yong-Jik. “Origins of Korean Nationalism: Political Movements and the Public Sphere.” *Social Criticism*, 1994.

Lee, Timothy S. “A Political Factor in the Rise of Protestantism in Korea: Protestantism and the 1919 March First Movement.” *Church History*, vol. 69, no. 1, 2000.

Masayuki, Nishi. “March 1 and May 4, 1919 in Korea, China & Japan: Toward an International History of East Asian Independence Movements.” *The Asia-Pacific Journal*, Japan Focus, vol. 5, no. 10, Oct. 1, 2007.

Pyle, Kenneth B. “Hiroshima and the Historians: History as Relative Truth.” *Asia-Pacific Review*, vol. 22, no. 2, 2015.

Record, Jeffrey. Rep. “Ends, Means, Ideology, and Pride: Why the Axis Lost and What We Can Learn from Its Defeat.” *Strategic Studies Institute*, US Army War College, 2017.

Robinson. “Nationalism and the Korean Tradition, 1896—1920: Iconoclasm, Reform, and National Identity.” *Korean Studies*, vol. 10, no. 1, 1986.
Rosenberg, Wolfgang. “Korea and Reunification.” *New Zealand International Review*, vol. 2, no. 1, 1977.
Savage, Timothy L. “The American Response to the Korean Independence Movement, 1910-1945.” *Korean Studies*, vol. 20, no. 1, 1996.
Schmidt, Jan and Naoko Shimazu, “A Historiographical Turn: Evolving Interpretations of Japan during World War I, 1914–2019.” Essay in *Writing the Great War: The Historiography of World War I from 1918 to the Present, edited by Christoph Cornelissen and Arndt Weinrich.* Berghahn Books, 2021.
Son, Jeong Mog. “A Study on Company Law.” *The Journal of Korean History*, vol. 45, 1984, pp.87-130.
Stent, Dylan. “A Century of Contention in South Korea.” *Asian Survey*, vol. 59, no. 5, 2019, pp.896.
Sullivan, Mark P. Rep. “Japan’s Decision to Surrender.” *Air UP*, 1995.
Uchida, Jun. “A Sentimental Journey: Mapping the Interior Frontier of Japanese Settlers in Colonial Korea.” *The Journal of Asian Studies*, vol. 70, no. 3, 2011.
Uchida, Jun. “The Public Sphere in Colonial Life: Residents’ Movements in Korea under Japanese Rule.” *Past and Present*, vol. 220, no. 1, 2013.
Van de Velde, James R. “Opinion: The Enola Gay Saved Lives.” *Political Science Quarterly*, vol. 110, no. 3, 1995.
Wales, Nym. “Rebel Korea.” *Pacific Affairs*, vol. 15, no. 1, 1942.
“The Yalta Conference.” *Office of the Historian, Foreign Service Institute, United States Department of State*, https://history.state.gov/milestones/1937-1945/yalta-conf Accessed Apr. 11, 2023.
Yellen, Jeremy A. “The Specter of Revolution: Reconsidering Japan's Decision to Surrender.” *The International History Review*, vol. 35, no. 1, 2013.
Youm, Kyu Ho. “Japanese Press Policy in Colonial Korea.” *Journal of Asian History*, vol. 26, no. 2, 1992.

Stratification and Power: A Study on Power and Status in Hierarchical Structures By Michael Zhu

Abstract

This paper analyzes several studies that discuss the dynamics between status and power in hierarchies to understand the components of social organizations that make them productive. It aims to reduce hierarchies to their most basic components and discuss their complex interactions. It also suggests several avenues of further research to better understand how hierarchies, and the way they are composed, can contribute to furthering the efficiency of our society. This project acknowledges that the research cited in this paper is overwhelmingly surveys of male dominated workplaces, and thus cannot capture the entire picture of workplace dynamics. With this fact in mind, we try to discuss the qualities we believe contribute to bolstering the productivity of a hierarchy.

Introduction

Complex Hierarchical organization is intrinsic to human society and culture. Hierarchies are so intrinsic to human nature that even in informal settings, such as friend groups, implicit rankings develop naturally. Professional settings, such as offices or organizations, employ explicit stratification to maintain order. To understand the fabrics and qualities of these hierarchies is to understand a very significant nature of human behavior. Thus, this research paper attempts to dissect the role of hierarchies in society by analyzing factors of hierarchical systems that contribute to increasing or decreasing workplace productivity.

In order to do so, we build on the preexisting Social Dominance Theory (SDT), first proposed by social psychologists Jim Sidanius, Erik Devereux, and Felicia Pratto in 1992, which hypothesizes a framework to which society inevitably organizes itself. SDT argues that social organizations are perpetuated by the existence of hierarchies, and thus by mechanics of inequality. In this paper, we plan to explore the nuances of SDT and try to understand how its components and their interactions can be used to explain the productivity and performance of a social organization.

We acknowledge the potential flaws and criticisms of SDT, notably the prominent rhetoric challenging the idea that all social interactions are inherently products of one's Social Dominance Orientation (SDO) (Wilson and Liu). Furthermore, the SDO scale seems to be a significantly weaker measure when analyzing women than men, making it an inaccurate predictor for social competence in gender-inclusive organizations (Wilson and Liu). These points of criticism are significant because they challenge the widely accepted notion of SDT and the prominent corporate subscription to the SDO scale. We further acknowledge that many of the studies referenced in this paper have conducted research on male dominated companies and therefore produce results that are skewed in an evolving workforce. Regardless, the findings from this research still try to further the present understanding of SDT and we suggest through this paper that the influence of power and status on a hierarchy's productivity is still significant.

Literature Review

We begin this paper with a section of definitions in order to clarify the proper terms that will be used in this paper. This section aims to better understand the definition of hierarchy, and the nuances that comprise its significance. Hence, we begin this section with the components, notably status, power, and control, and progress towards a more complete and encapsulating definition of hierarchy.

Status

Social status can be defined as perceived competence in a given social situation. This can be specified towards more quantifiable factors, such as educational experience and seniority, or less tangible factors such as leadership and charisma (Munuera et al.). Furthermore, this perceived competence, in many social environments, is influenced by one's identity markers, including but not limited to race, gender, or sexual orientation. Status is significant, playing a heavy role in the intrinsic understanding of one's hierarchical position. Therefore, seeing as elements of status are not only how peers judge each other, but also how employers judge prospective hires (Munuera et al.), status is especially influential in determining rankings in both formal and informal hierarchies. This concept of status is also instrumental in the definition of *prestige*, which is the control willingly given to an individual on account of one's status (Cheng and Tracey). Among coworkers, the more experienced and more charismatic will naturally gain more status, and thus prestige, over their fellow coworkers. For example,

Power

Social Power is an individual's control over critical resources. Examples of this can be often seen in workplaces, where managers have power over their branch by critically controlling pay, employment, and promotion opportunities (Blader and Chen). Social Power can also be seen in learning settings, notably classrooms, where teachers and professors demonstrate power over students by managing grades and research opportunities. Power is significant because it suggests the behavior of those in the hierarchy. Notably, individuals in the lower ranks are prompted to be subservient to those of higher rank simply because they want to be considered favorably in respect to these critical resources. Therefore, whereas status is an implicit difference in ranking, power is an explicit one.and influence, and is thought to be homologous with dominance hierarchical systems in nonhuman primates that result from agonistic contests" (Cheng and Tracey). This induction of fear is often perpetuated through the enforcement of power. In authoritarian societies, the government controls critical resources such as currency, freedom, and even life, in order to exact dominance over its people.

Control

Control is a product of control, status, and a myriad of other factors, and acts as the tangible representation of dominance and prestige. Control is demonstrated in a hierarchy when one is able to influence over someone else. This definition could be confused with power, which

also uses the concept of control in its meaning. Yet, power, or any other factor that plays into one's hierarchical ranking, does not necessitate control over people. A teacher may have control over a student's grade, but if the student was rebellious or uncaring, the control over that critical resource would mean nothing. Therefore, even though factors such as status or power are based on qualities of an individual, dominance is based on qualities of both the leader and the follower.

Favorable Workplace Outcomes

The term favorable workplace outcomes is used by Liliam M. de Menezes to describe traditional workplaces such as offices. In this paper, the definition of *workplace* will be used liberally, and to define organizations as a whole. Even though social environments such as school orchestras or sports teams are not traditionally considered "workplaces", they serve to work towards some defined goal: programming a concert or winning a championship. Less traditional workplaces are important to analyze because they, like traditional workplaces such as offices, are social organizations working towards common goals which employ hierarchies. Though the differences in structure and professionalism must be acknowledged, less traditional workplaces should be treated similarly to traditional workplaces.

This paper will define the *workplace* as a collection of individuals working towards a common goal or common goals. As a result, the definition of *favorable workplace outcomes* will vary significantly based on the common goals of the workplace. Yet, among most of these settings, a common theme seems to be associated with these favorable outcomes: performance. An employee's performance defines their merit. Hence, as seen by managers, an employee's performance and productivity is a favorable workplace outcome (M. de Menezes).

This definition relies on the assumption that a favorable workplace outcome is favorable from the perspective of managing roles exclusively. We recognize that what an employee finds favorable, such as job satisfaction or higher pay, may not be productive for the entire workplace.

Social Hierarchy

Social Hierarchies are orders of individuals separated by ranks in a social organization according to an individual's social value. Hierarchical ranking in this manner can be further dissected into *implicit* and *explicit* ranks (Magee and Galinsky, 2009). The terms *implicit* and *explicit* delineate whether a hierarchy has been formally defined. The term *social value* is used because an individual's hierarchical rank is dependent upon their potential value to the social organization. This concept of *social value* is composed of many factors, both quantifiable and less tangible, that interact to define an individual's relative ranking in a hierarchy. In different organizations with different cultures and values, the weight of *component factors* will also differ depending on the needs of the organization.

Formal Hierarchy

Larger social organizations often employ formal hierarchies and explicit ranking in order to maintain favorable workplace outcomes . Formal hierarchies are seen in professional settings

such as large corporate environments where a chain of command is established via explicitly ranking individuals. In formal hierarchies, although elements pertaining to favorable workplace outcomes are considered when the explicit ranking is defined, other status elements, such as race, gender, or sexual orientation, which do not directly indicate productivity, can lead to one's explicit ranking in a formal hierarchy. Furthermore, in Formal Hierarchies, elements of power (power is explicitly defined later), such as employment, pay, and promotions, have a greater influence on individuals in that hierarchy as power is often what maintains the structure of a formal hierarchy.

Informal Hierarchy

Smaller social groups with less complex organization typically form Informal Hierarchies, where social ranks are implicitly defined by individuals. In order to make this implicit stratification, individuals must be able to judge the social value of a group member through observation (Todorov et al.). As a result, there is comparatively more group agreement on individual's ranking compared to that of formal hierarchies (Magee and Galinsky) seeing as this implicit ranking is almost entirely based on qualities that are obvious to observation. For example, among peers in a group of friends, members of that group can agree on the implicit ranking of an individual that emerges as most outspoken and charismatic. Informal hierarchies are often seen in unprofessional social groups, such as friend groups, classrooms, or even as subsections of formal hierarchies: co-workers of equal explicit ranking collaborating on one project. In Informal Hierarchies, the component factors that play into one's social value will lean further towards status than power. An example of this can be observed in sports teams where other than the captains, team mates will naturally form an informal hierarchy based on status factors such as seniority and productivity.

Discussion

This section will try to analyze the complex interactions between previously mentioned component factors. These factors are important to further understand beyond their definitions because of their critical influence on members of a hierarchy.

Status vs. Power

Earlier, we defined status and power to fundamentally be in contrast. Power, being an explicit difference in ranking, differs significantly from status, being an implicit difference in ranking. As discussed previously, different organizations will see different component factors take precedence, and in the case of status vs. power, powerful leaders will dominate in environments with heavy emphasis on economic norms and practical benefits, whereas high status leaders will dominate in contexts where relational and reputational considerations take precedent (Steven L. Blader and Ya-Ru Chen Find Year). Yet, participants in hierarchies can not simply be categorized strictly by power and status. Teachers and professors, for example, are both high in status, considering their academic achievements compared to those they teach, and

high in power, seeing as they control the grade that their students receive. Typically, leaders and followers demonstrate varying levels of both status and power, which determines their position in a hierarchy as well as their effect. One of these effects is their collective effect on the empathy of leaders and followers.

In a study performed by Blader et al. (2013), Professor Steven L. Blader analyzed the effects of status and power on a participant's perspective taking. By placing an object on either the right or the left of a participant, taking note of whether the participant described the relative position of the object according to their perspective or that of the researcher, they were able to measure the willingness of the participant to take perspective. In their findings, Blader et al. determined that in high-power low-status environments, relative to high-status low-power environments, participants were far more likely to be egocentric by describing the object's position according to their own perspective. In contrast, participants in high-status low-power environments were far more likely to take perspective by describing the object's position in terms of the researcher's perspective. This makes sense because high status leaders are empowered by others who create their own perception of the leader's relative ranking - a process of perspective taking. Hence, high power leaders, who are empowered from their own control, are far less relatable than status leaders. This relationship between power and status is not only exemplified by a participant's perspective taking, but also a leader's perspective taking.

In a study conducted by Nathaniel J. Fast et al. in 2012, high-power male leaders were observed to be more likely to select demeaning tasks for their followers compared to high-status leaders, who are far more lenient. Interestingly, leaders with both high-power and high-status qualities did not demonstrate this same demeaning behavior that high-power low-status leaders did. Leaders with overwhelming power compared to status may feel the need to prove their status using power. Status leaders are implicitly accepted as high ranking. Hence, they do not need to use demeaning methods to simulate their status and prove their ranking, suggesting that either higher status bolsters perspective taking for high ranking individuals, or that an imbalance of power and status, with a high emphasis on power, tends to create a negative environment for lower ranking members of a hierarchy.

The current literature on Ideological Asymmetry suggests that as status increases, so does the legitimacy of a hierarchy. This section complicates that idea by saying how status has to match power in potency for a sustainable environment. In environments where powerful leaders do not support their ranking with equally high status, perspective taking is far less likely from both the leader and the follower, suggesting that an imbalance of power and status will lead to a more negative environment than hierarchies with overwhelming power and demonstrate the need for status in a hierarchy.

What determines Status in a social hierarchy?

As it was previously argued, status is shown to have significant importance when cultivating a positive environment in a hierarchy. Because status is an implicit understanding of one's social ranking, there are many intangible factors that determine one's status. Furthermore,

different factors have different weight in determining status in various social groups. For example, academic achievement would be more valued in a physics discussion panel than it would be in a friend group, where charisma may take precedence. Yet, there are two attributes that are generally most valued: competence and collective mindedness, products of the legitimizing myths theory (Imhoff and Bruder).

Individuals in a hierarchy must demonstrate their value to a workplace's goals, and competence is considered a direct predictor of productivity (Anderson and Cowan). Thus, those with abilities are valued in terms of the perceived competence of social status and are more likely to attain higher social rank (Blader and Chen). This is further evidenced by the fact that hiring practices and school admissions, though often plagued by nepotism (Lewin), are primarily determined by one's perceived competence. In settings where productivity is highly valued, so is perceived competence.

In tandem with competence, an individual must also be demonstrably willing to use this same competence in respect to the goals of the social organization(Anderson and Cowan). On a soccer team, a striker might have shown incredible explosiveness, ball control, and shot power during tryouts; but, if they do not apply these skills willingly for the team, whether that be exhibited through laziness in training or inconsistency in matches, they will not be considered collectively minded. Organizational motivation is significant, because even with provable ability, one is not considered important to a group unless that ability is demonstrated to be usable by the group. One only has to demonstrate commitment to gain status in a hierarchy. Hence, an individual may be able to feign perceived competence and commitment using different means. One of these ways is an individual's usage of the personality trait of dominance.

Trait dominance is different from the previously defined *dominance* as trait dominance speaks to one's personality while dominance correlates with one's use of power. Cowan and Anderson noticed a correlation between an individual's trait dominance and their perceived competence. In their study, they observed that individuals with higher trait dominance were more engaged in groups, evidenced by their greater inclination to speak, participate, and willingness to exhibit their abilities. Their dominant nature can create the false pretense of commitment to a group's success, as well as an inflated impression of competence. The perception of their competence was observed not only in situations where the trait dominant individual was equally as competitive as their peers, but even in situations where they were less so (Anderson and Cowan). Despite the risk of feigned commitment, trait dominance is also particularly important for a group.

In their study, Cowan and Anderson also observed that submissive individuals often exhibit hidden competence that are unseen by others. They further determined that trait dominance, or by extension the presence of some form of status, increased the chances that these talents were noticed. This further speaks to the importance of balancing status and trait dominance in a social organization as not only do they foster better work environments, but also directly contribute towards realizing favorable workplace outcomes by enhancing the contributions and productivity of individuals.

On Productivity and Performance

The ultimate goal of many workplaces is to produce, whether that be a sports championship or financial capital. Hence, the ultimate goal of this paper, which analyzes hierarchies in workplaces, is to determine what factors lead to productivity and performance. In terms of these ends, two factors play heavy roles in determining success: motivation and empowerment.

Motivation

Motivation is one primary factor that can boost a social organization's productivity. A motivated employee is one that is willing to use their competence for group minded goals and social organizations must be designed to best foster this motivation for their benefit. In the context of hierarchies, one way to induce motivation is through social mobility. The promise of advancement in a hierarchy is not only a promise of status and power, but also of social capital, money, and other tangible benefits. Therefore, hierarchies that have greater social mobility are more motivating for those reasons listed above. Yet, this begs the question: how does a social organization induce the appearance of social mobility in a hierarchy? Researchers Blader and Chen answer this question by noticing the ability of status based hierarchies to demonstrate this quality.

As a part of their study, Blader and Chen compared the effects of status and power hierarchies on employees in workplaces. In their research, they found that status hierarchies produce more competitive employees than power hierarchies. Status hierarchies were perceived to be more socially mobile, which drove individuals to be better motivated to prove their worth compared to power hierarchies. This observation points to a different facet of the importance of status in a hierarchy. Positive and motivating workplaces not only require leaders that demonstrate high, but also hierarchies that are structured around status.

Despite the positive aspects of social mobility, if not controlled, competition within an organization can have severely negative effects. Blader and Chen support this assertion with their study, noting that individuals may engage in competitive behaviors that are dysfunctional for the group's goals. Hierarchies that are too socially mobile, such as groups with elected positions, may encourage counter productive practices such as sabotage and nepotism. For example, a newspaper club at a highly competitive prep-school might see rampant toxicity and self destructive corruption as students compete for elected board positions that open every year. In this situation, the promise of board membership is coveted for its potential prestige in college applications, and the group appears socially mobile as new board members are elected every year with the changing senior class. When individuals are given promise and opportunity, they may engage in counterproductive behaviors.

Empowerment

Though many would argue that motivation and empowerment are synonyms, the primary difference between motivation and empowerment is that whereas motivation is intrinsic with

direct effects, empowerment is extrinsic with indirect effects. For example, whereas a motivated employee will directly feel the need to produce, an empowered individual in this context may feel more comfortable in the workplace itself, which may indirectly boost their productivity and performance. That being said, empowerment is thus a vital factor for social organizations to master, alongside motivation, in order to maximize their favorable workplace outcomes. One way to empower individuals in social organizations is by balancing status and power in a workplace.

Bunderson and Reagans studied the effects of status and power in learning environments and observed that when there was an imbalance in the two, learning efficiency was compromised. According to their study, in cases where the presence of status did not match that of power, the knowledge and perspective of those in power are more pronounced. If the findings of Bunderson and Reagans were extended beyond the restrictions of learning environments, they would prove that asymmetric status and power could prevent individuals in subservient roles from realizing their skills. These findings correlate with those from the study of Cowan and Anderson, who observed a similar result for informal hierarchies which lack trait dominance. This is significant because subservient individuals often possess hidden competencies that could benefit productivity and performance. These two studies, analyzed in tandem, prove that a lack of the presence of status in a social organization leads to counterproductivity both among peers and between dominant and subservient roles. On the other hand, balancing status can have the secondary effect of improving the workplace environment.

As mentioned before, status plays a vital role in helping individuals, both in managing and subservient roles, be more willing to take perspective. This means that balancing status in a workplace is able to better the job satisfaction of individuals. De Menezes notes in a study that satisfaction has a high correlation with many favorable workplace outcomes in several fields. Because the presence of status betters the environment of a workplace, by enforcing the status of those in positions of power, a social organization is able to better empower individuals to produce and perform.

Conclusion

In this paper, we complicate the preconceived understanding of Social Dominance Theory as we analyze hierarchies under the lens of status and power as major component factors. This research finds that by balancing status and power in hierarchies, an organization is able to better facilitate productivity and performance for individuals. The behavioral differences suggested by SDT must be balanced between status and power for better performance in a hierarchy. Successful organizations in the SDT model have leading figures who exhibit both status and power.

Yet, despite corroborating research, we must once again recognize that most of the sources recognized in this paper have conducted research and surveys limited to a non-representative population, and thus can only paint a fragmented picture of reality. With this being said, we suggest two avenues of further research. The first is a greater focus on dynamics

in the workplaces with predominantly non-male managers across various backgrounds. This is of great importance, as the majority of current research is done on and applicable to male dominated workplaces. The second proposal for further research is targeted towards application. Though, in this paper, we tried to better understand what makes hierarchies more productive, more research must be conducted to determine how these factors could be implemented in hiring practices, organization of human resources, promotions, and other components.

Works Cited

Anderson, Cameron, and Jon Cowan. "Personality and Status Attainment: A Micropolitics Perspective." *The Psychology of Social Status*, 2014, p. 99, https://dpl6hyzg28thp.cloudfront.net/media/978-1-4939-0867-7.pdf.

Blader, Steven L., and Ya-Ru Chen. "What's in a Name? Status, Power, and Other Forms of Social Hierarchy." *The Psychology of Social Status*, 2014, p. 71, https://dpl6hyzg28thp.cloudfront.net/media/978-1-4939-0867-7.pdf.

Cheng, Joey T., and Jessica L. Tracey. "Toward a Unified Science of Hierarchy: Dominance and Prestige Are Two Fundamental Pathways to Human Social Rank." *The Psychology of Social Status*, 2014, p. 3, https://dpl6hyzg28thp.cloudfront.net/media/978-1-4939-0867-7.pdf.

De Menezes, Lilian. "Job Satisfaction and Quality Management: An Empirical Analysis." *City, University of London Institutional Repository*, vol. 32, no. 3, 2011, pp. 308–28, https://doi.org/10.1108/01443571211212592.

Imhoff, R., and Martin Bruder. "Speaking (Un–)Truth to Power: Conspiracy Mentality as a Generalised Political Attitude." *European Journal of Personality*, vol. 28, Jan. 2014, pp. 25–43, https://doi.org/10.1002/per.1930.

Lewin, Tamar. "Study Finds Family Connections Give Big Advantage in College Admissions." *New York Times*, 8 Jan. 2011, https://www.nytimes.com/2011/01/09/education/09legacies.html.

Magee, Joe C., and Adam D. Galinsky. "Social Hierarchy: The Self‐reinforcing Nature of Power and Status." *The Academy of Management Annals*, vol. 2, no. 1, Apr. 2009, https://doi.org/10.5465/19416520802211628.

Munuera, Jérôme, et al. "Shared Neural Coding for Social Hierarchy and Reward Value in Primate Amygdala." *Nature Neuroscience*, vol. 21, Mar. 2018, pp. 415–23, https://doi.org/10.1038/s41593-018-0082-8.

Todorov, Alexander, et al. "Inferences of Competence from Faces Predict Election Outcomes." *Science*, vol. 308, no. 5728, June 2005, pp. 1623–26, https://doi.org/10.1126/science.1110589.

Wilson, Marc Stewart, and James H. Liu. "Social Dominance Orientation and Gender: The Moderating Role of Gender Identity." *British Journal of Social Psychology*, vol. 42, no. 2, Dec. 2010, pp. 187–98, https://doi.org/10.1348/014466603322127175.

Putting the Brakes on T Cell Exhaustion: Platforms for Chimeric Antigen Receptor (CAR) T Cell Regulation Lead the Way By Chungbi Yoon

Abstract

Chimeric antigen receptor (CAR) T cell therapy has shown excellent antitumor effects in treating hematological malignancies. However, applying this treatment to solid tumors has several challenges, including CAR T cell exhaustion. Maintaining CAR T cell effector activity and persistence while achieving therapeutic potency in solid tumors remains a crucial problem. Recent research suggests that regulatable CAR platforms can introduce transient rest to prevent or reverse exhaustion in solid tumors. Here, I examine the underlying mechanism of exhaustion and discuss new engineering advances that introduce transient rest, such as the use of drug-sensitive promoters in CARs, degradable CARs, synNotch CARs, and microenvironment-regulated CARs. These advances can ultimately mitigate exhaustion and improve clinical results for solid tumors.

Key terms*: chimeric antigen receptor, T cell exhaustion, solid tumor, transient rest*

Introduction

Cancer is a prevalent and significant global health issue. It is responsible for about 1 in every 6 deaths worldwide, surpassing the combined impact of AIDS, tuberculosis, and malaria. It is currently the second leading cause of death globally, following cardiovascular diseases (*American Cancer Society, 2023*). Uncontrolled and aberrant cell growth characterizes this type of disease, which develops due to mutations in a cell's genetic material (also known as its DNA). These modifications lead to uncontrolled and aberrant cell division, differentiation, and aberrant cell death. Unchecked cell growth caused by genetic mutations can result in the establishment of a tumor (*NCI, 2023*).

While normal cells express proteins on their surface (called antigens) which may be recognized by immune cells, mutations can lead to abnormal antigens, whose detection triggers the immune system to attack the mutated cells. These abnormal antigens are presented by major histocompatibility complex (MHC), which is located on mutated cells' cell membranes. These abnormal antigens can be classified into either tumor-specific or tumor-associated antigens (TAA). While tumor-specific antigens are only found on tumor cells, tumor-associated antigens are also found in normal cells but at lower levels.

T cells, a type of white blood cell, use their T cell receptor (TCR) to recognize MHC-presented antigens, including TAA **(Fig1A)**. T cells are cytotoxic cells that employ multiple mechanisms to eliminate tumor cells. One mechanism is the secretion of cytotoxic cytokines, such as perforin and granzymes, which induce cell death in tumor cells. These cytokines cause pores to form in the tumor cell membrane, allowing the entry of granzymes, leading to the activation of caspases and, ultimately, cell death.

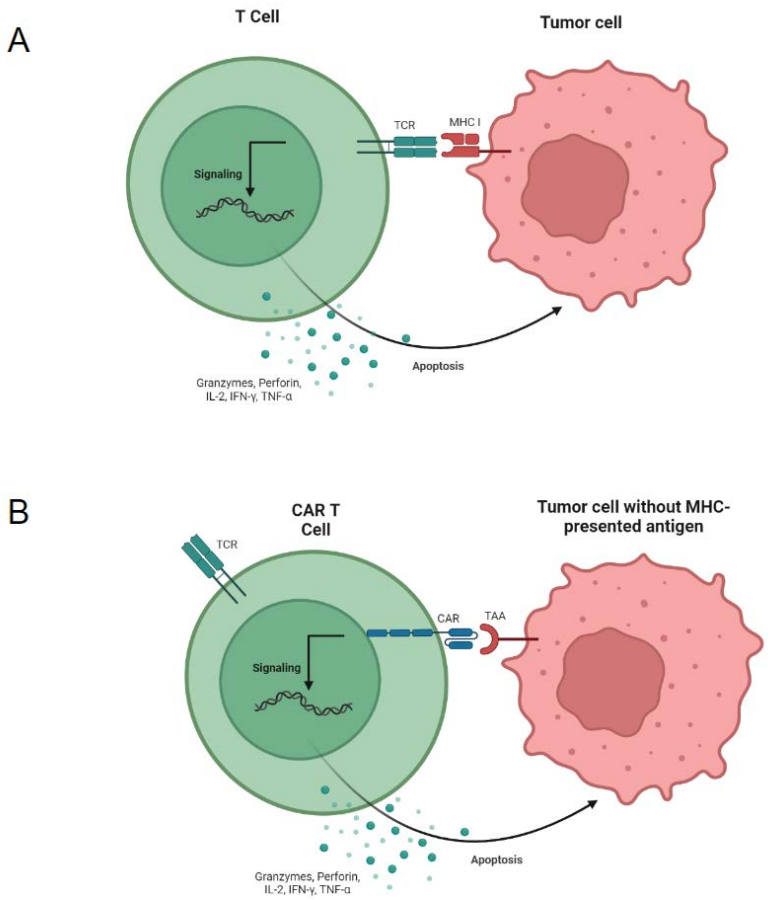


Fig 1: **(A)** T cell-mediated killing of tumor cells **(B)** CAR T cell-mediated killing of tumor cells (Created with Biorender.com)

However, some cancerous cells evade adaptive immunity through immune escape. This is made possible by either the immunosuppressive environment – an environment that suppresses immune responses and immune cell activity – or through further genetic mutations. These genetic mutations lead cancer cells to stop presenting conventional tumor-specific antigens on their MHC, thereby avoiding the detection by TCR of T-cells. Consequently, the proliferation of cancerous cells is unchecked even in the presence of the immune system. To address this problem, Eshhar *et al.* designed chimeric antigen receptor (CAR) T-cell therapy (1993). CAR T cells can recognize cancerous cells – even if they do not present conventional antigens on their MHC – by directly recognizing antigens on the cell membrane of the cancerous cell **(Fig1B)**. CAR T cells can be engineered to detect different types of tumor-specific antigens, tumor-associated antigens, or a combination of both. This recognition then allows them to release cytotoxic mediators that cause cancerous cells to undergo apoptosis.

This is made possible by the components of the CAR construct: an extracellular target antigen-binding domain, which binds to antigens of tumor cells; a hinge region, which properly

aligns the extracellular domain; a transmembrane domain, which anchors the CAR to the T cell; and one to four intracellular signaling domains, which promote CAR T cell signaling for T cell activation (Ahmad et al., 2022). The extracellular target antigen-binding domain consists of a single-chain fragment variable (scFv) capable of detecting the tumor-specific or tumor-associated antigens on the cell membrane of tumor cells. The CARs can be designed to target any antigen because the single-chain fragment variable is obtained from monoclonal antibodies designed and produced in the lab. Overall, CAR is a synthetic protein that is a combination of at least four distinct protein domains mentioned above.

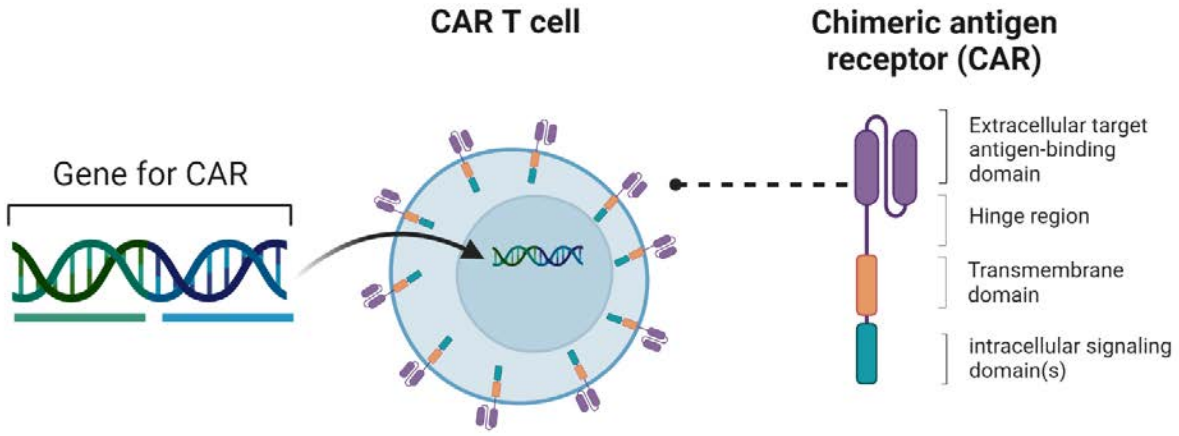


Fig 2: Chimeric Antigen Receptor (CAR) T cell (Created with Biorender.com)

The evolution of CAR-T cell therapy has progressed through five generations, culminating in the ongoing development of the fifth generation. Each generation differs in terms of the type and number of intracellular signaling domains. The first generation of CAR-T cells, which only had one intracellular signaling domain, demonstrated limited efficacy and necessitated supplementary administration of IL-2, a cytokine that promotes the proliferation and survival of T cells. The second generation introduced costimulatory domains to improve proliferation and sustain responses. The third-generation CAR-T cells integrated multiple costimulatory signaling domains to improve signaling strength. However, the expected enhanced efficacy compared to second-generation CAR T cells was not fully realized. The fourth generation, known as TRUCKs, incorporated an inducible cytokine expression cassette, achieving improved effectiveness while minimizing systemic toxicity. The fifth generation, which will be further discussed in this paper, incorporates regulatable CAR platforms where T-cell activation is regulated by the presence of antigens or small-molecule drugs (Tomasik et al., 2022).

Of all of the different CAR varieties, only the second-generation CAR T cells have been FDA-approved. These approved CAR T cells have shown promising results in the treatment of hematologic tumors. acute lymphoblastic leukemia (ALL), with response rates ranging from 70% to 90% in clinical trials. However, this therapy exhibits unpredictable and

incomplete responses against liquid tumors in the long term and little efficacy against solid tumors (Lee et al., 2015; Majzner and Mackall, 2019). These challenges involve T cell exhaustion, antigen heterogeneity, limited infiltration to tumor microenvironments, and immunosuppressive tumor microenvironments. Among the aforementioned challenges, T cell exhaustion is a major cause of impeded potency and persistence in CAR T cell treatment for solid tumors. To help contribute to overcoming this challenge, this paper will discuss the underlying mechanisms of CAR T cell exhaustion and recent engineering advances that have the potential to exponentially improve solid tumor treatment.

T cell exhaustion

T-cell exhaustion reduces persistence, proliferation, and cytotoxic function in CAR T cells. Furthermore, exhausted CAR T cells express more inhibitory receptors, increasing the likelihood of undesired immune checkpoints. Immune checkpoints are mechanisms whereby the immune response of immune cells can be stopped to prevent overactivation of the immune system and prevent damage to healthy cells. However, if T cells are exhausted, this mechanism may take place even when an immune response is needed. Once tumor-specific antigens on MHC are recognized, naive T cells – which are characterized by high levels of CD62L expression – in the body are activated and become effector T cells, which kill the tumor cells (Yang et al., 2011). These antigen-experienced effector T cells are characterized by CD44, a marker of T cell activation. After the tumor and, thus, antigens are cleared, most of the effector T cells undergo apoptosis. The rest can differentiate into memory cells. Once differentiated into memory cells, they gain the capacity to secrete multiple cytokines (such as interferon-γ (IFNγ), tumor necrosis factor (TNF), and interleukin-2 (IL-2)) and mount robust responses upon re-infection (*Nature Reviews Immunology, 2023*; Kaech and Cui, 2012). Importantly, memory T cells can persist in the absence of antigen stimulation for a long period because of their ability to self-renew, using interleukin-7 and interleukin-15). However, effector T cells can also differentiate into exhausted T cells instead of memory cells. Exhausted T cells exhibit a decline in cytokine polyfunctionality, proliferation, and production of IFNγ, chemokines, and TNF. Furthermore, these cells express more inhibitory receptors, such as lymphocyte activation gene 3 protein (LAG3), programmed cell death protein 1 (PD1), CD160, 2B4, and T cell immunoreceptor with immunoglobulin and ITIM domains (TIGIT) (Paley et al., 2012). The level of expression of CD127 and KLRG1 markers distinguishes between exhausted and memory cells: high expressions of KLRG1 and low expression of CD127 characterize exhausted cells, while the reciprocal level of expression characterizes memory cells.

As can be seen in Figure 3, T cell exhaustion occurs in three stages: progenitor exhausted, transitory exhausted, and terminal exhausted T cells. Based on the 60-day temporal analysis of naive T cells subject to the tumor microenvironment (TME), it was found that exhaustion can only be reversed in T cells at the early stages of exhaustion (Belk et al., 2022). Hudson *et al.* determined that transitory exhausted T cells can return to a progenitor exhausted cell state through an *in vivo* experiment where transitory exhausted T cells reversed exhaustion

and exhibited effector-like expression of surface markers such as chemokine receptor CX3CR1 (2019). Therefore, any CAR platform that mitigates T cell exhaustion will need to be responsive enough to induce transient rest and reverse T cell exhaustion before T cells are at the later stage of T cell exhaustion: terminal exhaustion.

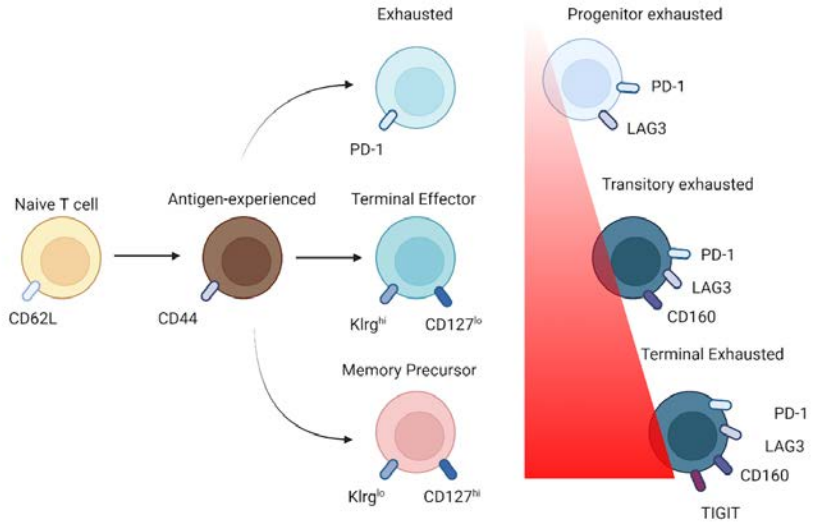


Fig 3: Stages of T cell exhaustion (Created with Biorender.com)

T cell exhaustion is often observed before and after the adoptive transfer of CAR T cells to the patient: T cell exhaustion may be observed for T cells after their removal from the patient or after they have been engineered to be CAR T cells (Finney et al., 2019; Fraietta et al., 2018). However, this paper focuses on T-cell exhaustion that occurs after CAR T cells have been administered to the patient. T-cell exhaustion after administration is caused by high tumor burdens in the patient and/or tonic signaling (Long et al., 2015; Lynn et al., 2019). High tumor burdens refer to persistent antigen stimulation (due to cancer persistence and incomplete antigen clearance), while tonic signaling refers to continuous low-level activation of CAR T cells in the absence of antigen stimulation. Tonic signaling occurs when CARs on a T-cell membrane aggregate independently of antigen stimulation.

Transient rest

T-cell exhaustion caused by tonic signaling can be mitigated by transient rest, the temporary inhibition of CAR T-cell signaling. This mitigation has the potential to improve antitumor activity and persistence, encouraging effector T cells to differentiate into memory cells (Weber et al., 2017).

Transient rest can be induced in CAR T cells by the regulation of CAR expression, where lower CAR expression leads to lower tonic signaling, and vice versa. This paper will discuss four recent platforms that are being developed to regulate CAR expression.

Platforms for CAR T cell regulation

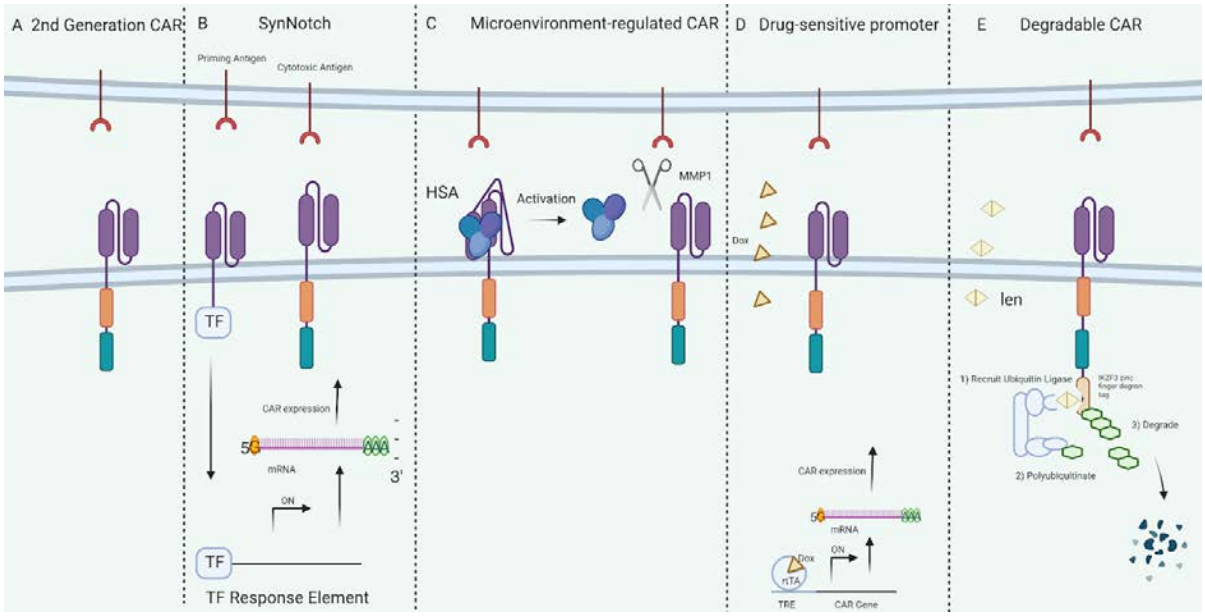


Figure 4: Regulatable Platforms for CAR Expression (Created with Biorender.com)

SynNotch CAR T cells

Exploiting logic gates when engineering CAR T cells also helps mitigate T cell exhaustion. Logic gates are principals from digital electronics that dictate what kind of output a device should make when it receives different combinations of signals. This can be applied in the context of CAR T cells by dictating what combination of antigens the CAR T cell needs to recognize for it to carry out a cytotoxic response. One of the engineering advances that employs logic gates for CAR T cells is the synNotch CAR circuit. This circuit involves the use of the Notch signaling pathway, where a receptor protein called 'Notch' plays a key role. This receptor protein consists of an extracellular domain, which recognizes ligands (such as Delta-like (DLL)-1, -3, -4, Jagged 1, and Jagged 2 ligands), and an intracellular domain which is a transcription factor (TF) of a protein. In a native Notch signaling pathway, when a ligand binds to the extracellular domain of this receptor, it is pulled up to expose a cleavage site between the extracellular domain and the intracellular domain. This site is then cut by enzymes to release the intracellular domain (transcription factor), which then moves to the cell's nucleus. Then, TFs – such as RBPJ (recombination signal binding protein for immunoglobulin kappa J region) and MAML (Mastermind-like proteins) -- alter the expression of a gene by attaching to the upstream promoter of that gene.

In a synNotch CAR circuit, the Notch signaling pathway is exploited to express multiple CARs with different antigen-specificities. A variety of antigens can be targeted by CARs. However, synNotch CAR circuits allow the T cell to be activated only when multiple antigens of the cancerous cell are recognized. In synNotch CAR circuits, the Notch extracellular domain is swapped for a CAR scFv against the first desired antigen. The Notch TF is swapped with one that promotes the transcription of the second CAR, which is a standard CAR. Thus, the first antigen detection effectively leads to the expression of another CAR that

is specific to another antigen. Once this CAR is expressed, it will detect whether another antigen specific to or associated with tumors is present. If it is present, then the second CAR will finally signal T-cell activation for the T-cell to begin cytotoxic activity.

In this way, a synNotch CAR circuit ensures that the T cell is only activated in the presence of multiple antigens, reducing the likelihood of off-tumor, on-target specificity. Axel *et al.* demonstrated that synNotch CAR T cells for mesothelioma and ovarian cancer exhibited greater persistence and less T cell exhaustion (2021). The mechanism behind this is that the second CAR is only expressed in the presence of the first antigen. This means that while there is no first antigen, there is no intracellular domain of the second CAR that can conduct tonic signaling. By preventing tonic signaling in the absence of the first antigen, the synNotch CAR T cell becomes less exhausted and becomes more inclined to differentiate into the memory-like phenotype. CD4+, CD8+, and synNotch CAR T-cells were all subjected to the first antigen but antigen-independent expansion – an indication of tonic signaling – was only observed in the first two types of cells. In another ex vivo T cell culture, it was found that at baseline antigen stimulation, CD4+ and CD8+ CAR T cells had a higher population of effector memory T cells; on the other hand, after antigen stimulation, synNotch CAR T cells had a higher population of T cell memory phenotypes. Despite these benefits of synNotch CAR T cells, the half-life of this particular platform was not illuminated (Hyrenius-Wittsten et al., 2021).

Microenvironment-regulated CAR T cells (MRS.CAR T)

While synNotch CAR T cells ensure that T cell activation only occurs in the presence of multiple tumor-associated antigens, the MRS.CAR T cell platform ensures that the activation only occurs in the tumor microenvironment. This is made possible by exploiting human serum albumin (HAS) and matrix metalloproteinase (MMP).

HSA is a protein that is present in the blood plasma of humans. It is the most abundant protein in the plasma and plays a crucial role in maintaining osmotic pressure, transporting various substances, and regulating pH levels in the blood. These proteins can bind to an HSA binding peptide (HSABP). In the context of CAR therapy, Wang *et al.* found that inserting an HSABP between the 5' terminal signal peptide and the scFv in a CAR construct would allow the CAR to be turned off in the presence of HSA: once the HSABP on the extracellular domain of the CAR construct binds to HSA, their interaction will lead to the CAR's conformational change, resulting in the scFv-antigen recognition site's occlusion (2023) In in vitro experiments, MRS.CAR T cells showed lower T cell exhaustion and, thus, a higher proportion of memory phenotypes than conventional CAR T cells. This suggests that the occlusion prevented tonic signaling from occurring. On the other hand, the removal of the HSA protein will reactivate the CAR. This part of the process involves MMPs. MMPs are a family of enzymes that play a crucial role in the breakdown and remodeling of the extracellular matrix (ECM). They degrade biological molecules whose cryptic biological functions are activated upon proteolysis. The points at which MMPs degrade these molecules are called MMP

cleavage sites. Wang *et al* also inserted an MMP1 cleavage site into the CAR construct so that the HSA protein can be released upon the interaction of the CAR T cell and MMP1 (2023). This is notable because tumor microenvironments are rich in MMP1. This finally leads to a CAR platform which is downregulated upon intravenous injection to the patient until it arrives in the tumor microenvironment rich with MMP1. Furthermore, a variety of tumors have upregulated levels of the MMP1 (which cleaves the occlusion in this CAR): nearly all tumors had upregulated levels of MMP1, MMP9, MMP10, MMP11, and MMP13. At least 10 different cancer types have upregulated MMP3, MMP7, MMP12, and MMP14. On the contrary, seven to nine different cancer types have considerably lower levels of MMP2, MMP7, MMP23B, MMP27, and MMP28 (Gobin et al., 2019).

This is a rather recent technology, and as of August 2023, there was only one MRS.CAR model developed. This model treats esophageal squamous carcinoma, a solid tumor in the esophagus, targeting B7-H3 as the antigen (Wang et al., 2023). Furthermore, the transduction efficiency was also highest in the MRS.CAR T cell. While this model requires the use of the technique used to produce recombinant proteins like the HSA binding peptide-MMP cleavage site-scFv protein (HSABP-MMPCS-scFv), the complexity and price of preparing MRS.CAR T cells are similar to those of conventional CAR T cells. This is because the engineering only requires the additional insertion of a small peptide domain. Most importantly, the anti-tumor function of this novel CAR platform was found to be on par with conventional CAR T cells.

However, the study suggests that there may be MMPs other than MMP1 that can cleave the MMP-cleavage site in the extracellular domain. This implies that MRS.CAR T Cells can be activated in regions with MMP even if they are not tumor microenvironments, leading to on-target, off-tumor toxicity.

Drug-sensitive promoters

CAR expression in T cells can be regulated by drug-sensitive promoters. The Tet-On system is one example of how this is done. It consists of a reverse tetracycline-controlled transactivator (rtTA), which is a modified version of the tetracycline-controlled transactivator (tTA). Upon adding tetracycline or its derivative, Dox (Dox), the rtTA protein is capable of binding to a tetracycline-responsive element (TRE) present in the target gene's promoter region (Das et al.). This binding leads to the transcription and expression of CAR. In this case, the transcription of the GOI leads to the synthesis of CAR on the T cell. On the other hand, in the absence of the antibiotic tetracycline/Dox, the rtTA protein does not bind to the TRE, impeding the expression of CAR.

For example, in 2016, researchers combined the second-generation CD19CAR with the third-generation Tet-On vector (referred to as Tet-CD19CAR) and then introduced this construct into primary CD8+ T cells using retroviral transduction (Sakemura et al., 2016). Flow cytometry revealed that this particular CAR platform required 100 ng/ml or more Dox for maximal CAR expression. Flow cytometric analysis also revealed that CAR was expressed

twelve hours after Dox was introduced, while CAR expression was repressed twenty-four hours after removal. In addition to this quick response to the presence or absence of Dox, this CAR platform was found to only show significant cytotoxic activity in the presence of Dox. While it was not significant, leaky activity, in the form of undesired CAR expression, was observed in the absence of Dox. Another limitation of this platform is that its transduction rate is relatively low: 5-10%.

Recent studies reveal some flaws in previous Tet-on systems (including the aforementioned CAR platform by Sakemura et al., 2016): rtTA proteins have the potential to bind to domains along the genome that have similar configurations with the TRE present in the target gene's promoter region (Whitsett and Perl, 2006; PMC, 2009; Morimoto and Kopan, 2009). This means that instead of the expression of CAR, other genes in the genome can be inadvertently transcribed. Furthermore, the expression of rtTA in lung cells was found to be toxic and a cause of emphysema (Sisson et al.).

To address this problem, a transactivator-free Tet-on system has been recently developed. This system uses a Tet repressor protein (TetR) and a chimeric insulator Is2 (Tristán-Manzano et al., 2023); TetR was developed from the bacterium *Escherichia* coli (E. coli), and Is2 was developed from chicken β -globin locus control region hypersensitive site 4 (HS4) and a synthetic scaffold/matrix attachment region (SAR2) (Benabdellah et al.). Instead of using the transactivator to regulate the expression of the CAR, the TetR is used. This means that this system works through a mechanism that is different from the aforementioned one. In the absence of the drug, TetR binds to the Tet operator (TetO) sequence, preventing the expression of the CAR. In the presence of the drug, however, TetR binds to the drug, which induces a conformational change. Then, TetR no longer can bind to TetO, allowing the expression of the CAR. TetR has a lower likelihood of binding to binding domains along the genome with similar configurations of the TRE because it does not contain a transactivation domain like the tTA. Further interrogation elucidated that there was no difference in the overall phenotype and function between T cells and this type of CAR T cells; that is, there was no damage done by the TetR through transcription of undesired genes. While the TetR regulates the expression of CAR, the Is2 insulator delineates the boundary between neighboring genes and the CAR. This prevents other regulatory elements proximal to the CAR from influencing its expression (Benabdellah et al.; Tristán-Manzano et al., 2023).

The responsiveness of this system was assessed by interrogating Jurkat cells, immortalized T cells used for research; it was found that within twenty-four to thirty-six hours of adding Dox, there was CAR expression in the Jurkat T cells (Tristán-Manzano et al., 2023). Furthermore, there was a near complete reversion of the response ninety-six hours after the Dox was removed (Tristán-Manzano et al., 2023). The same was observed for human primary T cells. Another benefit of this advancement is that it can be turned on with very small amounts of Dox (0.1 ng/ml) (Tristán-Manzano et al., 2023). In terms of transduction rates, it was found that up to 67% of human primary T cells interrogated are induced to be Dox-responsive (Tristán-Manzano et al., 2023).

When this system was incorporated into a third-generation CAR platform that targets CD19, leaky activity was also shown to be minimal (Tristán-Manzano et al., 2023). The antitumor activity was also controlled by the addition or removal of Dox (Tristán-Manzano et al., 2023). During *in vivo* experiments with mice that have lymphoma, there were near-background levels of CAR expression in the absence of Dox. Furthermore, mice treated with Dox exhibited lower tumor burden than those without, suggesting effective expression in the presence of Dox. Altogether, this implies that tonic signaling in CAR T cells can be tightly regulated as well since near-background levels of CAR expression also indicate near-background levels of tonic signaling. In comparison with previous CAR platforms, the amount of Dox needed to regulate CAR expression is also much lower, reducing the likelihood of the formation of antibiotic resistance in patients with long-term Dox treatment (Tristán-Manzano et al., 2023).

However, this CAR platform is not without its disadvantages: the repression of the CAR platform is relatively slow compared to other platforms upon Dox removal. This is because, after the removal of Dox, TetR takes time to accumulate and repress the promoter of the GOI. Oone of the caveats of this paper's methodology is that the exhaustion levels in CAR T cells engineered have not been interrogated, although it is widely accepted that lower CAR expression leads to lower tonic signaling, and therefore lower T cell exhaustion (Weber et al., 2021). In addition, the immunogenicity of this CAR platform has not been investigated.

Degradable CARs

While previous CAR platforms discussed were induced upon drug administration, degradable CARs work in the reverse direction: CAR expression is maintained in T cells until the administered drug triggers its degradation.

One way this can be done is through exploiting ubiquitination in T cells; this involves Cys2His2 (C2H2) zinc fingers and molecular glues such as thalidomide (and its analogs). The process of ubiquitination is as follows. Upon administration to the body, thalidomide and its analogs bind to C2H2 zinc fingers because they are recurrent degron motifs, which means that they are structural motifs found in many disease-related proteins that mark them for degradation. Thalidomide binds to cereblon (CRNN) in the CRL4CRBN E3 ubiquitin ligase; this brings together the CRBN and the target disease-related protein. This prompts the CRL4CRBN complex to add ubiquitin molecules to the target protein through a process called ubiquitination. Ubiquitin molecules are sequentially attached to the target protein, forming a chain. This ubiquitin chain serves as a "tag" that marks the target protein for degradation by the proteasome, a cellular organelle responsible for breaking down unwanted or damaged proteins.

Instead of degrading disease-related proteins, ubiquitination can be redirected to the CAR expressed on the T cell membrane. In other words, the addition of thalidomide and its analogs will trigger the ubiquitination of CARs. This will remove CARs and, thus, CAR signaling (including tonic signaling). This can be done by inserting degron domains into the CAR, allowing thalidomide and its analogs to bind to them. This type of CAR platform was

used to create an OFF-switch CAR that uses lenalidomide (an analog of thalidomide) (Jan et al.). In this CAR platform, the degron domain is IKZF3 (a lenalidomide-dependent degron developed from C2H2 zinc fingers).

While this CAR platform already exhibited its ability to be regulated, its regulation could be improved by increasing the efficiency at which its degron binds to the added small molecule drug. Jan et al. synthesized a degron, ZFP91-IKZF3, that showed higher efficiency in such binding (2021). Synthesized degrons like this are termed 'super degrons'. Because of this increase in efficiency, lower doses of lenalidomide could be used. In fact, this upgraded CAR platform required around 100-fold lower doses to degrade the CARs compared to CAR T cells IKZF3.

Using Jurkat cells, the kinetics of this CAR platform was interrogated. Results reveal that upon the addition of lenalidomide, half-maximal depletion of CAR platforms occurred within twenty minutes. On the other hand, the removal of the drug led to the half-maximal recovery of CAR expression within 3.6 hours. The results of *in vivo* experiments were consistent with those using Jurkat cells. CAR T cells with super-degrons showed five-fold lower CAR expression six hours after the addition of lenalidomide than CAR T cells without super-degrons. The CAR expression level before the drug administration was reached twenty-four hours after the removal of the drug. These findings demonstrate the most rapid kinetics in the regulation of CAR expression among the CAR platforms discussed in this paper (Jan et al., 2021).

In terms of antitumor activity, without the addition of lenalidomide, the super-degron CAR T cells (SD CAR) exhibited comparable levels of antitumor activity with conventional CAR T cells.

	Drug-sensitive promoters	Degradable CARs	SynNotch CAR	MRS.CAR
ON/OFF switch	ON switch	OFF switch	ON switch	ON switch
Leaky activity	Present	Low (half-maximal depletion in twenty minutes for in vitro)	None	Not determined in study
Responsiveness	Slow	Fast	Antitumor activity affected by presence of multiple antigens	Antitumor activity affected by MMP1 levels in the tumor microenvironment
Drug Dosage	High, may cause damage to gut microbacterium	Low and only administered for periods of transient rest	No dose required	No dose required
On-target, off-tumor toxicity	Not determined in study	Not determined in study	Relatively low due to higher specificity in terms of using logic gates	Further studies required to understand effect of MMPs other than MMP1 on MRS.CAR activation
Antitumor acitivty	Similar with conventional CARs	Similar with conventional CARs	Similar with conventional CARs	Similar with conventional CARs

Table 1: Comparison of CAR Platforms

In terms of T cell exhaustion, degradable CAR platforms that use super-degrons provide the most transient rest to exhausted CAR T cells. Degradable CARs that use an OFF

switch ensure that drugs are only administered when transient rest needs to be provided. This is significantly different from other CAR platforms that require the administration of drugs to be induced. In other words, the dosage of drugs for degradable CARs will be much lower than those of inducible CAR platforms. Furthermore, degradable CARs show rapid degradation and recovery, allowing the rest provided to CAR T cells to truly be 'transient'. In other words, unwarranted long periods of rest, which lets the tumor continue to grow, are less likely to occur when using degradable CARs. Because of its few limitations and robust mitigation of T-cell exhaustion, I think the degradable CAR is the most appropriate platform for solid tumors. It would be interesting to see it further developed and applied in clinical settings.

Conclusions and Future Directions

Although CAR T cell therapy is a new and successful technology against liquid tumors, it has limitations against solid tumors and leads to relapses of liquid tumors. Solutions to these limitations are currently being researched, leading to developments of regulatable CAR expression that can aid in overcoming T cell exhaustion by providing transient rest. The development of synNotch CARs, microenvironment-regulated CARs, drug-sensitive promoters in CARs, and degradable CARs shows promise in overcoming T-cell exhaustion.

Despite these advances in overcoming T cell exhaustion, CAR T cell therapy has other obstacles in treating solid tumors such as inadequate antitumor activity and limited infiltration of the dense tumor microenvironment. Of the four types of CAR T cells discussed in this paper, all showed similar levels of antitumor activity to untransduced T cells. One of the ways the antitumor activity of these CAR T cells can be enhanced is its synergy with CAR macrophages (M) in the innate immune system.

The innate immune system is the first line of defense of the body which supports the adaptive immune system. Unlike the adaptive immune system, its response to pathogens is nonspecific. Among the cells and proteins that comprise the innate immune system, M1 macrophages (a subtype of macrophages) play a key role in facilitating the infiltration of CAR T cells into the tumor microenvironment and enhancing CAR T-cell antitumor activity. Similar to T cells, macrophages can also be transduced to express chimeric antigen receptors, allowing them to be activated by specific antigens just as CAR T cells do. CAR M cells home to tumor microenvironments and release proinflammatory cytokines such as tumor necrosis factor-alpha (TNF-α), interleukin-1 beta (IL-1β), interleukin-6 (IL-6), and interleukin-12 (IL-12) (Arango Duque and Descoteaux); this leads to the inflammation of the dense extracellular matrix of the tumor, facilitating the infiltration of CAR T cells to the microenvironment. This is especially important for solid tumors in which CAR T cells have limited infiltration capacity.

The increase in antitumor activity is achieved through a synergy between CAR T cells and CAR-M cells: the release of inflammatory factors by CAR T cells causes the upregulation of costimulatory ligands CD86 and CD80 in CAR-M cells. These upregulated ligands on CAR-M cells can then activate the costimulatory domain in CAR T cells to amplify CAR T cell signaling, which in turn drives stronger antitumor activity. While this stronger signaling

may contribute to T cell exhaustion, it is interesting that the exposure of CAR T cells to CAR M cells actually mitigates T cell exhaustion. In an in vivo experiment where CAR T cells and CAR-M cells were both administered to a cell line with lymphoma, the expression of exhaustion markers (TIGIT, PD-1, LAG-3) on CAR T-cells decreased. Furthermore, the expression of CD62L decreased. This implies that CAR T cells exposed to CAR M cells developed a more memory-like phenotype (which has low CD62L expression and low expression of exhaustion markers). The exact mechanism whereby exhaustion markers and memory markers on CAR T cells are affected by the CAR-M cells is yet to be studied; however, these findings show promise in improving the antitumor activity of CAR T cell therapy by exploiting the synergistic effect of CAR T cells and CAR-M cells. Administering CAR M cells with regulatable CAR T cells can unleash the combined power of the adaptive and innate immune system.

Works Cited

Ahmad, Ubaid, et al. "Chimeric Antigen Receptor T Cell Structure, Its Manufacturing, and Related Toxicities; A Comprehensive Review." *Advances in Cancer Biology - Metastasis*, vol. 4, July 2022, p. 100035. *ScienceDirect*, https://doi.org/10.1016/j.adcanc.2022.100035.

Arango Duque, Guillermo, and Albert Descoteaux. "Macrophage Cytokines: Involvement in Immunity and Infectious Diseases." *Frontiers in Immunology*, vol. 5, Oct. 2014, p. 491. *PubMed Central*, https://doi.org/10.3389/fimmu.2014.00491.

Belk, Julia A., et al. "Epigenetic Regulation of T Cell Exhaustion." *Nature Immunology*, vol. 23, no. 6, June 2022, pp. 848–60. *DOI.org (Crossref)*, https://doi.org/10.1038/s41590-022-01224-z.

Benabdellah, Karim, et al. "Lent-On-Plus Lentiviral Vectors for Conditional Expression in Human Stem Cells." *Scientific Reports*, vol. 6, no. 1, 1, Nov. 2016, p. 37289. *www.nature.com*, https://doi.org/10.1038/srep37289.

Conditional Expression of Genes in the Respiratory Epithelium in Transgenic Mice - PMC. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2720111/. Accessed 11 Aug. 2023.

Cytokine Control of Memory T-Cell Development and Survival | Nature Reviews Immunology. https://www.nature.com/articles/nri1052. Accessed 10 Aug. 2023.

Das, Atze T., et al. "Tet-On Systems For Doxycycline-Inducible Gene Expression." *Current Gene Therapy*, vol. 16, no. 3, June 2016, pp. 156–67. *PubMed Central*, https://doi.org/10.2174/1566523216666160524144041.

Eshhar, Z., et al. "Specific Activation and Targeting of Cytotoxic Lymphocytes through Chimeric Single Chains Consisting of Antibody-Binding Domains and the Gamma or Zeta Subunits of the Immunoglobulin and T-Cell Receptors." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 90, no. 2, Jan. 1993, pp. 720–24. *PubMed*, https://doi.org/10.1073/pnas.90.2.720.

Finney, Olivia C., et al. "CD19 CAR T Cell Product and Disease Attributes Predict Leukemia Remission Durability." *The Journal of Clinical Investigation*, vol. 129, no. 5, May 2019, pp. 2123–32. *PubMed*, https://doi.org/10.1172/JCI125423.

Fraietta, Joseph A., et al. "Determinants of Response and Resistance to CD19 Chimeric Antigen Receptor (CAR) T Cell Therapy of Chronic Lymphocytic Leukemia." *Nature Medicine*, vol. 24, no. 5, May 2018, pp. 563–71. *PubMed*, https://doi.org/10.1038/s41591-018-0010-1.

Global Cancer Facts & Figures | American Cancer Society. https://www.cancer.org/research/cancer-facts-statistics/global.html. Accessed 7 Aug. 2023.

Gobin, Emily, et al. "A Pan-Cancer Perspective of Matrix Metalloproteases (MMP) Gene Expression Profile and Their Diagnostic/Prognostic Potential." *BMC Cancer*, vol. 19, no. 1, June 2019, p. 581. *BioMed Central*, https://doi.org/10.1186/s12885-019-5768-0.

Hudson, William H., et al. "Proliferating Transitory T Cells with an Effector-like Transcriptional Signature Emerge from PD-1+ Stem-like CD8+ T Cells during Chronic Infection." *Immunity*, vol. 51, no. 6, Dec. 2019, pp. 1043-1058.e4. *PubMed*, https://doi.org/10.1016/j.immuni.2019.11.002.

Hyrenius-Wittsten, Axel, et al. "SynNotch CAR Circuits Enhance Solid Tumor Recognition and Promote Persistent Antitumor Activity in Mouse Models." *Science*

Translational Medicine, vol. 13, no. 591, Apr. 2021, p. eabd8836. *DOI.org (Crossref)*, https://doi.org/10.1126/scitranslmed.abd8836.

Jan, Max, et al. "Reversible ON- and OFF-Switch Chimeric Antigen Receptors Controlled by Lenalidomide." *Science Translational Medicine*, vol. 13, no. 575, Jan. 2021, p. eabb6295. *PubMed*, https://doi.org/10.1126/scitranslmed.abb6295.

Kaech, Susan M., and Weiguo Cui. "Transcriptional Control of Effector and Memory CD8+ T Cell Differentiation." *Nature Reviews. Immunology*, vol. 12, no. 11, Nov. 2012, pp. 749–61. *PubMed*, https://doi.org/10.1038/nri3307.

Lee, Daniel W., et al. "T Cells Expressing CD19 Chimeric Antigen Receptors for Acute Lymphoblastic Leukaemia in Children and Young Adults: A Phase 1 Dose-Escalation Trial." *Lancet (London, England)*, vol. 385, no. 9967, Feb. 2015, pp. 517–28. *PubMed*, https://doi.org/10.1016/S0140-6736(14)61403-3.

Long, Adrienne H., et al. "4-1BB Costimulation Ameliorates T Cell Exhaustion Induced by Tonic Signaling of Chimeric Antigen Receptors." *Nature Medicine*, vol. 21, no. 6, June 2015, pp. 581–90. *PubMed*, https://doi.org/10.1038/nm.3838.

Lynn, Rachel C., et al. "C-Jun Overexpression in CAR T Cells Induces Exhaustion Resistance." *Nature*, vol. 576, no. 7786, Dec. 2019, pp. 293–300. *PubMed*, https://doi.org/10.1038/s41586-019-1805-z.

Majzner, Robbie G., and Crystal L. Mackall. "Clinical Lessons Learned from the First Leg of the CAR T Cell Journey." *Nature Medicine*, vol. 25, no. 9, Sept. 2019, pp. 1341–55. *PubMed*, https://doi.org/10.1038/s41591-019-0564-6.

Morimoto, Mitsuru, and Raphael Kopan. "rtTA Toxicity Limits The Usefulness Of The SP-C-rtTA Transgenic Mouse." *Developmental Biology*, vol. 325, no. 1, Jan. 2009, pp. 171–78. *PubMed Central*, https://doi.org/10.1016/j.ydbio.2008.10.013.

Paley, Michael A., et al. "Progenitor and Terminal Subsets of CD8+ T Cells Cooperate to Contain Chronic Viral Infection." *Science (New York, N.Y.)*, vol. 338, no. 6111, Nov. 2012, pp. 1220–25. *PubMed*, https://doi.org/10.1126/science.1229620.

Sakemura, Reona, et al. "A Tet-On Inducible System for Controlling CD19-Chimeric Antigen Receptor Expression upon Drug Administration." *Cancer Immunology Research*, vol. 4, no. 8, Aug. 2016, pp. 658–68. *PubMed*, https://doi.org/10.1158/2326-6066.CIR-16-0043.

Sisson, Thomas H., et al. "Expression of the Reverse Tetracycline-Transactivator Gene Causes Emphysema-like Changes in Mice." *American Journal of Respiratory Cell and Molecular Biology*, vol. 34, no. 5, May 2006, pp. 552–60. *PubMed*, https://doi.org/10.1165/rcmb.2005-0378OC.

The Integration of T Cell Migration, Differentiation and Function | Nature Reviews Immunology. https://www.nature.com/articles/nri3442. Accessed 10 Aug. 2023.

Tomasik, Jaromir, et al. "Next Generations of CAR-T Cells - New Therapeutic Opportunities in Hematology?" *Frontiers in Immunology*, vol. 13, 2022. *Frontiers*, https://www.frontiersin.org/articles/10.3389/fimmu.2022.1034707.

Tristán-Manzano, María, et al. "Lentiviral Vectors for Inducible, Transactivator-Free Advanced Therapy Medicinal Products: Application to CAR-T Cells." *Molecular Therapy. Nucleic Acids*, vol. 32, June 2023, pp. 322–39. *PubMed*, https://doi.org/10.1016/j.omtn.2023.03.018.

Wang, Lihong, et al. “A Novel Microenvironment Regulated System CAR-T (MRS.CAR-T) for Immunotherapeutic Treatment of Esophageal Squamous Carcinoma.” *Cancer Letters*, vol. 568, Aug. 2023, p. 216303. *DOI.org (Crossref)*, https://doi.org/10.1016/j.canlet.2023.216303.

Weber, Evan W., et al. “Transient Rest Restores Functionality in Exhausted CAR-T Cells through Epigenetic Remodeling.” *Science*, vol. 372, no. 6537, Apr. 2021, p. eaba1786. *DOI.org (Crossref)*, https://doi.org/10.1126/science.aba1786.

What Is Cancer? - NCI. https://www.cancer.gov/about-cancer/understanding/what-is-cancer. Accessed 7 Aug. 2023.

Whitsett, Jeffrey A., and Anne-Karina T. Perl. “Conditional Control of Gene Expression in the Respiratory Epithelium: A Cautionary Note.” *American Journal of Respiratory Cell and Molecular Biology*, vol. 34, no. 5, May 2006, pp. 519–20. *PubMed*, https://doi.org/10.1165/rcmb.F310.

Yang, Shicheng, et al. “The Shedding of CD62L (L-Selectin) Regulates the Acquisition of Lytic Activity in Human Tumor Reactive T Lymphocytes.” *PLoS ONE*, vol. 6, no. 7, July 2011, p. e22560. *PubMed Central*, https://doi.org/10.1371/journal.pone.0022560.

“Mass Shootings in America: The Mistaken Story” An Opinion Piece By Caroline Browning

Mass shootings have become a classic American tragedy over the past twenty-four years. When two students walked into a high school in Columbine, Colorado and shot and killed twelve students and one teacher, the country was in shock. The Columbine Massacre was an unprecedented act of violence we had no idea how to comprehend. We thought it would end there. Since the Columbine Massacre, America has seen approximately three-hundred and eighty-six school shootings.

As a country, we have developed digestible explanations to explain the violence we see. The most prominent is that mass shooters are victims themselves, of bullying, abuse, and malicious social isolation. This myth stems from our initial impression of the Columbine shooters. The public painted those criminals as bullied, lonely young men. We were wrong. Accounts now show those men to have been the bullies from the beginning. They were entitled bullies with a superiority complex who enjoyed tormenting their classmates with vandalism and violent threats. This is the mold that a typical mass shooter actually fits.

The High Self-Esteem Theory, developed by Baumeister, Smart, and Bolden, challenges the pre-existing assumption that low self esteem is a primary cause of violence. Instead, their research points to an inflated ego as an origin of cruelty (Baumeister). In fact, many mass shooters fit narcissistic traits, meaning their sense of self is completely exaggerated. In their eyes, they are superior beings. This ego; however, is fragile and takes a lot of attention, validation, and perceived admiration to maintain. To someone with such a heightened sense of self, others disagreeing with them, rejecting them, or not giving them the attention and “respect” they desire is a personal attack. They constantly see themselves as “the victim” when in reality they are far from it. This is where arrogance becomes dangerous. A person who ties their entire identity to their alleged supremacy will do anything it takes to uphold this belief and prove their power. It does not take a legitimate threat to convince them they’re being threatened.

Data shows that more mass shooters are white than every minority group combined, and the alarming majority are men. This doesn’t imply that white men have a natural affinity for violence or are innately egocentric. These statistics say a lot more about societal bias and our preconceived notions about what masculinity is. In white men, exaggerated value in competition, aggression, and ego are viewed as precursors to success. These are not traits that we encourage in ethnic minorities or women on such a sizable scale. Popular culture also enforces the belief that white men deserve the jobs, women, and whatever else that they want, which cultivates a dangerous sense of entitlement and righteousness.

Still, why are we seeing such staggeringly high rates of mass shootings following the Columbine Massacre? This is largely affected by the stories these shooters get. We put their names in bold, plaster their faces on televisions, release their ideas, and tell their stories. We may be painting them as villains, but we’re painting them nonetheless. Infamy is still fame, and to someone obsessed with perception, respect, and attention, that is the dream. Now, as each school

shooter is getting their "ticket to fame," we wonder why the cycle continues. This is why the cycle continues. It's an act of violence that assures everything someone so narcissistic and cruel could dream of.

The American shooter problem is not impossible to alleviate. This rush of violence can be slowed. Shifting the blame away from low-self esteem to inflated ego and bullying is one tool. We must give our attention to the bullies and the egomaniacs. Identifying those individuals in schools and communities will allow them to get the tools and support they need to lead a different life and alter their instincts and beliefs. This is especially useful for young people exhibiting these traits. The mean, righteous ones that remain combative are usually the ones to deny help, but often they need it most.

Then, targeting the most common demographic of shooters is another huge part of solving this problem. The fissures and damages in our society that allow misogyny and male entitlement to run rampant are dangerous. We must start teaching young boys that they do not need to be aggressive to be assertive or arrogant to be confident. We need to stop praising narcissistic and bullying behavior because it makes "a leader." Those traits do nothing for the men we nurture them in. The more as a society we enable violent, cruel men, the worse off we all are.

There's a hashtag that has circulated on social media as of late; #nonotoriety. This means that shooters' names, stories, and images do not deserve press. They do not deserve infamy. They deserve to fade into oblivion, because that is where they belong. In France, 2016, following the Bastille Day terrorist attacks, news organizations chose to stop releasing the images, propaganda, and names of the people perpetrating the violence to stop the effects of "posthumous glorification," as the press group *Le Monde* expressed in an editorial titled "Resisting the Strategy of Hate." This strategy is similar to the one used to cover suicides in many countries; limiting coverage and information to prevent copy cats. Some experts, in fact, believe that this is the most effective way to cover acts of violence without potentially glorifying them (Fenoglio). Though the instances in France are acts of terrorism, which is not always what mass shootings are motivated by, dramatic, overly informational media coverage can still be detrimental to ending the cycle of gun violence. While terrorists require attention to their cause and want to create fear in the population, mass shooters with narcissistic tendencies want to draw attention to themselves and get "respect" or fear from the masses. Therefore, there is some genuine merit to applying these strategies in covering mass shootings in the news. Perhaps we should turn our attention to the victims, or first responders on the scene. Minimize the "power" of the shooter, and maximize the strength of the people.

Works Cited

Baumeister, R F et al. “Relation of threatened egotism to violence and aggression: the dark side of high self-esteem.” *Psychological review* vol. 103,1 (1996): 5-33. doi:10.1037/0033-295x.103.1.5

Craw, Victoria. “Terror attacks: Expert argues media coverage can increase copycat violence | news.com.au — Australia's leading news site.” *News.com.au*, 27 July 2016, https://www.news.com.au/world/europe/expert-calls-for-terror-attacks-to-be-treated-same-way-as-suicide-when-it-comes-to-media-coverage/news-story/04d99ebfd2092d9be952928a92995360. Accessed 13 September 2023.

Fenoglio, Jérôme. “Attaque de Saint-Etienne-du-Rouvray : Résister à la stratégie de la haine.” *Le Monde*, 27 July 2016, https://www.lemonde.fr/idees/article/2016/07/27/resister-a-la-strategie-de-la-haine_4975150_3232.html. Accessed 13 September 2023.

Gun Violence Archive, https://www.gunviolencearchive.org/. Accessed 5 September 2023.

“Mass shootings by shooter’s race in the U.S. 2023.” *Statista*, 1 August 2023, https://www.statista.com/statistics/476456/mass-shootings-in-the-us-by-shooter-s-race/. Accessed 5 September 2023.

Brain Regions Activated in Different Types of Context-Dependent Memory: A Review By Abigail Bogdanovsky

Abstract

Encoding refers to the initial learning of information and storing it and retrieval is the recovery of the information from the storage. Oftentimes the memory or recall of an individual's experience is improved when that experience is encoded and then retrieved within the same context. This phenomenon is referred to as context-dependent memory (CDM; Hupbach et al., 2008). The four commonly referenced types of CDM are environmental context, state-dependent learning, mood-dependence/congruence, and cognitive context, which are outlined below. These four refer to different shared contexts between encoding and retrieval; more specifically, an environmental context could be a shared location (ex: learning and being tested in the same classroom), a state-dependent context would be a shared physical or emotional space (ex: learning and recalling information while being intoxicated), congruence would be a shared mood (ex: learning and being tested in a good mood), and a cognitive context would be a shared cognitive state (ex: recalling information in a certain language that were incurred in that language). Neurally, the hippocampus, prefrontal cortex and the specific regions of the parietal lobe have been implicated in CDM more broadly. However, the majority of experiments testing different types of CDM on different tasks are often experimented behaviorally rather than looking deeper into the physical brain regions the different types of CDM activate and rely on. To address how the different brain regions are involved between the four types of context, we will review studies of the subtypes of CDM and will consolidate them in this review. Our goal is to discern if all subtypes of CDM rely on the same brain regions or if there are variations across subtypes.

Introduction

Events that make up our memories often have a shared time, setting or internal state (emotion) - in other words, a shared context. Thus, context situates events that occur in one's daily life and plays a crucial role in the formation and recollection of memories by categorizing information for encoding (the process of forming memories) and cues for retrieval (the process of recovering memories).

For instance, imagine you have just graduated from high school and are thinking about your years from elementary school until now- are you able to remember the specific days of the 13 different years of school that you have attended? Do you have distinct memories for each of these days? Likely, no. As you speak to a friend, you recall moments from one of the classes that you shared - freshmen year biology class.

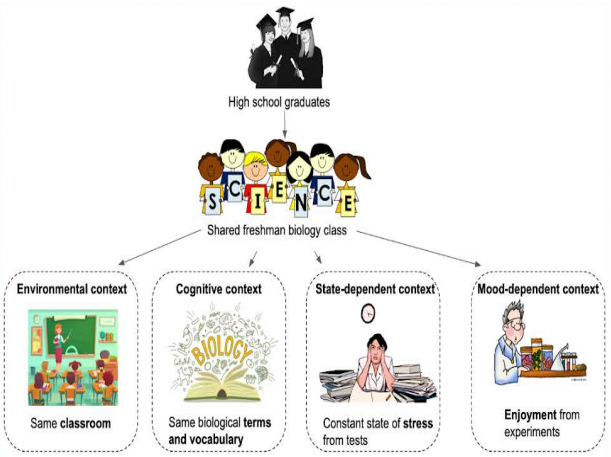


Fig 1: Example of Contexts + subevents.

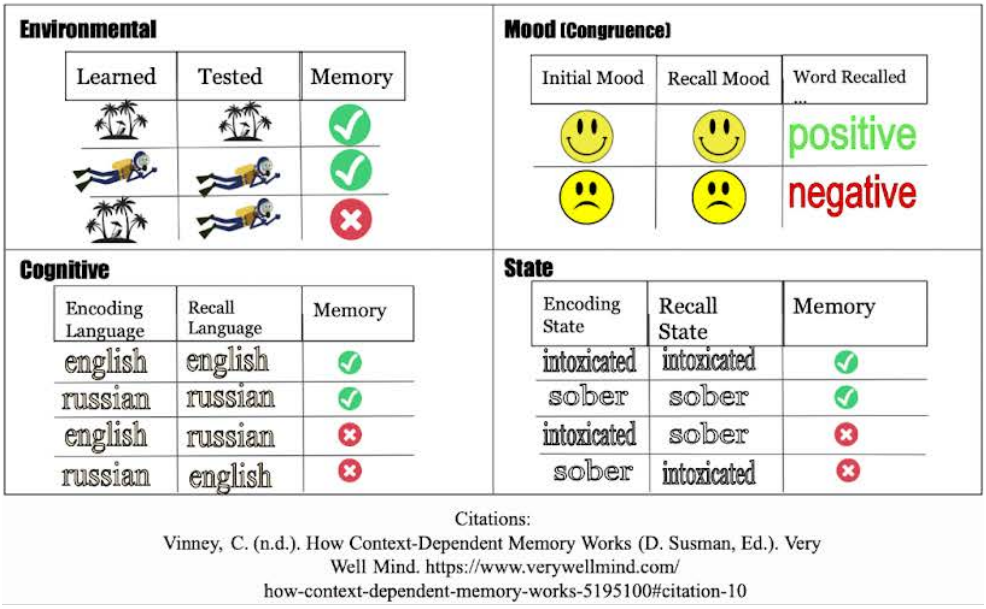


Fig 2: Types of Context-Dependent Memories and Summary of an Experiment Conducted Testing It.

Though you may recall the entertaining and fun classes and time-consuming projects more easily, the other hundreds of days spent in the class have a shared context — the same people sitting around you, the same teacher, the same classroom setting, and probably the same mood— and thus, are harder to remember (Fig 1).

All of these schooldays ("subevents") are likely grouped together based on the fact that they have a shared context and thus are retrieved from our memories as one singular event,

which is why it is hard for us to dissociate them. At the same time, a particularly exciting day would come with its own unique emotional state or location or activity (like a fun field trip), thus breaking away from the shared context, and existing as a new, dissociable "event" in our memories.

Context can come from many aspects of an event. This paper specifically focuses on the environmental, state, cognitive, and mood contexts (Fig 2). These four are the most commonly studied contexts when researching the topic of context-dependent memory (CDM). To see all articles included in this review, see section titled Supplementary Materials. CDM is the phenomenon that refers to the recollection of an individual's experience that can be improved when the experience is encoded and later retrieved within the same context. Such improved recollection is a part of one's everyday life.

Environmental dependent memory is when a memory can be enhanced when the setting of the period for encoding and retrieval are the same. For example, in the research of Godden & Baddeley, environmental context was studied by comparing memories on dry-land versus underwater (Godden & Baddeley 325 - 331). The experiment consisted of divers in several test groups that experimented with learning lists of words in different environments (underwater or dry land). They were then asked to recall the word lists in either the same setting as they had learned it in, or a different setting. The results of the experiment showed that subjects recalled the lists better if they had been asked to recall the lists in the same environment in which they learned them, confirming the presence of environmental context. A meta-analysis conducted by Smith and Vela (203 - 220) looked across many investigations testing environment CDM and confirmed that this form of context has clear effects on memory.

Cognitive dependent memory has been primarily studied in literature in relation to recall in different languages. Marian and Neisser conducted an experiment testing whether recall is improved when the language the recall was in was the same as encoding (Marian & Neisser 361 - 368). Subjects in the study were Russian immigrants and bilingual Russian and English speakers and were prompted to think of certain life events in response to prompted words. They found that the subjects were able to recall and remember more of their past if their past experiences were prompted in the same language that the experiences had occurred in. For example, the participants were able to better recall events that occurred in their old home in Russia if they were prompted to recall in Russian. A similar experiment was conducted with Japanese-Americans Matsumoto & Stanny. The results from this study demonstrated that cue words written in Japanese led to easier recollection of events from their past residence in Japan compared to cued words written in English (Matsumoto & Stanny 378 - 390). Ultimately, many studies exploring cognitive CDM corroborate the effect of language in memory function, but the majority of this work has been limited to language-based designs (see Conclusion).

State-dependent memory (SDM) correlates one's different physiological states during encoding and retrieval to the efficiency and accuracy of their memory. SDM is commonly tested with different types of substances such as alcohol, or other drugs to induce a certain state in the subject. A study conducted by Eich and colleagues had a group of subjects either smoke

marajiuana or a placebo (Eich J. E. et al. 408 - 417). The study found that the words learned during the intoxicated state were better remembered by the subjects if they were asked to recall them again in an intoxicated state. Goodwin and colleagues (191 - 198) found commensurate results when investigating drunk and sober subjects. The subjects were medical students tasked with memorizing a list while they were either in their assigned sober or drunk state. Twenty-four hours later the students were asked to recall the words in either the same state as they were in during the list memorization period, or in the opposite state. Similar to the results from the previous study, this study found that participants who were in the same state during encoding and recall performed better.

Mood dependent memory emphasizes the significance of emotion during the process of encoding and recalling memories. When studying mood-dependent memory (MDM), experimenters attempted to stimulate a certain mood in different participants. In the study, Eich ad Metcalfe (443–455), participants had to generate 16 target items based on certain criteria and given the first letter of a word, and read 16 of the target items while either feeling pleasant or unpleasant. Two days after the encoding of the generated by criteria target items, and the read target items, they tested the subjects retention through free-recall and several other methods. The subjects were either in the same pleasant or unpleasant mood as they had encoded in, or the opposite mood. The participants were either set in a happy or sad mood for encoding and were instructed to generate autobiographical moments. Two days later they were either recalling the memories from the same mood as they had been in previously, or the opposite mood. The results of the research support MDM, as memories were more likely to be recalled if the subject was in the same mood as during encoding.

The four types of CDM, environmental, cognitive, state, and mood, have all been extensively studied to determine their impact on memory. Years of research has shown that all types have the potential to be extremely impactful in their ability to improve recall. Overall, encoding a certain task in a specific context and recalling that task in the same context proved to aid in recall, therefore improving memory. While CDM is one example of how context can enhance memory, many other factors also contribute to context.

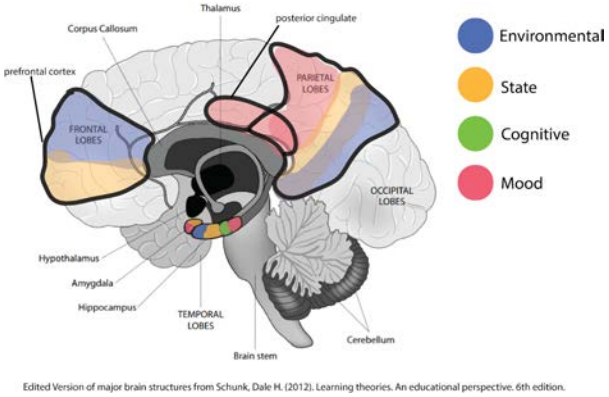


Fig 3: Neural basis of different types of CDM.

A typical view of context is that it sets up expectations or contingencies that can serve as ways of organizing information or as cues for retrieval. Context provides a rich set of retrieval cues, which can range from an emotional state to a physical space. The research investigating the four main types of CDM all use different tests, stimuli, and are generally conducted in very different ways. Despite these differences, the neural mechanisms responsible for the phenomena are mostly shared between the types. This paper will now address some of the shared anatomical structures that execute CDM (Fig 3).

Hippocampus

The hippocampus has been shown to play a role in the four types of CDM discussed. The hippocampus is a brain structure that is most recognized for its role in memory. The hippocampus is responsible for processing events and then helping retrieve the memory of the event. Given its relevance to the foundation of memories, it plays a key role in CDM.

The significance of the hippocampus to the function of memory was shown in a case report. An epilepsy patient had a bilateral medial temporal lobectomy which removed ⅔ of his hippocampus, and other brain structures, with the intent to abrogate seizure activity. After surgery, it was discovered that the patient had developed severe memory impairment. Recent events up to events from minutes ago were forgotten. It was the hippocampal lesions in this patient that clarified the involvement of the hippocampus in memory formation and retention.

The hippocampus is not necessarily needed for all types of memory, such as in dealing with old memories that have been semanticized (not an episodic memory anymore, such as memorizing a word meaning) or memories that are non-declarative (ex: riding a bike, brushing your teeth). But, in general the hippocampus is still crucial to the development of declarative memories (ex: recall of an event, including the processing of and formation of new memories from the event).

Parietal Cortices

Parts of the parietal cortices, particularly the somatosensory cortex, play a huge role in sensory understanding. This brain region utilizes and processes input from systems all around the body. Although involvement of the parietal cortices in episodic memory was previously unclear, it does have a role in memory. My review found evidence of the parietal cortices involvement in environmental, state, and mood CDM, but not cognitive CDM. Recent research has shown that certain parts of the parietal cortices are, in fact, heavily utilized during CDM. Studies were conducted on patients with parietal cortices lesions to examine the specific complications (Cabeza 613 - 625). One patient endured lesions in her left and right ventral parietal cortex (VPC) and noticed that she was still capable of remembering events in her life, but was unsure of where her memories came from. Other patients with VPC lesions reported feeling as though their recollection was not functioning properly, as their memories were not vivid or familiar to them. Another part of the study examined the differences between the VPC and dorsal parietal cortex (DPC) using neuroimaging and evidence from the lesions. There were significant differences

noted, such as functional differences regarding attention during encoding. Their functional differences help emphasize how parts of the parietal cortices are relevant to CDM. The research conducted with parietal cortex lesions has helped demonstrate how the parietal cortex may be activated during episodic memory and also contribute to encoding and retrieval during the CDM phenomena.

Prefrontal Cortex (PFC)

The prefrontal cortex is heavily involved in two out of the four types of CDM: environmental and state. The PFC is known for its role in carrying out executive functions such as goal setting, self control, staying focused, dealing with behavior, understanding consequences, and regulating personality. Its underlying functions are critical to storing memories as the PFC helps focus attention to encoding, and retrieval of memories.

A study conducted by, Zhang and colleagues found that the PFC, more specifically the left inferior frontal gyrus (IFG) was involved in CDM (Zhang, W et al. 579 - 593). The study utilized different pictures of faces overlaid on different backgrounds. The goal was to see if altering the picture background would affect memory in some way. The IFG showed stronger coupling, or a stronger feedback relationship, with regions involved selectively with face processing, when those faces were, in fact, remembered. The IFG helped generate and synthesize information from different areas of the brain that function in recognition and processing contexts. This study suggests that the PFC plays a significant role in CDM. Given the PFC's broad functions in working memory and emotion processing, further research is needed to determine the possible involvement of the PFC in mood and cognitive CDM.

Other Minorly Involved Neural Mechanisms

Memory is a complicated task that requires many different bodily systems to collaborate. Some brain regions are more heavily involved in the processes than others. Other brain regions involved in CDM are the midbrain, basal ganglia, posterior part of the cingulate gyrus, and the amygdala. They are relevant in at least one of the types of CDM, and could be relevant in more research that my review has not yet found. Below, I will highlight several relevant subfields of research, their corresponding neural underpinnings, and how they may relate to CDM.

Memory retrieval system

CDMs are forms of episodic memories, or memories of events. The posterior medial system (including countless parts of the brain such as the retrosplenial cortex, the mamillary bodies, anterior thalamic nuclei, presubiculum, parasubiculum, posterior cingulate, precuneus, angular gyrus, and ventromedial PFC) is a network of brain regions thought to support the encoding and retrieval of episodic memories and specifically, the retrieval of autobiographical memories (the memories of personal events) (Robinson 578 - 595). This system is also broadly attributed to functioning in attention, spatial navigation, memory, emotion, and understanding rewards. Although I've only found specific evidence for a subset of these regions, I would

suggest that *all* of them are crucial to the generalized process of CDM due to their heavy attribution to processing memories. .

Emotion/ mood

One type of CDM, mood CDM, heavily relies on one's personal emotions during the encoding and retrieval of memories. Two brain regions important for processing emotion include the amygdala and the basal ganglia. The amygdala is known for its relation in regulating emotions and processing stimuli that triggers certain emotions. Its prevalence in the process of emotional learning suggests that it may play an important and complex role in mood CDM.

Similarly, the basal ganglia is attributed to several different functions, such as motor control, executive functioning, behavior, emotion, and reward processing. The latter three play a huge role in the types of experimental tasks used in CDM studies, and in the formation of memories in general. The relation of the basal ganglia to the formation of memories means that it could be relevant to all of the types of CDM.

Movement

A key part of memory retrieval as seen with the CDM phenomena is having a repeated cue during encoding to aid the process of retrieval. Many different types of cues have been discussed in this paper, but an important cue that is part of everyday life is movement. Three relevant structures are involved in movement, including the basal ganglia, the midbrain, and the cerebellum.

While the basal ganglia is recognized for its significance in emotion and mood, it is also necessary for motor control. The basal ganglia is partially composed of three pathways that communicate with other brain areas to enact movement and prevent competing movements from interfering with the intended movement.

The midbrain is mostly commonly associated with the movement of the body and head; but when the hippocampus and midbrain interact, movements and choices are encoded into the hippocampus. This encoded moment is then essentially put into "storage" and can be later used to to help guide any future decision making that is relevant to the past encoded moment. .

The cerebellum is often associated with the phenomena of muscle memory, a type of procedural memory in which muscles are able to essentially remember a movement or are able to easily return to a certain strength due to repetitive actions or muscle stimulation. This procedural memory is more involved in environmental CDM experiments such as the Godden and Baddeley (325 - 331) which involved the exercise of scuba diving , but it could be applicable to other CDM experiments where movement is involved.

Conclusion

Context-dependent memory is a complex phenomena that utilizes many brain regions in order to make memory retrieval more efficient. There are 4 different subtypes of CDM, environmental CDM, cognitive CDM, state CDM, and mood CDM. They all are differentiated

by certain factors affecting the context of encoding and retrieval. The distinction in context (environmental, cognitive, state, and mood) leads to the involvement of different brain regions to enact the different types of CDM. This review was able to discover several variations in the involvement of different brain regions in different types of CDM (i.e., several brain regions that are "uniquely" involved in a subtype). Some of these variations were important to note with the parietal cortices and the prefrontal cortex. Though my review found many differential brain regions involved, it is expected that future work will uncover that there is more overlap than previously expected. For instance, the prefrontal cortex has not yet been reported to be involved in mood or cognitive CDM, which is surprising due to the PFC's involvement in cognitive functioning. The restriction of specific subtypes to certain brain regions may stem from the inclination of neuroscience research to adopt a region-of-interest approach. While this approach allows researchers to focus on the role of specific sub-regions, it poses challenges in identifying broader functions. By combining across multiple studies with this review, we can start to take a whole-brain approach.

CDM has been studied for several decades, but environmental CDM has been studied more in depth compared to the other types of CDM. Environmental is the easiest CDM to explore due to the flexibility of testing in different settings; meanwhile cognitive, state, and mood CDM are not able to be manipulated as easily as the environment. Finding different environments to investigate is an easy feat, meanwhile manipulating a person's mood, or cognitive, or mental state is a lot more difficult to achieve experimentally. Currently, research regarding state CDM has only been able to manipulate a person's state by intoxicating them, while cognitive CDM has only been experimented with using different languages for the memory tests. In order for our understanding of CDM to progress, we need to find more complex and diverse ways to study these types of CDM.

Works Cited

Akiko Matsumoto & Claudia Stanny (2006) Language-dependent access to autobiographical memory in Japanese-English bilinguals and US monolinguals, Memory, 14:3, 378-390, DOI: 10.1080/09658210500365763

Cabeza, R., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2008). The parietal cortex and episodic memory: an attentional account. *Nature reviews. Neuroscience*, *9*(8), 613–625. https://doi.org/10.1038/nrn2459

Eich, E., & Macaulay, D. (2000). Are real moods required to reveal mood-congruent and mood-dependent memory?. *Psychological science*, *11*(3), 244–248. https:/.doi.org/10.1111/1467-9280.00249

Eich, E., & Metcalfe, J. (1989). Mood dependent memory for internal versus external events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(3), 443–455. https://doi.org/10.1037/0278-7393.15.3.443

Eich, J. E., Weingartner, H., Stillman, R. C., & Gillin, J. C. (1975). State-dependent accessibility of retrieval cues in the retention of a categorized list. *Journal of Verbal Learning & Verbal Behavior, 14*(4), 408–417. https://doi.org/10.1016/S0022-5371(75)80020-X

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. British Journal of Psychology, 66(3), 325–331. https://doi.org/10.1111/j.2044-8295.1975.tb01468.x

Goodwin, D. W., Crane, J. B., & Guze, S. B. (1969). Alcoholic "blackouts": A review and clinical study of 100 alcoholics. *The American Journal of Psychiatry, 126*(2), 191–198. https://doi.org/10.1176/ajp.126.2.191

Marian, V., & Neisser, U. (2000). Language-dependent recall of autobiographical memories. Journal of Experimental Psychology: General, 129(3), 361–368. https://doi.org/10.1037/0096-3445.129.3.361

Robinson, J. A. (1976). Sampling autobiographical memory. *Cognitive Psychology, 8*(4), 578–595. https://doi.org/10.1016/0010-0285(76)90020-7

Smith, S.M., Vela, E. Environmental context-dependent memory: A review and meta-analysis.*Psychonomic Bulletin & Review* **8**, 203–220 (2001). https://doi.org/10.3758/BF03196157

Zhang, W., van Ast, V. A., Klumpers, F., Roelofs, K., & Hermans, E. J. (2018). Memory Contextualization: The Role of Prefrontal Cortex in Functional Integration across Item and Context Representational Regions. *Journal of cognitive neuroscience*, *30*(4), 579–593. https://doi.org/10.1162/jocn_a_01218

Overexpression of ATP Synthase in E.Coli By Jude Thomas, Meta Foster, and David Crowley

Abstract

Rising shortages of food are becoming a concern due to climate change. Conventional agriculture methods will not sustain the increasing population, and more worryingly, climate change will wreak havoc on the agriculture industries. As ATP Synthase is the enzyme that converts food(glucose) into energy(ATP), enhancing the protein could allow humans to produce more ATP with less food, leading to a decrease in shortages and other problems. Thus, the purpose of this study is to investigate if overexpression of ATP Synthase in E. coli cells could lead to increased energy production with the same amount of food. This is because an increase in the expression of those genes could create more ATP Synthase proteins. This would allow the body to harvest energy faster, and more effectively with increased ATP Synthase. Using the DNA sequences from the National Library of Medicine, a plasmid was created with ligation. The plasmid was then transformed into E.coli cells on plates, and controls without the plasmid, and with the plasmid but without the gene were created. The plate without the gene but with the plasmid showed the most growth, while plates with subunit B and subunit beta showed no evidence of any recombinant cells that accepted the subunits. Therefore, overexpression of ATP Synthase subunits B and Beta provided no positive visible effect on the growth rates of the cell and hampered the cells as transformants with the subunits were not seen on the plates.

Introduction

With the advent of COVID-19 and the economic downturn, food shortages are increasingly common. Countries such as Djibouti, Honduras, and Sierra Leone are bearing the brunt of these shortages, and around the world, more than 345 million people face food insecurity. In the ongoing crisis for sustainability, alternative food sources have been brought up as a solution to the problem. Crickets (Murugu) and other exoskeletal organisms have been considered viable options, however, rarely has one considered gaining energy straight from the source. Adenosine Triphosphate(ATP), known as the molecular unit of currency provides the energy for humans to conduct their bodily functions(Dunn). Cellular Respiration is the process by which the body converts oxygen and (2*R*,3*S*,4*R*,5*R*)-2,3,4,5,6-Pentahydroxyhexana l(Glucose) into approximately 32-40 ATP molecules(Chaudhry). However, processes in cellular respiration consume 2 ATP to produce that large amount of ATP molecules, which could theoretically be improved. ATP Synthase, the enzyme that produces ATP molecules, is composed of two regions F1 and F0, which altogether have 9 subunits(Junge). The Beta subunit is found in the F1 complex and contains the active sites where ADP and phosphate are synthesized into ATP. The b subunit is found in the F0 complex and provides the structural foundation for the protein allowing for rotation and proton translocation(Junge). Overexpressing the Beta subunit of ATP and the b subunit of ATP could allow for more H+ to flow down the concentration gradient and join more ADP and phosphate thereby creating more energy. This is because of the addition of

more ATP Synthase proteins being produced by the E. coli cell. The cell will use resources to continue expressing ATP Synthase for an increased quantity than in normal cells (Lehner). However, the cell will likely make these resources back up with the increase in ATP production. The benefit of the findings in E.coli, if found possible, could apply to humans and allow future researchers to solve the sustainability crisis.

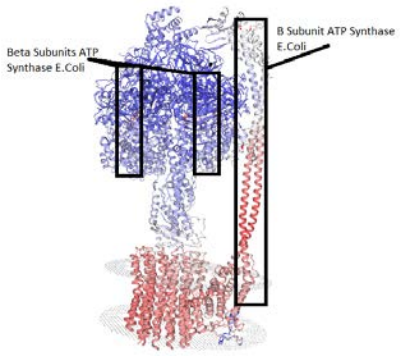


Figure 1. ATP Synthase tertiary structure with selected subunits- B, and Beta highlighted as to their location

Materials and Methods

Primers were designed using the NCBI Primer Design tool to complement the DNA sequences received from the National Library of Medicine. The forward primer for the beta subunit was as follows 5'-GACTACGGATCC**ATGGCTACTGGAAAGATTGT**-3' while the reverse primer was 5'-TAGCATTTCGAATTAAAGTTTTTTGGCTTTTTCCAC-3'. The primers for the b subunit were as follows with the forward primer first, 5' -GACTACGGATCC**GTGAATCTTAACGCAACAAT-3',** and **5'-TAGCAT**TTCGAA**TTACAGTTCAGCGACAAGTT-3'** for the reverse primer. (Both primers for the Beta Subunit had between 30-50% GC and TM levels that are 72℃ and 68℃ respectively. The primers for the B subunit had between 35-45% GC level and TM levels that were 71℃ and 69℃ respectively). The cloning plasmid used was Puc-19. Running a PCR to amplify the DNA of each subunit separately, Fig 2. represents the sequence of temperatures added to the thermocycler. A digest was run on gel, and then the DNA vector was ligated into the puc-19 plasmid, which was transformed. The preparation mixture for PCR was created using 25µl Promega's PCR Master Mix along with 5µl of the primers and 8µl of the DNA sequence and 12µl of water. Tubes were labeled and put into the thermocycler. Annealing and Denaturation temperatures were set to those shown in Fig 2.

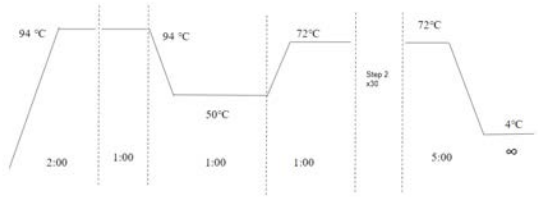


Figure 2. Annealing and Denaturation Temperatures set in the Thermocycler

The purification was done using the sigma NA 1020 GenElute PCR Clean-Up Kit. The ligation was done using 10µl of 2x buffer, 1µl of Quick Ligase from NEB, 1µl of puc19, 2 µl of the b or beta gene, and 6 µl of water. A tube of NEB 5-alpha Competent E.coli cells on ice was thawed out and mixed, then 50µl of cells were pipetted into a transformation tube on ice. 1-5µl containing 100pg to 1µg of plasmid DNA was added to the cell mixture, the tube was flicked a few times to mix cells and DNA. The mixture was placed on ice for 30 mins, and after, heat shocked at exactly 42℃ for exactly 30 seconds. The mixture was placed on the ice for 5 mins, and then 950µl of room temperature SOC was pipetted into the mixture. The mixture was placed at 37℃ for 60 mins and rotated at 250 rpm. The plates were warmed to 37 ℃ , and the cells were mixed thoroughly by flicking the tube and inverting, then many 10-fold serial dilutions were performed. The dilution was spread 50-100µl onto an amp/x-gal selection plate and incubated overnight at 37℃. This was done for every plasmid and subunit mixture.

Results

Figure 4b and Figure 4c are the x-gal/amp selection plates with the b and Beta subunits overexpressed into E.coli. Although the plasmid was ligated into the E.coli cells and contained either the b subunit or Beta subunit, there were no cell colonies that contained white pigmentation. There were however, 2 blue colonies on Beta x-gal/amp plate, and 14 blue colonies on the b x-gal plate. Specifically, for the plasmid puc 19, recombinants are characterized by their white pigment while normal colonies have a blue pigmentation. Figures 3, 4d, and 4e are controls meant to identify if the transformation procedure was performed successfully, and if any contamination occurred. Due to the sizable population of blue colonies in Figure 3d, the transformation of the plasmid mostly occurred successfully, while the fact no colonies appeared in Figure 3, shows that no contamination occurred in the x-gal/amp plates. Figure 3e is the E.coli cells that were inserted with plasmid puc-19 that were cut but genes were not inserted.



Figure 3. Control(E.coli cells alone without transformation)

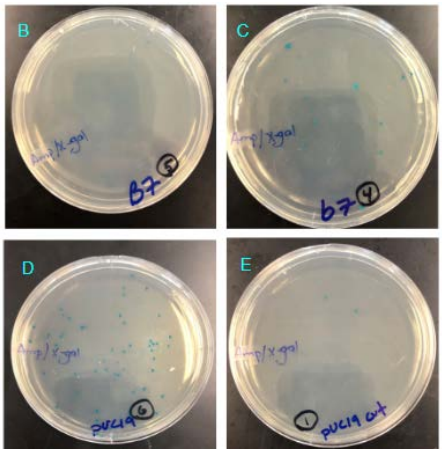


Figure 4. Experimental Trials showing the x-gal/amp plates with control and experimental results

Discussion

Based on the results above, the most likely explanation was that the overexpression of subunits b and beta did not provide increased subunit yield and instead negatively affected the cell's ATP Synthase enzyme. The overexpression of genes likely "poisoned" the cell, by depleting the cell resources, and using those for the overexpression instead of required cellular functions. According to Lenher, "high concentrations of enzymes and other proteins can harm cells in several ways, for example by activating or overloading specific biological pathways, disrupting regulation, or by aggregating together". Thus, the same process most likely occurred with the transformed cells. In a similar study with the goal of overexpressing ATP Synthase subunit Epsilon, the results identified that overexpressing ATP Synthase led to a decrease in E-cadherin, a protein important in slowing down cancer progression. Although no white colonies were found on the plates, several potential additions could allow for fruitful experiments. Instead of just using one subunit in each plate, all the subunits of both the F1 subunit and F0 subunit could be ligated together into the E.coli cell. This would provide the cell with ATP Synthase with the additional structural and functional components, allowing ATP to be overexpressed in the fully functional utility. Additionally, a different plasmid could have been used for the ligation such as pet 31b or pSCU. Another change to the experiment would be to change the organism from E.coli, as bacterial structure is different from mammalian structure. A mammalian cell

could allow for easier transfer to humans if the experiment proves successful. Despite the results of the experiment, overexpression of ATP should not be discounted as a way of gaining additional ATP from the cell.

Works Cited

Chaudhry R, Varacallo M. Biochemistry, Glycolysis. 2022 Aug 15. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan–. PMID: 29493928.

Dunn J, Grider MH. Physiology, Adenosine Triphosphate. 2023 Feb 13. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan–. PMID: 31985968.

Junge, Wolfgang. "ATP Synthase and Other Motor Proteins." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 9, 1999, pp. 4735–37. *JSTOR*, http://www.jstor.org/stable/47289. Accessed 7 June 2023.

Murugu DK, Onyango AN, Ndiritu AK, Osuga IM, Xavier C, Nakimbugwe D, Tanga CM. From Farm to Fork: Crickets as Alternative Source of Protein, Minerals, and Vitamins. Front Nutr. 2021 Aug 10;8:704002. doi: 10.3389/fnut.2021.704002. PMID: 34447775; PMCID: PMC8382788.

Benedetta Bolognesi, Ben Lehner (2018) Protein Overexpression: Reaching the limit eLife 7:e39804 https://doi.org/

Huang, Yan-Jiun et al. “ATP Synthase Subunit Epsilon Overexpression Promotes Metastasis by Modulating AMPK Signaling to Induce Epithelial-to-Mesenchymal Transition and Is a Poor Prognostic Marker in Colorectal Cancer Patients.” *Journal of clinical medicine* vol. 8,7 1070. 21 Jul. 2019, doi:10.3390/jcm8071070

Why People Obey the Law: A Critical Review of Existing Research by Makayla P. Wu

Abstract

Understanding the various reasons why people obey the law can benefit policymakers and promote a more law-abiding society. Reasons are separated into internal and external factors. Internal factors largely rely on an individual's personal beliefs. The paper discusses existing research surrounding three main internal factors: perceived legitimacy, morality, and legal socialization. These internal factors can interact with external factors such as nudging and deterrence to result in law-following. Most research suggests that internal factors are more significant compared to external factors when used to predict lawfulness. However, deterrence-based policies have shown to be effective in deterring some types of crime, and more research is needed to understand which factor is most significant for different areas of crime.

Introduction

From elementary school rules to the Constitution and all the rules and laws in between, they are present to prevent harm by regulating citizens. Why people obey the law is a complex topic that is a result of internal and external factors. Understanding how different factors interact to lead to rule-following can help authorities and society together create a more cohesive population.

Most of the research is either on the macro level or empirical, since this topic relates to society as a whole and it is impossible to control the environment to only study certain factors. The following research mentioned are all published within the last 30 years and remain relevant to today's society. Internal factors include perceived legitimacy and procedural fairness, morality, and legal socialization. External factors are factors that come from authority or other external figures from society that cause people to follow the laws. These include nudging and deterrence. These factors will be further discussed and evaluated in the following paper.

1. Internal factors

Internal factors are intrinsic motivations to follow the law, relying on an individual's attitude and morals. The most influential factors include perceived legitimacy, and morality shaped through legal socialization.

A. Legitimacy and procedural fairness

Legitimacy is established when citizens recognize the right of authorities, such as law enforcement, to exert power and accept it (Jackson, 2020). Citizens choose to follow institutions with legitimacy because they regard the institution's authority as appropriate and just (Tyler & Darley, 2000). To create a legitimate authority, it is necessary to be consistent and fair to its community members so there is a feeling of trust and respect.

Legitimacy in the justice system has been linked with citizens' perceived sense of just punishment and established truth, but most importantly, citizens' perception of fairness and whether or not it is established (Sevier & Tyler, 2014). By courts achieving what they were established to do, a sense of legitimacy and procedural fairness ensues. Procedural fairness builds legitimacy, which is important in achieving compliance and respect, as well as having authority's power justified by the population (Tyler & Jackson, 2013). In other words, procedural fairness exists when authority treats people with respect and fairness, leading to legitimacy. In a system with legitimacy, people obey laws and authority because they think it's their duty as citizens to obey the authorities of that system.

Without legitimacy, the right and justice of authority is questioned, making enforcement difficult. Legal cynicism is when the law and the agents of its enforcement are viewed as illegitimate and incapable of correct functioning, which impedes collective efficacy (Kirk & Matsuda 2011). Legal cynicism is a cultural mechanism that has been positively associated with an increase in violence (Kirk & Papachristos, 2011). When a community views authority as unavailable to assist, they feel as if resolving the conflict themselves is the best option.

On the surface, legitimacy and procedural justice are why people follow the law. However, a person seeing law enforcement as a legitimate institution can be rooted in reasons such as personal involvement, interactions with enforcement, and a person's morals.

B. Morality (Moral alignment)

Moral alignment is a person's perceived right and wrong which affects their perception of the law. Morals can be developed through procedural fairness and representing community values (Jackson, 2012). Laws are meant to be a reflection of the collective morals of the community.

Authorities can shape morals by creating a front of consistency and fairness so that citizens perceive enforcement as morally upheld and respectable. Laws that align with a population's morals would be followed even if they were not written laws (Tyler & Darley, 2000). People establish their morals through their community and society as well as moral socialization. Through moral socialization, people learn what is considered right and wrong intermixed with their values (Tyler & Trinkner, 2017).

People's behaviors are influenced by their personal morals. However, enlisting community members to participate in creating laws can ensure that the law accurately reflects the morals of society. If there are discrepancies between a person's moral alignment and the law, compliance becomes difficult.

Diversity in the community makes it difficult to ensure that every member of the community identifies with the law morally. For example, some people believe the death penalty is morally wrong, while others believe that it is justified. With moral gray areas, enforcing a law that most members of the public do not see a moral issue with is very difficult, as breaking the law would be widespread since people feel no intrinsic motivation to obey. Law and morality

have a connection, and there is more law enforcement could do to tie laws back to their moral origins and make the community see how it is a reflection of their morals.

C. Legal socialization

Legal socialization is the process where people come to develop their understanding and attitude about laws and law enforcement (Tyler & Trinkner 2017). Largely formed through childhood or other early encounters with law enforcement, legal socialization is a fragile but necessary process for achieving perceived procedural fairness and legitimacy, and trust that leads to compliance with the law. Key phases in legal socialization include one's family, school, or community idea of the law.

Legal socialization is pivotal in law enforcement because it provides the foundation of how people perceive the law. Children learn their morals about what is right and what is wrong simultaneously as they see how authority figures interact with their community (Tyler & Trinkner, 2017). Poor legal socialization promotes distrust of law enforcement and dissonance between one's morals and the law.

It is important to promote legal socialization by allowing input and participation. When children are socialized to hold values of obedience towards authority figures, they are more likely to trust authority figures and abide by the law. Legal socialization incorporates one's values into one's identity.

Legal socialization appears both formally and informally, so it is important to ensure that it is a positive process. Legal socialization extends to judges to justly and swiftly carry out consequences (Tyler & Jackson, 2013). Though it can be difficult to ensure positive legal socialization experiences, more outreach to communities that have a high level of distrust can change their attitude towards the law and promote more compliance. Studies have shown that diversity in policing can improve interactions between law enforcement and community members (Ba et al., 2021). This diversity can lead to positive legal socialization and trust, then a perceived legitimacy.

2. External factors

While internal factors rely on a person's feeling of internal obligation to follow the law, external factors are reasons directly caused by the authority's actions. This can be through legislation, police presence, or subtle intervention.

A. Nudging as a way to promote law-following

Nudging is the practice of authorities consciously shaping the population in a particular direction. A significant part of nudging is choice architecture, the framing of choices that nudging can be presented through. Authorities shape people's behavior by using their own biases to influence themselves. It is important to note that the nudges are not forced but can be viewed as a type of manipulation (Sunstein, 2014). Nudges are initiatives that "affect people's behavior"

without significant consequences if rejecting the nudges or substantial gain for following the nudges (Sunstein, 2019).

It is generally difficult to object to nudges and choice architecture as a whole because one may not even be aware that they are being nudged in a certain direction (Sunstein, 2014). Points defending choice architecture and nudging include that "nature itself nudges" (Sunstein, 2015) which shows that some nudges are inevitable and natural. Both cannot be avoided and is pointless to eradicate it. Nudges can alter the choice architecture and are relevant to why people follow the law because it is an external factor that doesn't rely on a person's morals and instead relies on the outside factors that cause a person to follow the law.

Authorities can utilize nudges to promote law-following by altering choice architecture and framing the preferable outcome as more favorable. For example, a nudge in Germany that made people automatically enrolled in a green energy program but were still given to full option to opt out proved to be an effective nudge towards green energy (Thaler & Sunstein 2008). Nudges can steer people in a set path or direction, but people who feel the need can steer away from that path and reject the nudge.

B. Deterrence

Deterrence-based policies assume that humans are rational and are deterred from crime by the punishment that may be given. A relevant theory of crime control is the rational choice theory, which states that humans are rational and logically choose to commit crimes because the benefits outweigh the consequences. Classical deterrence theory states that policing and swift and severe punishment may effectively deter criminal acts due to reduced motivation for gain and attractiveness of crime (MacCoun, 1993).

Deterrence strategies are relevant to why people follow the law. The strategy in itself is to scare or deter people due to fear of punishment into following the law. Sometimes, even the presence of authority is enough to deter crimes (Cook, 2012). This creates an external motivation for following the law, largely based on fear or avoidance of punishment.

Systems can use deterrence-based policies by increasing punishment and policing. When the punishment is certain, it can produce substantial deterrent effects. With specific criminal acts that may be on the rise, deterrence-based policies may help control significantly. Jurisdictions prone to gang violence, overt drug markets, and repeat offenders have been shown to be reduced with focused deterrence strategies (Braga & Weisburd, 2012). There is a statistically significant association between imprisonment rates and crime rates (Cook, 2012), which shows that increased incarceration does deter crime to an extent.

Deterrence-based policies may be more harmful and less effective compared to other strategies. Though a court appearance does have some deterrent effects, it allows for the stigmatizing of individual offenders (McGrath, 2009). Once someone is labeled as an offender, it leaves room for emotional reactions and reoffending. A law-abiding society should desire to follow the law, instead of being motivated by fear (Tyler & Darley, 2000). The deterrent theory builds off of rational choice theory, but people are simply not rule-based in all situations and

more often make decisions based on emotions (Clore, 2011). Many crimes are not premeditated, and even the ones that do are likely more focused on the probability and strategy for success and gain instead of researching the likely consequences. To be able to successfully deter crime through deterrence-based policies would ensure that the punishment is clearly applicable to a specific crime and publicized enough that people are aware.

Conclusion

Despite attempts to use external factors to motivate people to follow the law, those factors are not able to make as significant an impact on communities as internal reasons can. Internal reasons such as moral alignment, perceived legitimacy, and legal socialization take more consistency and time to build, however, it would be more effective and sustainable.

External factors such as using the deterrent theory to guide legislation can build bad relationships between the community and law enforcement. In addition, nudges are not as sustainable. Organizing a campaign to nudge people in a certain way takes up significant government resources and efforts, all for a small nudge that people may easily ignore.

Specific research on which factors have the most influence on a person's choice to disobey the law in different crimes could make an impact on creating legislation. For example, the majority of homicides are spontaneous. The deterrent theory would have a different impact on people who commit homicide and people who use illegal drugs. Research that differentiates the leading factors for disobeying the law for each crime could help authorities know which strategy to prioritize in creating legislation.

Acknowledgment

I would like to express my deepest appreciation to Dr. Carlos De la Rosa Xochitiotzi for guiding me through the entire process of this paper. This paper would not be possible without the resources and support given to me through Indigo Research. I attest that the ideas, graphics, and writing in this paper are entirely my own

Works Cited

Ba, Bocar, et al. "The role of officer race and gender in police-civilian interactions in Chicago." *Science*, vol. 371, no 6530, 2021, pp 696-702. https://doi.org/10.1126/science.abd8694. Accessed September 1, 2023.

Braga, Anthony A., and Weisburd, David L.. "The Effects of Focused Deterrence Strategies on Crime: A Systematic Review and Meta-Analysis of the Empirical Evidence." *Journal of Research in Crime and Delinquency*, vol. 49, no. 3, 2012, pp. 323-358. *Sage Journals,* https://doi.org/10.1177/0022427811419368. Accessed March 3, 2023.

Clore, Gerald L.. "Psychology and the Rationality of Emotion." *Modern Theology,* vol. 27, no. 2, 2012, pp. 325-338. *Wiley Online Library,* https://doi.org/10.1111/j.1468-0025.2010.01679.x. Accessed Augst 27, 2023.

Cook, Justin, et al. "The Deterrent Effect of Imprisonment." *Controlling crime: Strategies and tradeoffs,* edited by P. J. Cook, J. Ludwig, J. McCrary, University of Chicago Press, 2012, Pages 43–94.

Jackson, Jonathan, et al. "Why do People Comply with the Law? Legitimacy and the Influence of Legal Institutions." *The British Journal of Criminology*, vol. 52, no. 6, 2012, pp. 1051–1071. *Oxford Academic,* https://doi.org/10.1093/bjc/azs032. Accessed March 2, 2023.

Kirk, David S., and Mauri Matsuda. "Legal Cynicism, collective efficacy, and the ecology of arrest." *American Society of Criminology*, vol. 43, no. 2, 2011. *Wiley Online Library,* https://doi.org/10.1111/j.1745-9125.2011.00226.x. Accessed August 24, 2023.

Kirk, David S., and Andrew V. Papachristos. "Cultural Mechanisms and the Persistence of Neighborhood Violence." *American Journal of Sociology*, vol. 116, no. 4, 2011, pp. 1190-1233. *The University of Chicago Press,* http://doi.org/10.1086/655754. Accessed August 15, 2023.

MacCoun, Robert J.. "Drugs and the law: a psychological analysis of drug prohibition." *Psychological Bulletin*, vol. 113, no. 3, 1993, pp. 497-512. *American Psychological Association,* https://doi.org/10.1037/0033-2909.113.3.497. Accessed June 14, 2023.

McGrath, Andrew. "Offenders' Perceptions of the Sentencing Process: A Study of Deterrence and Stigmatisation in the New South Wales Children's Court." *The Australian and New Zealand Journal of Criminology*, vol. 42, no. 1, 2009, pp. 24-46, *Sage Journals,* https://doi.org/10.1375/acri.42.1.24. Accessed May 26, 2023.

Sevier, Justin., and Tom R. Tyler. "How Do the Courts Create Popular Legitimacy?: The Role of Establishing the Truth, Punishing Justly, and/or Acting Through Just Procedures." *Albany Law Review*, vol. 77, no. 3, 2014, pp. 1095-1137. https://ir.law.fsu.edu/articles/123. Accessed April 14, 2023

Sunstein, Cass. R.. "Behaviorally Informed Policy: A Brisk Progress Report." *SSRN Electronic Journal,* 2019. http://dx.doi.org/10.2139/ssrn.3461781. Accessed May 26, 2023.

Sunstein, Cass. R.. “The Ethics of Nudging.” *SSRN Electronic Journal,* 2014. http://dx.doi.org/10.2139/ssrn.2526341. Accessed May 26, 2023.

Tyler, Tom. R., and John M. Darley. “Building a Law-Abiding Society: Taking Public Views about Morality and the Legitimacy of Legal Authorities into Account When Formulating Substantive Law.” *Hofstra Law Review*, vol. 28, no. 3, 2000, pp. 707-741. http://scholarlycommons.law.hofstra.edu/hlr/vol28/iss3/5. Accessed April 14, 2023.

Thaler, Richard, and Cass R. Sunstein. *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press, 2008. https://doi.org/10.1017/S1474747209990175

Tyler, Tom. R., and Jonathan Jackson. “Popular Legitimacy and the Exercise of Legal Authority: Motivating Compliance, Cooperation, and Engagement.” *Psychology, Public Policy, and Law*, vol. 20, no. 1, 2013, pp. 78–95. *American Psychology Association,* https://doi.org/10.1037/a0034514. Accessed April 14, 2023.

Tyler, Tom. R., and Rick Trinkner. *Why children follow the rules: Legal socialization and the development of legitimacy.* Oxford University Press, 2020. https://doi.org/10.1093/acprof:oso/9780190644147.001.0001

An Investigation of the Labor Movements in the US and China By Jianing Chen

Abstract

The goal of this paper is to examine the historical development and the contemporary development of the conditions workers face in both the United States and China, the two biggest economies of the world. An analysis of the labor system in both countries will be provided as well as how the workers are currently navigating a post-pandemic world. This paper also seeks to look into the future and draw conclusions regarding what the future holds for Chinese and American workers.

Introduction

The labor movement has had a long and controversial history globally. Since the start of industrialization and since the first factories were opened in Europe, workers have always used different tactics to prevent exploitation. They often fought for their interests through forming unions, associations of workers that push for better working conditions, high wages, or simply improved treatment. When necessary, union representatives would negotiate new contracts with the employer and present them with demands or grievances. This tactic of collective bargaining has allowed workers to gain an edge over their employers. When absolutely necessary, meaning that negotiations are off the table, unions lead workers on strike, a display of dissatisfaction with the employer through the stoppage of work. Strikes are often successful and workers are able to get the terms they bargained for; however, there is still a huge possibility that the employer might forcefully stop the strike or force a compromise between the parties. It is safe to say that unions and strikes have been key to the labor movement since the very beginning.

History of Unions and Strikes in the United States

Unions and strikes have had a long history in America. Ever since colonial times, workers have stopped their work to protest unfair treatment (Kiger, 2023). Craftsmen, artisans, and merchants also formed an array of associations to protect their own interest and avoid getting taken advantage of. As a collective, members decided on the prices of goods and devised plans to decrease competition. Though not truly unions yet, these were likely the early form. When the industrial revolution hit America in the mid-1800s, numerous factories were built up and millions of Americans were attracted by this new lifestyle. These newcomers rushed to urban areas where big corporations provided an unprecedented amount of opportunities unheard of in the rural areas.

Despite the clear benefits of finding work in factories, there were also clear downsides, such as unregulated working hours, horrid working conditions, and low wages barely enough to support a family. These factors led to some of the biggest strikes in history, such as the Great Southwest Railroad Strike of 1886, which involved 200,000 railroad workers who were dissatisfied with the working conditions and long hours (Investopedia, 2022). A strike of even larger magnitude took place in 1894; In the Pullman Strike, 250,000 factory workers under the

leadership of the American Railway Union protested against long hours and low pay (Investopedia, 2022).

These strikes would only grow in magnitude in the 20th century, with strikers doubling in the United Mine Workers Strike of America of 1946. In some cases, the workers were able to obtain solid results; in other cases, the employers used violent tactics to fight against the gathering strikers. Nevertheless, it is safe to say that the tactic of striking has always had great influence in the labor movement and is responsible for achieving some "bread and butter" goals for workers. Despite the fact that unions have fought hard for worker's rights, union membership has been in decline for the last few decades, with the number of members peaking in the 50s (The White House, 2022). This was due to the US government and presidents suppressing the growth of unions such as the infamous 1981 PATCO strike, which resulted in President Ronald Reagan firing over 11,000 air traffic workers (Barera, 2021). Moreover, globalization and cheaper foreign labor further decreased labor union's bargaining abilities, causing the downward slope of membership for the past few decades.

Contemporary American Workers Rights and Labor Laws

After decades of the labor movement and through hard work of activists, the US has a relatively secure set of labor laws that protect employees from exploitation and harmful practices. This includes the right to protest and strike when faced with unfair wages, hours, and working conditions; in addition, workers also have the right to workers support and worker benefit. Employment laws also protect workers from discrimination based on gender, race, religion, sexual orientation, and national origin. Though American labor laws seemingly grant workers with a sufficient system of labor protection, a study done by Oxfam in 2023 shows that the United States ranks among the lowest in terms of worker treatment in developed countries (Henderson 2023). This is due to American corporate laws which puts big corporations above workers; essentially, due to capitalism and government's inability to protect workers, the US is behind in many areas in terms of worker's treatment when compared to other developed countries such as Australia and New Zealand. In areas such as wage policies, workers protection, and right to organize, the US overall ranks at the dead end among other developed countries.

Moreover, it appears that there are numerous loopholes American employers get to exploit in the current economy. Hiring part-time workers is a great strategy for employers to minimize their own responsibilities such as providing workers with fair benefits and protection. In addition, wage inconsistency is another huge issue facing part-time workers since they are commonly paid minimum wage and can be replaced easily. This means that workers are oftentimes unable to support themselves or their families; even worse for these employees, part-time jobs rarely provide social security or healthcare benefits. Under part-time occupation, workers are also unable to join unions, meaning they cannot organize strikes against their employers to battle for better treatment. When employees complain about these conditions, they are simply replaced by someone who is willing to withstand such treatments.

Labor laws also face the problem of loose enforcement. The National Labor Relations Board, though claims to be an agency that protects worker's rights, has diminished many groups' ability to unionize under the Trump administration. An event that demonstrates this is when the NLRB identified Uber drivers as independent contractors instead of a collective of employees hired by Uber; this successfully stopped the drivers from the possibility of using collective bargaining to obtain gains, begging the question- does government agency truly protect disadvantaged workers or do they favor big corporations?

Recent Strikes and What They Mean for the Future of the American Labor Movement

Despite decades of decline of unions, it seems like in the post-pandemic era, strikes have appeared more often on the news than ever. Unions have been making a comeback as well. For the past few months, the Hollywood actor and writer strike has taken over social media sites such as Youtube, Twitter, and TikTok. Thousands of people working in the entertainment industry and allies are currently protesting in the streets of Los Angeles with the majority of them being members of SAG-AFRA, an American actors union. It has been reported that approximately 11,500 writers and 160,000 actors have been on strike against Hollywood production companies starting May of this year (Stewart, 2023). These protests are happening due to shifts occurring post-pandemic, including shorter TV seasons, potential usage of AI technology, and larger work gaps. All of these factors are essentially taking jobs away from actors and writers as they are being paid less and less, while the production companies are making billions of dollars. This is yet another classic case of corporate greed getting out of control and mistreating employees to gain more revenue.

The actor and writer strike is a representation of the larger labor movement happening in the post-Covid world. As the workers navigate an unstable workplace, they face numerous challenges such as rising inflation and increasing cost of living. It is clear as day that the pandemic has completely reshaped the system which has sustained itself for decades since the development of globalization. According to recent statistics, around 1.1 million American workers have been taken out of the workforce due to Covid (Meadway, 2023). More Americans are also suffering from the long-term effects of the virus and unable to work. Due to experiencing such turmoil both physically and psychologically, it is not surprising that modern workers are demanding their employers to improve the pay and working conditions. The unstable global market, shrinking working population, and a heightened sense of awareness are all factors that give power to the workers.

These strikes are a huge indicator that we are living in the age where the workers are finally realizing the value they bring to the table. Moreover, the new generation, namely Gen Z, is questioning and doubting the ideology of capitalism as a core operation system for the US. Though capitalism has gotten America to where it is today, there appears to be a shift in general opinion of the ideology; a large group of the new generation on the left believes that capitalism exploits the workers and only puts money in the pocket of the wealthy. This movement of awareness is happening all across popular social media platforms, especially TikTok, where the

dissatisfied youth show how the rising cost of living and housing has impacted them negatively. More individuals are becoming aware of corporate greed, which is exemplified by the big production companies SAG-AFRA members are striking against. The population is becoming more informed than ever.

History of Unions and Strikes in China

When it comes to unions and strikes, China does not have a long history to offer. While Europe and North America were industrializing rapidly in the 19th and 20th century, China remained largely agrarian, meaning that the most common occupation at that time was farming. Before the forming of "New China," there are few records of unions fighting for worker's rights in China due to its limited industrialization. However, there are depictions of such strikes in popular media; in the drama "Thunderstorm" (Lei Yu) adapted from the famous play written by Chinese dramatist Cao Yu. The drama depicts 1920s /30s China before the Japanese invasion and displays a worker's strike at a corporate factory; this proves that scattered and small-scale strikes did occur before the rapid industrialization under the Communist Party. A great example is the labor unrest in 1930s Shanghai, which happened due to economic depression and high living cost (Lin, 2021). Strikes of such magnitude would become rare in China as the economy recovered and employers accepted more of workers' demands.

During the Japanese invasion, production and manufacturing had stopped at large. From the start of Japanese occupation to roughly 1945, factories were either abandoned by their owners or taken over by Japanese forces. True industrialization would take place in the 1950s under the leadership of the Communist Party. However, strikes actually became less common in the period of rapid industrial growth because most factories were state-owned and state-ran; under this restrictive regime, protests, riots, and strikes are usually put down by force, explaining why we see so little uprisings in China. In addition, these events are rarely reported in the media due to Chinese censorship. Thus, a vast majority of the population are unaware of labor unrest or strikes unless it is happening in close proximity.

The Flawed Labor System in Contemporary China

Since the 1990s, China has introduced numerous legislations that determined the rights of the workers, the employers, and the balance between the two groups. This occurred as China moved towards a market-economy and joined the global economy. As China reached its peak in passing and amending labor laws in the 2000s, a problem had become apparent, and there is a perfect example to illustrate this (CLB, 2023). Out of all the legislations that were passed up until that point, the 2008 Labor Contract Law became the most polarizing as it guaranteed the worker's right to a detailed employment contract and a severance fee when they are laid off by the employer. This sounded like great news for Chinese workers because it protected them from employer misconduct; however, it turned out that the employers had found loopholes that allowed them to avoid the law by hiring new workers and firing old workers. Even with the amendment of the law in 2013, the issue was still not solved. Employers could always find a way

to exploit loopholes and squeeze as much profit as possible, showing the lack of enforcement of labor laws.

The current union system in place in China is flawed as well and cannot bargain for its workers effectively. Compared to labor unions in the US, which function as independent bodies, unions in China have considerably less influence over the labor system. A main reason for this phenomenon is that all trade unions in China are regulated under the All-China Federation of Trade Unions (ACFTU), meaning that independent unions are seen as illegal. The ACFTU contains a hierarchy with provincial unions at the top and grassroot unions at the bottom. Several problems arise from this situation; when a dispute arises, a local union would have to consult a high level body instead of advocating for the best interests of the employees. In addition, most of these local or grassroot unions do not yield much influence and are therefore inadequate when the workers need protection from unfair labor practices. In the end, corporations will continue to exploit this inadequacy of the ACFTU to its advantage. It is safe to say that the ACFTU serves the government rather than the workers that it claims to represent. Though Party leaders have for years asked the ACFTU to address its inefficiency and inability to properly represent the workers, there is simply more said than done.

Recent Strikes in China and Their Significance

The recent strikes in China are centered around the manufacturing sector, which is ironic because it has been the most buzzing sector economically for the past two decades due to foreign companies exploiting China's loose labor laws to make more profit. However, changes are slowly transforming this sector as manufacturing shifts globally. Recently, large parts of international manufacturing have been moved from China to South Asian countries, where the cost of labor is even lower; this leaves many manufacturing workers in despair due to delayed wages and closed factories. This slowing of international trade has led to a sharp increase in the number of strikes in 2023, with experts predicting a doubling of the number from 2022 (Zhou, 2023). The heightened strikes can be reflected by the videos of thousands of dissatisfied workers protesting in front of factories , which can be found on Doujin, a popular Chinese social media app. In the post-Covid era, Chinese workers face many of the similar challenges that American workers face; this includes the skyrocketing cost of living and stagnant wages. However, a positive trend that is occurring among Chinese workers is a similar sense of awareness as they become more willing to strike and protest unfair labor practices. Utilizing the power of collective bargaining and social media has allowed some success and led to compromises for the Chinese workers.

The Future for Workers in US and in China: The Awakening

There is still a long way to go for both American and Chinese workers. Recent strikes in both countries demonstrate the dissatisfaction of workers and the urgent need for reform. The bad news is that the instability in the global economy as well as the rising cost of living do not seem to end any time soon. People are struggling to make ends meet while working full-time

jobs due to the stagnant wages in both countries; it will be almost impossible for the new generation of workers to afford houses like the older generations; employers are able to exploit workers without consequences due to the loopholes in the system. However, the good news is that more people are realizing that they themselves are not to blame; the root of the issue is the system in which individual workers have barely any control over. Corporate greed and governmental corruption leave the workers as the ultimate victims. Workers are also becoming aware of the power of collective bargaining and their edge over the employers in the current economy; both the US and China have been seeing a surge in strikes along with an increase of American union membership. Chinese workers have also learned to use social media to document the injustice they face and advocate for worker's rights. More people are waking up to the reality that they are simply being treated unfairly and deserve more. Ultimately, the workers of both countries are on the right path to taking back their power, though it will likely take decades.

Works Cited

Barera, Michael. “The 1981 PATCO Strike.” *University of Texas at Arlington Libraries*, UTA Libraries , 2 Sept. 2021, libraries.uta.edu/news-events/blog/1981-patco-strike.

Henderson, Kaitlyn. “Where Hard Work Doesn’t Pay off - Country Labor Index 2023.” *Oxfam America*, Oxfam, 3 May 2023, www.oxfamamerica.org/explore/issues/economic-justice/country-labor-index-map/.

Kiger , Patrick. “10 Major Labor Strikes Throughout US History.” *History.Com*, A&E Television Networks, 3 May 2023, www.history.com/news/strikes-labor-movement.

Lin, Lisa. “Booms and Strikes in 1930s Shanghai .” *Economic History*, LSE Economic History Blog, 8 Oct. 2021, blogs.lse.ac.uk/economichistory/2021/10/08/booms-and-strikes-in-1930s-shanghai/.

Meadway, James. “Covid Reshaped the Global Labour Market – These Strikes Are Just the Start.” *openDemocracy*, openDemocracy , 17 Feb. 2023, www.opendemocracy.net/en/oureconomy/covid-strikes-workers-power-change-global-labour-market/.

Stewart, Emily. “It’s Not Just Famous Actors and Big-Name Writers the Hollywood Strikes Are Hurting.” *Vox*, Vox, 22 Aug. 2023, www.vox.com/money/2023/8/22/23840473/writers-strike-actors-wga-sag-workers-economy-impact.

Team, The Investopedia. “The 10 Biggest Strikes In U.S. History.” *Investopedia*, Investopedia, 26 Nov. 2022, www.investopedia.com/the-10-biggest-strikes-in-u-s-history-4773384.

“The State of Our Unions.” *The White House*, The United States Government, 5 Sept. 2022, www.whitehouse.gov/cea/written-materials/2022/09/05/the-state-of-our-unions/.

“Workers’ Rights and Labour Relations in China.” *China Labour Bulletin*, China Labour Bulletin , 10 July 2023, clb.org.hk/en/content/workers%E2%80%99-rights-and-labour-relations-china.

Zhou, Marrian. “Factory Strikes Flare up in China as Economic Woes Deepen.” *Nikkei Asia*, Nikkei Asia, 28 Aug. 2023, asia.nikkei.com/Business/Business-trends/Factory-strikes-flare-up-in-China-as-economic-woes-deepen.

Modified A* Algorithm With Octile Distance By Eunchan Lee

Abstract

Path planning is a fundamental concept in logistics, involving finding optimal paths while avoiding obstacles. This paper presents a modification of the A* algorithm by incorporating the octile distance heuristic to enhance efficiency and reduce memory usage. The proposed algorithm introduces the octile distance, which incorporates both diagonal movement and horizontal and vertical movement. This heuristic enables early termination of search and reduces unnecessary exploration. Results from experiments using benchmark datasets are presented, showing that the modified A* algorithm consistently demonstrates improvements in memory usage and search efficiency across different map types and sizes compared to traditional A* algorithm. Performance is evaluated based on the number of iterations required to solve the problem. These findings demonstrate that the modified approach substantially improves memory efficiency compared to the traditional A* algorithm. This research contributes to the field of path planning by addressing the limitations of the traditional A* algorithm and providing a more efficient alternative for various applications in logistics.

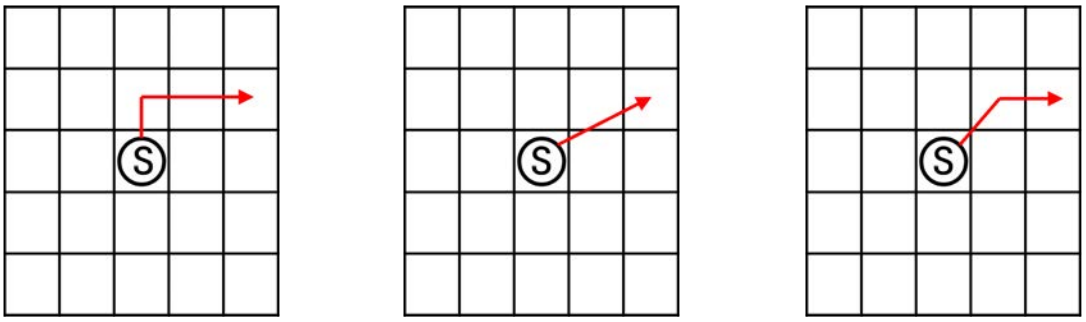
1. Introduction

Path planning is a fundamental concept in logistics that involves finding an optimal path from a starting point to a target point in a given environment. The primary objective of path planning is to compute a path that satisfies certain criteria, such as minimizing distance, and time, while avoiding collisions with obstacles (Karur et al. 2). Because of its ability to efficiently navigate through complex environments, path planning has been applied in various fields such as industrial robots, aerial drones, and artificial intelligence (Duchoň et al. 2). In the field of logistics, path planning plays a crucial role in route optimization and risk management. For example, for mobile robots, path planning helps them to navigate smoothly in a congested setting and the algorithm takes account of the robot's computation capabilities, the setting's kinematics and dynamics, and information availability (Karur et al. 2). Path planning is also an essential technique in game development. In games, non-player characters, also known as NPCs, need to navigate around the game world, and path planning algorithms help them plan routes to avoid obstacles and reach destinations (Nieuwenhuisen et al. 1).

Path planning algorithms typically take into account the environment's characteristics, such as the geometry of obstacles, their positions, and the robot's physical constraints and capabilities. A widely used path planning algorithm is the A* algorithm (Hart et al. 3). However, one drawback of the A* algorithm is that it utilizes too much memory space. The reason for this is that the A* algorithm maintains a record of visited states in a data structure known as the open list, and this can lead to extremely high memory usage in complex and large environments (Duchoň et al. 3). In scenarios with complex environments, the algorithm's memory requirements can be a significant limitation. To address this limitation, they conducted various different types of research about extensions of A* such as the Basic Theta* and Phi*. In that research, the problem they found with the basic A* algorithm was that it could only search for

paths relying on the links between the closest possible cells. However, this could slow down the algorithm as a lot of open space might exist between the connected cells and those cells might not be in zigzag style. Therefore, they came up with the Basic Theta* and Phi* algorithms that search the cells at every angle. This allowed their algorithm to only search for the cells that robots can pass. Furthermore, another alternative method to A* algorithm was developed in one research (Warren 3). In that research, the problem with the A* algorithm being extremely slow with a large number of cells was acknowledged. Therefore, the research utilized trial vectors that span several cells and performed a loose search on the fine grid. The research found out that the search algorithm got drastically faster, with a solution time of under 100 milliseconds for three-dimensional problems.

When searching for cells, the A* algorithm utilizes the *g*-value, which represents the current cell's distance or time from the starting point, and the *h*-value, heuristic value, which represents the estimated distance or time from the current cell to the target point. In the traditional A* algorithm, it used the Manhattan distance(Paul 1) and Euclidean distance in order to calculate the *h*-value (Hart et al. 3). However, in a grid environment, the calculated *h*-value differs from the actual value. Thus, we sought to use the octile distance. Octile distance is another heuristic distance used in path planning algorithms that provides an estimation of the distance between cells by considering both horizontal, vertical, and diagonal movements. It divides the plane into eight octants and considers the different combinations of horizontal, vertical, and diagonal movement. Each distance metric is shown in Fig 1. By utilizing the octile distance, the algorithm will check if it is possible to reach the target point in one octile distance, and by doing so, the algorithm will end early and be faster and more efficient.



(a) Manhattan (b) Euclidean (c) Octile

Fig 1: Different distance metrics

2. Methodology

Path planning is a fundamental concept in logistics, which entails identifying a viable route for a mobile entity such as a robot or vehicle to move from an initial point to a desired destination while avoiding obstacles and complying with various constraints. The ultimate goal of path planning is to ascertain the most efficient path that minimizes specified costs.

A uniform grid environment, where all grids have equal sizes, was used to test the algorithm. Each grid is defined as a cell. An agent is able to move both horizontally and vertically as well as diagonally. If there is an obstacle in horizontal or vertical paths, the agent moves diagonally. If diagonally adjacent sides are both blocked, the agent cannot move diagonally and therefore moves horizontally or vertically. All obstacles' locations are known. In this experiment, speed and acceleration are not considered and only the length of the path is taken into account in the calculation. It means that there is no cost when the agent changes direction and the agent can move or stop with no delay.

To test the algorithm, we employed a uniform grid environment with equally-sized grids referred to as cells. Agents in this environment can move both horizontally and vertically, as well as diagonally. If there is an obstacle obstructing the horizontal or vertical path, the agent will opt to move diagonally. In the event that both cardinally adjacent sides connecting with cell in diagonal direction are blocked, the agent cannot move diagonally even if there is no obstacle in agent's diagonal direction. For example, north and east direction are blocked, the agent cannot move to northeast direction. Importantly, all obstacle locations are predetermined and known.

In this particular experiment, we do not take into consideration factors such as speed and acceleration; rather, we solely focus on calculating the length of the path. Consequently, there is no associated cost when the agent changes direction, and the agent can transition or halt without any delay.

In this section, we will describe the modification of the A* algorithm using octile distance. To begin, in a uniform grid map, we initialize a starting point, a target point, and obstacles. All information regarding path planning is readily available. Throughout the following discussion, we will refer to the points as cells.

We create two key data structures: an open list, and a closed list. The open list contains cells that have not been fully explored, while the closed list contains cells that have been fully explored. After this initialization, we insert the starting point into the open list. The iteration continues until the size of the open list reaches 0 or the algorithm's stopping criteria are met. During each iteration, the first cell in the open list becomes the current cell. If there are multiple cells in the open list, the current cell is selected using specific criteria, which will be described below. The current cell in the first iteration is, of course, the starting point.

Once the current cell is determined, it is removed from the open list and inserted into the closed list. Importantly, each cell can only exist in either the open list or the closed list, but not both. A search is then performed where the eight adjacent cells to the current cell are examined. However, if an adjacent cell cannot be reached from the current cell due to obstacles, that particular cell cannot be searched. In this context, the current cell is referred to as the parent, and the maximum of eight cells searched from the current cell are known as the parent's children. It is crucial to understand that in order to reach a child via the shortest distance, one must go through the parent. Each cell has only one parent.

When a child is created, it is assigned g-value and h-value. The g-value represents the actual distance from the starting point to the current cell, while the h-value signifies the

estimated distance from the current cell to the target point. The *g*-value for the starting point is set to 0, and the *h*-value for the target point is also set to 0. A straight path is assigned a value of 1, while a diagonal path is assigned a value of the square root of 2. The *g*-value for a child is calculated by adding the Euclidean distance between the child and the parent to the *g*-value of the parent. Specifically, the *g*-value for a child in the diagonal direction from the starting point is $\sqrt{2}$, and in the straight direction, it is 1.

After the search, a maximum of 8 adjacent cells from the current cell are added to the open list. During any iteration, if a child's location *(x, y)* matches that of any cell in the closed list, the child is removed. Similarly, if a child's location *(x, y)* matches that of any cell in the open list, a comparison is made to determine which is the better parent for that child. As previously mentioned, a child can have only one parent. We refer to the same cell in the open list as the previous cell. If the *g*-value of the previous cell is less than or equal to that of the child, the child is removed. Conversely, if the previous cell's *g*-value is greater, the previous cell's parent is updated to be the child's parent, and its *g*-value is adjusted to match the child's. A smaller *g*-value indicates that the cell can be reached via the shortest path from the starting point to that cell.

After adding children to the open list, the algorithm proceeds to determine if it is feasible to reach the target point using octile movement from the children. A successful move implies that there are no obstacles along the octile path from the children's cells to the target point. To evaluate this, we calculate the absolute difference between the child's x and y location and the target point's x and y location, which we refer to as *x-diff* and *y-diff*. Octile movement can be divided into two categories: 1) When *x-diff* is not the same as *y-diff*: This category has six distinct shapes. 2) When *x-diff* is the same as *y-diff*: This category has one shape. Each shape is depicted in Fig 2, where (a) to (f) represent not same diff shapes, while (g) represents the same diff shape. In the figure, "c" denotes a cardinal move, and "d" denotes a diagonal move. For instance, "(a) d-c" signifies a path from the start point to the goal point that involves a diagonal move followed by a cardinal move at the end. If either x-diff or y-diff is 0, the path consists solely of cardinal moves, as illustrated in (e) and (f).

During this iteration, the children added to the open list draw a path from their location to the target point without considering obstacles. Subsequently, we validate the feasibility of this path by considering obstacles. If any obstacle is present along the drawn octile path, that path is discarded. Otherwise, the target point can be added to the open list, with its parent being the child from this iteration. Since one iteration can have a maximum of 8 children, up to 8 target points with different parents and *g*-values can be generated.

The modified A* algorithm filters these target points, ultimately retaining only one cell with the same location as the target point. Similar to the previous process, if any cell with the same location as the new target point, which is reachable from the children, is already in the open list and the new target point has a higher *g*-value, it is removed. If the new target point is already in the closed list, it is also removed. However, there is no need to check the closed list, the reason for which is explained below.

In the traditional A* algorithm, this process does not exist; it only searches adjacent cells in each iteration. In open spaces with no obstacles, the modified A* algorithm can reach the target point in just one iteration, whereas the traditional A* algorithm's progress depends on the distance between the starting point and the target point, as shown in Fig 3. After the first iteration, the traditional A* algorithm only searches three cells. In contrast, in the modified A* algorithm, the same cells as in the traditional A* algorithm, as well as the target point, remain in the open list. The modified A* algorithm then selects the target point as the next current cell based on the *f*-value. Among the cells in the open list, a cell located in the northeast direction from the starting point and whose parent is the cell with the same *f*-value is chosen. These *f*-values are the lowest. In this situation, we choose the target point with the lower *h*-value as the next current cell, as the algorithm terminates upon selecting the target point. By incorporating octile movement, the modified A* algorithm requires fewer iterations than the traditional A* algorithm.

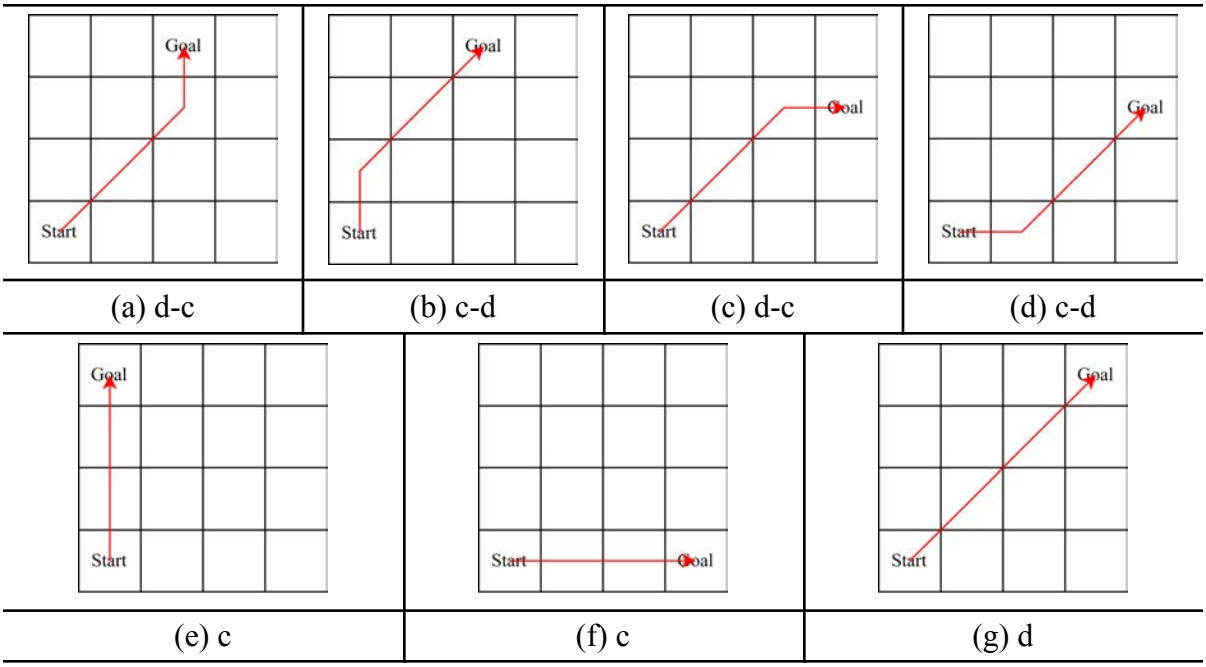
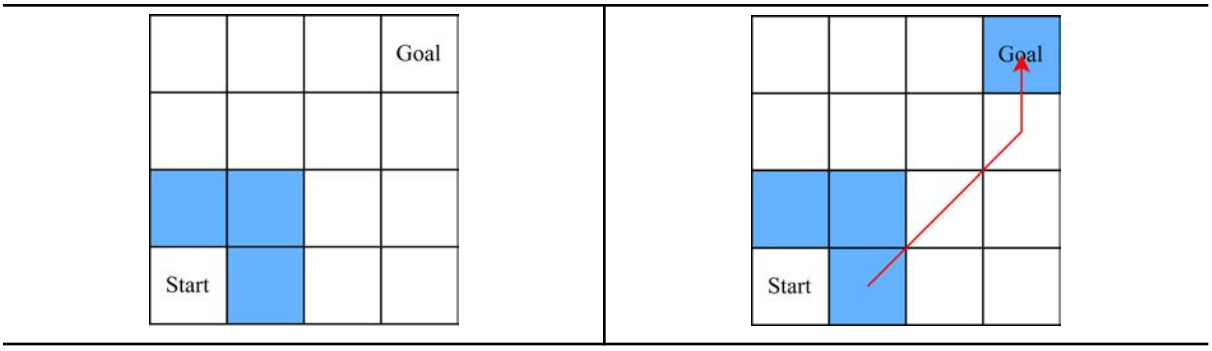


Fig 2: Shapes of octile path



(a) traditional A* algorithm	(b) modified A* algorithm

Fig 3: Open list(painted cells) after first iteration

Each cell in the open list represents a potential candidate for the next iteration. The open list is then sorted in ascending order based on the *f*-value, which is calculated by summing the *g*-value and *h*-value. However, when there are multiple cells with the same *f*-value, the cell with the lower *h*-value is prioritized in the sorting process. A lower *h*-value indicates that the cell is closer to the target point than others. Therefore, it has a greater likelihood of having an octile movement path to the target point while considering obstacles. Cells that are closer to the starting point have a lower probability of reaching the target point using an octile path due to obstacles.In the sorted open list, the cell with the lowest *f*-value becomes the new current cell for the next iteration, and it is moved from the open list to the closed list. If this cell happens to be the target point, the stopping criteria are met, and the algorithm concludes. This explains why we don't need to check the closed list when a new cell with the same location as the target point is generated.

Once the algorithm concludes, an optimal path can be extracted. Every visited cell has a parent (except for the starting point), and by tracing back through the parent cells from the target point to the starting point, the optimal path is constructed. The length of the optimal path is equal to the *g*-value of the target point, which in this case is also the *f*-value.

For your reference, the pseudo-code, flowchart, and an example of the path can be found in Algorithm 1, Fig 4, and Fig 5, respectively.

Algorithm 1 Pseudo-code of A*algorithm with octile distance

```
procedure AstarWithOctileDistance(start, target point)
    Open_List = [ ], Close_List = [ ]
    Add Start to Open_List
    while Open_List is not empty do
        current_node = first node of Open_List
        Add current_node to Close_List
        if current_node == target point
            break
        end if
        Search adjacent 8 nodes at current_node
        Sort Open_List ascending order by f-value
    end while
    return current_node.g-value
end procedure
```

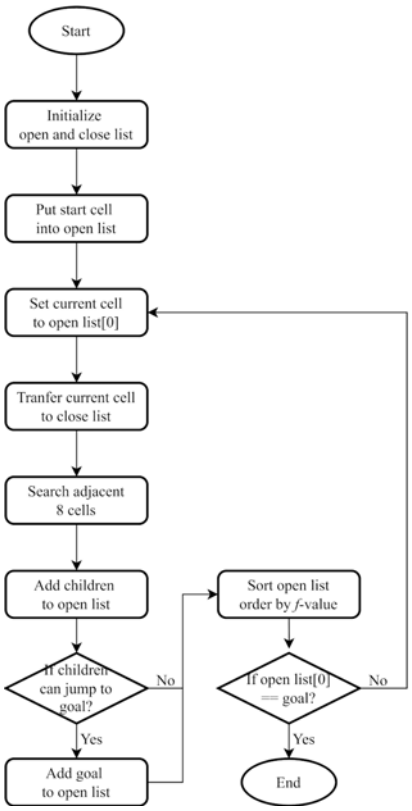


Fig 4: Flowchart of modified A* algorithm

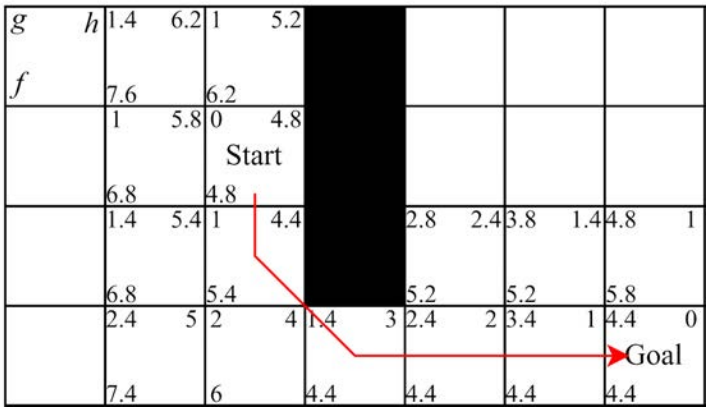


Fig 5: Example path of modified A* algorithm

3. Result

The purpose of this experiment is to compare the performance of the traditional A* algorithm and the modified A* algorithm. The modified A* algorithm must have the same path length as the traditional A* algorithm. Having the same path length means that both methodologies ensure optimization. To measure performance, we determine the performance of

the two algorithms based on the number of iterations required to solve the problem. For the experiment, we will use some of the benchmark datasets used in previous studies (Stern et al. 5). All of these benchmark datasets are lattice structures and consist of a total of 24 maps. Each map contains 25 separate scenarios. Each scenario consists of several single problems, and the total number of problems in individual scenarios varies depending on the map. Each problem includes information about the starting point, the target point, and the length of the optimized path, as shown in Table 1. In this experiment, a total of 10 maps will be utilized, and the first scenario of each map will be employed. Table 2 provides information about the maps, and Fig 6 displays the map's layout. The experiment is conducted with Intel(R) Core i5(TM) CPU (2.6 Ghz), RAM 6GB, and Windows 10, and the language used is Python.

Table 1: Example of problem in scenario

Map name	Height	Width	StartX	StartY	GoalX	GoalY	Answer
Berlin_1_256.map	256	256	220	92	194	65	45.38477631

Table 2: Description of each map

Type	Map	Height x Width	Size
Open	empty-8-8	8x8	Small
	empty-16-16	16x16	Small
Open+obstacles	random-32-32-10	32x32	Medium
	random-32-32-20	32x32	Medium
	random-64-64-20	64x64	Medium
Maze	maze-32-32-2	32x32	Medium
Room	room-32-32-4	32x32	Medium
	room-64-64-8	64x64	Medium
	room-64-64-16	64x64	Medium
City	Berlin_1_256	256x256	Large

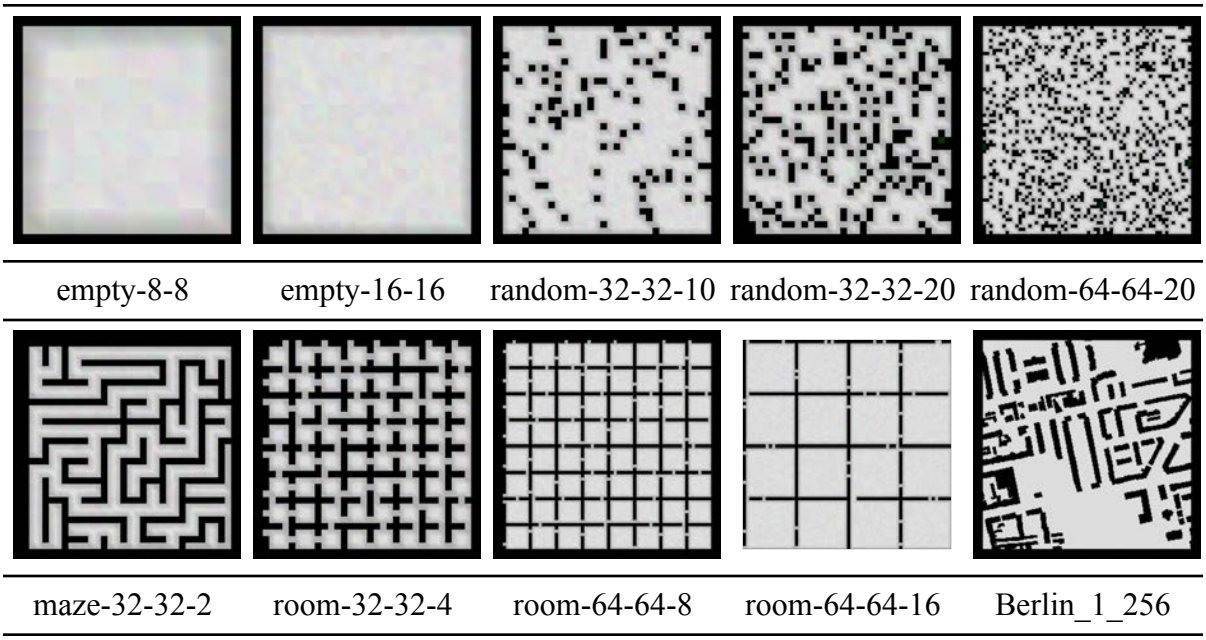


Fig 6: Configuration of each map

3.1. Open

Open refers to an environment with no obstacles. Maps belonging to this environment are ‘empty-8-8’ and ‘empty-16-16’. The results of the experiment are shown in Table 3. The results show that the average path length of the two methodologies was the same in empty-8-8. However, having the same average value does not imply that the path length is the same for all problems. To support this, statistical analysis confirmed a correlation coefficient of 1, indicating that all problem results were the same. Similarly, other problems yielded the same statistical analysis results. Therefore, we will present only the average path length for all future problems. The null hypothesis of the T-test for comparing the performance of the methodologies is 'there is no difference in the average memory size of the two methods' and the confidence level (α) is 0.05. It was confirmed that the null hypothesis was rejected for all experimental results ($p\text{-value}<0.05$). When comparing memory sizes, the modified A* algorithm's memory size decreased by approximately 66%, averaging 2, whereas the traditional A* algorithm had an average memory size of 6. Similarly, in the case of 'empty-16-16,' the average path length remained the same. In this map, the memory size of the modified A* algorithm decreased by approximately 90%, averaging 2, while the traditional A* algorithm had an average memory size of 21. In the case of an open map, the modified A* algorithm always took 2 iterations. This is because, in an open map, there are no obstacles between the starting point and the target point, allowing the target point to be reached from the beginning. This represents the most significant difference between the existing A* algorithm and the approach used in this study.

Table 3: Result of empty-8-8

empty-8-8	A* algorithm	Modified A* algorithm	Improvement

Avg. path length	3.591	3.591	0
Avg. iteration (memory used)	5.875	2.000	65.96%

Table 4: Result of empty-16-16

empty-16-16	A* algorithm	Modified A* algorithm	Improvement
Avg. path length	9.174	9.174	0
Avg. iteration (memory used)	20.766	2.000	90.369%

3.2. Open + Obstacles

Open+obstacles refers to an environment where obstacles are randomly placed in empty spaces. Maps belonging to this type are 'random-32-32-10', 'random-32-32-20' and 'random-64-64-20'. The difference between 'random-32-32-10' and 'random-32-32-20' lies in the number of obstacles, with the latter having more. The difference between random-32-32 and random-64-64 is the size of the map, 32x32 for the former and 64x64 for the latter. The results of the experiment are shown in Table 4. Looking at the experimental results, the average path length of the two methodologies was the same in random-32-32-10. Comparing the memory size, the modified A* algorithm’s memory size decreased by about 86% with an average memory size of 8, while the traditional A* algorithm had an average memory size of 57."In 'random-32-32-20,' the average path length of the two methodologies was the same. The modified A* algorithm's memory size decreased by approximately 60%, with an average memory size of 20, while the traditional A* algorithm had an average memory size of 51. "Furthermore, 'random-64-64-20' also had the same average path length for the two methods. The modified A* algorithm's memory size decreased by approximately 54%, with an average memory size of 80, while the traditional A* algorithm had an average memory size of 175. Open+obstacle did not exhibit a significant difference in memory size reduction compared to the open environment. Hence, it can be concluded that the modified A* algorithm performs well in empty spaces or maps with randomly distributed obstacles. For maps of the same type, it was confirmed that there was no memory reduction based on size.

Table 5: Result of random-32-32-10

random-32-32-10	A* algorithm	Modified A* algorithm	Improvement
Avg. path length	17.602	17.602	0
Avg. iteration	56.978	8.100	85.784%

(memory used)			

Table 6: Result of random-32-32-20

random-32-32-20	A* algorithm	Modified A* algorithm	Improvement
Avg. path length	18.320	18.320	0
Avg. iteration (memory used)	50.630	20.370	59.767%

Table 7: Result of random-64-64-20

random-64-64-20	A* algorithm	Modified A* algorithm	Improvement
Avg. path length	40.354	40.354	0
Avg. iteration (memory used)	175.055	79.785	54.422%

3.3. Maze

A maze is an environment resembling a maze. This category includes 'maze-32-32-2'. In the Maze-type map, there are dead ends, which can lead to an increase in the number of searches. Based on the experimental results, the average memory usage of the existing A* algorithm was 216, while the modified A* algorithm averaged 205, indicating a difference of approximately 5%. This difference is significantly distinct from that of 'random-32-32,' a map of the same size. Due to the complex environment, the difference in the average number of executions between the two methodologies in the maze map is not substantial. Similar to the previous experiments, the average path length of the two methodologies remains the same.

Table 8: Result of maze-32-32-2

maze-32-32-2	A* algorithm	Modified A* algorithm	Improvement
Avg. path length	41.826	41.826	0
Avg. iteration (memory used)	216.461	205.761	4.943%

3.4. Room

Similar to a maze, the room type is a familiar environment. As the name suggests, the entire space has been divided into a number of identical spaces, creating passageways that can be

traversed. Like maze, there are dead ends throughout the map, which can increase the number of searches. Maps of this type are 'room-32-32-4', 'room-64-64-8' and 'room-64-64-16'. The difference between 'room-32-32' and 'room-64-64' lies in the total space size, with the former measuring 32x32 and the latter measuring 64x64.Room-64-64-8 and room-64-64-16 have different room sizes. The former consists of a total of 64 small rooms and the latter has 16 large rooms. Examining the experimental results, it was observed that the average path length of 'room-32-32-4' remained the same, while the memory size was 101 for the traditional A* algorithm and 89 for the modified A* algorithm, indicating a difference of approximately 12%." In the case of 'room-64-64-8,' the average path length remained the same. The memory size of the modified A* algorithm decreased by approximately 3%, averaging 745, while the traditional A* algorithm had an average memory size of 768. Lastly, the average path length of 'room-64-64-16' was the same. The memory size of the modified A* algorithm decreased by approximately 5%, averaging 1,198, while the traditional A* algorithm had an average memory size of 1,266. In the case of the room type, the total space size influenced the difference in memory size. There was a variance in the average memory size reduction between 'room-64-64-8' and 'room-64-64-16,' but in the room layout, it is possible to reach the target point immediately upon entering the passageway between the rooms. Room-64-64-16 contains each room of a larger size, so it has fewer passageways compared to room-64-64-8. Hence, 'room-64-64-16,' which can reach the passageway to a room with a target point more rapidly, exhibits a more significant difference in average memory usage.

Table 9: Result of room-32-32-4

room-32-32-4	**A* algorithm**	**Modified A* algorithm**	**Improvement**
Avg. path length	23.445	23.445	0
Avg. iteration (memory used)	100.846	88.777	11.968%

Table 10: Result of room-64-64-8

room-64-64-8	**A* algorithm**	**Modified A* algorithm**	**Improvement**
Avg. path length	58.929	58.929	0
Avg. iteration (memory used)	767.748	745.429	2.907%

Table 11: Result of room-64-64-16

room-64-64-16	**A* algorithm**	**Modified A* algorithm**	**Improvement**

Avg. path length	78.490	78.490	0
Avg. iteration (memory used)	1,266.163	1,197.745	5.404%

3.5. City

The 'City' environment is modeled after cities from around the world. For this study, a map representing Berlin, Germany, was utilized with dimensions of 256x256. In contrast to the previously mentioned types, this map is the largest in size. Examining the experimental results, it was found that the average memory usage of the traditional A* algorithm was 5,295, while the modified A* algorithms averaged 4,733, indicating a decrease of approximately 11%.

Table 12: Result of Berlin_1_256

Berlin_1_256	**A* algorithm**	**Modified A* algorithm**	**Improvement**
Avg. path length	188.505	188.505	0
Avg. iteration (memory used)	5,295.245	4,732.699	10.624%

4. Discussion

In this study, the path planning problem was solved with the modified A* algorithm by applying the concept of octile distance. In the introduction, differences from existing studies were analyzed by comparing them with previous research, and the contribution of this study was outlined. The estimated distance to the target point was calculated more accurately by utilizing the octile distance to the *h*-value instead of Manhattan distance. Experiments confirmed that the memory size, a drawback of the traditional A* algorithm, can be efficiently reduced by utilizing jump movements towards the target point.

Experiment – The experiment was conducted using Python. Using the benchmark dataset used in previous studies, the performance of the traditional A* algorithm and the modified A* algorithm was compared. Ten maps were selected from the benchmark dataset, and they fall into five categories (Empty, Empty+Obstacles, Maze, Room, City). Each map has a grid structure of the same size, and only fixed obstacles exist. Also, each map has one scenario. The scenario consists of several problems, each of which stores a starting point and a target point. The experiment assumes a single agent, which can move in both straight and diagonal directions If there is an obstacle, the agent cannot move in that direction, and it cannot move diagonally if obstacles are present on both adjacent sides of the diagonal cell. Since both algorithms guarantee the optimal path, performance was measured not in terms of path length but as the number of iterations required to reach the target point for each problem. As a result of the experiment, it

was confirmed that the modified algorithm performed better than the traditional algorithm on all of the maps. It was observed that the difference was particularly significant in the 'Open' and 'Open + Obstacle' scenarios.

Limitations and Future Studies – The modified A* algorithm with octile distance performed better than the traditional A* algorithm on all maps, but the difference was small on certain types of maps (Room, City). If the difference in the number of iterations is small, the modified A* algorithm reevaluates the octile path to the target point for each child, potentially increasing the execution time. Therefore, future studies should 1) identify an environment where the modified A* algorithm has a minimal impact and devises an improved methodology accordingly, and 2) decreases processing time through the parallelization process.

Conclusion – As automation technology is introduced to industrial sites due to advances in technology, unmanned robots such as AGV are being introduced to logistics sites. In the case of AGVs used in logistics warehouses, they must reach the desired location at the shortest distance without hitting the rack. For this, the path planning method is used and a quick calculation is required. A* algorithm, which is mainly used for path planning, has the disadvantage of slow speed due to its large amounts of calculation. The application of the methodology proposed in this study is expected to enhance response times between agents and the system, allow for the simultaneous operation of a larger number of agents, thanks to a reduction in computational overhead.

Works Cited

Duchoň, František, et al. "Path planning with modified a star algorithm for a mobile robot." *Procedia engineering* 96 (2014): 59-69. https://doi.org/10.1016/j.proeng.2014.12.098

Hart, Peter E., Nils J. Nilsson, and Bertram Raphael. "A formal basis for the heuristic determination of minimum cost paths." *IEEE transactions on Systems Science and Cybernetics* 4.2 (1968): 100-107. https://doi.org/10.1109/TSSC.1968.300136

Karur, Karthik, et al. "A survey of path planning algorithms for mobile robots." *Vehicles* 3.3 (2021): 448-468. https://doi.org/10.3390/vehicles3030027

Nieuwenhuisen, Dennis, et al. "Automatic construction of roadmaps for path planning in games." *International Conference on Computer Games: Artificial Intelligence, Design and Education*. 2004.

Paul E. Black, "Manhattan distance", Dictionary of Algorithms and Data Structures [online], https://www.nist.gov/dads/HTML/manhattanDistance.html. Accessed 10 Sept. 2023.

Stern, Roni, et al. "Multi-agent pathfinding: Definitions, variants, and benchmarks." *Proceedings of the International Symposium on Combinatorial Search*. Vol. 10. No. 1. 2019. https://doi.org/10.1609/socs.v10i1.18510

Warren, Charles W. "Fast path planning using modified A* method." *[1993] Proceedings IEEE International Conference on Robotics and Automation*. IEEE, 1993. https://doi.org/10.1109/ROBOT.1993.291883

Exoplanet Detection with Decision Trees By Sriram Loganathan

Abstract

Exoplanets can be detected through the observations of brightness and movement of the stars they orbit. In the past, machine learning algorithms have been able to classify possible candidates using specific techniques such as analyzing large samples of data and automating the otherwise tedious process of classification. In our research, we train a decision tree algorithm on datasets containing confirmed exoplanets, candidates, and false positives from the Kepler Mission in the NASA Exoplanet Archive. From this training, we build a decision tree classification model with a 94.12% accuracy at classifying exoplanets when training on confirmed exoplanets, candidates, and false positives, and a 99.78% accuracy when training only on confirmed exoplanets and false positives. Alternatively, when training a decision tree regression model to predict Kepler Object of Interest KOI) scores, we obtain a loss of 0.04. The decision tree algorithm is a viable option in classifying and detecting exoplanets, as displayed by its effectiveness.

Introduction

Machine learning has a large potential in revolutionizing exoplanet detection; it could possibly be the first step to discovering alien life in our universe. As the search for such life continues, the integration of machine learning into astronomy serves to increase our capabilities by offering a new path to navigate data. The decision tree algorithm can provide astronomers with a more systematic approach to analyze data and extract relevant key details to identify potential exoplanet candidates. The objective of this paper is to demonstrate the effectiveness of machine learning and decision tree models as valuable tools, utilizing their ability to analyze extensive datasets, detect patterns, and make predictions, to enhance the discovery of exoplanets.

Exoplanets, also known as extrasolar planets, are planets located outside our solar system. The proposal of exoplanets has a rich history that dates back to 400 BCE. The concept was first proposed by the Greek philosopher Democritus, who suggested that there might be worlds similar to ours in the universe[1]. However, it wasn't until the 20th century that technology was advanced enough to make such statements. In 1992, astronomers at the Arecibo Observatory in Puerto Rico made a groundbreaking discovery, detecting and confirming the existence of exoplanets[2]. This discovery provided direct proof that planets could exist outside of our solar system and opened new opportunities for research and advancement.

In the following years, the field of exoplanets experienced a surge of advancements and discoveries. The development of more sophisticated telescopes, space missions, and analytical techniques enabled astronomers to detect exoplanets with increasing precision and to uncover a wider range of planetary systems. Most notably, the Kepler space telescope, launched in 2009 by NASA, played a pivotal role in this progress. Kepler's mission was committed to observing an extensive number of stars in a fixed field of view, monitoring their brightness for subtle dips caused by transiting exoplanets[3]. Prolonged observations on the same field allowed Kepler to

capture multiple exoplanet transits, especially those with longer orbital periods. Its highly sensitive detectors and sophisticated data processing algorithms ensured reliable candidates. And with such precision, Kepler led to the identification of thousands of exoplanet candidates.

As astronomers continue to refine their methods and explore newer opportunities, the use of machine learning in exoplanet research holds the key to uncovering more profound insights into the diversity of exoplanetary systems, and their habitability potential, hidden within the collected data. This application represents a large stride towards uncovering new worlds, their enigmatic compositions, and potentially answering one of humanity's most profound questions: Are we alone in the universe?

Exoplanet Detection Methods

Exploring the universe to identify exoplanets involves many methods, each serving to detect distinct signatures left by these celestial objects, spanning from direct imaging to the observation of light curves. In this section, we examine the techniques used by astronomers to discover exoplanets and learn more about their properties. From the radial velocity method, which detects the wobble induced by orbiting exoplanets, to the transit method, which observes the periodic dimming of a star caused by a passing planet, we unravel the physics behind exoplanet detection. Additionally, we examine the phenomenon of gravitational microlensing, where the gravity of a star or planet acts as a cosmic lens, magnifying the light of background objects. By understanding these methods, we lay the foundation for comprehending the data that machine learning algorithms will process, opening the door to a deeper exploration of exoplanetary systems.

At the forefront of exoplanet detection lies the radial velocity method, which unveils the otherwise nearly hidden movement between stars and their orbiting planets. This method converts light into data, revealing the presence and properties of exoplanets through the electromagnetic spectrum. When an exoplanet orbits its host star, both celestial bodies are bound by mutual gravitational forces. As the planet exerts its subtle tug, the star reciprocates with a slight shift in its movement, oscillating around their common center of mass. This leaves behind a Doppler shift to the star's emitted light[4]. The formula for radial velocity using the Doppler Shift effect is given by:

$$\frac{\Delta\lambda}{\lambda} = \frac{v}{c},$$

where

$\Delta\lambda$ represents the change in wavelength of the star’s spectrum,
λ is the original wavelength of the star’s light,
v is the radial velocity of the star caused by the presence of an exoplanet, and
c is the speed of light.

When light is analyzed, it appears as a spectrum of lines, each corresponding to specific wavelengths. The Doppler effect, a fundamental principle in the field of optics in physics, is introduced. As a star wobbles due to the gravitational influence of an exoplanet, its light shifts either toward the blue or red end of the spectrum. This phenomenon, known as blueshifting and redshifting, signifies motion approaching or receding, respectively[5]. These shifts convey a wealth of information, enabling astronomers to determine the velocity and other details of a star's motion.

A method that has been more recently popularized is the transit method. This method involves observing the change in brightness of a star as an exoplanet transits in front of it[6]. The depth of this transit (ΔF) can be described using this equation:

$$\Delta F = (R_{planet}/R_{star})^2,$$

where

R_{planet} is the radius of the exoplanet and
R_{star} is the radius of the star

As the exoplanet orbits its star, these recurring transits generate a series of fluctuations in the star's luminosity, forming a light curve. A light curve is a set of data that can hold the information of an exoplanet such as its orbital period and distance from its star[7].

Decision Trees

In the discovery of exoplanets, machine learning emerged as a compass, helping astronomers process and analyze massive amounts of data. These algorithmic techniques methodically dissect data, mimicking behavior similar to that of a human.

A decision tree is a supervised machine learning algorithm. Beginning at the root node, the algorithm selects the most informative feature and splits the data into subsets. This process is repeated recursively for each resulting subset, creating a branching structure that resembles a tree[8]. The algorithm determines the optimal splitting points using criteria such as the cross-entropy loss for classification tasks[9], given by

$$H(P, Q) = -\sum_{x \in X} P(x) \log(Q(x))$$

where

$H(P, Q)$ represents the cross-entropy between two probability distributions, P and Q,
$P(x)$ represents the probability of an event x occurring according to the distribution P,
$Q(x)$ represents the probability of the same event x occurring according to the distribution Q.

$\sum_{x \in X}$ indicates that the calculation is performed over all possible events x in the set X.

or mean squared error for regression tasks[11], given by:

$$\text{MSE} = \frac{1}{|D|} \sum_{i=1}^{|D|} (y_i - \hat{y_i})^2$$

where

$|D|$ denotes the number of samples in the dataset D

y_i represents the actual target value (also called the ground truth) for the ith sample in the dataset D.

$\hat{y_i}$ represents the corresponding predicted value to y_i, given the current state of the model. (We convert words or string values into numeric values using the one-hot encoding method[10])

As the tree grows, it captures hierarchical relationships within the data, with each internal node representing a decision based on a specific feature, and each leaf node containing the final prediction or decision. To make predictions for new data, the algorithm follows the decision path from the root to a leaf, based on the feature values, and outputs the corresponding prediction or class label.

The interpretability of decision trees is one of their major strengths. The decision path from the root to a leaf node represents a clear and easily understandable set of rules, making it straightforward to interpret and explain the reasoning behind predictions. However, this transparency can also lead to overfitting, where the tree captures noise in the training data[8]. To mitigate this, various strategies are employed, such as limiting the depth of the tree, setting a minimum number of samples required for a node to be split further, and pruning branches that contribute little to predictive accuracy.

Gradient Boosting Classifier/Regressor

Gradient boosting is a powerful and widely-used ensemble learning technique for both classification and regression tasks. It combines the predictive power of multiple weak learners, usually decision trees, to create a strong and accurate model. The core principle behind gradient boosting is the iterative construction of a sequence of models, each correcting the errors of the previous ones. The process begins by creating a simple base model, typically a shallow decision tree, which serves as the initial prediction. Subsequent models are then built in a stepwise manner, where each new model focuses on reducing the errors made by the combined predictions of the existing models[12].

The key innovation of gradient boosting lies in its optimization strategy. It minimizes a specific loss function, which quantifies the difference between the predicted values of the combined ensemble and the actual target values. Instead of just approximating the target directly, each subsequent tree aims to correct the errors made by the preceding trees[12]. In each iteration, the algorithm identifies the direction (gradient) in which the loss function has the steepest descent, and constructs a new model to fit this residual error which involves the formula given by

$$\text{minimize} \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i))$$

where

L is the loss function such as the Mean Squared Error(MSE),
$F_{m-1}(x_i)$ represents the ensemble's predictions up to the $(m-1)$ iteration for the ith sample,
γ is the learning rate which determines the contribution of the new weak learner,
$h_m(x_i)$ is the prediction of the mth weak learner for the ith sample.

With each iteration, a new tree m is constructed, not with the objective of directly predicting the target, but to compute the residual or error of the preceding tree $m-1$'s prediction. This means that if the first model made specific prediction errors, the subsequent model aims to 'correct' these inaccuracies. The final prediction is obtained by summing up the predictions of all the individual models, weighted by a learning rate that determines the contribution of each model to the final result. This refinement and technique results in robust and accurate model making gradient boosting a powerful tool in machine learning.

Results

In our model training and evaluation utilizing the scikit-learn library[13], our objective is to optimize a GradientBoostingClassifier by accessing a range of random hyperparameter combinations for tuning. The hyperparameters considered include:

- N_estimators: The number of boosting stages to be run (number of trees used).
- learning_rate: The step size at each iteration while moving towards a minimum of the loss function.
- max_depth: The maximum depth of the individual estimators.
- max_features: The number of features to consider when looking for the best split.

The optimal hyperparameters are selected based on their performance on a validation set.

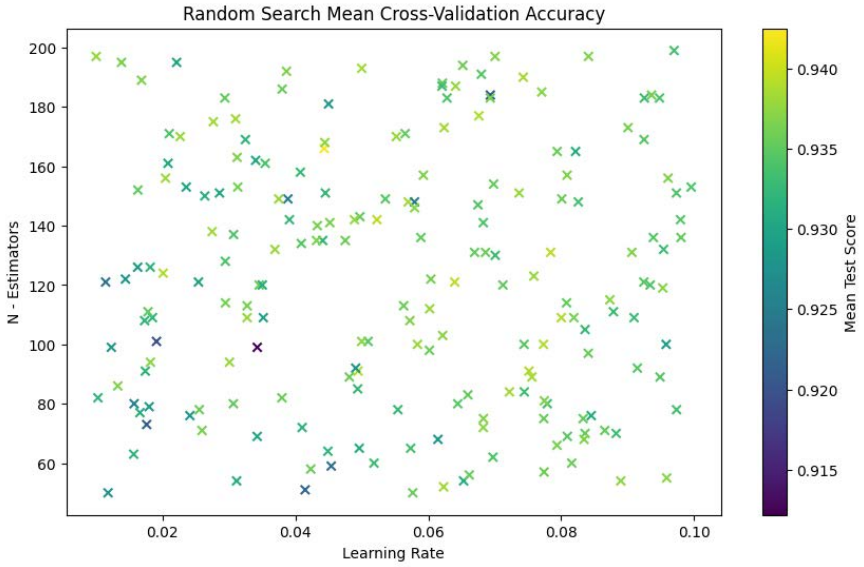


Fig 1: Random search mean cross-validation accuracy chart, varying the hyperparameters n-estimators (number of trees) and learning rate.

Our GradientBoostingClassifier achieves an accuracy of 94.12% on the testing dataset containing confirmed exoplanets, candidates, and false positives. Despite the low variance between the training and testing datasets, there is avoidable bias, in that exoplanet candidates are all either exoplanets or false positives in reality, but have just not been determined as either one yet. When training only on confirmed exoplanets and false positives, the accuracy is 99.78% .

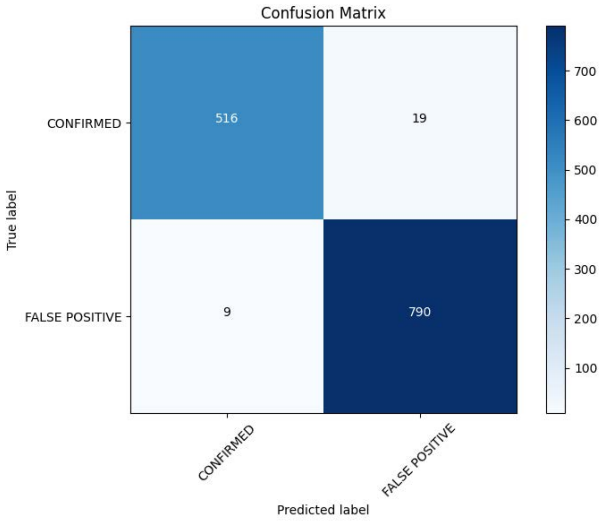


Fig 2: Confusion Matrix Distinguishing Confirmed Exoplanets From False Positives

In addition to classifying exoplanets and false positives, we seek to predict the Kepler Object of Interest (KOI) score. The KOI score, calculated from the Kepler space telescope's observations, quantifies the likelihood that a given celestial signal hints at the existence of an

exoplanet. Scores range from 0 to 1, with values closer to 1 suggesting a higher probability that the observed signal corresponds to an exoplanet, while values closer to 0 indicate a higher likelihood of the signal being a false positive or stemming from other cosmic phenomena.

Given the continuous nature of the KOI score, which ranges between 0 and 1, a classification approach would be unsuitable. Instead, we opt for a regression tree, a machine learning model adept at predicting continuous outputs. Upon training our regression tree on the dataset, we achieved a Mean Squared Error (MSE) of 0.04. This means that, on average, our predictions deviate from the actual values by 0.2 units, representing a 20% error with respect to the full scale.

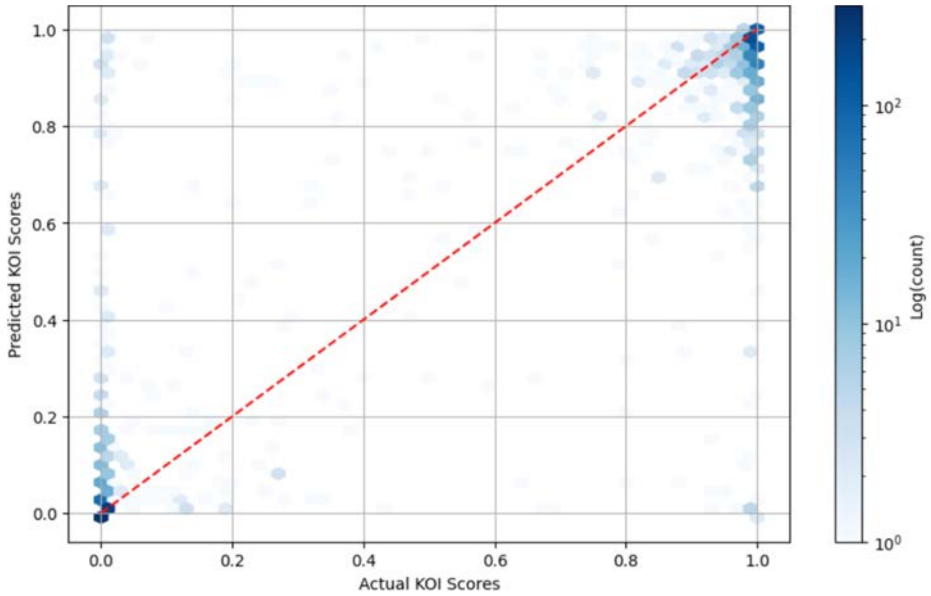


Fig 3: Hexbin Density Plot of Actual vs. Predicted KOI Scores

Works Cited

1. Gagnon, Jean, et al. “IAL 18: Exoplanets & General Planetary Systems.” UNLV Physics, 1997, https://www.physics.unlv.edu/~jeffery/astro/ial/ial_018.html.
2. Britannica, The Editors of Encyclopaedia. "Arecibo Observatory". Encyclopedia Britannica, 1 Aug. 2023, https://www.britannica.com/topic/Arecibo-Observatory.
3. Dooling, Dave. “Kepler.” *Encyclopædia Britannica*, Encyclopædia Britannica, inc., 2009, www.britannica.com/topic/Kepler-satellite.
4. Richmond, M. (2001). A connection between radial velocity and distance. http://spiff.rit.edu/classes/phys240/lectures/expand/expand.html
5. Dobrijevic, D., & Howell, E. (2022, January 14). Redshift and blueshift: What do they mean? https://www.space.com/25732-redshift-blueshift.html
6. Rauf, J. (2021). Looking for Exoplanets. https://www.uc.edu/content/dam/refresh/cont-ed-62/olli/21-fall/exoplanets4.pdf
7. Richmond, M. (2014). Important parameters of an eclipsing system. What can we learn from light curves? http://spiff.rit.edu/classes/phys373/lectures/light_curves/light_curves.html
8. Stanford Online. (2020, April 17). Lecture 10 - Decision Trees and Ensemble Methods | Stanford CS229: Machine Learning (Autumn 2018) [Video]. YouTube. https://www.youtube.com/watch?v=wr9gUr-eWdA
9. Koech, K. 2020, August 20. Cross-Entropy Loss Function. Towards Data Science. https://towardsdatascience.com/cross-entropy-loss-function-f38c4ec8643e
10. Yadav, D. 2019, December 9. Categorical encoding using Label-Encoding and One-Hot-Encoder. Towards Data Science. https://towardsdatascience.com/categorical-encoding-using-label-encoding-and-one-hot-encoder-911ef77fb5bd
11. 2008. Mean Squared Error. In: The Concise Encyclopedia of Statistics. Springer, New York, NY. https://doi.org/10.1007/978-0-387-32833-1_251
12. Kurama, V. 2020, March 29. Gradient Boosting for Classification. https://blog.paperspace.com/gradient-boosting-for-classification/
13. Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.

Performance of OpenCV HOG Model for Various Image Quality Changes By Soumik Sinha

Introduction

Pedestrian detection is a crucial task for computer vision specifically in the application of self-driving cars and road safety. Accurate and robust pedestrian detection models are essential for ensuring the safety and efficiency of a range of applications such as autonomous vehicles and surveillance systems. However, real-world situations often provide challenging conditions. One example is image quality. Variations in image resolution, lighting, blurring, and contrast can pose a threat to the safety of these systems. This research paper aims to investigate the robustness and adaptability of a computer vision model in detecting pedestrians.

The paper focuses on the Histogram of Oriented Gradients (HOG) algorithm, a widely-used technique for pedestrian detection, implemented in the OpenCV computer vision library [1, 2]. The HOG is a computer vision technique used to understand the shapes and textures in images. It does this by analyzing and finding the directions of changes in image brightness, essentially by looking at how the pixel values change from one region to another. HOG divides an image into small parts, calculates the orientations of these brightness changes in each part, and then creates a unique description of these patterns which can be used to capture information about the shapes and textures within the image.

In this research, I explore how well a pedestrian detection model based on HOG, as implemented in OpenCV, performs under different image quality conditions. By varying image quality factors like brightness, contrast, and blur, this paper aims to understand how robust the model is in scenarios where image conditions may not be ideal.

Methods

The data I use is from Kaggle containing pictures, diagrams, and traffic [3]. Two sets of pictures are there: pictures with pedestrians and pictures without them. I have found 40 images that would work best for the project. The data is not very clean as it had various stock images of cars and pedestrians and also had some images taken from a higher height meaning that the angle of the picture wouldn't be accurate for autonomous cars. Additionally in the "no pedestrian" data set, I have found images containing people on scooters and bikes however those would be considered pedestrians in my project. I find better images throughout the dataset that look as if they came from a self-driving car camera with real traffic and pedestrians on the road. I also make sure that all of the images I choose have similar image qualities in terms of brightness, contrast, and blur so that I could accurately test the influence of those factors on the model.

As mentioned before I am using the OpenCV package for the HOG algorithm. Additionally, I am using it to alter the characteristics of the images in the data. The HOG algorithm detects the pedestrians themselves and outputs the regions. To check if there are any pedestrians, I use an if statement to check if there are regions outputted by the HOG model.

To check the performance of the HOG model when altering image brightness, contrast, etc. I take a baseline accuracy measurement. This is done by looping through the images in the data, having the HOG model run on the image, and checking whether it correctly or incorrectly guessed the presence of a pedestrian.
I find that the baseline accuracy for the HOG model is just found by calculating the percent of images that the model guesses correctly out of the 20 images with pedestrians and the 20 without.

Discussion

The expectation of the model performance when altering the data is that the HOG model will perform worse. With different levels of brightness, contrast, and blur the model would be less able to detect pedestrians as it wasn't specifically trained on them. However, I expect the model wouldn't be too far from the baseline accuracy and may only get a few images wrong.

The OpenCV HOG model performs fairly well for images that haven't been altered. It correctly classifies 35 of the 40 images, giving it an accuracy of 0.875. The model performs very well for the images without pedestrians as it gets all 20 of them correct. It never mistakes anything else for a pedestrian. However, it doesn't do as well as detecting pedestrians in the images which is much more important. It correctly identifies 15 of the 20 images as having a pedestrian.




Both of the above images are examples of when the model correctly identifies and highlights a pedestrian on the street.




However, these 2 images show the limitations of the model as it does not detect any pedestrians here.

Brightness Comparison

The accuracy of the model changes after the brightness is both increased and decreased. When the brightness is increased the model's accuracy drops to 0.65 when classifying the images with pedestrians and 1.0 when classifying images without them. The model is able to correctly detect pedestrians in many of the bright images as it did detect the unaltered ones except for a few. For example in the following images, for the one on the left (unaltered) the model correctly sees a pedestrian; however it doesn't for the one on the right (increased brightness). This may be because the pedestrian blends into the road too much when the rightness is increased.



The accuracy after lowering the image brightness drops to 0.5 for images with pedestrians and 1.0 for images without. For example, in the following images, the pedestrians are undetected by the model when the brightness of the image is lower.



In general, it seems like not altering the brightness is better for the model. The model works slightly better with higher brightness than lower brightness overall; however I notice an image where the opposite is true. On the right is the brightened image which the model failed to detect; however, when the brightness is decreased, the model works better. This may be because the original image is right and lowering the brightness allows the person to blend it less with the background.



Contrast Comparison

The accuracy of the model, when the contrast is increased, is 0.7 for images with pedestrians and 0.9 for images without pedestrians. The accuracy of the model, when the contrast is decreased, is also 0.7 for images with pedestrians and 1.0 for images without pedestrians. While the lower contrast generally performs better than the higher contrast, there are cases where higher contrast does better than lower contrast. The following example shows a higher contrast image (left) in which the model works better than the low contrast image (right). This is an image for which the baseline model isn't able to detect that there is a pedestrian.



However, adding more contrast also results in the model detecting pedestrians in images where there aren't any like in the example below. This may be because the higher contrast makes it so that the model mistakes the traces and edges that it finds for people.



Blur

As expected the blurry images result in a decreased performance of the model. The model had an accuracy of 0.6 for pedestrians and 0.95 for those without pedestrians for the blurred images. While I do get quite a few of them correct, it misses some that the baseline gets. Below is an example of this. Baseline (left) found the pedestrian, while blurry (left) did not.



However, it would be more important to guess the images with pedestrians correctly. There is much more danger in getting images with pedestrians incorrect than getting the images without them incorrect. More importance should be placed on getting a higher accuracy for just the pedestrian images.

In general altering the images results in a lower performance. In situations with the real world, situations like bad weather, or night images would result in a lower performance. However, on a case-by-case basis, some images do better when altered. For example, increasing contrast helps in some images, perhaps because the HOG model is able to detect the changes in pixel intensity easier. To improve the model, changes could be made to the images. For example, if there is a dark background, the model might benefit by increasing the brightness or vice versa. If the image has a very high contrast and mistakes cars and other objects for people, it would be better to lower the image contrast.

In real-world scenarios, there is a possibility that the change in image quality will improve from the baseline model and can be used in such algorithms to improve the performance of those models.

Conclusion

How does the OpenCV HOG model perform with changes to the image quality, specifically its brightness, contrast, and blur, when detected by pedestrians on the road? In general, the accuracy of altering those image characteristics decreases the performance of the model. The model performs better without any alterations. However, in some cases, it can improve the model.

Works Cited

Mallick, Satya. “Histogram of Oriented Gradients Explained Using Opencv.” *LearnOpenCV*, 30 Nov. 2021, learnopencv.com/histogram-of-oriented-gradients/.

“Home.” *OpenCV*, 27 Sept. 2023, opencv.org/.

Tejasvagarwal. “Pedestrian No Pedestrian.” *Kaggle*, 9 Dec. 2017, www.kaggle.com/datasets/tejasvdante/pedestrian-no-pedestrian.

The Comorbidities of Iron Deficiency Anemia and Psychiatric Illness: Analyzing Global Data to Bridge the Gap in the Literature By Sophia Smallwood

Abstract

Iron deficiency anemia is a condition marked by hemoglobin levels of <12.0 gm/dl in women and <13.5 gm/dl in men, and ferritin (protein where iron is stored in the blood) levels of <50 ng/mL. Healthy levels of hemoglobin are generally regarded as the range between 13.8 gm/dl to 17.2 gm/dl in men, and 12.1 gm/dl to 15.1 gm/dl in women. Healthy ferritin levels are widely seen as >100 ng/mL in both men and women. Rampant in pregnant and menstruating women, the condition is known for its physiological manifestations, including, but not limited to, extreme fatigue, shortness of breath, increased heart rate, and unexplained weakness in the body. The psychiatric aspects of the disease, though, remain rather under researched in the fields of hematology and psychiatry. *Executive functions and psychiatric disorders in adults with iron deficiency anemia* is a study that was conducted in Turkey and published in Cukurova Medical Journal, and is one of the few sets of data we have regarding co-occurring psychiatric disorders in iron deficiency anemia patients at the moment. The purpose of this study was to compare psychiatric and cognitive functioning in patients with iron deficiency anemia (IDA) with healthy patients, which determined whether or not any psychiatric disorders were co-occurring. Upon evaluation with the DSM-5 Clinican's Version, it was determined that psychiatric disorders were 29.4% more common in IDA patients (43%) than in healthy controls (13.6%). It was suggested that patients with IDA should be closely monitored by a physician for any comorbid health consequences. Despite the promising results of the original study in creating a bridge between IDA and its psychiatric comorbidities, it lacks discussion and data in areas including the role of iron supplements and intravenous iron in the co-occurrence of IDA and psychiatric illness, the reality behind accessibility to intravenous iron, limitations of the study in terms of data collection over a 6-month-period, and a very small amount of studies done on adults with IDA in the realm of psychiatric disorders.

The role of iron supplements and intravenous iron in the comorbidities of IDA and psychiatric illness

The role of iron supplements and intravenous iron in the coexistence of IDA and psychiatric illness has been left somewhat overlooked in this study. The most common way to treat IDA is through oral iron supplementation, followed by intravenous (IV) iron. The study mentions the way "...people with IDA who take iron supplements have a noticeably lower risk of developing psychiatric illness." Additionally, the study mentions an increase in cognitive functions and working memory after "iron replacement therapy" (oral supplementation or intravenous iron course of treatment not specified). Though mentioned, information concerning treating IDA is left rather vague in the original study. What does being "properly monitored" by a physician actually mean? Sit around and wait for something bad to happen, or proactively treat the issue? The topic surrounding treatment needed to be more integrated in this study, and I think

the study would have been of more meaning had there been a group of patients with IDA who were treated with supplementation of some sort involved in the experiment. *Psychiatric disorders risk in patients with iron deficiency anemia and association with iron supplementation medications: a nationwide database analysis*, a study conducted in Taiwan, further dabbled in the link between psychiatric illness and IDA patients who were treated vs. not treated. Both the "results" and "discussion" tabs of the article gloss over the outcomes of implementing treatment in the IDA group. It was concluded that "iron supplementation in IDA subjects was associated with a significantly lower risk of psychiatric disorders compared with no iron supplementation in IDA patients" after adjusting the experiment for confounding factors. However, a notable finding was that "even IDA patients receiving iron supplementation had a higher incidence of psychiatric disorders compared to the non-IDA group," which holds potential to demonstrate the possible severity of the condition as it relates to psychiatric illness, and thus, the importance of treatment as the need becomes apparent.

The reality behind the accessibility of intravenous iron

The importance of treatment of IDA through a form of iron replacement therapy remains constant, as demonstrated in the second study; however, evaluating the place of intravenous (IV) iron in the clinical setting is not considered. IV iron is great in theory and shows remarkable results in terms of immediate increases in ferritin and hemoglobin levels. IV iron is faster-acting than oral supplements, which can take months as opposed to days/weeks to kick in. The problem here lies with the reality behind actually accessing intravenous infusions. The first resort to treating IDA is through oral iron supplements, which can have brutal digestive side effects including diarrhea, constipation, nausea, and stomach pain. More commonly than not, hematologists will not infuse iron unless hemoglobin is at a certain (low) level, but the flaw of this standard is that ferritin levels are not considered here. Ferritin levels alone are responsible for iron deficiency and all of the symptoms that go with it, but without hemoglobin levels of <12.0 gm/dl in women and <13.5 gm/dl in men, the diagnosis of anemia can not be given, and therefore, IV iron is not often given as an option. It doesn't make a lot of sense to consider IV iron as a feasible solution to managing IDA to prevent or decrease the prevalence of psychiatric illness when the modern clinical standards do not allow for sickly patients, who have all of the same symptoms as IDA patients, but have "normal" hemoglobin levels, to have access to them. *Role of iron in the treatment of attention deficit-hyperactivity disorder* is a study conducted in Spain where patients (ages 6-14) who were ents) (4mg/kg/day) for three months. Of the nine subtypes with ADHD, the treatment was successful in the total control of symptoms in five of them, partially effective in three, and ineffective in one patient, as opposed to being completely ineffective in seven out of eight subjects without ADHD. The problem with ferrous sulphate as a treatment for IDA (and therefore decreasing risk of psychiatric illness, as previously explained) is that three months is too long for those, especially menstruating women, with low levels of ferritin, but whose hemoglobin levels aren't "low enough" to be given an iron infusion. The IV

iron treatment works. We need it to be accessible to those who need them as fast as possible, and don't have three months to wait.

Limitations of the study in terms of data collection over a 6-month-period

Another thing to consider about the original study is that the design of the experiment consisted of a limitation in terms of data collection over a 6-month-period. 71% of the IDA group in the original study was made up of women, and while age is not specified beyond the fact that everyone is 18+, we can assume that a good portion of these women are currently menstruating. Ferritin levels fluctuate throughout the course of the menstrual cycle due to the loss of iron through periods, which builds back up, and in the case of IDA patients (especially women who bleed heavily), not fast enough for the body to function normally. Many factors are unknown, including the flow, length, and general nature of any of these women's periods. It is also unknown whether or not any of these subjects are currently taking contraceptives or are pregnant; which would also create a fluctuation in ferritin levels. I suggest that there be a study that observes these patients over a longer period of time for the results to be more generalizable, due to the mere fact that cycles are rather unpredictable sometimes, and there are factors (stress, exercise, etc.) that may cause a shift in period length, timing, and heaviness of flow. *Iron Deficiency in Menstruating Adult Women: Much More than Anemia*, another study conducted in Spain, found that "...among women of reproductive age, menstrual blood loss is the most common cause of ID and IDA." The research suggests that "...high menstrual blood losses…have been consistently associated with poor iron status in women, emphasizing the importance of recognition and diagnosis of ID in this particularly vulnerable female population." These findings highlight the importance of closely monitoring the role of menstruation in studies dealing with IDA and women. The impact menstruation has on iron levels is very significant, as backed by the literature, and is subject to change over certain confounding factors.

Limitation in the field due to a very small amount of studies done on adults with IDA in the realm of psychiatric disorders is apparent. The value of the original study remains great, as it is one of the only studies in the field that connects psychiatric illness with IDA in adult subjects. This study has integrated research from elsewhere and dissected data from loosely connected studies dealing with IDA or psychiatric illness, tying global data together. During my initial research, my hope was to find studies regarding women specifically as they relate to IDA and its link with mental disorders, but unfortunately, there seems to be a gap in the literature when it comes to these topics combined. I think that overall, this represents a deeper issue than the literature itself, and a flaw in the medical field as we see it today. As society progresses, we can expect that research in women's health will advance further and hope that contributes to women being treated more fairly in the medical system by having their symptoms listened to, attended to, and taken seriously. The hope is that this research, which is newly published (2023) will raise awareness in the fields of hematology and psychiatry as they become more integrated with one another in the future.

Works Cited

Fernandez-Jimenez MC, Moreno G, Wright I, Shih PC, Vaquero MP, Remacha AF. Iron Deficiency in Menstruating Adult Women: Much More than Anemia. Womens Health Rep (New Rochelle). 2020 Jan 29;1(1):26-35. doi: 10.1089/whr.2019.0011. PMID: 33786470; PMCID: PMC7784796.

Lee HS, Chao HH, Huang WT, Chen SC, Yang HY. Psychiatric disorders risk in patients with iron deficiency anemia and association with iron supplementation medications: a nationwide database analysis. BMC Psychiatry. 2020 May 11;20(1):216. doi: 10.1186/s12888-020-02621-0. PMID: 32393355; PMCID: PMC7216322.

Soto-Insuga, V., et al. “Utilidad Del Hierro En El Tratamiento Del Trastorno Por Déficit de Atención E Hiperactividad.” Anales de Pediatría, The Spanish Association of Pediatrics, 1 Oct. 2013, www.analesdepediatria.org/es-utilidad-del-hierro-el-tratamiento-articulo-S1695403313000805.

Yılmaz, Y. , Terzi, H. , Taşova, B. "Executive functions and psychiatric disorders in adults with iron deficiency anemia" . Cukurova Medical Journal 48 (2023): 394-401 <https://dergipark.org.tr/en/pub/cumj/issue/76399/1252790>

In the Name of Profit and Compassion: Increasing Employee Engagement By Serena Wu

Abstract

Economic fortune is the aim of nearly every organization. It is not only influenced by macroeconomic factors in the greater economy, but also some which can be controlled by the organization or employees. This research paper posits that directing organizations' resources and policies towards increasing employee engagement is not only an empathetic gesture that leads employees to associate their jobs with positive emotions, but that is also a lucrative pursuit. The paper is divided into several sections that discuss specific avenues through which employee engagement can be increased: leadership, creativity, and positive work relationships. Each section draws connections between these three variables and employee engagement, and includes strategies for increasing employee engagement based on organizational psychology research.

Introduction

Success - every business's primary goal. Yet, many companies simply focus on one aspect of success: profit. While numbers are a pretty foolproof way to gauge growth or decay, this article contends that the true holy grail should be employee engagement. Increasing employee engagement is positively correlated with improved employee physical health, employee and customer satisfaction, employee performance, and higher rates of organizational citizenship behavior - i.e., acts of loyalty, helpfulness, or service to colleagues (Harter et al.). Further, employee engagement is negatively correlated with rates of stress and turnover. Increasing employee engagement through cultivating positive leadership, revitalizing creativity, and nurturing positive connections is a fantastic solution to both maximizing profit and employee happiness. An example of successful execution of these tactics is the case with Griffin Hospital (Cameron, 10-13). After a period of chaos and turmoil, previous Vice President Charmel, who was forced to resign, was brought back and reinstated as the company's President and CEO. Although he was forced to make layoffs due to the poor financial situation, Charmel continued exerting a positive influence over his workers, emphasizing positive social connections and virtuous ideals of trust, compassion, and integrity. For example, when one employee was diagnosed with terminal cancer, staff from various branches of the hospital generously donated their vacation days to that employee. In fact, that employee was able to live out their last few months without working, but continued to receive their salary until the day they passed away. This company can be characterized by positive deviance, where workers and the company as a whole thrive, rather than just survive, in productivity, profit, effectiveness, and worker well-being. As a result of this positive deviance, employees became more creative about how to improve their facility and policies. One notable product of this innovation was the idea of installing double beds in the maternity ward so that the parents could sleep in the same bed instead of one partner sleeping in an uncomfortable chair. The Griffin Hospital is the perfect example of how a workplace can be reformed for the better through scientifically supported methods of improving employee engagement.

Leaders and Employee Engagement

Out of any individuals within a company, leaders play the most monumental role in increasing employee engagement. It is important to remember that leaders are not just teachers, but they are also sources of inspiration for their employees (Tarricone and Luca, 640). In fact, research demonstrates that intelligence is not a significantly important factor for predicting leaders' competence (Kohs and Irle). Arguably, a leader's most important role is creating the optimal environment for their employees to thrive and learn from each other (Hill et al.).

Two of the most effective leadership styles that help create this positive environment are transformational leadership and authentic leadership. Both leadership styles are based around the principles of positive organizational psychology, which aims to create happy and productive workplaces ("What Is Leadership?"). Both transformational and authentic leadership styles emphasize inspiring workers and building a healthy climate so that the team can thrive as a whole. These leaders set a high standard for morality and empathy and are willing to make personal sacrifices for their employees, creating a greater sense of organizational commitment. Organizational commitment is a feeling of loyalty that employees have for each other as well as to the organization. They show this loyalty in the form of organizational citizenship behaviors (OCBs), which can be described as actions that employees take that go above and beyond what workers are expected to do to help their colleagues and company. While transformational and authentic leadership styles share many similarities, transformational leaders focus more on inspiring followers through high standards of ethics and morals and motivating rhetoric (Lord et al., 8), while authentic leaders inspire their followers through their authentic personalities and genuine behaviors that align with their values or mission (Reichard).

Another approach to building and managing teams that has many potential benefits is called strengths-based leadership. Strengths-based leadership is often described as an individual leadership style. However, it can also be viewed as a strategy to use in conjunction with other leadership styles (Mind Tools Content Team). This strategy emphasizes building a well-rounded team made up of uniquely talented individuals by focusing on discovering and further developing employees' special gifts. This is a fantastic way to keep employees engaged since each is working within their specialties and is given tasks that interest them. One potential drawback of using only this strategy is that it can make people feel trapped within their one role. So, as leaders, it is important to keep a line of open communication with workers to ensure that employees are fully engaged and able to explore alternative work responsibilities.

Creativity and Employee Engagement

Another variable that affects employee engagement is creativity (*How to Enhance*). Creativity fosters a sense of absorption where employees are engaged and engrossed in their work, leading to increased productivity and efficiency (Schaufeli et al.). While some individuals may have higher levels of innate creativity, they can further develop innovative traits through a few channels. Some ways that leaders can encourage innovation is through designing tasks that give room for their employees to be creative, making employees feel valued, and giving effective

feedback. A tip for ensuring that feedback is constructive without being too reprimanding is being aware that all feedback is inherently biased and, therefore, framing criticism as one's personal opinion, rather than objective truth (*How to Enhance*). Although it may be difficult, finding a balance between having a high-pressure environment and giving employees agency over their work can optimize productivity while discouraging procrastination (Seppälä and Cameron). This is because a consistently high pressure and high stakes environment creates more stress, decreases productivity, and damages sustainable employee engagement.

Another factor that affects creativity is the makeup of a team (West and Sacramento, 30-31). Diversity is one factor that helps leaders to build creative and innovative teams. This can mean demographic diversity (e.g., race, gender, culture) as well as skill set diversity (e.g., educational background). Different levels of diversity within a team allows for team members to share multiple perspectives or varying experiences, so the employees can build on each other's ideas. Thus, diversity within teams is one important goal that leaders should consider while hiring. More traits to look for while hiring to build highly creative teams are openness to new ideas, which can indicate a creative mindset, and strong analytical skills, which help employees evaluate which innovative ideas are worth pursuing. Therefore, having a hiring process with clear goals and direction is very important for helping leaders to hire a suitable pool of employees that are not only competent but also compatible and innovative.

Positive Work Relationships and Employee Engagement

While employees may sometimes feel powerless over their engagement since they cannot rewrite policies or create teams, one thing that all employees have control over is creating high-quality connections with their co-workers. When looking to establish these positive relationships with colleagues, employees must approach this task with genuine intentions (Dutton and Worline). Strong social bonds are not possible if both parties are not truly interested in building them. A sense of reciprocated genuine interest in each other can come from any type of interaction, whether it be short or long, work-related or personal, etc. Either way, these interactions are the building blocks of a successful and happy bond. Some simple ways to enrich high-quality connections at work are to reach out to others when they are in need, even in seemingly insignificant ways, such as holding a door for someone. Another way to build high-quality connections is simply showing people that they are valued by showing support, giving verbal validation, or making small gestures like wishing someone a happy birthday. Acts that build up trust amongst co-workers (e.g., following through on promises related to deadlines) also enrich co-worker bonds. Lastly, to create high-quality connections at work, research shows that employees should "play," or simply have fun together. For example, in a story featured in a lecture given at the University of Michigan, to make her daily duties more enjoyable, a hospital custodian re-defined her typical routine of cleaning the facility to include "play" by making the patients laugh. Therefore, at work, people can practice building strong relationships between co-workers, leaders, or even customers by re-thinking their roles and routines to include these positive interactions. Building these social connections at work is a fantastic way to increase

employee engagement, since people will naturally be more excited to work when they are amongst peers with whom they feel comfortable.

When many employees foster high-quality connections with co-workers, this builds an organizational culture characterized by high social capital, which increases employee engagement (Heffernan). Leaders and workers alike can create this climate. When organizations have high social capital,workers feel comfortable communicating with their bosses and peers, which increases creativity, innovation, productivity, and psychological safety. One reason for these beneficial outcomes is that social capital makes people feel safe to share their ideas, and that they will be met with respectful, interesting debate, if they are met with disagreement. Social capital is also beneficial to organizations because it teaches teams resilience, which further strengthens their bond. This begins a positive cycle of growth that reaps many benefits for individuals and the entire organization.

Conclusion

When used effectively, these methods of revitalizing employee engagement by implementing positive leadership, curating conditions for increased innovation and creativity, and building a culture that encourages social capital and high-quality relationships can reap benefits in profit and employee satisfaction for the organization as a whole. The workplace has untapped potential as a source of happiness, and this enjoyment is by no means unattainable. With time, the utilization of these tactics can usher in an age of personal as well as financial success for any organization willing to reform.

Works Cited

Cameron, Kim "Preface." *Positive Leadership: Strategies for Extraordinary Performance*, by Kim S. Cameron, Berrett-Koehler Publishers, 2012, pp. ix-15. Accessed 17 July 2023.Dutton, Jane, and Monica Worline. “Putting High Quality Connections into Practice - Positive Links Speaker Series.” Youtube.com. 20 Feb. 2023, Ann Arbor, Ross School of Business, https://www.youtube.com/watch?v=G-jVvvCr3Bw. Accessed 31 July 2023.

Harter, James K., et al. “Business-Unit-Level Relationship Between Employee Satisfaction, Employee Engagement, and Business Outcomes: A Meta-Analysis.” *Journal of Applied Psychology,* vol. 87, no. 2 (2002): 268–79. https://doi.org/10.1037/0021-9010.87.2.268.

Heffernan, Margaret. "The Secret Ingredient That Makes Some Teams Better Than Others." *TED*, 5 May 2015, ideas.ted.com/the-secret-ingredient-that-makes-some-teams-better-than-others/. Accessed 26 July 2023.

Hill, Linda A., et al. "What Makes a Great Leader?" *Harvard Business Review*, 19 Sept. 2022, hbr.org/2022/09/what-makes-a-great-leader. Accessed 19 July 2023.

How to Enhance Employee Engagement and Commitment. Performance by Marcus Buckingham and Julie Devoll, Harvard Business Review, 2022. *Harvard Business Review*, hbr.org/webinar/2022/05/how-to-enhance-employee-engagement-and-commitment. Accessed 25 Aug. 2023.

Kohs, S.C, and K.W Irle. "Prophesying Army Promotion." *Journal of Applied Psychology*, vol. 4, no. 1, Mar. 1920, pp. 73-87. https://doi.org/10.1037/h0070002. Accessed 15 July 2023.

Lord, Robert G., et al. "Leadership in Applied Psychology: Three Waves of Theory and Research." *Journal of Applied Psychology*, vol. 102, no. 3, 2017, pp. 434-51, https://doi.org/10.1037/apl0000089. Accessed 29 Sept. 2023.

Mind Tools Content Team. "Strengths-Based Leadership." *MindTools*, www.mindtools.com/alktajz/strengths-based-leadership. Accessed 18 July 2023.

Reichard, Becky. "Leadership Theories." *Prezi*, 1 Sept. 2021, prezi.com/fvus6aq8iugj/leadership-theories/. Accessed 18 July 2023.

Schaufeli, Wilmar B., et al. "The Measurement of Engagement and Burnout: A Two Sample Confirmatory Factor Analytic Approach." *Journal of Happiness Studies*, vol. 3, no. 1, 2002, pp. 71-92, https://doi.org/10.1023/a:1015630930326.

Seppälä, Emma, and Kim Cameron. "Proof That Positive Work Cultures Are More Productive." *Harvard Business Review*, 1 Dec. 2015, hbr.org/2015/12/proof-that-positive-work-cultures-are-more-productive. Accessed 14 Aug. 2023.

Tarricone, Pina, and Joe Luca. "Successful Teamwork: A Case Study." *HERSA*, July 2022, pp. 640-46, www.researchgate.net/publication/228348698_Successful_teamwork_A_case_study. Accessed 29 Sept. 2023.

West, Michael A., and Claudia A. Sacramento. "Flourishing in Teams: Developing Creativity and Innovation." *Creative Management and Development*, 2006, pp. 25-44, https://doi.org/10.4135/9781446213704.n3. Accessed 29 July 2023.

"What Is Leadership? Understanding Different Leadership Styles." *The National Society of Leadership and Success*, www.nsls.org/understanding-different-leadership-styles. Accessed 11 July 2023.

WWI and Women's Fashion by Emma Lu

Abstract

In England, the period between 1900 and 1925 witnessed significant transformations in both women's societal standing and her closet. Previous study has often focused on the progression of women's fashion from Victorian and Edwardian England straight into 1920s flapper gowns. However, a key player in this style transformation was World War I, which invited women to participate in roles beyond the domestic sphere, prompting more practical styles of dress. This essay examines World War I's impact on feminine fashion in Britain—specifically evaluating the evolution of men's influence on women's clothing—and ultimately concludes that the war helped significantly reduce womenswear's adherence to male preferences. This lack of male influence would allow dresses to become far more freeing and sensible in the decades following the war, a feature that was embodied by the rise of flapper dresses in the 1920s.

Introduction

Women's fashion can serve as a unique reflection of society during specific periods of time. For instance, during wartime, dresses tend to become simpler due lack of materials and time, and nowhere is this trend more clear than in Great Britain during the First World War. Contrasting the vibrant Victorian and Edwardian fashions of the century prior, the typical attire of English women dulled into a subdued and functional style between 1914 and 1918 to accommodate working women on the homefront. While fashion technically declined during these years, it is precisely this step back that created a revolution in the female wardrobe in the years following. The critical need for labor in World War I prompted men to surrender their authority—including dress authority—over women and bring them into the workforce, which warranted the simplification of women's wardrobe and ultimately laid the foundation for empowering flapper dress trends in the 1920s.

Fashion Before the War

To examine the Great War's influence on women's fashion, it is first necessary to scrutinize clothing trends and gender roles in Britain from before the war. The decades leading up to WWI are perhaps some of the most iconic in Western fashion history, housing both the Victorian era (1837 to 1901) and Edwardian era (1901 to 1910). During these times, women's fashion was highly dependent on the male perspective and reflective of unequal gender roles in English society. The unnatural, restrictive, and predominantly male-influenced nature of Victorian and Edwardian fashion trends reveal womenswear's lack of feminist influences in the 19th and early 20th centuries.

In the Victorian era, the fashionable silhouette was defined by a massive, bell-shaped skirt. These skirts could reach six yards in circumference (Yarwood) and called for six full layers of petticoats (Bowman). The sheer size and weight of gowns hindered women's movement in

their daily lives. So why wear a garment so supremely uncomfortable? These burdensome feminine dress codes can be traced to men's objectification of women. Victorian men saw fit to use their wives and daughters as ornaments to showcase wealth, which resulted in the popularization of extravagant but uncomfortable gowns for many high-class women (Ford). It is important to realize the role of men in deciding the magnitude of these skirts, as it reveals how womenswear was greatly dictated by men's wills. Victorian women dressed according to men's wishes, and womenswear of this time is reflective of unequal gender norms in English society.

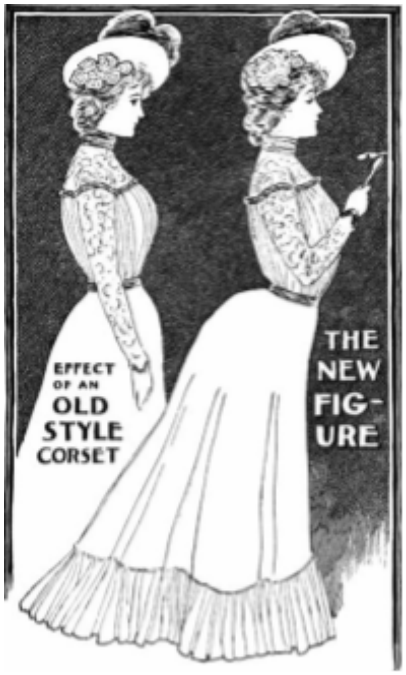


Fig. 1. *Illustration from 1900 Ladies Home Journal*

By the turn of the century, traditional Victorian ideals of a submissive, domestic wife were replaced with slightly looser and more progressive ways of thinking. This gave rise to a new ideal woman—the Gibson Girl—who would act as a model for many fashions of the decade. The Gibson Girl was flirtatious, sensual, and assertive with her suitors (Valente). However, despite the looser feminine ideals, women's clothing remained impractical and catered towards the male gaze. This phenomenon is well-exemplified by the popularity of the S-shaped corset in the early 1900s'. Influenced by the Gibson Girl's figure, Edwardian women began coveting an S-shaped silhouette, which manifested in a new corset style that pushed the bust forward and bum back (see Figure 1). This corset artificially altered a woman's figure, demonstrating a lack of acceptance for the natural female shape. Furthermore, it was uncomfortable, and limited women's mobility, continuing the trend of restrictive feminine garments through the Edwardian era (Fields). It is also important to note that the image of the Gibson Girl was created not by a woman but by a man—Charles Dana Gibson (Kuspit). Thus, the S-shaped corset represents women's efforts to comply with a man's vision of the perfect woman, and cannot be said to speak for women's personal preferences in clothing. This demonstrates a continuity in men dominating women's fashion, as had been observed in Victorian England.

Additionally, hobble skirts came into style briefly in the early 1910s. These skirts featured "an extremely narrow hemline around the ankles that inhibited walking" (Fields). Again, this trend represents a continuation of ridiculous dress codes in feminine fashion. The garment was so binding and inconvenient, in fact, that it was dangerous for women to simply walk in these skirts. A middle-aged woman named Ethel Hawksley Linley once fell while walking in a hobble skirt, "broke her ankle, and subsequently died in hospital of the shock" (Bowman). The extremity of the consequences of wearing the hobble skirt—which was defined by death in the case of Miss Linley—highlights how truly absurd the style was. It is also important to note that the hobble skirt is often credited to designer Paul Poiret, a man (Bowman).

Similarly to the Gibson Girl, the hobble skirt reveals how women's dress at this time was predominantly designed by or largely influenced by men. So, whilst the early 20th century witnessed a shift away from the opulent, heavy skirts of Victorian England, Edwardian women still had very little say in what they chose to wear. Whether they did so intentionally or not, women from the 19th to early 20th century often dressed in accordance with men's wills and imaginings, not their own.

During the War

When war broke out in 1914, fashion quickly became the least of concerns. With the desperate conditions of WWI, society no longer had the luxury to restrict women or their clothing; women were integrated into the war effort, taking up jobs that promoted novel, practical dressing styles.

With the war consuming the majority of people's livelihoods, men no longer had the energy or means to idealize women in the manner they had in Victorian and Edwardian England. Instead, necessity propelled men to recruit women into the war effort. Figure 2 is a British recruitment poster from WWI, wherein three women are depicted in the image and "Woman Clerks Wanted at Once" is printed boldly in the title. The desperate tone and wording of "wanted at once" reveal how the dire circumstances of WWI necessitated women's involvement in various war efforts. The stark contrast between this poster and the Gibson Girl exemplifies how women's value shifted from being beautiful to useful. Additionally, during this time women began working jobs that were traditionally reserved for men, including as physicians, factory workers, and bus drivers ("Women in WWI"). As men were sent to the frontlines of war, the vacancies they left in domestic labor created opportunities for women to step in. As a result, within just a few years, the war prompted a movement towards gender equality in the workforce, and women's dress reform would soon follow suit. How could it not? Women needed mobility to work effectively and safely in their jobs—especially in places such as munitions factories where workers were expected to operate heavy machinery (Edwards). If there failed to be a swift normalization of more functional women's attire, it was not just women who would feel the consequences, but rather the entire war effort. Thus, WWI propelled dress reform at a pace rarely seen before.



Fig. 2. *Women Clerks Wanted*, 1914/1918

Additionally, the war promoted the conservation of materials and fabrics, further enforced by fabric rationing (Edwards). Clothes were made simpler with less fabric, and styles

like the trench coat—its name inspired by WWI's trench warfare—became fashion staples (Cole). A clear contrast can be observed between the trench coat (see Figure 2) and popular styles in earlier decades. The trench coat fails to sexualize women and is also practical to move in as opposed to the aforementioned S-shaped corset and hobble skirt. The popularization of the trench coat represents how necessity from war inhibited the continuation of impractical, male-influenced garments in women's fashion. As such, while the First World War technically limited the resources and time available for fashion, it actually instigated a positive change in women's clothing, which was further supported by women's increased importance on the homefront in various jobs and professions.



Fig. 3. *Day Suit*, 1920

After the War

By the time the war came to a close in 1918, women had experienced various new freedoms in both career and clothing. The introduction of utilitarian garments during WWI in conjunction with women's mounting sway in society opened the door for continued progressive changes in women's attire, which culminated in the emergence of 1920s' flapper fashion.

First, it is quite incredible to see how WWI's influences on women's fashion extended far beyond just the wartime period. For example, the brown suit-coat shown in Figure 3 "represents the quintessential style of the 1920s and marks a transitional moment in fashion, as the silhouette changed" ("Day Suit"). This suit exhibits very few similarities with women's fashion from Victorian or Edwardian England, indicating a clear shift away from those styles. On the other hand, multiple elements of this suit parallel those of the trench coat, such as the loose boxy silhouette and short hemline. This resemblance hints at the existence of lasting influences from the war's utilitarian closet, which would eventually evolve into flapper dresses by the 1920s. Flapper dresses featured a rectangular shape, drop-waist, short hemline, and often no sleeves (see Figure 4). The dresses were light and easy to move in, and marked huge milestones for women's dress reform in various ways: women cast aside their corsets and, for the first time in centuries, displayed their bare legs in public. The latter in particular was a significant change; Western women had not been able to acceptably show their legs for multiple centuries prior to WWI (Edwards). The war played a considerable role in driving forward this



Fig. 4. *Evening Flapper Dress*, 1926

reform, as it catalyzed the initial emergence of more lightweight and convenient designs in women's fashion.

Another reason for the rise of the flapper dress can be found in the job market. Following the war's end, women became concerned that they would be re-delegated to the domestic sphere. To fight against this possibility, women began dressing in a more masculine manner, trying to announce that "women's bodies were fit for public view and physically capable too" through their dress choices (Ford). This notion influenced many fashion trends of the decade and is one of the core reasons for the flapper dress' drop-waist and flat-chested look, which many people labeled as "boyish" (Yarwood). Women's desire to fight for their jobs would not have emerged without an initial taste of work, which was presented to them due to the war. Thus, the rise of flapper fashion is linked to the freedoms that women first enjoyed during the wartime period.

Beyond acting as women's entrance to the domestic labor force, WWI's progression towards workforce equality also inspired a level of women's empowerment and feminist thinking, which can be witnessed when flapper fashion is examined in context with men. Many men were not supportive of the flapper trend, even attempting to implement dress codes that limited skirts to a maximum of twelve inches off the ground. Some articles blamed the "unfeminine" style for causing divorces and used it to justify acts of domestic violence (Ford). Nevertheless, flapper fashion gained momentum, becoming a trend that truly belonged to women. Its widespread popularity despite male criticism is indicative of women's confidence in post-war society. Especially when compared to pre-war trends such as the S-shaped corset and hobble skirts, which were largely resultant from the musings of men, the flapper dress is a significant landmark in women's journey towards feminine empowerment.

However, the flapper dress was not all positive. To achieve a more boyish silhouette, some women bound their chests with large strips of cloth (Ford). Despite the advances that WWI brought to women's fashion, society had yet to fully embrace the natural female figure. The road to dress reform still had a ways to go, but at least WWI launched the movement onto its path.

Conclusion

World War I served as an effective catalyst for women's dress reform, dragging the female closet from restrictive machineries to freer flapper gowns. During the process, men lost a significant amount of influence over women's fashion, and their influence would continue to diminish through the 20th century. As women began fighting not just for dress reform, but also for women's suffrage and equal rights, their clothing became a tool for feminism, giving rise to a variety of unique styles that would eventually join the iconic flapper dress in the historical lineup of empowering feminine fashion.

Works Cited

Bowman, Karen. *Corsets and Codpieces: A History of Outrageous Fashion.* Skyhorse Publishing Company, Incorporated, 2016.

Cole, Daniel James, and Nancy Deihl. *History of Modern Fashion: From 1850.* Laurence King Publishing, 2015.

Day Suit. The Metropolitan Museum of Art, New York. https://jstor.org/stable/community.14796334.

Edwards, Lydia. *How to Read a Dress: A Guide to Changing Fashion from the 16th to the 21st Century*. Bloomsbury, 2022.

Evening Dress. The Metropolitan Museum of Art, New York https://jstor.org/stable/community.14813388.

Fields, Jill. "'Fighting the Corsetless Evil': Shaping Corsets and Culture, 1900-1930." *Journal of Social History*, vol. 33, no. 2, 1999, pp. 355–84. *JSTOR*, http://www.jstor.org/stable/3789627. Accessed 25 Sept. 2023.

Ford, Richard Thompson. *Dress Codes: How the Laws of Fashion Made History*. Simon & Schuster Paperbacks, 2022.

Illustration from 1900 Ladies Home Journal of Victorian to Edwardian corsets. 2012. Collector's Weekly, https://www.collectorsweekly.com/articles/everything-you-know-about-corsets-is-false/. Accessed 29 Sept. 2023.

Kuspit, Donald B. "Charles Dana Gibson's Girl." *Jahrbuch Für Amerikastudien*, vol. 7, 1962, pp. 183–87. *JSTOR*, http://www.jstor.org/stable/41155010. Accessed 24 Sept. 2023.

National Service : Women Clerks : Wanted at Once for Service in France with the British Army : For Conditions of Service and Application Forms : Apply Women's Section National Service Department, : St. Ermin's Westminster, S. W. I. University of Minnesota Libraries. https://jstor.org/stable/community.12312385.

Valente, Jose . "Charles Dana Gibson." American History, ABC-CLIO, 2023, americanhistory.abc-clio.com/Search/Display/255547. Accessed 24 Sept. 2023.

"Women in WWI." *National WWI Museum and Memorial*, 4 Feb. 2022, https://www.theworldwar.org/learn/women.

Yarwood, Doreen. *Fashion in the Western World: 1500-1990*. Drama Book Publ., 1992.

Evaluating the Correlation of Different NBA Factors on Team Success By Ishan Katpally

Abstract

National Basketball Association (NBA) teams and their success are affected by many different parts of the game. This paper aims to evaluate what part of NBA basketball is most correlated to a team's success. We test with statistics that represent some of the biggest parts of the game, including defense, offense, shooting, passing, and rebounding. From the 2003-2022 seasons, we analyze the correlation for each of these statistics to performance by computing Cramer's V, which quantifies the correlation on a scale of 0 to 1, with 0 meaning no correlation and 1 meaning complete correlation. We find that defense has been the most correlated to a team's success over the past 20 years with an average value of 0.546, but offense and rebounding have been increasingly correlated in recent years while defense has been going down in correlation.

Introduction

The National Basketball Association (NBA) is one of the most popular sports leagues in the US and is gaining popularity internationally as well. In each NBA season, teams play a regular season that consists of 82 games in order to determine the 16 teams playing in the playoffs. The playoffs are then carried out in a single elimination bracket, with each round being a best of 7 games, and the winner is crowned champion of the NBA.

While the premise of NBA basketball might seem simple, there are many parts of the game that add to its complexity. There isn't one concrete formula that teams follow in order to win a championship. Teams are constantly rethinking their strategy, making changes to their roster from season to season. Some teams may be more focused on having more offensive production, while others may value defense more. Given the amount of variables that have the potential to affect NBA teams, a question arises: what part of NBA basketball is most correlated to making a successful NBA team?

The aim of this paper will be to analyze the results and statistics from previous NBA seasons and test the correlation between a team's skill set in one aspect of the game and how successful they were that season. Among the different parts of the game, some of the most noticeable ones are how much more a team can score compared to other teams, how much more a team allows points compared to other teams, shooting efficiency, the ability to pass and take care of the ball, the ability to rebound, and how much a team has already won with their roster in the regular season. In order to quantify the correlation between performance in each of these statistics and performance in the NBA, the Cramer's V value will be computed for each of these statistics from 2003-2022. This involves computing the Chi-Square value using the scipy library in Python and using this value to compute Cramer's V. Cramer's V will have a value between 0 and 1, with 0 meaning no correlation and 1 meaning complete correlation. Averaging this value over the past 20 years for each statistic will show the correlation of each of these statistics to performance in the NBA.

Methodology

To find out the area of play that most influences success in the NBA, certain statistics need to be tested for correlation for success throughout past years in the NBA. The first is Offensive Rating, which measures a team's overall offensive capabilities with the amount of points the team scores per 100 possessions (National Basketball Association). Similarly, Defensive Rating measures a team's overall defensive capabilities with the amount of points the team allows per 100 possessions (National Basketball Association). The next one is Assist/Turnover Ratio (AST/TO), which takes the total number of assists from a team and divides it by the total number of turnovers (National Basketball Association). This statistic measures a team's capability to create offense by passing the ball but also factors in ball safety. If a team tries to do a lot of passing plays but turns the ball over more often than making a successful play, it is more hurtful to the team's offense and overall success, and this is reflected in the AST/TO stat, as the higher amount of turnovers will make the value go down. The next statistic that will be tested is Rebounding Percentage, which is calculated by dividing the total rebounds from the team by the total rebounds from all the team's games (National Basketball Association). Rebounding is a possible factor in NBA games because it limits points for the opposing team from offensive rebounds and second-chance points, and it also gives more offensive opportunities for your team by providing more possessions. Another statistic that will be tested is the Effective Field Goal Percentage (EFG%), which is calculated by taking the total field goals made, adding half of the 3 point field goals made, and dividing by the total field goal attempts. This statistic measures overall shooting efficiency from a team, but it also accounts for a team's 3 point shooting (National Basketball Association). With the rise in 3 point shooting in today's NBA, accounting for a team's 3 point shooting is becoming increasingly important, so EFG% is an accurate measure of a team's overall shooting efficiency. The final statistic will be the amount of wins a team got in the regular season. Due to the way the playoffs are set, a team is usually required to win close to half of their regular season games to have a chance of winning the championship, and it is impossible for a team to win if they do not win enough in the regular season. It follows that regular season wins should have a high correlation to a team's success, so this statistic will be used as a comparison point with the rest of the statistics. Together, these statistics constitute a well-rounded analysis for what is correlated to an NBA team's success. All of the statistics were taken from the official NBA Advanced Statistics page (National Basketball Association).

To measure the correlation between each of these statistics and success in the NBA, we used Cramer's V. Cramer's V is a value between 0 and 1 that measures the correlation between two categorical variables, with a value between 0 and 0.2 considered a weak correlation, between 0.2 and 0.6 a moderate correlation, and above 0.6 a strong correlation (International Business Machines). In this case, the two variables were the end result variable, or how well the team performed in the season, and the stat ranking variable, or how good a team's performance in the tested stat compares to the other teams. Therefore, in order to use this measure of correlation, the variables need to be made into categorical variables. The two variables were the end result of a

team in the season and how the team ranks in the given statistic compared to every other team. To make these variables into categorical variables, the end result variable had 3 categories and the stat ranking variable had 6 categories. For the end result variable, the 3 categories were top 4, made playoffs, and missed playoffs. If a team made it to the conference finals or finals, it was in the "top 4" category. If a team made the playoffs but did not make it as far as the conference finals, it was in the "made playoffs" category. If a team missed the playoffs altogether, it was in the "missed playoffs" category. With the team ranking variable, the 6 categories were rank 1-6, rank 7-12, rank 13-18, rank 19-24, and rank 25-30. Some of the seasons have only 29 teams playing, so in these seasons, the 6th category was rank 25-29 instead. If a team ranked from 1-6 in the tested statistic when compared to the other teams in the league, they were in the "rank 1-6" category, and the same logic applies for the other five categories.

To compute Cramer's V, an intermediate value, Chi-Square, was first computed. For each cell in the 3x3 grid we have created, we add (observed value - expected value)^2 / expected value. Adding this value up over all the cells in the grid, we get the Chi-Square value. From there, we calculated Cramer's V by dividing by the total number of teams, dividing by min(r-1, c-1), where r is the number of rows and c is the number of columns, and taking the square root. This gave a value between 0 and 1. 0 indicates that there is no correlation between the two variables, and 1 indicates that there is a complete correlation.

To find this value efficiently for many different years and stats, we used the scipy.stats library in Python (Virtanen *et al.*, 2020). This library has a function called chi2_contingency that can take in a two-dimensional array and calculate the Chi-Square value for it. Thus, to calculate the Chi-Square value, a 3x6 two-dimensional array was used, with the columns representing the end result variable and the rows representing the stat ranking variable. Each cell in this array was a number representing the number of teams that correspond with the variables. For example, if two teams that finished the season in the top 4 were between ranks 7-12 in the tested stat, then the cell that corresponds to the "top 4" column and the "ranks 7-12" row had a value of 2.

To populate this two-dimensional array, the final results of the season were kept in a file, where the top 4 teams are in the first 4 lines, the teams that made the playoffs are in lines 5-16, and the teams that missed the playoffs are in the remaining lines. The teams were then made into a separate order that represents the performance in the tested stat (eg: rank 1 is listed first, rank 30 is listed last). From there, we can iterate through this list. For each team, we found what category it is in for the end result variable, and we kept track of what the rank is so we could figure out which category it is in for the stat ranking variable. From there, we simply indexed into the array in the appropriate spot and added one to the current value. Once the entire array was populated appropriately, the chi2-contingency function gave a Chi-Square Value.

Once the Chi-Square value was calculated, we calculated Cramer's V with the process described above. Divide by the total number of teams, divide by min(rows-1, cols-1) which is 2 in our case, and take the square root. This process resulted in a number between 0 and 1 that quantifies the correlation between the tested stat and a team's success in the NBA. A result of 0 means there is no correlation, and a result of 1 means there is a complete correlation.

For each of the 6 statistics, the value of Cramer's V was calculated from the years 2003-2022, yielding 20 years total. From there, we evaluated the correlation to success by taking the average over the 20 year period, and we also analyzed any possible trends or changes over time to the result.

Results

Table 1 shows average quantities for the Cramer's V values from the 2003-2022 seasons. A Cramer's V value below 0.2 is considered a weak correlation, between 0.2 and 0.6 a moderate correlation, and above 0.6 a strong correlation. The average value for Offensive Rating is 0.526, indicating a moderate to strong correlation between a high offensive rating and performing well in the season. The average value for Defensive Rating is 0.546, once again indicating a moderate to strong correlation and slightly higher than that of the Offensive Rating. Regular Season Record has an average value of 0.756, showing a strong correlation. The value for Assist/Turnover Ratio is lower at 0.454, and the value for rebound percentage is even lower at 0.421, showing moderate correlations for both statistics. Effective Field Goal Percentage has a value of 0.508, which is a moderate to strong correlation, though not as strong as Offensive Rating or Defensive Rating.

Average Cramer's V values over the 2003-2022 seasons

Statistic	Average Cramer's V Value
Offensive Rating	0.526
Defensive Rating	0.546
Regular Season Record	0.756
Assist/Turnover Ratio	0.454
Rebound Percentage	0.421
Effective Field Goal Percentage	0.508

Table 1. Average value of Cramer's V from 2003-2022 for each of the 6 tested statistics

Figure 1 depicts the progression of Cramer's V values for Offensive Rating. The values appear to be gradually increasing in the more recent seasons. The lowest value is in the years 2004 and 2005, while the highest value is in 2016, followed closely by other recent years in 2018, 2019, and 2022.

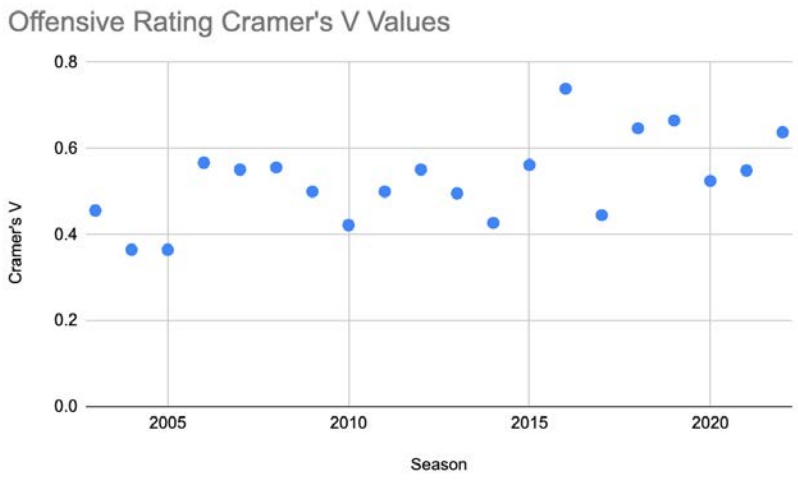


Figure 1. Yearly progression of Cramer's V values for Offensive Rating

Figure 2 shows the progression of Cramer's V values for Defensive Rating. From 2008, where the value was at its highest, the values are gradually decreasing in recent years and going back to a similar level as that of the years before 2008. There are also some lower values in recent years like 2015, 2017, and 2021.

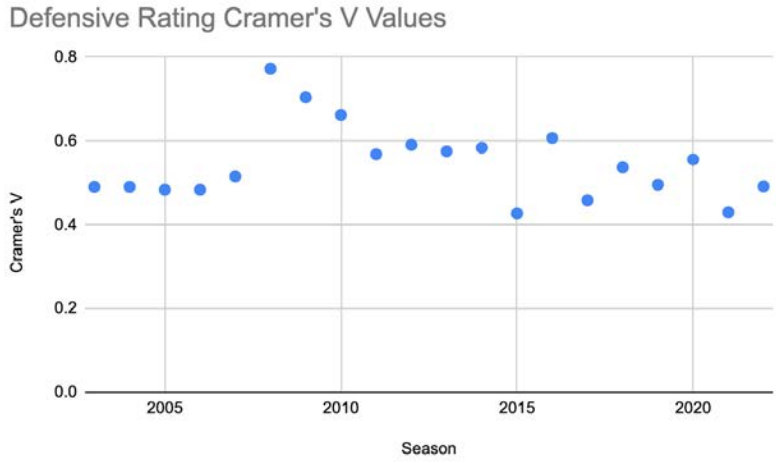


Figure 2. Yearly progression of Cramer's V values for Defensive Rating

Figure 3 shows the yearly Cramer's V values for the Regular Season Record statistic. The values for this statistic have been consistently around the 0.75 mark, with no major low or high outliers.

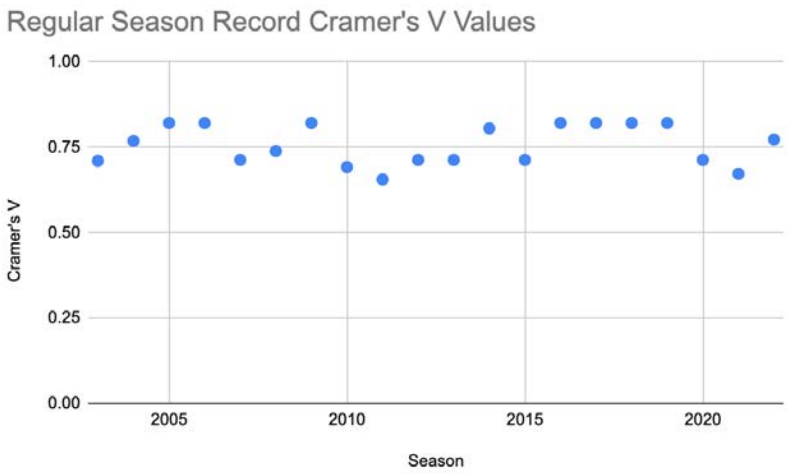


Figure 3. Yearly progression of Cramer's V values for Regular Season Record

Figure 4 highlights the yearly values of Cramer's V for the Assist to Turnover Ratio. The values are mostly around the 0.4 to 0.5 range, but there are several outliers. Most notably, 2013 had a value around 0.71, and 2012 had a value around 0.3. The values do not seem to follow a particular upwards or downwards trend over the years.

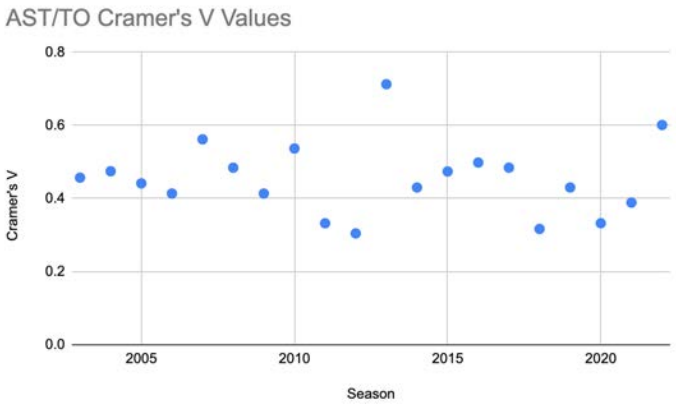


Figure 4. Yearly Progression of Cramer's V values for Assist to Turnover Ratio

Figure 5 depicts the yearly progression of Cramer's V values for the Rebound Percentage. The earlier years feature inconsistency, with both high and low values showing up. The values do not seem to follow a clear trend in the earlier years, but there is an upwards trend in recent years starting from around 2017.

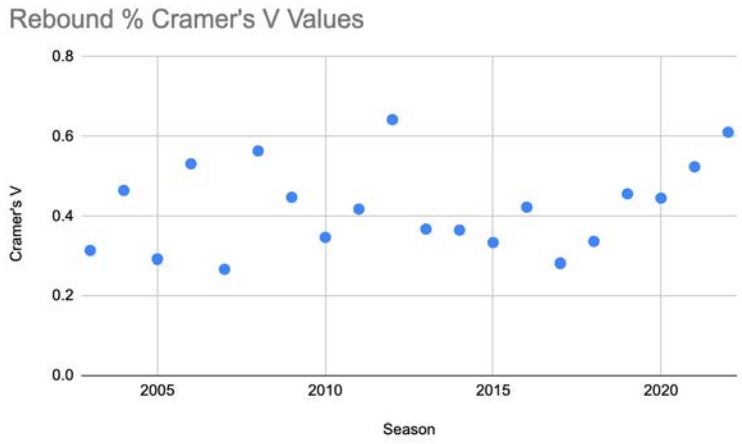


Figure 5. Yearly progression of Cramer's V values for Rebound Percentage

Figure 6 shows the yearly Cramer's V values for Effective Field Goal Percentage. The values are consistently between 0.4 and 0.6, with no values falling outside this range. There is also no clear pattern or trend with the values, as they are only varied within the smaller range of 0.4 to 0.6.

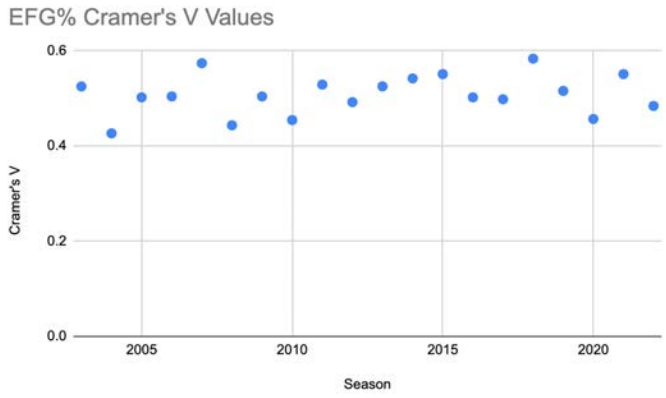


Figure 6. Yearly progression of Cramer's V values for Effective Field Goal Percentage

Discussion

Out of the 6 statistics, the regular season wins statistic acts as a point of comparison. Teams need to win a certain amount to receive a high enough seeding in the league to make it to the playoffs. If a team were to rank last in wins in the regular season, it would be impossible for them to win the championship. Therefore, winning enough games in the regular season is required to win a championship, and the teams that do not win enough are guaranteed to be in the "missed playoffs" category. This relationship suggests that regular season wins should have a high correlation to success. This assertion is supported by the average Cramer's V value for

regular season wins of 0.756. We can evaluate the impact of the other statistics by comparing their Cramer's V values to that of regular season wins.

We can see that the value for regular season wins is considerably higher than the values for the other 5 statistics, meaning that none of them have a high correlation to success. In other words, being excellent in only one of these fields is not enough to have a good chance of winning the championship. To be successful, teams need to cover all parts of the game well.

One of the more notable observations is that the value for defensive rating is slightly higher than the value for offensive rating, meaning that over the past 20 years, defense has been slightly more correlated to success than offense. However, when examining the progression for each of these statistics, we can see that in recent years, the correlation of defense has gone down when compared to previous seasons, especially 2008-2014. On the other hand, the correlation of offense has been going up in recent years. This may indicate a rise in offensive production in recent years that would cause offensive teams to find more success, increasing the correlation of offense to success. Although defense has had a higher correlation overall for the past 20 years, offense is set to take on a higher correlation in the future.

Of the remaining 3 statistics, effective field goal percentage has the highest Cramer's V value, showing a higher correlation for being able to score efficiently rather than distributing and taking care of the ball or being able to rebound over the other team. Even with the higher volume of 3 point shooting in recent years, it has stayed relatively consistent over the past 20 years. On the other hand, the assist to turnover ratio and rebounding percentage have been less consistent, with outliers on both ends of the spectrum, high and low. One could say that due to this inconsistency, passing, taking care of the ball, and rebounding might be less correlated than shooting, defense, and offensive production as a whole, and the average values for AST/TO and rebounding percentage support this. However, specifically for rebounding percentage, the value has been going up in recent years. It is hard to say definitively whether this represents a trend for the future or just more inconsistency, but an answer may come up with more data from future years.

With the data we have, we can answer the original question posed of what is most correlated to a successful NBA team. On average for the past 20 years, defense has been the most correlated to a successful team over offense, and out of shooting efficiency, passing and taking care of the ball, and rebounding, shooting efficiency has been the most correlated to a team's success. However, in recent years, offense and rebounding have been increasingly correlated, so this answer may be subject to change with more analysis in the future. Although it cannot be said that any of these parts of the game are direct causes of a team's success, their correlation values may be valuable assets in determining the likelihood of success for a given team in future seasons.

Conclusion

This paper aimed to evaluate which parts of NBA basketball are most correlated to a team's success in the NBA. Through statistical analysis of the 2003-2022 NBA seasons, defense was found to be the most correlated to a team's success, and out of shooting, passing, and

rebounding, shooting was found to be the most correlated out of the 3. By looking at the year-by-year progression for each of these statistics, defense had a slight decline in correlation in recent years while offense and rebounding had a slight increase in recent years, which may be a possible trend for future seasons.

There are numerous possible continuations for this analysis. The first would be to expand the analysis to more past years, going back to the 1990s, 1980s, and even 1970s. This would possibly give different average values for each statistic as well as uncover any similarities or progressions from past NBA seasons to present NBA seasons. Another continuation would be to test more statistics. There are many advanced statistics that the NBA keeps track of, and testing different ones might give a different perspective on the research question and might show different patterns over the years.

Acknowledgments

I would like to sincerely thank Won Park from the University of Michigan for guiding me throughout the creation and development of this research paper.

Works Cited

International Business Machines. (n.d.). Cramér's V. In ibm.com documentation. Retrieved July 20, 2022, from https://www.ibm.com/docs/en/cognos-analytics/11.1.0?topic=terms-cramrs-v

National Basketball Association. (n.d.). NBA Advanced Stats. In nba.com statistics page. Retrieved July 20, 2022, from https://www.nba.com/stats/help/glossary

National Basketball Association. (n.d.). NBA Advanced Stats. In nba.com Advanced Stats page. Retrieved July 20, 2022, from https://www.nba.com/stats/teams/advanced

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, *et al*. (2020) SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods, 17(3), 261-272.

The Effect and Solutions of Climate Change in Orange County By Amudhan Gurumoorthy

Abstract

Across Orange County, in Southern California, climate change has become a serious problem. Natural disasters, including fires, floods, and tsunamis, have always been a threat to Orange County, and intense heat and air pollution are now growing as major issues, with over 4000 wildfires striking the region each year. However, the government is not taking enough action to stop these threats, and without a strong effort, the climate will continue to get worse.

In this research paper, we will take a look at the problems with the environment in Orange County. We will first examine how the climate is affecting Orange County, through natural disasters, lack of biodiversity, and more. Then, we will look at different climate groups in Orange County, such as the Climate Action Campaign and the Surfrider Foundation, and what they are doing to stop this threat. After that, we will examine the steps the government is taking to prevent climate change, and whether it is working. Finally, we will look at some solutions both the government and individuals can take to stop this threat. These solutions can be as simple as joining climate groups and creating new laws, but also as complex as forming a full new climate action plan.

1: Introduction

The climate has become a significant threat to most of the world, and Orange County, California is no exception. Fires, floods, and air pollution are just a few of the problems that are harming this region. Through the constant warnings of natural disasters and environmental changes, the Orange County climate has become very frightening. This leads to an important question: knowing the drastic impact climate change will have on life in Orange County, is enough being done to address it?

Since I live in Orange County, I have always felt the personal impact of this issue. My entire life, I have lived with the dangers of natural disasters and other effects of the changing climate. Most significantly, I have been forced to evacuate my home due to an extreme fire that ravaged very close to me. I can still remember the fear I experienced while I was stuck in my car, bracing myself for the worst. Fortunately, none of these threats have had any significant impact on me so far, but I know that unless something is done, the climate will continue to get worse, and these disasters will happen more and more frequently.

Many people realize the dangers of climate change as well. However, despite all of these environmental problems, government-enacted laws and civilian organizations are not doing enough. Some groups are taking a stand to help fight this topic, but unless significant changes are made, this dangerous threat will continue harming the region. While many people in Orange County think climate change is a problem, most people, including city officials, are not willing to step up and confront this issue, and consequently, if this does not stop, Orange County's environmental crisis will continue to get worse.

First, however, it's important to understand how climate change is impacting the rest of California, and after that, zoom in on Orange County.

1.1: Climate Change across the Southwest Region

Similar to Arizona and Nevada, California is a part of the southwestern region of the United States. There have been many impacts across this entire region, leading to a lot of significant natural disasters. According to the Fourth National Climate Assessment, the Southwest has experienced major temperature rise, increasing by as much as 3℉ from 1960-2016 (NCA 2018). This has led to major droughts throughout California, causing many major food shortages. According to the National Climate Assessment, agricultural irrigation accounts for nearly three-quarters of water use in the Southwest region, which grows half of the fruits, vegetables, and nuts and most of the wine grapes, strawberries, and lettuce for the United States (NCA 2018).

Additionally, this heat has led to a significant increase in wildfires across the entirety of California. Across the last five years, California has experienced over 4.800 wildfires burning through 390,000 acres per year *(Cal Fire* 2023). That is over thirteen wildfires a day. The destructive impact of these wildfires is immense, and they will continue to grow in the future. (Projected increase in wildfires – data, as well as their destruction)

However, despite these huge issues, little action has been taken to stop these threats, as some Californians still believe that climate change is not a real threat. In fact, one Congressman, Tom McClintock, in California has stated that he doesn't believe in climate change. He thinks that "the climate has been changing for four and a half billion years", questioning the role humans play causing climate change (Worland 2018). Climate change across California has been more prevalent than ever, yet many people like McClintock are not doing anything to stop it.

1.2: An Introduction to Orange County

Orange County is located on the southern side of California. It is located south of Los Angeles and north of San Diego, right on the coast of the Pacific Ocean. Some of the biggest cities in Orange County include Anaheim, Irvine, Santa Ana, Fullerton, Huntington Beach, and more. This leads to Orange County having a population of 3.17 million people, the third most populous county in all of California. The importance of maintaining the climate in the region is huge, but sadly, many efforts are not being taken to address this.

2: Climate Change in Orange County

As stated before, Orange County has been heavily affected by climate change, through natural disasters and more. However, it's important to know the exact effects so that appropriate measures can be taken to stop these threats.

2.1: Natural Disasters in Orange County

Orange County is a region of many natural disasters. Firstly, since Orange County is right on the coast of the Pacific Ocean, sea level rise will become a main issue. As the ocean creeps

closer and closer to land, our beaches will erode until the ocean washes into buildings. Furthermore, if the ocean level rises high enough above the current mark due to immense precipitation, many cities in Orange County could face rapid flooding; “Flash floods can strike any time with little or no warning turning Orange County streets and freeways into rivers within seconds” (ReadyOC 2021). Additionally, earthquakes have always been a grave threat in the Orange County region. If a strong earthquake occurs in the Pacific Ocean that causes a massive wave to rise, tsunamis will wreck the region. Major hazard zones include the area between the Balboa Peninsula and Pacific Coast Highway, as well as Newport’s Back Bay. (Nguyen 2021).

Finally, and most importantly, wildfires have been a constant threat to Orange County. Similar to California, wildfires are the most frequent disasters across Orange County, with about 250 fires burning around 1,000 acres each year. Each of these fires is caused by the absorption of CO2 and other greenhouse gasses into the atmosphere and can burn fiercely through the region.

All these environmental disasters can be catastrophic, leading to the destruction of houses, the forced evacuation of people, and the loss of human life. Figuring out a way to prevent or reduce the damage of these natural disasters is critical to sustaining the region.

2.2: Climate in Orange County

Along with all of these natural disasters, the climate is influencing Orange County in many different ways. First of all, as mentioned before, extreme weather has become a huge issue in Orange County. In places like Anaheim and Santa Ana, temperatures hit record-breaking highs of 106- and 103-degrees Fahrenheit, respectively, in the spring of 2022 (Antonios 2022). Additionally, it began to snow in Orange County on March 1, 2023, a rare sight which has not been seen for about 10 years. This extreme weather has led to droughts and food shortages, affecting Americans all across the region.

Additionally, air pollution is starting to become a greater issue. While air pollution has not yet reached its peak, due to the release of pollutants from the exhausts of motor vehicles, particularly diesel trucks, air pollution is slowly growing in the region. Particulate matter is the most alarming pollutant, as it is slowly rising to a moderate level of pollution in the sky. If we do not prevent this, air pollution will continue to get worse, leading to mass amounts of particulate matter in the sky, such as dust, dirt, soot, or smoke, that can go deep into a person’s lungs and heart.

2.3: Loss of biodiversity

There are also many other effects of human-induced climate change across the county. First of all, Orange County has been heavily growing in its population. This forces other plants and animals to change the way they live, decreasing the biodiversity within the region. For example, a study done on the seaweed in Orange County revealed that “during the past fifty years, the human population increased greatly throughout southern California, resulting in extensive modifications of the coastal landscape” (Murray et al 2008.). This change in the coast has resulted in new formations of seaweed, leading to “the abundances of disturbance-resistant seaweeds have increased and of large, fleshy species have decreased” (Murray et al 2008.). This

phenomenon does not only apply to seaweed, though; other species of plants and animals have also been forced to change or leave the region due to increases in human population, creating a lack of biodiversity within Orange County.

2.4: Human Ignorance

All of these changes have led to significant problems, but one of the biggest problems is the lack of human involvement. People in Orange County are aware of these threats, yet they are not doing enough to stop them. In order to prevent more climate change from taking place, we need to take a stand now.

The topic of climate change has led to a huge divide within Orange County, with many groups advocating for preventing it, and other groups advocating against it. In the next section, we will look at some of those groups, seeing their justification for fighting for or against it and looking at some of the things they have done.

3: Groups for vs groups against

Despite the risks stated above, there is still a lot of conflict in Orange County about climate change. According to a 2020 survey by Chapman University, 73% of people believe that climate change is real, an increase of 19% from 2010. Additionally, 66% of people are worried about climate change, and 63% of people believe climate change will affect generations to come. Groups are forming both for and against this climate initiative, trying to plead their case on why climate change should be acknowledged or neglected.

3.1: The Climate Action Campaign

One of the largest groups advocating for fighting climate change is the Climate Action Campaign. Originally started in San Diego, this group saw the importance of maintaining the climate in high-risk cities in Orange County, like Santa Ana and Anaheim. They are dedicated to creating a zero-carbon Orange County. Through their efforts, they helped Irvine commit to carbon neutrality by 2030; helped establish Orange County Power Authority, which was the region's first Community Choice Energy Program; and helped implement climate action plans for many cities in Orange County (Climate Action Campaign 2018). They are the biggest group currently operating in Orange County fighting this grave issue, and they promise to continue to fight climate change until new policies are enacted.

3.2: Other Climate Activist Groups

In addition to the Climate Action Campaign, there are a few other groups who are advocating for a better climate within Orange County. One of these groups is the Surfrider Foundation, a group dedicated to protecting the coasts from coastline erosion. According to Steganie Sekich-Quinn, the coastal preservation manager for the Orange County Surfrider chapter, "it's important for the group to focus on the ocean and coast because it doesn't get much attention when it comes to climate change" (Brazil 2022). They have helped work with the city

of San Clemente to create a coastal resiliency plan. They are also trying to push for better solutions in places like Capistrano Beach, which only currently has boulders and sandbags to deal with rising tides.

Another group that is working hard against climate change is the Orange County Environmental Justice group. Instead of focusing on the broader community, this group has been focusing more on protecting those from the environment within lower-income regions. They have made the greatest impact on Santa Ana, where they discovered lead-contaminated soil across poor neighborhoods within the region (Brazil 2022). Their efforts have helped Santa Ana become much more climate-friendly.

3.3: Climate Action Rejectors

While many groups are fighting back, there are still people who are wary of fighting this issue. Some of these people just do not believe that climate change is real. According to a Yale survey, 77% of Californians think climate change is real. However, in Orange County, only 66% of people think that climate change is a serious threat, a much smaller value than that of California (Chapman 2020). This shows how some people living in this region will not fight back because they do not think the issue exists.

However, other people recognize the existence of climate change, but they do not think it is worth fighting against it. For some people, they think that fight would be more expensive to be worth it. Others simply think that it is not an issue they are willing to strongly fight against. Finally, some people just don't know how to get involved in this fight.

Of course, there is no one that is openly expressing this, keeping these climate thoughts to themselves. However, that still does not change their beliefs. While there are a lot of climate action advocates, it's important to recognize that this other side exists as well, and these people must also step up to protect our climate.

4: Government efforts to stop climate change

While individual groups can help stop climate change, ultimately, the government has to create public support to solve this issue. Some cities in Orange County are creating their own climate plans to stop climate change, doing things like monitoring vehicle use and funding environmental protection movements. Some of these plans are working very well, but a lot of these plans are falling behind. In this section, we will take a look at some of these climate change plans, examining what different cities are doing to stop this huge issue.

4.1: Irvine

Greenhouse gases are the biggest facilitators of global warming. One of the most dangerous greenhouse gases is carbon, and much of the carbon in the atmosphere is released through human actions. Irvine, California, has decided to commit to a no-carbon plan by 2030, fifteen years earlier than California's no-carbon by 2045 plan. To do this, Irvine created their

Climate Action and Adaptation Plan, also known as the CAAP. This Irvine organization will help build towards the creation of a climate-smart city in the future.

However, this Climate Action and Adaptation Plan has ultimately done little to address the issue of climate change. They have recommended that action should be taken, promising to take some action themselves, but currently, no measures have been taken. In order for the CAAP to work, they must not only set goals, but actually accomplish those goals.

To do this, the CAAP has begun meetings to discuss how they will make Irvine a no-carbon city. At their city council meeting on June 27th, 2023, the Irvine committee developed many methods to achieve this no-carbon goal. Some of these include increasing renewable energy and eliminating natural gas use in new construction, expanding bicycling and walking paths while creating low-CO2 emission vehicles, and implementing new waste reduction measures (Sean and Chi 2023). However, since it was so recent, none of these methods have been implemented yet, and it remains unclear whether these solutions will actually work.

4.2: Santa Ana

Another city in Orange County, Santa Ana, has a completely different approach to the climate problem. Unlike Irvine, Santa Ana faces a heavy amount of social imbalance, with many low-income neighborhoods throughout the city. As stated before, environmental issues will harm the lower class much more than the higher class, so stopping climate change is a social justice issue as well as an environmental issue. Therefore, the Santa Ana government is trying to be more proactive in this fight, trying their hardest to ensure the divide between poor and rich neighborhoods does not deepen further.

In December 2015, Santa Ana created a Climate Action Plan. trying to prevent climate change within the region from becoming worse. In this plan, they have explained the goals and measures they are taking, the way they are implementing those measures, and solutions a person in Santa Ana can take in their own home (Santa Ana 2015). This climate action plan has been updated a few times, but not fully remade into a new plan. However, it provides a basic foundation, showing the commitment the Santa Ana government has towards creating a safer environment.

Additionally, Santa Ana has used a significant amount of money to solve these issues. They have spent about $2.6 million per year on their tree programs, trying to reduce the amount of CO2 in the atmosphere (Hicks 2021). In its 2024 budget, Santa Ana plans to extend their environmental efforts by allocating an additional $750,000 to create a Climate Action Plan (Climate Action Campaign 2023). This is a huge step towards creating a region of environmental and social justice. However, it remains to be seen whether the Climate Action Plan will help fight climate change productively.

4.3: Huntington Beach

While it is not the biggest city, Huntington Beach is one of the most significant cities in Orange County when it comes to climate change. According to many groups, including the

Climate Action Campaign, Huntington Beach has been the most successful city in Orange County in the fight against this issue by far.

One thing that Huntington Beach has done well is their focus on implementation and monitoring. For example, the government has tried to reduce vehicle miles traveled by people, reducing the amount of CO2 produced in 2012 by 1% through monitoring vehicle usage in their city alone (Huntington Beach 2017).

Additionally, Huntington Beach has also kept up very well with the state. Their Climate Action Plan, known as the Greenhouse Gas Reduction Plan, is the only Orange County CAP that meets state greenhouse gas emission targets. They are following state recommendations, unlike any other Orange County city (Climate Action Campaign 2020).

Huntington Beach still has a few issues to solve, such as updating their CAP from 2017 and building electrification. However, currently, they are doing well, and they should serve as a model for all other cities in Orange County.

4.4: Other Cities

In addition to these three cities, there are many other cities in Orange County with their own laws and regulations. Fullerton has been praised for its great walking and biking plans, developing a committee that sees the regulation of these plans. Anaheim developed their first Climate Action Plan in May 2020, focusing on renewable resources, greenhouse gas reduction, water conservation, and energy efficiency. Other cities in Orange County have also developed plans to stop climate change, some being more successful than others. However, as seen in the next section, a lot more still needs to be done.

5: Are the Government Actions Helping?

As seen in the previous section, a lot of government actions have been taken and are planning to be taken. However, we still have not measured how effective these strategies are, and whether they are impacting our climate currently and in the future. That is what we will look at in this section; how much have these government actions helped?

5.1: Orange County Climate in the Present

In 2020, the Climate Action Campaign was also curious on the success of the Orange County government. After conducting their own research, they created a report on this topic, giving a grade to each city to show how effective the city's policies are against climate change. The results were shocking. According to their report, the highest ranked city, Huntington Beach, scored a 40/100. No other cities scored above 30/100.

There are many reasons for this. According to this climate group, "Only six (6) of 35 Orange County municipalities have adopted a CAP" (Climate Action Campaign 2020). That's about 17% of all of the county's regions, which is a ridiculously small amount. Additionally, OC cities are failing to plan for climate impacts by not working toward zero carbon.

This reveals how little every city in Orange County has done to fight against climate change up to the present. Currently, the issue of climate change is just as bad as it was five years ago, if not worse. A few bits of progress have been made, but ultimately, just not enough has been done against this issue.

The Climate Action Campaign recognized that there were some beneficial steps being taken in the present. They "praised Fullerton for its walking and biking plans, Santa Ana for its commitment to building affordable housing near transit and Huntington Beach for its commitment to monitoring progress on greenhouse gas reduction" (Climate Action Campaign 2023). However, more significant measures should be taken to prevent this issue.

5.2: Orange County Climate in the Future

Many issues are still expected to occur in Orange County. According to the 6th National Risk Assessment on Hazardous Heat, Orange County is expected to see the greatest percentage increase of hot days in 2053 (First Street Foundation 2022). This demonstrates how when compared to other regions, Orange County is still not preparing for the future. Cities like Irvine have set no-carbon goals by 2030, but the plans they have set will not reach that goal.

Fortunately, a lot of cities have begun to make some progress. They are currently discussing the environment a lot in their city council meetings, recognizing the severity of climate change and identifying how they can stop it. While there may be some backlash from officials who are against spending money on this issue, cities have still tried to fight this issue and save the future. But unfortunately, it isn't enough.

Ultimately, everyone needs to do their part to save Orange County from environmental problems. In the next section, we will take a look at some of the solutions people, both in and out of the government, can take.

6: Solutions to Climate Problems

Climate change in Orange County has been a big issue since climate change started. However, there are plenty of solutions that can be taken to try to fight this issue. These solutions are not just for the government; there are a lot of steps that can be taken at the individual level as well to raise awareness on this issue.

6.1: Government Solutions

Across the government, many different laws and policies can be enacted to help stop this issue. Establishing cooling centers is one way that some cities can address rising temperatures, especially when it comes to some of the most vulnerable groups of people like the elderly and homeless. (Hicks 2021). They can also promote using renewable energy sources, moving to green energy to help prevent more greenhouse gases from going to the atmosphere. Most importantly, each government needs to review and update their climate action plans. A city cannot just create something and stick with it; as time changes, the plan should change with it.

Another thing city governments should do is to take the good things other cities in Orange County are doing and apply it to their own city. For example, as talked about before,

Huntington Beach's monitoring of vehicle use is very well done, while Santa Ana is adding a lot more money to the climate budget. Other cities like Anaheim and Irvine can learn from these laws, applying them to their own cities to help against the fight against climate change.

Finally, governments can also try to promote public support on this issue. They have a lot of power, as citizens will look up to their leaders. If the government is able to show everyone the dangers of climate change, it will inspire a lot more people to fight this challenge. The government can do a lot, but motivating the citizens will get a lot more work to be done.

6.2: Individual Solutions

The government has a lot of power to create solutions to environmental problems. However, in each city, the government is only composed of a few people. There are still many people in these cities that are encouraged to stop climate change, but they have no idea where to start. Fortunately, there are a few solutions that can be taken to begin to fight this dangerous issue.

First of all, a person can do their part by joining climate action groups. Some of these include the Climate Action Campaign, Surfrider Foundation, and Orange County Environmental Justice, as mentioned above, but there are also plenty of smaller groups that are always looking for more members. If none of these groups are appealing, a person can just create their own group and find members to join it. Working together with a group makes it much easier to spark change than just working as an individual.

This also leads to another important part of stopping this climate problem in Orange County; making your voice heard. If enough people group together, they can get their city's government to notice, forcing them to address this climate issue, and helping spark change. The government can definitely persuade the individual, but it works the other way, too.

Finally, a person can reflect on what they do in their own lives and determine how they are causing environmental damage. For example, it is very common to leave a faucet on for much longer than it needs to be, or to shower longer than needed. This leads to wasted water. Additionally, instead of traveling by car to places that are close by, it would be more environment-friendly to travel by bike or walking. If each person in Orange County reflects on what they can change, and then, actually change it, they can make a difference. Using all of these techniques will tremendously benefit the entire region.

7: Conclusion

Climate change in Orange County is still a huge threat today. In July 2023, a tropical storm hit Orange County, the first one since 1939. Increases in human population have led to a loss of biodiversity across the region, and although the extent of this loss has not been determined yet, remaining an unanswered question for further research, the decrease in biodiversity is still very significant.

Some people, like climate action groups, have tried to stop these problems from getting worse. However, ultimately, not enough is being done. Governments have begun creating climate

action plans to prepare for the future, and the success of the implementation of these plans require further research.

If we want to solve this issue and try to decrease the impacts that climate change is having, we need everyone to act together. Each person needs to play their own role in stopping climate change.

This doesn't just apply to Orange County, too. While Orange County is heavily impacted by climate change, across the world, many places are facing climate problems as well. If every person in the world realizes the impact of climate change, we will be able to make a difference. There are many solutions we can take, like those listed above and more, to save our planet.

Climate change is a huge issue. However, if everyone works together, we can fix this problem and sustain our Earth for centuries to come.

Works Cited

Anaheim Public Utilities. *Sustainable Electric & Water Initiatives GREENHOUSE GAS REDUCTION PLAN*. 2020.

Antonios, Caitlin. "106 Degrees in Anaheim Tops List of Heat Records Broken." *Orange County Register*, 9 Apr. 2022, www.ocregister.com/2022/04/08/heat-records-broken-again-in-several-southern-california-cities-but-cool-down-to-come/. Accessed 24 Aug. 2023.

Ben Brazil. "These Groups Are Helping Change Climate Policy in Orange County." *Daily Pilot*, 25 Feb. 2022, www.latimes.com/socal/daily-pilot/entertainment/story/2022-02-24/these-groups-are-helping-change-climate-policy-in-orange-county. Accessed 22 July 2023.

California Department of Homeland and Fire Protection. "Statistics | CAL FIRE." Www.fire.ca.gov, 22 Aug. 2023, www.fire.ca.gov/our-impact/statistics.

California Department of Public Health. "Climate Change and Health Profile Report Orange County." *Www.cdph.ca.gov*, Feb. 2017, www.cdph.ca.gov/Programs/OHE/CDPH%20Document%20Library/CHPRs/CHPR059range_County2-23-17.pdf

CDC. "Stay Safe during an Earthquake | Natural Disasters and Severe Weather | CDC." *Www.cdc.gov*, CDC, 24 Sept. 2020, www.cdc.gov/disasters/earthquakes/during.html.

Climate Action and Adaptation Plan. "Climate Planning." *City of Irvine*, 24 June 2021, www.cityofirvine.org/environmental-programs/climate-planning. Accessed 22 July 2023.

Climate Action Campaign. "ORANGE COUNTY | Climate Action Campaign." *CAC*, 2021, www.climateactioncampaign.org/orange-county. Accessed 22 July 2023.

Climate Action Campaign. "Orange County First Edition Report Card." *Orange County CAP Report Card*, 2022, www.climateactioncampaign.org/_files/ugd/91c4c2_11c30731160543b5aafdb585d293932c.pdf.

Climate Action Campaign. "Santa Ana Takes Bold Action in Addressing Climate and Justice Inequalities." *CAC*, 20 June 2023, www.climateactioncampaign.org/post/santa-ana-takes-bold-action-in-addressing-climate-and-justice-inequalities#:~:text=Santa%20Ana%20recently%20made%20a. Accessed 23 Aug. 2023.

Crumby, Sean, and Oliver Chi. *REQUEST for CITY COUNCIL ACTION MEETING DATE: TITLE: RECOMMENDED ACTION EXECUTIVE SUMMARY COMMISSION/BOARD/COMMITTEE RECOMMENDATION ANALYSIS*. 27 June 2023.

Hicks, Angelina. "What Are Orange County Cities Doing to Prepare for Climate Change?" *Voice of OC*, 15 July 2021, voiceofoc.org/2021/07/what-are-orange-county-cities-doing-to-prepare-for-climate-change/.

Huntington Beach. *Environmental Resources and Conservation Introduction and Purpose*. 2 Oct. 2017.

ICLEI–USA. *Santa Ana Climate Action Plan*. Dec. 2015.

Jill Replogle. "Orange County's Climate Plans Are "Decades Behind."" *LAist*, 26 Jan. 2023, laist.com/news/climate-environment/orange-county-climate-plans-decades-behind. Accessed 22 July 2023.

Marlon, Jennifer, et al. "Yale Climate Opinion Maps 2019 - Yale Program on Climate Change Communication." Yale Program on Climate Change Communication, 23 Feb. 2022, climatecommunication.yale.edu/visualizations-data/ycom-us/.

Murray, S. N., et al. "LONG-TERM CHANGES in ROCKY INTERTIDAL SEAWEED POPULATIONS in URBAN SOUTHERN CALIFORNIA." *Journal of Phycology*, vol. 37, no. 0022-3646, 24 Sept. 2008, pp. 37–37, https://doi.org/10.1111/j.1529-8817.2001.jpy37303-93.x. Accessed 5 Apr. 2019.

Nguyen, Lilly. "Updated Tsunami Hazard Maps for Orange County Highlight At-Risk Areas." *Los Angeles Times*, 12 July 2021, www.latimes.com/california/story/2021-07-11/new-tsunami-hazard-maps-orange-county-california-geological-survey#:~:text=Hazard%20zones%20include%20the%20area. Accessed 24 Aug. 2023.

Ready OC. "ReadyOC: Top 5 Disasters for Orange County." *ReadyOC*, readyoc.net/prepare/top5.html#tab-3. Accessed 22 July 2023.

Smoller, Fred, and Michael A. Moodian. "2020 Orange County Annual Survey." *Orangecounty_annualsurvey*, 2020, www.chapman.edu/wilkinson/political-science/_files/2020_orangecounty_annualsurvey.pdf.

U.S. Global Change Research Program. "Fourth National Climate Assessment." *Globalchange.gov*, vol. 2, no. 4, 31 Dec. 2017, pp. 1–470, nca2018.globalchange.gov/.

Worland, Justin. "How Some Environmentalists Hope to Make Climate Change an Issue in the Midterms." *Time*, 15 Oct. 2018, time.com/5413384/california-wildfires-climate-change/.

Using AI Models to Predict the Occurrence of Invasive Species Based on Habitat Preferences By JanetLou Guan

Abstract

Invasive species are posing significant ecological, socio-economic, and human health threats to our society. It is imperative to accurately predict the potential occurrences of invasive species in order for management to concentrate efforts on prevention, early detection, and swift response. This study aims to employ various AI algorithms to forecast the likelihood of invasive species occurrences based on habitat preferences. I compared the prediction accuracies of three AI algorithms: Random Forest, Logistic Regression, and Gaussian Naive Bayes. Our analysis utilized data collected from twelve lakes located in the Adirondack region of Upstate New York. The outcomes of our study reveal that the Gaussian Naive Bayes model exhibited markedly higher accuracy levels compared to both the random forest and the logistic regression model. These findings highlight the effectiveness of the Gaussian NB model in predicting invasive species occurrences, underscoring its potential as a valuable tool for proactive management and conservation efforts.

Key words: Random Forest; logistic regression; Gaussian naive bayes; machine learning; algorithm; probability; occurrence; habitat preferences

1. Introduction

An invasive species is a non-native species whose introduction causes or is likely to cause economic, environmental, or human health harm, or threatens to disrupt the current balanced ecosystem (Executive Order 13751). It can be any kind of living organism: plant, insect, fish, bacteria, fungus, or even an organism's seeds or eggs. It can infiltrate new environments through various means. Many invasive species are introduced into a new region accidentally and a lot of times, by human beings. The era of globalization has significantly amplified both long-distance travel and trade, thereby escalating the incidence of non-native flora and fauna being brought into different ecosystems worldwide.

Invasive species can impact both the native species living within an ecosystem as well as the ecosystem itself. They can change the food web, compete with native organisms for limited resources, cause the extinction of native plants and animals, reduce biodiversity (e.g. Vilà et al., 2011), and alter ecosystem functioning (e.g. Pejchar & Mooney, 2009). Numerous studies have shown the negative effect of invasive species on native biodiversity. Some of them concentrated on the genetic level such as how the breeding process (specifically, males and females of two different species were bred together) affected biodiversity (Largiadèr 2008; Kumschick et al. 2015). Other studies have demonstrated the negative effect from the ecosystem level (Lazzaro et al. 2020; Viciani et al. 2020).

According to the National Wildlife Federation, approximately 42 percent of threatened or endangered species are at risk due to invasive species. In the United States, the annual estimated

economic and health-related costs of invasive species have been reported at more than $21 billion (United States geological survey).

Aquatic invasive species (AIS) have inflicted significant ecological harm on freshwater ecosystems (Ricciardi and MacIsaac 2000, Cucherousset and Olden 2011), underscoring the need for a more proactive approach to invasive species management (Leung et al. 2002, Pagnucco et al. 2015). The identification and effective management of invasive species relies on an accurate prediction of locations and environment favorable to non-native species for them to survive, establish, reproduce, and spread (Kramer et al. 2017).

However, ecological data are often high dimensional with nonlinear and complex interactions among variables, and with many missing values among measured variables (Sabat-Tomala et al. 2020). Traditional statistical approaches can encounter difficulties in extracting meaningful analyses from such data. Linear statistical methods, such as generalized linear models (GLMs), in particular, may prove inadequate for revealing patterns and relationships that can be unveiled by more advanced techniques (De'ath and Fabricius 2000). Among the statistical techniques frequently utilized in ecology, classification procedures stand out as some of the most widely adopted, finding applications in tasks such as remote sensing-based vegetation mapping (Steele 2000) and species distribution modeling (Guisan and Thuiller 2005). In recent years, classification trees (Breiman et al. 1984) have gained widespread popularity among ecologists due to their straightforward interpretability, exceptional classification accuracy, and capacity to characterize complex interactions among variables. The main objective of this study is to construct several different machine learning models and use these models to classify the known species and unknown species (classified as "other"). I then use these classified labels to analyze the population for each category. Based on the population change on the temporal dimensions data, we will be able to predict whether or not new invasive species have occurred in a particular region. The three models we selected were Random Forest, Logistic Regression, and Gaussian Naive Bayes.

2. Data and Methods–2.1. Area of Interest

The study area included twelve lakes in the Adirondack region in Upstate New York (Figure 1).

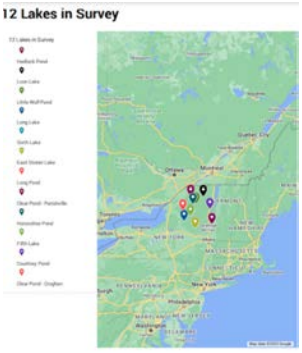


Figure 1. Area of interest

2.2. Species

Four species in the twelve lakes were chosen in this research based on their spreading speed and management agency concerns. These four species include two plants and two invertebrates: Variable Leaf Milfoil, Eurasian Watermilfoil, Spiny water fleas, and Asian clam.

Variable Leaf Milfoil (Myriophyllum Heterophyllum) (VLM) is a submerged aquatic plant with fine, feather-like leaves whorled around a main stem. VLM is native to the Southeastern and Midwestern United States. It was first observed in New England in Bridgeport, Connecticut in 1932 and is now widely distributed across the New England States. Variable milfoil grows in both still and flowing waters in a variety of substrates at depths from 1 to 5 meters.

Variable leaf milfoil can cause numerous ecological, cultural and economic impacts. VLM spreads quickly via fragmentation and can easily displace beneficial native aquatic plants. Dense beds of variable leaf milfoil degrade water quality for numerous species of fish and wildlife. In large mats, dissolved oxygen levels can be reduced to zero, making the area completely uninhabitable to game fish. Thick growths of VLM can impede fishing, swimming and boating, thus indirectly impacting tourism and the economic activity of lake towns.

Eurasian Watermilfoil (Myriophyllum spicatum) is a submersed, rooted aquatic plant native to Europe, Asia, and northern Africa. It was first reported in the United States in the 1880s (Eiswerth et al. 2000). The plants are rooted at the lake bottom and grow rapidly creating dense beds and canopies. They typically grow in water 1 to 4 meters (3.2 to 13 feet) deep, but have been found in water as deep as 10 m (32.8 ft). Stem densities can exceed 300/m2 (359/yd2) in shallow water.

The widespread of Eurasian Watermilfoil could cause both ecological and economic damages. The introduction of Eurasian watermilfoil can result in native macrophyte diversity and abundance declines. Eurasian watermilfoil beds form dense canopies at the water surface thereby reducing light penetration early in the season before native macrophytes have reached their full growth, shading them out and slowing/reducing growth potential. Milfoil-infested lakes tend to have reduced fish spawning areas and lowered fish growth rates. Besides, the negative impacts on wildlife and fish populations in water bodies with high densities of Eurasian watermilfoil and the difficulty of motor boating and swimming in infested areas result in recreation-oriented financial losses and the depreciation of shoreline property values.

Spiny Waterflea (Bythotrephes longimanus) is native to Europe and Asia. The species was unintentionally introduced into the United States' Great Lakes through the discharge of contaminated cargo ship ballast water. They were first discovered in Lake Huron in 1984; established in all of the Great Lakes by 1987 (Cullis 1988). Spiny waterfleas live in freshwater habitats and prefer cold temperatures, but can tolerate both brackish and warm water. Spiny waterfleas spread by attaching to fishing lines, downriggers, anchor ropes, and fishing nets and hitch a ride to other bodies of water.

Spiny waterfleas negatively impact native fish populations, aquatic habitats, and sports fishing. Spiny waterflea can clog eyelets of fishing rods and prevent fish from being landed. They also prey on native zooplankton, including Daphnia, which are an important food source

for native fishes. In some lakes, spiny waterfleas can cause the decline or elimination of some species of native zooplankton.

Asian Clam (Corbicula fluminea) is native to the fresh waters of eastern and southern Asia. It was likely introduced to the West Coast of North America around 1930, initially assumed to have been imported as a food source for the immigrating Chinese population (USACE ERDC 2007). Live Asian clams were first detected in US waters in 1938 in the Columbia River, Washington; the species quickly spread across the continent and is currently found in 44 states. The Asian clam is an invasive freshwater clam that prefers sandy lake bottoms and can be found at the sediment surface or slightly buried. Asian clams multiply rapidly and populations can easily reach high densities in freshwater. Asian clams are filter feeders, which means that they take in lake water and strain out algae. At high densities, Asian clams can out-compete other native filter feeders (such as fish, mussels and aquatic insects) for available food.

2.3. Data

I used data from Adirondack Research Invasive Species Mapping on LILA BC (Labeled Information Library of Alexandria: Biology and Conservation). The data set contains interpolated lake characteristics data of twelve lakes in the Adirondack region, Upstate New York, including depth, substrate hardness, and vegetation presence. This data is useful for calculating the probability of occurrence of other biological organisms that have habitat preferences related to factors such as water depth, vegetation, and substrate.

I obtained 5,769,518 instances of data, collected from a period of 24 months between 2018 to 2019. Of this data, there were 73,525 instances of Variable-Leaf Watermilfoil, 18,193 instances of Eurasian watermilfoil, 1,627 instances of Spiny Waterflea, 149 instances of Asian clam, and 984 instances of other unidentified species. All remaining instances were NaN values which did not include any species.

In order for the data to be usable in logistic Machine learning algorithms, we preprocessed the raw data and cleaned it by removing all NaN values in the dataset. I also dropped survey dates, removed duplicate data points, and used encoding techniques to convert non-numerical data to numerical data. The data was then randomly split into training and testing sections with 70% of the data used for training and 30% of the data used for testing. By training our models on data that only included instances of species, we were able to obtain predictions with higher accuracy, as the models would predict NaN values had they been trained with NaN values included.

This data set contains the following interpolated lake characteristics data of twelve lakes: Distance from shore (feet), Depth (feet), Bottom Hardness (relative units), Vegetation (relative units, indicating vegetation height), Longitude, and Latitude.

2.4. Algorithms

I selected three machine learning algorithms and compared their prediction accuracies: Random Forest (RF), Logistic Regression, and Gaussian Naive Bayes. With these models, the

features (location input) we used were: Distance from shore, Depth, Bottom Hardness, Vegetation, Longitude, and Latitude. The labels (species output) that the model tried to predict were: Variable-leaf watermilfoil, Eurasian watermilfoil, Spiny waterflea, Asian clam, Other (any other species), and NaN (no species).

Random Forest Classifier

Random forests or random decision forests is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees.

Random Forest algorithm is a supervised learning algorithm used in classification and regression problems, based upon decision trees. It is a way of constructing and then averaging multiple decision trees, as well as correcting the overfitting that often occurs when using a decision tree because of which, it generally yields more accurate results. Random forests are frequently used in business models due to their ability to make predictions based on a wide range of data.

A decision tree algorithm follows a hierarchical, tree-like model in order to make predictions. Based on significant features, trees can also be split into subtrees. A tree consists of several parts: root, internal, and leaf nodes, as well as branches. Root nodes are the starting points of the tree. Internal nodes are used to make decisions based on specific criteria and have multiple branches. Leaf nodes are the final output of decisions at the bottom of the tree. Branches are connections in the tree between nodes. In order to make predictions, the model starts from the root node and follows the branches and leaf nodes, making decisions based on criteria in the tree.

Logistic Regression:

Logistic regression is an algorithm based on the logistic function that predicts the relationship between two data factors. This algorithm is particularly useful in predicting two or more values using features that do not have a linear relationship.

It is frequently used in machine learning, especially within the medical field. Previously it has been used in medical scales to assess the severity of a patient's condition, or in detecting the risk of diseases.

The equation for logistic regression is:

$Y = 1/(1 + e^{-(\beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n)})$

Where Y is the value we are predicting, the x values are the inputs—feature vectors, and each β is a coefficient.

Using the logistic function

$P(x) = 1/(1+e^{Y})$

For multi-target models, variables such as X, Y, etc. are used similarly to how they appear in a linear regression model. The logistic function can be used as an activation function to predict the probability of the label.

There are mainly three types of Logistic Regression:

Binomial: Binomial logistic regression has a dichotomous dependent variable. It means there can be only two possible types of the dependent variables, such as 0 or 1, Yes or No, etc.
Multinomial: Multinomial logistic regression extends the approach for situations where the independent variable has more than two categories. Multinomial logistic can be applied to either ordered or unordered outcomes, such as "cats", "dogs", or "sheep".
Ordinal: In contrast with multinomial, ordinal logistic is only for ordered outcomes, such as "low", "Medium", or "High".

Gaussian Naive Bayes Classifier

Gaussian Naive Bayes (GNB) is a classification technique used in Machine Learning based on the probabilistic approach and Gaussian distribution. Gaussian Naive Bayes assumes that each parameter (also called features or predictors) has an independent capacity of predicting the output variable. Bayes' theorem states the following relationship, given class variable y and dependent feature vector x_1 through x_n

$$P(y|x_1, \ldots, x_n) = (P(y) * P(x_1, \ldots, x_n|y))/P(x_1, \ldots, x_n)$$

Using the naive conditional independence assumption that

$$P(x_i|y) = P(x_i|y, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$$

we can use the Maximum A Posteriori (MAP) estimation to estimate P(y) and $P(x_i|y)$. The former is then the relative frequency of class y in the training set. After using the database to train each respective model, results were calculated regarding the accuracy of the model's predictions. The accuracy of the models was calculated using the mean square error (MSE). The MSE is calculated through taking the absolute value of the square of the differences between each predicted value and the correct value, finding the sum, then dividing it by the total number of values. The equation for the MSE is Using the MSE, we could find the percentage of our models' predictions that was correct. Our models were tested with unbalanced (sometimes called imbalanced) data. Unbalanced data indicates that the amount of data points accessible for each class varies. For example, if there are two classes, balanced data means 50 percentage points for each class. Slight imbalance is not a concern for most Machine Learning approaches. As a result, if one class has 60% of the points while the other has 40%, there should be no noticeable performance reduction. Only when the imbalanced datasets of machine learning are extreme (i.e. 90% for one class and 10% for the other), would typical optimization parameters or performance metrics be ineffective and require adjustment. The dataset used in our research was considered unbalanced. For example: there were 73,525 instances of Variable-Leaf Watermilfoil, but only 149 instances of Asian clam.

3. Results

The prediction accuracy of all three AI algorithms is listed in Table 1 and Table 2.
Table 1: Comparison of results on dataset including all NaN values

	Random Forest	Logistic Regression	Gaussian NB
Unbalanced	95.20%	36.36%	29.18%
Balanced	35.11%	50.56%	81.23%

Table 2: Comparison of results on dataset excluding all NaN values

	Random Forest	Logistic Regression	Gaussian NB
Unbalanced	99.35%	98.12%	96.14%
Balanced	99.89%	80.80%	95.16%

All three models were trained both unbalanced (with equal weight on all labels) and balanced (with greater weight on non NaN labels). In each scenario, each was trained first with all NaN values included, then with all NaN values excluded. In comparison, including all NaN values, random forest yielded the highest unbalanced accuracy, followed by Logistic regression, then by Gaussian NB. However, random forest yielded the lowest balanced accuracy, whereas Gaussian NB yielded the highest balanced accuracy. Excluding all NaN values, random forest yielded the highest unbalanced accuracy, followed by Logistic regression, then by Gaussian NB. Random forest also yielded the highest balanced accuracy, followed by Gaussian NB, then by Logistic regression. At the same time, occurrences of different species were analyzed according to specific features in order to analyze where in a region each species is most likely to inhabit. Graphs were created using Python's Matplotlib and Seaborn libraries. Figure 2 and 3 below showed occurrences of each of the four species and the "Other" new species based on distance from shore and depth, respectively. The graphs were created using Python's Matplotlib and Seaborn libraries. Matplotlib allowed for the creation of various graphs when analyzing the data used in the study. Through Matplotlib, histogram graphs were generated, based on how many times each species appeared with each feature.

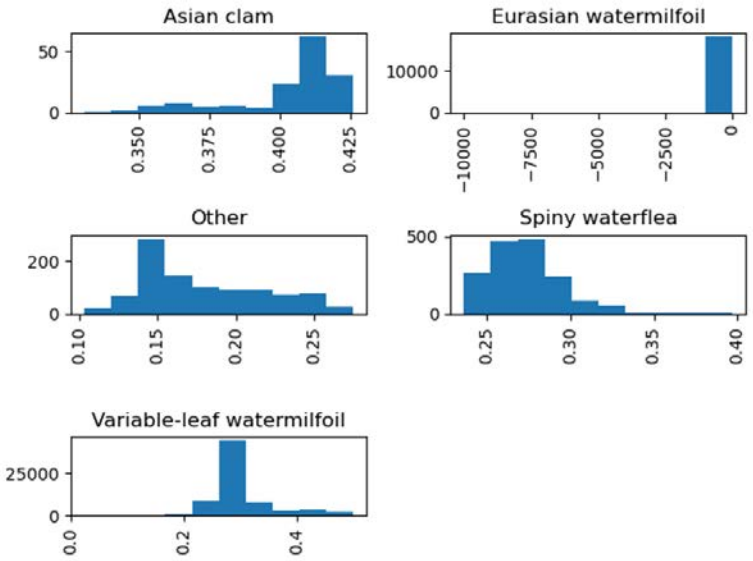


Figure 2. Occurrences of species based on distance from shore (the x-axis represents the distance from shore in meters, while the y-axis represents occurrences of species.)

Figure 2 showed that species such as the Asian Clam were found closest to the shore, with most occurrences appearing less than 0.5 meters from the shore. The Spiny Waterflea had the most occurrences appearing between 0.25 and 0.3 meters from shore, while most occurrences of other, non-specified species in the data occurred at about 0.15 meters. Most instances of the Eurasian Watermilfoil and the Variable-Leaf Watermilfoil occurred the farthest from shore at about 0.5 meters.

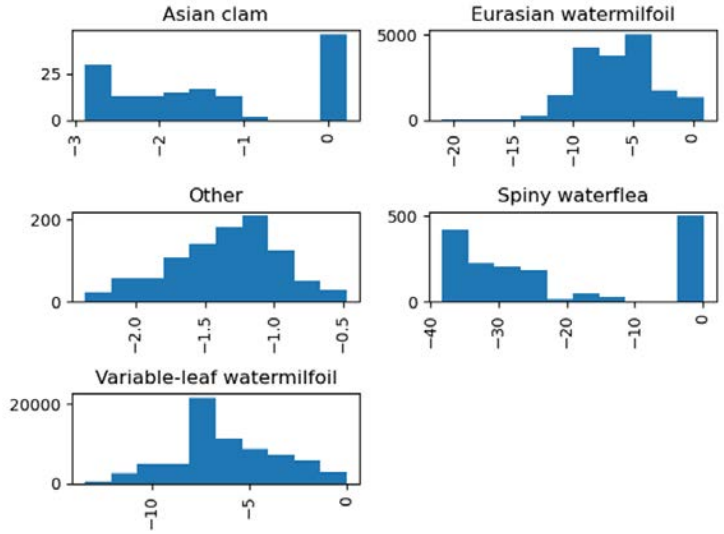


Figure 3. Number of occurrences of species based on depth (the x-axis represents the depth in meters, while the y-axis represents occurrences of species.)

Figure 3 showed that most Asian Clam and spiny waterflea were found around a depth of 0. Most instances of the Eurasian Watermilfoil occurred around a depth of -5 meters, while

instances of the Variable-Leaf Watermilfoil occurred at a depth between -5 and -10 meters. Other, non-specified species were found to mostly occur at a depth of around -1 meters. Seaborn allowed for the creation of scatter plots and area charts, color coded based on the five species types present in the data (seen in the key) and organized based on the regional features of the data (labeled on the y-axis).

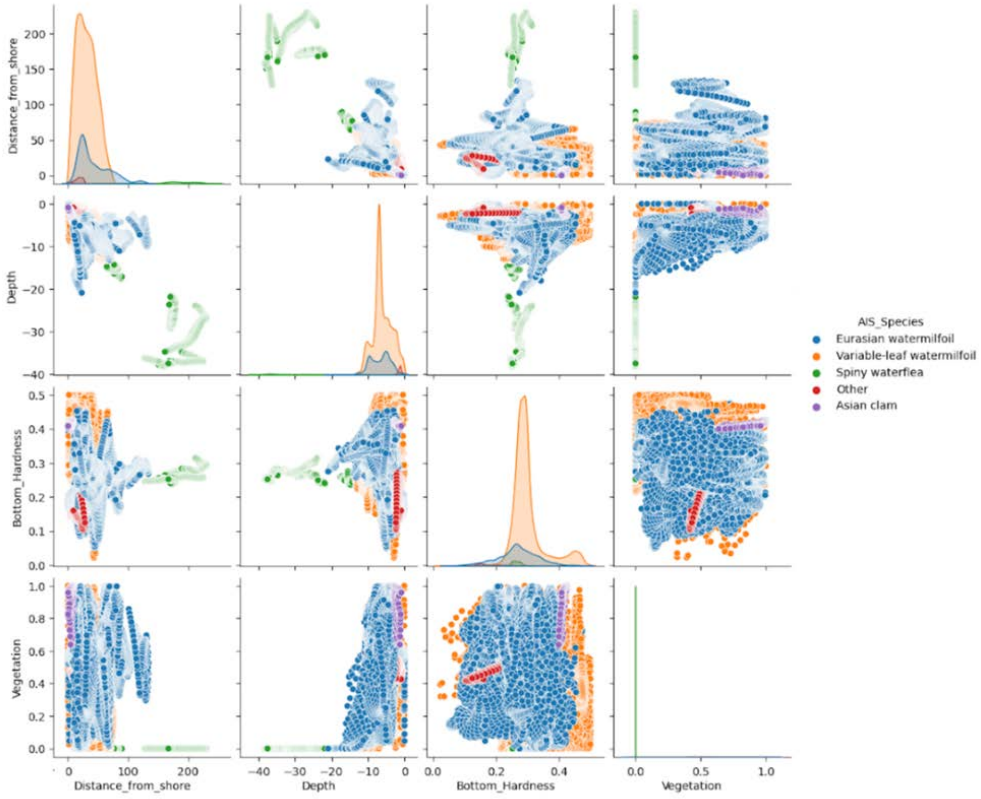


Figure 4. Visual representation using Seaborn (where distance and depth is measured in meters, hardness is measured in Joules/cubic meters, vegetation is measured in volume/cubic meters, and longitude and latitude are measured in degrees.)

From Figure 4 we can observe the relationship between specific variables and occurrences of the five species types observed in the study. The Eurasian watermilfoil (blue) appeared the most in the figure, and most instances occurred at greater depths. It was also observed to occur closer to the shore, with most instances found less than 100 meters from shore. The variable-leaf watermilfoil (orange) occurred mostly closer to the shore at areas with greater bottom hardness; however, there were several occurrences further from shore as well. The species appeared to be the most spread out. The spiny waterflea (green) consistently appeared furthest from the shore compared to the other four species, over 100 meters from shore. Alongside that, it is also found in areas of lower vegetation. The Asian clam (purple) was seen to appear in areas of high vegetation, as well as closer to the shore. It appeared minimally varied, with occurrences all exhibiting the same features. Other species (red) seemed to have limited

variation in their spread. They mostly appeared under 50 meters from the shore and in areas with depth near 0 and less bottom hardness. Points on the graphs showing longitude and latitude can be mostly seen along two lines (-75 and -73.5 degrees longitude and 43.4 and 43.8 degrees latitude) due to the twelve lakes being located closely along those longitudes and latitudes. As a result, not much information can be obtained solely by looking at the longitude and latitude.

4. Discussion

Being able to accurately predict the probability of occurrence of invasive species in a particular region is very important for developing efficient management and prevention strategies. In this study, three AI models were used to predict the probability of occurrence and their accuracies were compared.

Based on the results gathered from each model, the random forest model performed with the highest unbalanced accuracy at 99.35%, with NaN values excluded, and 95.20% with NaN values included. Logistic regression and Gaussian NB models' unbalanced, NaN-excluded accuracies were slightly lower at 98.12% and 96.14%, respectively; furthermore, their balanced models were also lower at 80.80% and 95.16%, respectively. Thus, the random forest model consistently performed slightly better than both the logistic regression and Gaussian NB models. Despite this, it is important to note that a model that could predict accurately with NaN values included would be the most useful in forecasting occurrences of invasive species, as there may be numerous locations in a region where no species occurs. Looking at the accuracies of the three models trained with NaN values included, we can see that, while the random forest model predicted with a high unbalanced accuracy of 95.20%, its balanced accuracy was much lower at 35.11%. The logistic regression model predicted much lower as well, with a 36.36% unbalanced and 50.56% balanced score, respectively. The Gaussian NB model performed with the lowest accuracy unbalanced at 29.18%; however, it performed with the highest balanced accuracy at 81.23%.

Balanced accuracy score is a further development on the standard accuracy metric where it's adjusted to perform better on imbalanced datasets which is the case for our research. The way it does this is by calculating the average accuracy for each class, instead of combining them as is the case with standard accuracy.

Balanced accuracy scores are better signifiers of accuracy than unbalanced as they are less prone to overfitting or underfitting (which occurs when the models become too specialized and make biased predictions favoring a label over others in the dataset). An unbalanced accuracy score with our database, for example, would overfit and put more emphasis on predicting a species such as variable-leaf watermilfoil rather than the Asian clam due to there being far more instances of the former in the data: 73,525 instances of variable-leaf watermilfoil versus 149 instances of Asian clam. Therefore, a balanced accuracy score would be more useful in prediction than an unbalanced one.

As a result, I look at the best performing model that has the highest balanced accuracy score with all NaN values included. This makes Gaussian NB the ideal method with an 81.23%

accuracy among the three tested models, when making species predictions based on the locational data provided.

5. Conclusion

The objective of this study has been to forecast the likelihood of invasive species occurrence through the utilization of various AI models. I constructed three models using random forest, logistic regression, and Gaussian NB. In the study, it is evident that AI models serve as robust instruments for accurately predicting the probability of invasive species occurrences. Among the triad of models evaluated, Gaussian NB stands out with the greatest potential as an invaluable resource for steering proactive management and conservation endeavors.

6. Limitations and Future Work

My study is not without limitations:

(1) Machine learning depends on labeled data, but accessing such data in biology and conservation is a challenge. Due to using a single database, there are limitations in the way our models have been trained. For instance, one algorithm may perform better for a smaller dataset than another; and

(2) all our species study areas are confined to the Adirondack region in Upstate NY. For future work, we would apply algorithms in more species and in more geographical conditions with differently distributed features; and

(3) the data was gathered in a two-year time span from 2018 to 2019, making it limited when taking into account how populations change over a longer period of time. Taking population numbers in a small-time frame means our models may make unreliable predictions when used to predict species several years from now.

Works Cited

Kramer, Andrew M., Annis, Gust, Wittmann, Marion E., Wittmann, William L. Chadderton, Rutherford, Edward S., Lodge, David M., Mason, Lacey, Beletsky, Dmitry, Riseng, Catherine, Drake, John M.. 2017. Suitability of Laurentian Great Lakes for invasive species based on global species distribution models and local habitat https://doi.org/10.1002/ecs2.1883

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. Classification and regression trees. Wadsworth and Brooks/Cole, Monterey, California, USA.

Cucherousset, J., and J. D. Olden. 2011. Ecological impacts of nonnative freshwater fishes. Fisheries 36: 215–230.

De'ath, G., and K. E. Fabricius. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81:3178–3192.

Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8: 993–1009.

Kumschick, S., M., Gaertner, M., Vilà, F., Essl, J. M., Jeschke, P., Pyšek, A., Ricciardi, S., Bacher, T. M., Blackburn, J. T. A., Dick, T., Evans, P. E., Hulme, I., Kühn, A., Mrugała, J., Pergl, W., Rabitsch, D. M., Richardson, A., Sendek, M., Winter. 2015. Ecological Impacts of Alien Species: Quantification, Scope, Caveats, and Recommendations. BioScience. 65:55–63. doi: https://doi.org/10.1093/biosci/biu193. [Crossref], [Web of Science ®], [Google Scholar]

Largiadèr C. R. 2008. Hybridization and Introgression Between Native and Alien Species. In: Nentwig W. (eds) Biological Invasions. Ecological Studies (Analysis and Synthesis), vol. 193. Springer, Berlin, Heidelberg. doi:https://doi.org/10.1007/978-3-540-36920-2_16. [Crossref], [Google Scholar]

Lazzaro, L., R., Bolpagni, G., Buffa, R., Gentili, M., Lonati, A., Stinca, A.T.R., Acosta, et al. 2020. Impact of invasive alien plants on native plant communities and Natura 2000 habitats: State of the art, gap analysis and perspectives in Italy. Journal of Environmental Management. 274:111140. doi: https://doi.org/10.1016/j.jenvman.2020.111140. [Crossref], [PubMed], [Web of Science ®], [Google Scholar]

Leung, B., D. M. Lodge, D. Finnoff, J. F. Shogren, M. A. Lewis, and G. Lamberti. 2002. An ounce of prevention or a pound of cure: Bioeconomic risk analysis of invasive species. Proceedings of the Royal Society of London Series B: Biological Sciences 269: 2407–2413.

Pagnucco, K. S., G. A. Maynard, S. A. Fera, N. D. Yan, T. F. Nalepa, and A. Ricciardi. 2015. The future of species invasions in the Great Lakes-St. Lawrence River basin. Journal of Great Lakes Research 41: 96–107.

Pejchar, L., & Mooney, H. A. (2009). Invasive species, ecosystem services and human well-being. *Trends in Ecology and Evolution*, 24, 497–504 https://doi.org/10.1016/j.tree.2009.03.016

Ricciardi, A., and H. J. MacIsaac. 2000. Recent mass invasion of the North American Great Lakes by Ponto-Caspian species. Trends in Ecology & Evolution 15: 62–65.

Sabat-Tomala, Anita, Edwin Raczko, and Bogdan Zagajewski. 2020. Comparison of Support Vector Machine and Random Forest Algorithms for Invasive and Expansive Species Classification Using Airborne Hyperspectral Data. *Remote Sensing* 12, no. 3: 516. https://doi.org/10.3390/rs12030516

Viciani, D., M., Vidali, D., Gigante, R., Bolpagni, M., Villani, A.T.R., Acosta, et al. 2020. A first checklist of the alien-dominated vegetation in Italy. Plant Sociology. 57(1):29–54. doi: https://doi.org/10.3897/pls2020571/04. [Crossref], [Google Scholar]

Vilà, M., Espinar, J. L., Hejda, M., Hulme, P. E., Jarošík, V., Maron, J. L., … Pyšek, P. (2011). Ecological impacts of invasive alien plants: A meta-analysis of their effects on species, communities and ecosystems. Ecology Letters, 14, 702–708. https://doi.org/10.1111/j.1461-0248.2011.01628.x

Borderline Personality Disorder: Dissecting Gendered Stigmas, Historical Context, and Media Impact By Victoria Lai

Intro

Borderline Personality Disorder (BPD) is a complex mental health condition that presents unique individual challenges and exerts significant societal impacts. Notably, its prevalence is strikingly high among women, with approximately 75% of cases affecting females (American Psychiatric Association, 2013). This gender disparity raises questions about causation and correlation, potentially tracing it back to its historical roots. Early symptoms of the disorder have been documented in medical literature, spanning thousands of years. The emergence of official research on BPD, which scholars believe to have evolved from hysteria, was diagnosed by Sigmund Freud and many other researchers. These diagnoses marked a turning point in understanding hysteria, albeit one that portrayed it negatively; despite the progress made since the Victorian era, when unsubstantiated treatments aimed to cure "spontaneous uterus movement" or hysteria, misconceptions about the modern DSM-5 criteria of BPD persist in modern society.

This paper addresses the critical issue of the lack of education about borderline personality diagnosis and the perpetuation of stereotypes in the media, which further stigmatize individuals with BPD. These misrepresentations intensify the gendered stigma surrounding BPD and exacerbate the overall stigma associated with mental illness. Recent studies reveal that individuals with BPD encounter higher levels of discrimination and prejudice than those with other mental health diagnoses (Sheppard et al., 2023). The criteria for diagnosing individuals along the BPD spectrum encompass numerous symptoms, often leading to misdiagnosis and complications. These symptoms include disturbances in self-identity, such as an unstable self-image, dissociation, harsh self-criticism, heightened interpersonal sensitivity, abrupt shifts between idealization and devaluation of others, oscillating between over-involvement and withdrawal in relationships, erratic mood fluctuations, experiences of depression and anxiety, impulsive behaviors, and episodes of hostility (DSM-5; American Psychiatric Association, 2013).

Through this paper, I hope to advocate for more accurate and informed portrayals in the media and public discourse to contribute to the destigmatization of BPD and raise awareness about this complex mental health condition. I will explore how this stigmatization can lead to self-stigma among individuals diagnosed with BPD, examining its impact on their interpersonal relationships. Addressing these issues requires a multifaceted response that includes improved education, spreading awareness, and promoting more responsible media portrayals and open conversations about mental health. By critically analyzing the multifaceted nature of BPD, uncovering the historical and societal contexts that have shaped perceptions, and advocating for more accurate and empathetic portrayals of this disorder, we can spread more information and awareness about this nuanced mental health condition in hopes of destigmatization.

The Impact of Emotional Dysregulation on Interpersonal Relationships in BPD

BPD is a complex and debatable diagnosis, with uncertain evidence regarding its cause. Nonetheless, shared traits among individuals with BPD suggest that genetic and environmental factors, including emotional vulnerability, childhood abuse, and insecure attachment, may play a role (Klein et al., 2022). The symptoms of BPD significantly hinder daily functioning and form the central point of this discussion. This section overviews BPD symptoms and diagnostic criteria, examining the disorder's repercussions on everyday life and interpersonal connections. It highlights the significance of emotional dysregulation and self-destructive behaviors within BPD's framework. BPD manifests through a spectrum of symptoms characterized by intense emotional and behavioral fluctuations.

The Diagnostic and Statistical Manual of Mental Disorders (DSM-5) characterizes diagnostic criteria for BPD as disturbances in identity, self-image, and relationships. Key features include unstable self-perception, acute abandonment fears, impulsivity, mood volatility, and a chronic sense of emptiness (American Psychiatric Association, 2013). Complicating matters, those with BPD are at elevated risk of self-harm and suicide, with up to 10% of patients facing this terrible outcome (Paris et al., 2019). This combination of symptoms creates an inner turmoil that impedes daily management and interpersonal bonds. Emotional dysregulation within BPD enhances impulsive and self-destructive actions such as self-harm, suicidal ideation, gambling, irrational choices, and substance misuse. Notably, 73% of BPD patients meet the criteria for a lifetime substance use disorder (Grant et al., 2008), exacerbating their mental health challenges. However, it is vital to acknowledge that these actions stem from the intense agony and despair experienced at the moment, which compels them to adopt toxic coping mechanisms. Such impulsive conduct further strains interpersonal relationships, fostering perceptions of instability and prompting potential ostracization due to perceived difficulty and unpredictability, which triggers BPD patients' core wound of the fear of abandonment. The persistent feelings of emptiness and identity instability within BPD hinder the pursuit of long-term goals and cohesive self-development. The comorbidity of BPD with other major mental disorders exacerbates its complexity and susceptibility to misdiagnosis. A study scrutinizing BPD misdiagnosis through structured interviews found significantly increased odds of prior bipolar misdiagnosis among BPD patients (Ruggero et al., 2011). Almost 40% of individuals with BPD were incorrectly diagnosed with bipolar II (Healthline).

Additionally, disordered eating behaviors are commonly expected, with prevalence rates up to 53%, linked to BPD's emotional regulation difficulties, alexithymia—the struggle with identifying and describing one's emotions, and an externally oriented thinking style (Sifneos, 1973). These trends underscore the urgency of improved education and training for mental health professionals in managing this complex condition. Central to BPD lies emotional dysregulation, evident in intense responses to minor triggers and the struggle to manage emotions effectively. This fundamental characteristic is closely intertwined with how the disorder affects relationships and social interactions. People with BPD frequently develop intense emotional dependencies on a "favorite person" (FP), characterized by intense and frequently insecure attachments. This

favorite person role encompasses friends, romantic partners, or family members. Within these relationships, BPD individuals oscillate between two extremes, shifting from idealization to devaluation, thus encapsulating the dual nature of craving intimacy and fearing rejection. The emergence of interpersonal aggression, particularly in response to social rejection cues, compounds relationship challenges by disrupting effective interpersonal functioning (Jeong et al., 2022).

BPD individuals are trapped in uncontrollable emotions, propelling them to focus on their F.P. increasingly. Paradoxically, this fixation causes their F.P. to gradually lose faith in sustaining the relationship and meeting the BPD individual's needs. This recurring pattern triggers a profound emotional response within individuals with BPD, tapping into their core wounds and activating their deep-rooted fear of abandonment. Consequently, their reaction to this fear manifests as clinginess and an urgent pursuit to avoid rejection. Ironically, these actions intended to secure the relationship often yield the opposite outcome, further straining the connection. The uncertainty of emotional responses and behaviors highlights the complexity of relationships for individuals with BPD. The emotional turbulence from oscillating between idealization and devaluation confuses interactions, frequently culminating in conflicts. The heightened sensitivity to cues of rejection, combined with challenges in emotion management, fuels a cycle of interpersonal difficulties. As individuals with BPD grapple with their fear of abandonment while attempting to preserve the relationship, their actions inadvertently contribute to the distancing they fear. The dynamics surrounding favorite person attachments within BPD show the contradictory aspects of the disorder, from the individual's yearning for intimacy to their subsequent tendency to push their favorite person away through shifting emotions. This impacts their relationships and social interactions, which perpetuate instability and isolation. A greater understanding of these interconnected facets raises awareness and humanizes the challenges posed by this disorder.

The Gendered Historical Progression and Stigmatization of BPD.

The striking similarities in symptoms between BPD and hysteria have prompted speculation regarding a potential evolutionary or close connection between these disorders. Historically, hysteria has been associated primarily with individuals possessing a uterus, and ancient civilizations, such as the Egyptians and Greeks, believed that the womb could influence the body's overall health (McGill, 2017). This historical perspective attributing various health issues to the uterus has complicated diagnosing and treating BPD and hysteria. The historical link between uterine factors and health complications has added complexity to the challenges individuals with these disorders face. The shared symptoms, including emotional instability, impulsivity, and relationship difficulties, have frequently led to diagnostic and treatment confusion. Researchers and clinicians have grappled with distinguishing between these conditions, as evidenced by statements like, "Borderline personality disorder sometimes exhibits certain hysterical symptoms. In our clinical practice, we encounter patients for whom differentiating between BPD and a hysterical disorder in terms of psychopathological aspects is

challenging" (Tatsuo Oshima, MD, Ph.D., 2000). However, some researchers propose that "borderline personality disorder simply replaced hysteria, reflecting contemporary values regarding women's behavior." This perspective characterizes BPD as a psychiatric diagnosis often perceived as "feminized" due to its higher prevalence among women, with a ratio ranging from 3:1 to 7:1 (Becker, 2000). The criteria for diagnosing BPD frequently encompass symptoms associated with traditionally considered "feminine qualities" (Jimenez, 1997).
The historical context surrounding understanding and treating disorders like hysteria and BPD has significantly shaped the challenges individuals face today.

In the past, when ethical interventions and solid scientific evidence were lacking, especially women with these disorders endured unfounded and often cruel treatments. The false belief that these conditions were exclusively linked to femininity led to stigmatization, witchcraft accusations, and misguided therapies, such as ritualistic burnings (R. Porter, 1988). Even Sigmund Freud's proposed solution, suggesting that women should "get married and have sex," centered around women's sexuality and reflected the biased perspectives of the time, perpetuating stigma associated with conditions stereotypically associated with femininity (McGill, 2017). This gendered narrative persisted over time, with hysteria evolving to reflect shifts in societal perceptions. Initially attributed solely to the "frustrated uterus" affecting widows and virgins, hysteria was a disorder exclusively ascribed to women or even linked to the absence of male genitalia. Over time, hysteria evolved into a medically recognized condition affecting a person's brain (Bogousslavsky, 2011).

In the 19th century, it remained common yet misunderstood in Europe, posing significant challenges to the medical field. French neuroscientist Jean-Martin Charcot played a pivotal role in unraveling hysteria through his innovative approach, "the great neurosis," involving comprehensive patient observations. Charcot identified critical features like "convulsions, contortions, fainting, and impairment of consciousness," which he termed "hystero-epilepsy," linking the condition to disruptions in neurons and challenging prevailing notions (Waraich & Shah, 2017). Charcot's perspective shifted the understanding of hysteria from being solely a female affliction to a neurological condition that could affect individuals, regardless of gender or traditionally masculine roles, such as railway engineers or soldiers. This historical transformation of Hysteria from a gender-bound concept into a recognized medical diagnosis highlights a recurring pattern in which women's mental health has been misconstrued and mischaracterized. The clinical descriptions of individuals with BPD, which often switch between extreme dysfunction and manipulative charm, reflect this enduring pattern, deeply influenced by ingrained misogynistic beliefs (Masland et al., 2022; Robert et al. Center, 2022). Such portrayals serve as derogatory labels and contribute to concrete stigmas that disproportionately affect women.

The Impact of Labeling on Mental Health and the Stigma Surrounding BPD

Within the broader context of discussing the impact of labeling on mental health and the stigma surrounding BPD, it is essential to recognize a compelling link between trauma,

especially abuse, and the diagnosis of BPD in women. Notably, 81% of individuals diagnosed with BPD have a history of past abuse, and many women diagnosed have been sexually abused in childhood (Ussher, 2011). Paradoxically, the BPD diagnosis can unintentionally divert attention from the abuse, instead emphasizing the victim's perceived "disordered personality." Many diagnosed with BPD often feel their traumatic experiences are downplayed despite substantial evidence linking such incidents to the disorder. This connection highlights the need to consider how labeling and stigmatization can have far-reaching consequences on mental health. Sly Sarkisova, a psychotherapist specializing in trauma, characterizes BPD as "a label that is often misused and applied, especially to women, to pathologize them for emotional expressions of suffering." noting that BPD is sometimes misused as a label to pathologize emotional responses to suffering (Shimo, Alexandra 2019). Many psychologists share concerns that BPD diminishes women's suffering by affixing pejorative psychiatric labels to them, diverting attention from a society riddled with traumatic experiences for women. Clare Shaw asserts that "the real issue is the pervasive abuse of women and girls in this society coupled with the continuing silencing and invalidation of women's experiences." (C. Shaw, 2018). This disproportionate gender bias also manifests in the process of diagnosis. Carol S. North's work in 2018 demonstrates that women are diagnosed with BPD between three and seven times more frequently than men. This can be attributed, at least partly, to the modern criteria for BPD, which encompass symptoms that align with stereotypical "feminine qualities," including internalized negative emotions like depression and emotional volatility. Clinicians describe some women with BPD as pathologically dysfunctional, often labeling them as "difficult" or "problematic" (Masland et al., 2022), while others depict them as attractive, successful, and charming manipulators who seduce others, including therapists, into destructive relationships. These depictions serve as pejorative markers and literal stigmas, predominantly affecting women. Tragically, these traits that society associates with BPD are often perceived as making women "both damaged and dangerous." The historical archetype of the "demanding, angry, aggressive woman" labeled as "mentally disordered" perpetuates the gendered nature of this condition (Jimenez et al., 1987).

Many researchers and psychologists believe diagnosing BPD can inadvertently downplay the gravity of women's suffering with the BPD label, which is heavily associated with stigma. Dr. Joel Dvoskin, a forensic psychologist affiliated with Georgetown Law, believes that action must be taken to remove the stigma associated with the term "BPD" in diagnosis. He emphasizes the real impact of the "BPD" label when applied to an individual, highlighting that "not all mental health diagnoses promote treatment" and singles out BPD as a diagnosis that "significantly harms people." He critiques the term because it often results in sub-standard treatment for those diagnosed with the disorder. For instance, mental health professionals frequently label the unwanted behaviors of BPD clients as "manipulative" and in need of punishment. He candidly recalls his early career experiences when he accepted the BPD label without critical examination, using it to explain a range of disruptive behaviors observed within prison settings. He encountered male and female inmates who exhibited patterns of seeking and rejecting intimacy

and dependency, overreacting to seemingly minor emotional triggers, and occasionally engaging in self-injurious behaviors. "These people, I was told, engaged in these behaviors because they were 'Borderlines.'' And we knew that they were 'Borderlines' because they performed these behaviors. This tight circle of logic had the advantage of letting me, as a prison psychologist, completely off the hook when I failed to help them because 'Borderlines,'' I was told, were difficult or impossible to treat." (Joel A. Dvoskin, Ph.D., A.B.P.P.1996) Through this perspective, Dr. Dvoskin sheds light on the broader issue of labeling, demonstrating how the diagnosis of BPD can become a catch-all explanation for complex behaviors, resulting in lowered expectations for successful interventions and potentially perpetuating a cycle of stigmatization and inadequate care.

Within mental health, language holds profound significance beyond mere semantics. Words like 'disorder' can hinder a deeper understanding and shape thought and practice, discouraging individuals from seeking help and fostering self-stigmatization, where people internalize negative beliefs about themselves. Researchers have found that individuals with personality disorders, especially those with BPD, tend to stigmatize themselves more than individuals without such disorders. The heightened self-stigmatization experienced by individuals with BPD is believed to be rooted in the fact that they "expect and perceive social rejection more intensely than the general population" (Grambal et al., 2016). Their heightened sensitivity to social rejection can lead to increased self-stigmatization. Individuals with BPD may internalize negative societal perceptions and judgments, further impacting their self-esteem and self-worth. Moreover, researchers believe that “individuals with BPD also exhibit negative and unstable self- and other evaluations compared to healthy individuals.” These factors could be associated with a higher level of self-stigma and shame due to the nature of BPD, given that “persistent challenges in social interactions, intense and unstable interpersonal relationships, and exaggerated efforts to avoid abandonment constitute core features of BPD" (Grambal et al., 2016).

Furthermore, these components of stigma, including labeling, stereotyping, separation, status loss, self-shame, and discrimination, have significant implications in the context of mental health disorders like BPD. In the case of BPD, individuals can find themselves labeled and stereotyped due to their condition, which can lead to feelings of self-shame (Link & Phelan, 2001). Society's tendency to separate and stigmatize those with mental health disorders, including BPD, perpetuates this self-shame and contributes to the broader stigma surrounding mental health. This is why spreading awareness and educating the masses about BPD and mental health is vital to challenging societal norms and attitudes contributing to stigma. Understanding the multifaceted nature of stigma, as outlined by Link and Phelan (2001), underscores the importance of addressing these issues on both an individual and societal level to create a more compassionate and inclusive society.

The gendered aspects of BPD extend beyond their impact on women's perspectives on mental health and have profound implications for our society. Despite the progress in moving away from the constraints of a patriarchal society that once limited women to domestic roles,

specific gender disparities persist, particularly in mental health. It is crucial to recognize that the gendered stigma surrounding BPD affects not only women but also men, albeit in different ways. For women, the stigma surrounding BPD often revolves around characterizations of being overly emotional, complicated, or manipulative. These labels hinder their willingness to seek help and perpetuate harmful stereotypes about women's mental health.

Conversely, the gendered stigma associated with BPD may manifest differently for men, potentially leading to underdiagnosis or misdiagnosis. Men with BPD may encounter challenges seeking support due to societal expectations of traditional masculinity, discouraging vulnerability and emotional expression.BPD is an exemplar, illustrating how inequalities persist in our society despite societal advancements. The stereotype associating BPD with femininity reinforces societal norms that hinder progress, possibly contributing to the fact that the DSM-5 reports 75% of diagnoses as women. The perception that BPD predominantly afflicts women creates a significant predicament for men grappling with this disorder. The notion that BPD and mental health support are associated with femininity or weakness may deter men from seeking help, as they fear being labeled as "feminine" or "weak" (Karim et al., 2019). This reluctance to seek assistance exacerbates their condition and perpetuates the stigma surrounding BPD, ultimately undermining their mental well-being.

The complex dynamics we observe in today's society can be traced back to historical accounts that shed light on the challenges men face. For instance, the long-held belief that Hysteria was exclusively a "woman's disease" and intricately tied to the essence of femininity, as articulated by Laycock (1840) and Otto Weininger's assertion in 1903 that "hysteria is the organic crisis of the organic mendacity of woman" (Bronfen, 1998), evolved to include the nervous system but still contributed to the enduring stigma around men accessing mental health support. This historical context plays a crucial role in shaping contemporary dynamics deeply entrenched in stigmatizing men who seek help for their mental health concerns (Sagar-Ouriaghli et al., 2019). Another historical example involves British neurologist Sir John Russell Reynolds in the 18th century, who perpetuated gendered perceptions by attributing men and boys with Hysteria as having a "feminine constitution." Similarly, Emile Batault's analysis of male Hysteria in the late 1800s revealed prevailing perceptions of these men as "coquettish and fearful... preferring ribbons and scarves to labor" (Gilman et al., 1993). These stereotypes hindered the accurate diagnosis of Hysteria in men and aligned with the conventional masculinity of the time. These historical examples vividly illustrate how societal attitudes and gender norms have long stigmatized men seeking assistance for mental health concerns. This profoundly ingrained stigma, rooted in the association of emotional vulnerability with femininity, has far-reaching implications for men's mental well-being.

Today, men are often conditioned to suppress emotions and maintain an image of strength and resilience. Consequently, seeking help for mental health issues is perceived as taboo and incongruent with traditional masculinity (McKenzie et al., 2018). This dynamic, compounded by the misconception that BPD is a predominantly feminine disorder, creates an environment that prevents men from reaching out for support. These societal stigmas continue to exert influence,

and the historical negative perception surrounding Hysteria also resonates with the present reputation of BPD.

It is important to emphasize that although there are negative consequences for men stemming from the gendered perceptions of BPD, women remain the prime targets of these stereotypes and are the ones mainly affected. Women continue to live with the stigma associated with BPD, facing the double challenge of dealing with the disorder and its societal bias. While men may encounter barriers in seeking help, women often endure the weight of being labeled as unhinged, manipulative, or esoteric due to their BPD diagnosis. To address these issues, our society needs to break down these gendered perceptions along with more unbiased evaluation tools and enhanced education to promote an environment that encourages both men and women to prioritize their mental health without fear of judgment. While acknowledging the challenges faced by men in seeking mental health support, it is crucial to remember that the overarching impact of BPD's gendered stigma primarily falls upon women, and addressing this gender bias should remain a central focus in our efforts to improve mental health awareness and support for all individuals.

The Media's Perpetuation of Mental Health Stigmatization and the Role of Accurate Portrayals

In contemporary society, the 'Hot Crazy Matrix' has garnered scholarly attention due to its widespread presence in the media and its substantial impact on societal perceptions of attractiveness. Originally introduced as an internet-derived theory, this matrix has shaped popular culture and public awareness through gendered stereotypes. This theoretical framework posits a correlation between a woman's perceived attractiveness level and her emotional stability, or " craziness." According to the matrix, as a woman's perceived attractiveness increases on a 10-point scale, there is an associated rise in her perceived emotional volatility. Researchers such as Blanchard and their colleagues assert that specific traits are commonly associated with borderline personalities positioned on the 'crazy' end of this spectrum.

Consequently, the theory suggests that women must navigate a delicate equilibrium between being deemed attractive and not excessively so, as an overabundance of attractiveness may render apparent emotional instability less acceptable to men. This unique perspective raises intriguing questions regarding the threshold at which 'craziness' becomes acceptable or even appealing, particularly in women characterized by certain perceived attractiveness levels, including those with BPD (Blanchard et al., 2021). These considerations extend to the fetishization and romanticization of women with BPD,

reinforcing a gendered archetype and emphasizing the imperative of educational initiatives aimed at dispelling the misconceptions perpetuated by theories like the “Hot-Crazy Matrix.”

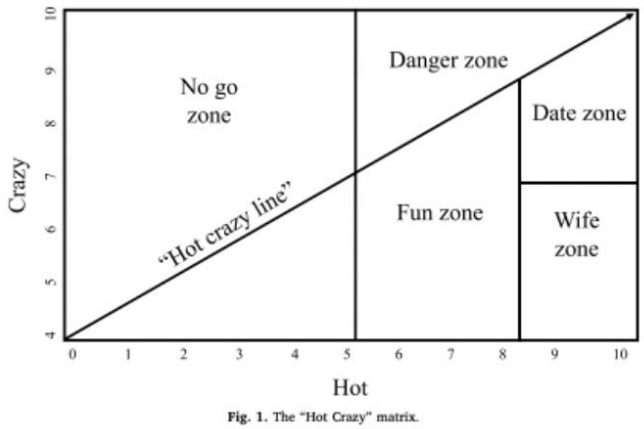


Fig. 1. The "Hot Crazy" matrix.

Figure 1: Example image illustrating the "hot crazy matrix" theory, derived from https://scienceintegritydigest.com/2021/02/12/the-hot-crazy-matrix-paper/

In the 21st century, the media's pervasive presence and substantial influence have significantly shaped public perceptions of mental health disorders, often perpetuating stereotypes and misconceptions. This influence extends beyond news media, encompassing the entire entertainment industry. Researcher Kubrak's (2020) meta-analysis, which examined the impact of films on young people's attitudes, highlights the profound reach of the entertainment industry in molding beliefs and opinions, particularly on topical social issues. Kubrak's work emphasizes the capacity of movies to significantly impact specific areas, such as gender and ethnic stereotypes, thereby effectively transforming societal perceptions and generating new opinions on various subjects. She suggests that movies, as an immersive experience driven by narrative transportation, are responsible for their significant impact on viewers' beliefs and attitudes, including those related to mental health (Kubrak, 2020). When movies portray mental health positively, they foster empathy and understanding among viewers. Conversely, negative portrayals can reinforce stereotypes and biases, perpetuating harmful attitudes and stigma. This insight underscores the responsibility of the entertainment industry to depict mental health issues accurately and sensitively, given its potential to influence societal perceptions and attitudes.

Expanding upon Kubrak's insights, Green and Sestir (2017) introduce the concept of transportation theory to explain the immersive power of film narratives. Transportation theory describes how individuals mentally immerse themselves in the world of a narrative, shifting their focus from the real world to the fictional one. This immersion results in a strong sense of absorption, encompassing emotional and cognitive responses to the narrative content that closely mirrors reactions to real-world events. Those transported into a narrative may also experience vivid mental imagery, further blurring the line between fiction and reality (Green & Sestir, 2017). This state of narrative transportation underscores the persuasive impact of storytelling in films, as influenced individuals often exhibit increased beliefs, attitudes, and behaviors consistent with the story's content. This concept underscores the potency of storytelling through film, as it can effectively shape public perceptions and contribute to both positive and negative societal attitudes, depending on how mental health issues are portrayed on screen.

The media's influence on shaping perceptions becomes especially evident when examining the portrayal of BPD and other cluster B personality disorders. Dr. Michael E. Thase, a distinguished professor of Psychiatry and director of the Penn Medicine Mood and Anxiety Program, highlights the enduring misunderstandings surrounding BPD, partially attributing this to its complex diagnostic history and its often unfavorable depiction in media (Wust, 2018). Dr. Thase notes that movies and T.V. shows tend to represent individuals with BPD as overly colorful, theatrical figures with unpredictable mood swings and potentially dangerous behavior, particularly when angered. One notable example is the portrayal of the 'bunny boiler' character by Glenn Close in "Fatal Attraction," which epitomizes this broad and exaggerated representation. He emphasizes that such portrayals should not be viewed as accurate reflections of BPD but rather sensationalized depictions that "fail to capture the complexity of this genuine mental health condition" (Wust, 2018). These examples show that the perpetuation of misconceptions through media misrepresentations contributes to the stigmatization of BPD, highlighting the necessity for more precise and nuanced portrayals in media outlets.

In filmmaking, directors and screenwriters grapple with challenges when attempting to convey the complex nature of BPD within the confines of a two-hour movie. The demand for dramatic portrayals and entertainment value often precedes the need for accuracy. The film analysis conducted by Alexis Ian Rainey on 'A Textual Analysis of Mental Illness in Film' offers valuable insights into the portrayal of mental illness in cinema, and she critiques various films depicting mental illness and analyzes them, specifically in the film "Girl Interrupted." Girl Interrupted presents an illustrative example of BPD on screen, portrayed theatrically. While the film is based on author Susanna Kaysen's real-life experiences during her 18-month stay at a mental hospital in the late 1960s, the director James Mangold, known for his dramatic works, adapted this movie, which resulted in creative discrepancies between the book and the film (IMDb). While the film draws inspiration from author Susanna Kaysen's real-life experiences with BPD, it is essential to note that Susanna's symptoms, as portrayed in the film, deviate from an official diagnosis of BPD at the time. This discrepancy underscores how the film misrepresents BPD by depicting actions that do not align with the diagnosis.

Furthermore, in Rainey's (2015) analysis, she notes her observations on the character of Susanna, which is quite noteworthy. Susanna, the protagonist at one point in the movie, arrives at the hospital with diagnostic paperwork labeling her as 'depressed' and 'promiscuous,' along with a note of a suicide attempt. What is particularly interesting is that Susanna is the only character who eventually leaves the hospital in a seemingly mentally stable state, capable of resuming life outside the hospital walls. This state exceeds the mental well-being of other patients. While not diagnostically accurate, this portrayal of Susanna significantly normalizes her character, making her more relatable and accessible to the audience. This observation underscores the film's potential to destigmatize BPD and mental illness by presenting characters who challenge stereotypes and evoke empathy from the viewers. (Rainey 2015). The movie "Girl Interrupted" exemplifies how filmmakers may resort to dramatic portrayals to capture the essence of a mental health disorder, even at the cost of sacrificing accuracy for entertainment purposes.

Consequently, these inaccurate portrayals in movies contribute to misconceptions about mental health, as the exaggerated nature of films primarily designed for entertainment perpetuates stereotypes and hinders public understanding (IMDb). By highlighting Dr. Thase's insights and the challenges filmmakers face, it becomes apparent how media depictions of BPD can diverge significantly from the reality of the condition, ultimately impacting societal perceptions and attitudes. These portrayals influence how the general public understands BPD and underscore the importance of responsible and accurate depictions to combat stigmatization and promote empathy and understanding.

Filmmakers resort to dramatization for various reasons. Firstly, drama and exaggeration are fundamental elements of storytelling and entertainment, designed to captivate audiences and maintain engagement. Given the intricacies of mental health disorders like BPD, which are characterized by subtle nuances, filmmakers may feel compelled to accentuate specific traits or behaviors to make the narrative more compelling. Henry J. Steadman, Ph., an expert in criminal justice and behavioral health, corroborates this notion by noting that mental illnesses are often portrayed in films and media with "exaggerated, inaccurate, and comical images used to depict individuals with psychiatric disorders, along with providing incorrect information about mental illness" (Steadman et al., 1981). Television, including films, plays a significant role in teaching people about behaviors, social conventions, and rules of conduct. It frequently reinforces social norms regarding how individuals with psychiatric disorders should be treated. This phenomenon aligns with Albert Bandura's (1977) social learning theory, which suggests that people learn through direct experience and observation (Saul Mcleod,2023). Without personal experiences with individuals with mental illnesses, many people rely on media portrayals to form their perceptions of those with mental health conditions.

Addressing these issues necessitates a concerted effort by the entertainment industry to responsibly and accurately depict mental health issues. Furthermore, it calls for greater public awareness and media literacy to help viewers distinguish between cinematic and theoretical stereotyped portrayals and the real-life experiences of individuals with mental health conditions. Ultimately, the paramount goal should be the promotion of empathy and understanding through accurate portrayals to reduce the stigma surrounding mental illness.

Conclusion

In conclusion, Borderline Personality Disorder is inextricably linked with pervasive societal stigma, giving rise to a web of complex challenges that individuals living with this disorder navigate. The complex interaction of historical factors, societal misconceptions, and internalized self-stigma significantly shapes the experiences of individuals affected by BPD. This dynamic is rooted in the historical perception of BPD as primarily affecting women, thus contributing to gendered stigmatization that continues to influence our collective consciousness. Although more pronounced for women, these gendered stereotypes involve negative perceptions such as being overly emotional, manipulative, or unstable, which can negatively impact men’s willingness to access mental health resources. When individuals receive a BPD diagnosis, these

stereotypes can trigger self-stigma, shifting the focus from potential underlying trauma and adversity to perceived inherent personality flaws. This shift exacerbates self-stigmatization as those affected may feel that their suffering is dismissed or misattributed.

Moreover, society significantly perpetuates BPD stigma through a lack of education and awareness regarding this complex mental health condition. Frequently sensationalized and inaccurate media portrayals have further entrenched these stereotypes, making it exceedingly challenging for individuals with BPD to find understanding and support within their communities. Despite the presence of legitimate diagnostic criteria for BPD in the DSM-5, it remains misunderstood, misdiagnosed, and heavily stigmatized. This stigmatization also extends to men, raising questions about whether societal values and expectations of individuals contribute to the observed gender disparity in BPD diagnoses. This gendered nature of diagnosis criteria, exemplified by instruments like the Beck Depression Inventory (BDI) for depression, introduces an additional layer of complexity to mental health disparities. The BDI's potential to inadvertently inflate the presence of depressive symptoms in women due to its alignment with traditional expressions of depression symptoms can distort conclusions about gender disparities in mental health. Confronting the historical and societal biases that have contributed to the stigmatization of BPD is crucial for cultivating a more inclusive and empathetic society. As a society, we must address the gender disparities in mental health diagnoses to advocate for the well-being of individuals affected by BPD and mental illness and bring us a step closer to destigmatizing mental illness. Achieving this goal necessitates a concerted effort to educate the public and media about the realities of this condition, promote accurate portrayals in the media, and implement comprehensive educational initiatives to dispel misconceptions and cultivate empathy. By engaging in these efforts, we can contribute to breaking the cycle of stigmatization and establishing an environment where individuals with BPD receive acceptance and active support on their path to recovery and well-being.

Works Cited

Ahmedani, B. K. (2011). Mental Health Stigma: Society, Individuals, and the Profession. Journal of Social Work Values and Ethics, 8(2), 41–416.

Ali, A., & Adshead, G. (2022). Gender bias in the diagnosis of depression: A systematic review. *International Review of Psychiatry*, *34*(2), 97–107. doi:10.1080/09540261.2021.1903147

American Psychiatric Association. (2013). Borderline Personality Disorder Diagnostic Criteria 301.83 (F60.3) In Diagnostic and Statistical Manual of Mental Disorders (5th ed.). https://doi.org/10.1176/appi.books.9780890425596.CautionaryStatement

P. Antolin, "Challenging Borders: Susanna Kaysen's Girl, Interrupted as a Subversive Disability Memoir." *European Journal of American Studies*, vol. 15, no. 2, June 23 2020, https://doi.org/10.4000/ejas.16051.

Armstrong, J., & Nacu, S. (2023, February 10). Opinion: What is wrong with "Girl Interrupted." The Cougar Press. URL Feb.10

Boone, T., Reilly, A. J., & Sashkin, M. (1977). SOCIAL LEARNING THEORY Albert Bandura Englewood Cliffs, N.J.: Prentice-Hall, 1977. 247 pp., paperbound. *Group & Organization Studies*, *2*(3), 384–385. https://doi.org/10.1177/105960117700200317

Berenson, K. R., Downey, G., Rafaeli, E., Coifman, K. G., & Paquin, N. L. (2011). The rejection-rage contingency in borderline personality disorder. *Journal of Abnormal Psychology*, *120*(3), 681–690. https://doi.org/10.1037/a0023335

Bik, Elies. "The Hot-Crazy Matrix Paper." *Science Integrity Digest*, February 12. , scienceintegritydigest.com/2021/02/12/the-hot-crazy-matrix-paper/.

Blanchard, A. E., Dunn, T. J., & Sumich, A. (2021). Borderline personality traits in attractive women and wealthy, attractive men are relatively favored by the opposite sex. Personality and Individual Differences, 169. doi:10.1016/j.paid.2020.109964

Bogod, E. (2010). What is in a Name? Borderline Personality Disorder Label Creates Stigma. Community Mental Health Worker, In Recovery from BPD. Retrieved from https://icmha.ca/wp-content/uploads/2015/01/Bogod2010BPD.pdf

Bronfen, E. (1998). *The knotted subject: Hysteria and its discontents*. Princeton University Press.

Dorfman, N., & Reynolds, Joel Michael (forthcoming). The New Hysteria: Borderline Personality Disorder and Epistemic Injustice. *International Journal of Feminist Approaches to Bioethics*.

Dvoskin, Joel A, (1996). "The Abuse of Psychiatric Diagnosis in Prisons." www.vachss.com/guest_dispatches/ dvoskin.html.

Eaton, N. R., Krueger, R. F., Keyes, K. M., Skodol, A. E., Markon, K. E., Grant, B. F., & Hasin, D. S. (2011). Borderline personality disorder comorbidity: Relationship to the internalizing-externalizing structure of common mental disorders. Psychological Medicine, 41(5), 1041-1050. doi:10.1017/S0033291710001662

Farrell, M. D. (2012). The portrayal of schizophrenia in contemporary mass media. Psychiatric News, 47(5), 27-29.

Feingold A. (1992). Gender differences in mate selection preferences: a test of the parental investment model. *Psychological Bulletin*, *112*(1), 125–139. https://doi.org/10.1037/0033-2909.112.1.125

Ferguson, Sian. "Signs of Bipolar Misdiagnosis." *Healthline*,1 Feb. 2023, www.healthline.com/health/bipolar-disorder/Bipolar-misdiagnosis.

Gadassi, R., Snir, A., Berenson, K., Downey, G., & Rafaeli, E. (2014). Out of the frying pan, into the fire: mixed affective reactions to social proximity in borderline and avoidant personality disorders in daily life.

Journal of Abnormal Psychology, *123*(3), 613–622. https://doi.org/10.1037/a0036929

Gilman (1994) L. *The Case of Sigmund Freud: Medicine and Identity at the Fin De Siècle.* Baltimore, Johns Hopkins Univ. Press, 1994. press.jhu.edu/books/title/2522/case-sigmund-freud.

Green, Melanie & Sestir, M. (2017). Transportation Theory. 10.1002/9781118783764.wbieme0083.

Grambal, A., Prasko, J., Kamaradova, D., Latalova, K., Holubova, M., Marackova, M., Ociskova, M., & Slepecky, M. (2016). Self-stigma in borderline personality disorder - cross-sectional comparison with schizophrenia spectrum disorder, major depressive disorder, and anxiety disorders. Neuropsychiatric Disease and Treatment, 12, 2439–2448. doi:10.2147/NDT.S114671

Henry J. Steadman, A situational approach to violence, International Journal of Law and Psychiatry,

5(2)(1982),https://doi.org/10.1016/0160-2527(82)90004-8.

Karim, Hiba & Shirazi, Elham & Nohesara, Shabnam & Sadeghi, H.M. & Saeb, A. & Alavi, Kaveh. (2019). Comparison of gender roles in male and female patients with borderline personality disorder (BPD) with control group and its correlation with severity of clinical symptoms. Revista Latinoamericana de Hipertension. 14. 1-7.

Kubrak T. (2020). Impact of Films: Changes in Young People's Attitudes after Watching a Movie. *Behavioral sciences (Basel, Switzerland)*, *10*(5), 86. https://doi.org/10.3390/bs10050086

Grambal, A., Prasko, J., Kamaradova, D., Latalova, K., Holubova, M., Marackova, M., Ociskova, M., & Slepecky, M. (2016). Self-stigma in borderline personality disorder - cross-sectional comparison with schizophrenia spectrum disorder, major depressive disorder, and anxiety disorders. *Neuropsychiatric disease and treatment*, *12*, 2439–2448. https://doi.org/10.2147/NDT.S114671

Jeong, H., Jin, M. J., & Hyun, M. H. (2022). Understanding a Mutually Destructive Relationship Between Individuals With Borderline Personality Disorder and Their Favorite Person. Psychiatry Investigation, 19(12), 1069–1077. doi:10.30773/pi.2022.0079

Jimenez, M. A. (1997). Gender and Psychiatry: Psychiatric Conceptions of Mental Disorders in Women, 1960-1994. *Affilia*, *12*(2), 154–175. https://doi.org/10.1177/088610999701200202

Karim, S., Arshad, M., & Matin, A. (2019). The gendered nature of borderline personality disorder and mental health: A comprehensive analysis. *Journal of Gender Studies*, *28*(6), 698–712. doi:10.1080/09589236.2018.1562422

Klein, P., Fairweather, A. K., & Lawn, S. (2022). Structural stigma and its impact on healthcare for borderline personality disorder: A scoping review. *International Journal of Mental Health Systems*, *16*(1), 48. doi:10.1186/s13033-022-00558-3

Lambert, V. (2020). Pathological Femininity and the Western Construction of Selfhood: A Genealogy of the Borderline Subject. Retrieved from https://researchspace.auckland.ac.nz/bitstream/handle/2292/53466/Lambert-2020-thesis.pdf?sequence=5

Lampropoulos, G., Anastasiadis, T., Siakas, K., & Siakas, E. (2022). The impact of personality traits on social media use and engagement: An overview. International Journal on Social and Education Sciences (IJonSES), 4(1), 34-51. https://doi.org/10.46328/ijonses.264

Link, B. G., & Phelan, J. C. (2001). Conceptualizing stigma. *Annual Review of Sociology, pp. 27,* 363–385. https://doi.org/10.1146/annurev.soc.27.1.363

Mcleod, S. Albert Bandura's Social Learning Theory. *Simply Psychology* (2023), www.simplypsychology.org/bandura.html.

Munson, K. A., Janney, C. A., Goodwin, K., & Nagalla, M. (2022). Cultural Representations of Borderline Personality Disorder. *Frontiers in Sociology*, p. 7, 832497. https://doi.org/10.3389/fsoc.2022.832497

New, A., Rot, M., Ripoll, L., Perez-Rodriguez, M. M., Lazarus, S., Zipursky, E., ... Siever, L. (2012).

Empathy and Alexithymia in Borderline Personality Disorder: Clinical and Laboratory Measures. Journal of Personality Disorders, 1-16. https://doi.org/10.1521/pedi_2012_26_037

North, C. S. (2015). The Classification of Hysteria and Related Disorders: Historical and Phenomenological Considerations. *Behavioral Sciences*, pp. 5, 496–517. doi:10.3390/bs5040496

Ohshima, T. (2001). Borderline personality traits in hysterical neurosis. *Psychiatry and Clinical Neurosciences*, *55*(2), 131–136. https://doi.org

Sansone, R. A., & Sansone, L. A. (2010). Fatal attraction syndrome: stalking behavior and borderline personality. *Psychiatry (Edgmont (Pa.: Township))*, *7*(5), 42–46.

Alaska, Candice. (2020). "Why We Need Responsible Representation in the Film Industry." *National Alliance for Mental Health*, December 16, 2020, www.nami.org/Blogs/NAMI-Blog/December-2020/“Why-We-Need-Responsible-Representation-in-the-Film-Industry.

A.I. 2015 *Stigma on the Screen: A Textual Analysis of Mental Illness*

in Film. U of Georgia, MA thesis. getd.libs.uga.edu/pdfs/Rainey_alexis_i_201505_ma.pdf.

Freeman, Ronald Roesch, Richard J., (1989). Mental disorder and the criminal justice system: A review, International Journal of Law and Psychiatry, 12(2-3),105-1150160-2527,https://doi.org/10.1016/0160-2527(89)90002-2.

Sagar-Ouriaghli, I., Godfrey, E., Bridge, L., Meade, L., & Brown, J. S. L. (2019). Improving Mental Health Service Utilization Among Men: A Systematic Review and Synthesis of Behavior Change Techniques Within Interventions Targeting Help-Seeking. *American journal of men's health*, *13*(3), 1557988319857009. https://doi.org/10.1177/1557988319857009

Shaw, C. & Proctor, G. (2005). I. Women at the Margins: A Critique of the Diagnosis of Borderline Personality Disorder. Feminism & Psychology 15. 483–490. 10.1177/0959-353505057620.

McKenzie, S. K., Collings, S., Jenkin, G., & River, J. (2018). Masculinity, Social Connectedness, and Mental Health: Men's Diverse Patterns of Practice. *American journal of men's health*, *12*(5), 1247–1261. https://doi.org/10.1177/1557988318772732

Sheppard, H., Bizumic, B., & Calear, A. (2023). Prejudice toward people with borderline personality disorder: Application of the prejudice toward people with mental illness framework. *International Journal of Social Psychiatry*. https://doi.org/10.1177/00207640231155056

Boozer, Jack, editor. *Authorship in Film Adaptation*. University of Texas Press, 2008. *JSTOR*, https://doi.org/10.7560/702851. Accessed September 14, 2023.

A.(2019) "Are Sexual Abuse Victims Being Diagnosed with a Mental Disorder They Do not Have?" *The Guardian*, March 27, 2019, www.theguardian.com/lifeandstyle/2019/mar/27/Are-sexual-abuse-victims-being-diagnosed-with-a-mental-disorder-they-dont-have.

Waraich, M., & Shah, S. (2018). The life and work of Jean-Martin Charcot (1825-1893): 'The Napoleon of Neuroses.' *Journal of the Intensive Care Society*, *19*(1), 48–49.

The Influence of the Shape of the Cam Pulley on Leg Extension Machines By Junghyun Koh

Abstract

The optimal design of leg extension exercise machines is crucial in effectively targeting and activating the quadricep muscles during exercise. This research paper investigates the shape effectiveness of the cam pulley on leg extension machines on the tension in the user's quadriceps. This study surveyed three subjects for anatomical measurements, collected data using two different selectorized leg extension machines, and calculated the force and torque. The results provide insights into the function of the non-circular shape of the cam pulley in designing leg extension machines. This keeps the force on the shin pad consistent by applying more force during certain periods throughout the course of leg extension. With this insight, a design of a leg extension machine accounting for the differences in each individual's leg weight is proposed.

Introduction

Training quadriceps with a leg extension (LE) machine is important because quadricep muscles are critical for daily movement by helping to extend the leg at the knee joint and stabilize the knee by holding the patella (kneecap). A study by Scarborough (1999) shows the importance of quadricep strength in elderly patients. The patients with the greatest quadriceps strength were more stable when conducting chair rise exercises. In fact, the authors concluded that quadriceps strength had a significant correlation with maximum upper body vertical linear momentum as well as maximum upper body anteroposterior (A/P) linear momentum (Scarborough 1999). This demonstrates the strength of the quadricep muscle increases the subjects' speed in performance. Thus, the quadriceps are significant for conducting athletic movements, such as jumping, running, and kicking, as well as basic mobility movements such as walking, sitting up from a chair, and climbing stairs.

Moreover, a study by Jonsson and Alfredson (2005) revealed that eccentric quadriceps training can relieve knee pain. In this study, the authors used 15 subjects who suffered from chronic pain from the proximal patellar tendon. These subjects were asked to train their quadriceps on a decline board every day for 12 weeks. The results showed reduced tendon pain during activity and improved function. This suggests that not only does quadriceps training increase efficiency when performing basic mobility such as walking, running, and standing up, but it also can help relieve knee pain. Therefore, quadriceps are significant for rehabilitation as well.

Many selectorized leg extension machines (SLE) contain cam pulleys (Fig. 1). On an SLE, the cam pulley is usually located next to the seat. With the upward thrusting motion of the leg on the shin pad (SP), the cam pulley (CP) pushes down on the cable that is connected to the weights. This enables the person to load weights on their legs to effectively train their lower

body. Unlike conventional pulleys that are circular, these cam pulleys are often non-circular, such as being triangular with round edges or oval (Fig. 2).

While several studies have investigated the design of LE machines, none have specifically shown how the shape of the CP on the LE machines affects the tension in the quadriceps. This study will explore the function and shape of the cam pulley and test how the tension in the quadriceps changes over the course of LE.

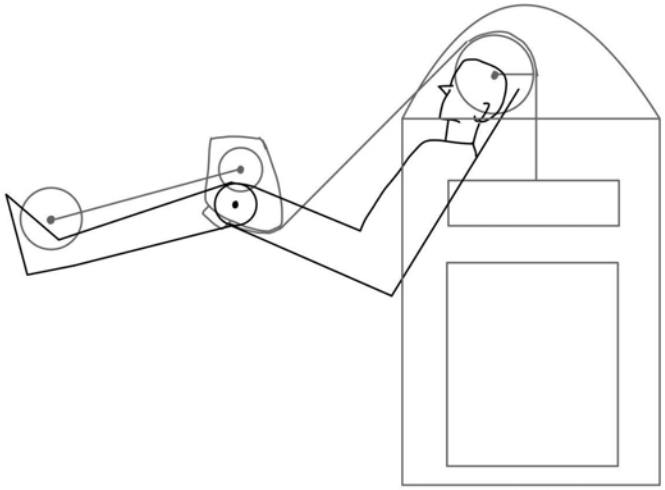


Figure 1: Selectorized leg extension machine that contains a cam pulley

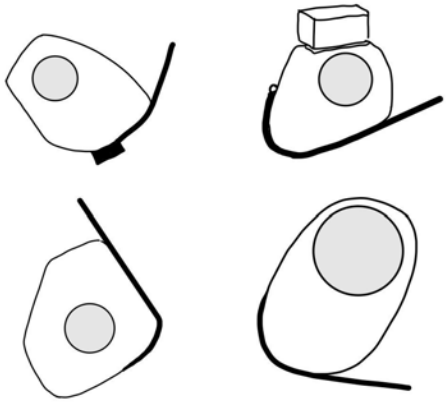


Figure 2: Unlike conventional pulleys that are circular, the cam pulleys on many selectorized leg extension machines are not circular. They are often irregularly shaped: triangular with round edges or oval-shaped.

Methods and Materials–*Subjects*

For morphometric data collection, three participants (two females and one male) served as subjects. Through a series of surveys, each subject's weight, distance from the foot to the patella, and distance from the patella to the knee joint were measured (Tab. 3). The participants

were not required to have been trained on how to use the LE machine because only their bodily measurements were used for data analysis.

Data Collection

This experiment used two types of SLE machines, A and B SLE machines with varying shapes of CPs (Fig.3). First, the force of the shinpad (SP) on a leg (F_{SL}) without any weight added was measured using an electronic luggage scale (ELS). The ELS was tied around the SP and was elevated, allowing the ELS to scale the F_{SL} at three angles: 30°, 60°, and 90°, which is the angle of the tibia rotation from the starting position. The rope of the ELS (rELS) was held at 90º relative to the SP to measure the maximum normal force exerted by the SP. Next, the average of the measured F_{SL} values in three trials at each angle was calculated. Additional data for the calculation was collected from a human cadaver study (Clauser et al. 1969). According to the result of that study, the approximate mass of a human calf and foot combined is 3.805 ± 0.12 kg, and the approximate length from the knee joint to the center of mass (CM) of a human tibiale is 21.67 ± 0.3 cm. A value of 9.8 m/s^2 was used for the gravitational acceleration constant.



Figure 3: Selectorized leg extension machines with varying shapes of cam pulleys

Rigid Bodies	ABB	Forces & Torques
Shin pad+Cam pulley	S	$\Sigma F = F_{LS} + F_{GS} + F_{CS} + F_{MS} = 0$ $\Sigma \tau = \tau_{LS} + \tau_{GS} + \tau_{CS} = 0$
Leg+Foot	L	$\Sigma F = F_{SL} + F_{LS} = 0$ $\Sigma \tau = \tau_{LS} + \tau_{GL} + \tau_{LP} = 0$
Tension-only bodies		
Quadriceps	Q	$\Sigma F = F_{FQ} + F_{PQ} = 0$
Patellar tendon/ligament	P	$\Sigma F = F_{LP} + F_{QP} = 0$
Cable	C	$\Sigma F = F_{SC} + F_{WC} = 0$
Other bodies		
Machine+Torso+Femur	R	Fixed reference
Weight	W	$\Sigma F = F_{CW} + F_{GW} = 0$
Circular pulley	U	$\Sigma F = F_{CU} + F_{GU} + F_{MU} = 0$
Not bodies		
Gravity	G	

Table 1: Force body table showing forces present in the course of LE in static equilibrium

Center of Mass of Tibial	T
Knee Joint	K
Foot	F
Distance from center of mass of tibial to knee joint	D_{TK}
Distance from patella to knee joint	D_{PK}
Distance from foot to patella	D_{FP}

Table 2: Subscript definitions

Calculation

With the average F_{SL} values at each angle, measurements were collected from each participant, using the approximate mass of a calf and foot of 3.805 ± 0.12 kg, and the approximate length from the knee joint to the center of mass (CM) of the tibia 21.67 ± 0.3 cm, the tension in the quadriceps was calculated. The force of the SP on F_{SL} without any weight added was substituted into a derived equation using the fundamental physics principles of force and torque to find the tension in the quadriceps. At each angle, three trials of measurements were made for accuracy, and there were a total of three measured values for each angle. To calculate the tension in the quadriceps, the average of these three values was calculated (Tab. 4.1, 5.1). Next, the average values were plugged into the derived equation. The equation was derived under the consideration that the system is at static equilibrium (Equation 1) where T_{LF} is the tension in the leg+foot system or the quadriceps, F_{LS} is the force of the leg+foot system on the shinpad, D_{FP} is the distance from foot to the patella, F_{GL} is the force of gravity of the leg+foot system, D_{TK} is distance from center of mass of tibial to knee joint, and D_{PK} is the distance from patella to knee joint.

$$T_{LF} = \frac{sin\theta\left(F_{LS} \cdot D_{FP} + F_{GL} \cdot D_{TK}\right)}{D_{PK}}$$

Equation 1: The derived equation for the tension in the quadriceps equation during the course of LE

At static equilibrium, the sum of the forces in a system equals 0 N and the same is true for the sum of the torques. The same calculation was made for each angle and each SLE machine. As a result, there were a total of three values, which represented tension in the quadriceps for each SLE machine. The same process was repeated for each participant (Tab. 4.2, 5.2).

Results

			Distance from the patella to the knee joint	Distance from foot to patella
Participant	Age (year)	Weight (kg)	D_PK (cm)	D_FP (cm)
1	48	54	6.5	50.5
3	18	50	5.5	45.8
2	39	75	7.2	44.5

Table 3: Measurements from subjects included age, weight, distance from the patella to the knee joint, and distance from the foot to the patella

SLE Machine A:

Angles (Degrees)	Average F_SL (N)
30	82
60	74
90	75

Table 4.1. F_{SL} of SLE Machine A

SLE Machine A	30º (N)	60º (N)	90º (N)
Participant #1	380.7973562	604.172581	706.77502
Participant #2	309.597891	492.1773479	575.5857819
Participant #3	414.9857664	659.4102959	771.2142964

Table 4.2. Tension in quadriceps when using SLE Machine A

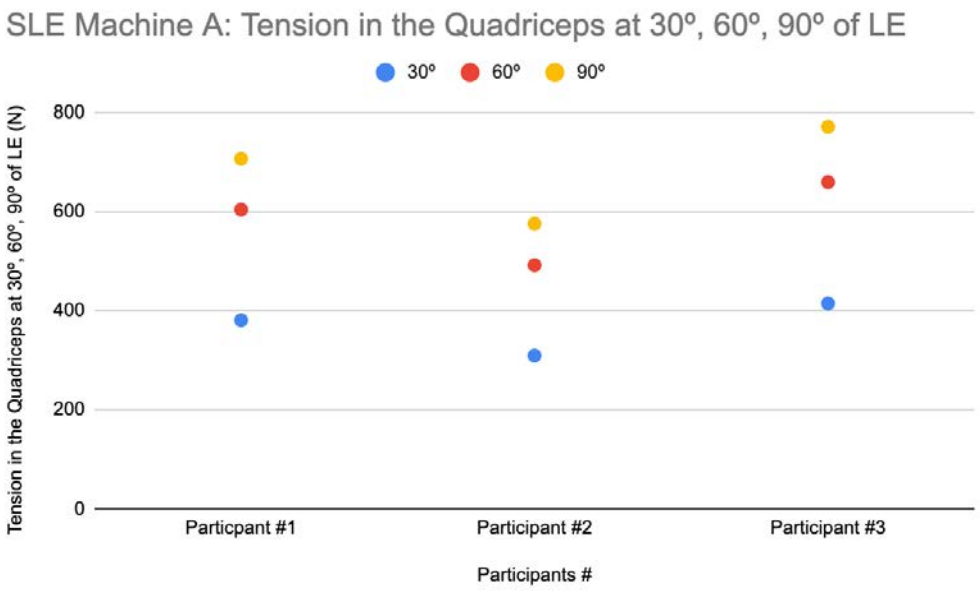


Figure 4: Graph of SLE Machine A: Tension in the Quadriceps at 30º, 60º, 90º of LE

SLE Machine B:

Angles (Degrees)	Average F_SL (N)
30	130
60	120
90	112

Table 5.1: F_{SL} of SLE Machine B

SLE Machine B	30º	60º	90º
Participant #1	568.4786638	918.6960869	992.2942508
Participant #2	458.9015715	742.38616	802.7211986
Participant #3	616.1476936	996.525139	1077.241569

Table 5.2: Tension in quadriceps when using SLE Machine B

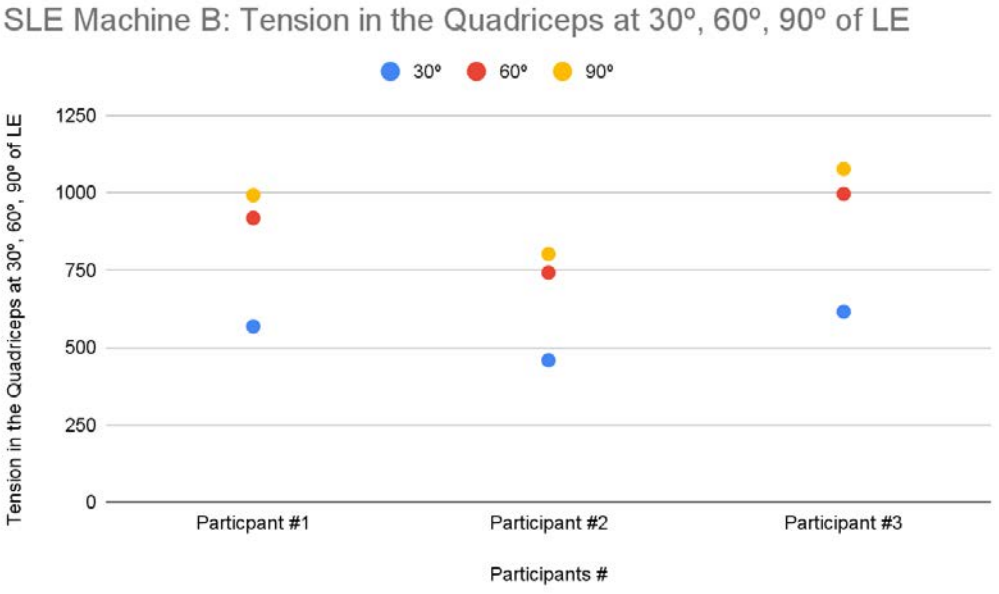


Figure 4: Graph of SLE Machine B: Tension in the Quadriceps at 30º, 60º, 90º of LE

Over the course of an LE, the angle of the moving arm and the F_{SL} and F_{GL} should change. For example, the torque from both F_{SL} and F_{GL} increases as the angle of extension (θ) increases. Torque is equivalent to the perpendicular force multiplied by the length of the moment arm. However, as the θ changes, both F_{SL} and F_{GL} no longer exert force perpendicularly to their moving arms. Thus, the torque changes by the factor of sin(θ). At 90º, the torque on the leg-foot system is the greatest, while at 30º, the torque on the leg-foot system is the smallest out of the range of data collected (Fig. 5).

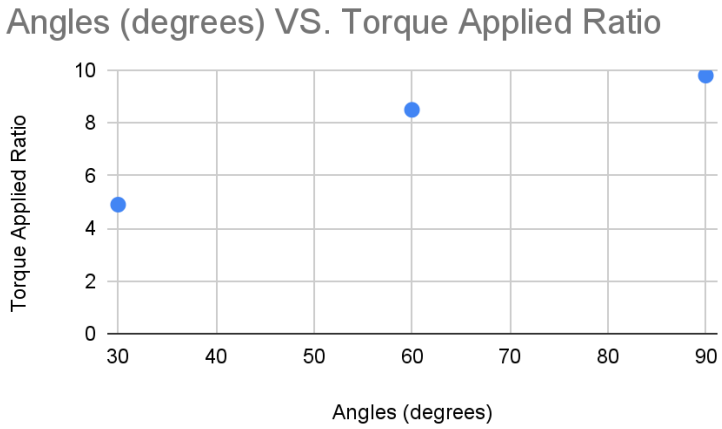


Figure 5. Graph showing the relationship between the angle of extension of the leg and the torque on the leg+foot system

The ratio of the torque on the leg-foot system should be 50% at 30º, 87% at 60º, and 100% at 90º (Fig. 5). Thus, the measured F_{SL} should also have values following this ratio over the course of using the LE. However, the measured F_{SL} values did not follow this ratio. For example, in SLE machine A, the F_{SL} values at 30º, 60º, and 90º, were 82 N, 74 N, and 75 N, respectively (Tab. 4.1). Instead, the F_{SL} values were similar at all three angles. Both measured F_{SL} values from SLE machines A and B were similar. The F_{SL} values at 30º, 60º, and 90º, in order, were 130 N, 120 N, and 112 N (Tab. 5.1).

The differences among the F_{SL} values across a range of LE angles from both SLE machines A and B suggest that an athlete performing on the LE machines would have similar tension in the quadriceps over the course of the workout. If the CP was circular, the measured F_{SL} values should vary by the ratio, 0.5:0.87:1. However, the actual values from both SLE machines A and B follow the approximate ratio of 1.0:0.9:0.9, varying only slightly from each of their mean values. In fact, it is the non-circular shape of the CPs of the SLE machines that prevents such significant variances in measured F_{SL} values. The shape of the cam pulley is designed to add extra force in the first 0º to 60º of LE to maintain similar force, and thus, torque, on the leg-foot system throughout the course of LE.

Discussion

This study provides evidence that the shape of the CP modifies the force of the SP on the leg to maintain constant force throughout the course of an LE, although the findings are limited to measurements obtained from two SLE machines and calculations. Further studies should investigate several SLE machines, each with different CP shapes. By exploring the function of the non-circular CP in SLE machines, this study provides insight into potential cam shape designs that allow the force on the leg to be nearly constant throughout the LE. And also identifies a means to discover a design that not only maintains consistent force of the shin pad on the leg but also ensures a consistent tension in the quadriceps throughout the LE.

For future studies to ensure consistent tension in the quadriceps through LE, the machine should account for the change in torque on the leg due to gravity. Depending on the weight of the athlete's leg, the force of the SP will be felt differently due to gravity on a leg at any one angle. Thus, when athletes with different leg weights perform an LE, the athlete with greater leg weight would have more tension in their quadriceps even though the same amount of weight is added by the machine.

Furthermore, to maintain a set tension in the quadriceps of athletes with all leg weights, the CP should keep its shape and include rubber or memory foam padding around its rim. Depending on the weight of the leg or the force of the leg on the SP, the rubber or foam padding will prevent force on the athlete's leg by lessening the momentum in the SP by the applied force. The rubber or foam padding will work for most athletes with different leg weights because the padding adjusts the force of the SP on a leg with its property of dampening the force exerted without having the individual to adjust the LE machine.

An advantage of having such a design of an LE machine is establishing consistent force of the SP for all athletes. Such design of LE machines with this CP makes them more accessible, regardless of the experience of the athlete with the machine use. Such CP designs will prevent the need for individual adjustment of the machine while maintaining consistent tension in the quads for most users. Free-weight quadricep exercises require trial and error depending on the athlete's proficiency level because they require mindful adjustment in form, weight, and power distribution. Compared to free weights, this design of the LE machine will be more efficient and easy to use for beginners.

Works Cited

Clauser, C. E., Mc Conville, J. T., & Young, J. W. (1969). Weight, volume, and center of mass of segments of the human body. Ntrs.nasa.gov. Retrieved from https://ntrs.nasa.gov/citations/19700027497

Jonsson P, Alfredson H. (2005). Superior results with eccentric compared to concentric quadriceps training in patients with jumper's knee: a prospective randomized study. British Journal of Sports Medicine, 39:847-850. DOI: 10.1136/bjsm.2005.01863

Moxley Scarborough, D., Krebs, D. E., & Harris, B. A. (1999). Quadriceps muscle strength and dynamic stability in elderly persons. Gait & Posture, 10(1), 10-20. DOI: 10.1016/S0966-6362(99)00018-1

Using Regulatory T-Cells to Treat Type 1 Diabetes in Humans By Jaisal Vohra

Introduction

Type 1 Diabetes (T1D) is an autoimmune disease that results from the destruction of insulin-producing beta cells in the Islets of Langerhans of the pancreas. This destruction is caused by T cells and leads to insulin deficiency (Chen et al.). Insulin regulates blood glucose levels by facilitating the uptake of blood sugar into cells, which the cells can then use for energy. Without insulin, sugar builds up in the bloodstream, leading to hyperglycemia (CDC). Two HLA genes involved in antigen presentation, DR3 and DR4-DQ8, as well as around 60 non-HLA genes, are associated with the development of T1D (DiMeglio et al.); HLA genes code for surface proteins that are used to identify body cells and foreign cells (Nordquist and Radia). Also, viral infections, including enterovirus infections, might be linked to the development of T1D (DiMeglio et al.).

Regulatory T Cells (Tregs) play a key role in preventing autoimmune diseases, including T1D, by regulating peripheral immune tolerance (Visperas and Dario). Immune tolerance involves mechanisms that regulate the immune system and prevent it from attacking the body's own cells. There are two types of immune tolerance - central immune tolerance and peripheral immune tolerance. Central immune tolerance takes place in the thymus. It involves mechanisms that destroy T cells that have an extremely high affinity for self-antigens, as that would lead to autoimmunity. Also, it destroys T cells that have a low affinity for the human leukocyte antigens (HLA), as that would prevent T cells from binding to other cells. Peripheral immune tolerance, on the other hand, involves mechanisms outside the thymus that regulate the immune system, and this involves Tregs (Erdem et al.).

Defective Tregs lead to a range of autoimmune diseases, including T1D. There are two types of Tregs, thymic-derived tTregs, which develop in the thymus, and peripherally-derived pTregs, which develop in the periphery. tTregs are the primary type of defective cell in T1D. tTregs arise from helper T-Cells (CD4 T Cells) that have a high affinity for self-antigens, an example of central immune tolerance. They travel to peripheral tissues during times of inflammation, suppressing the immune response caused by immunocytes like effector T cells (Teffs) (Visperas and Dario). Tregs do this by stopping the differentiation of CD4 T Cells as well as other immune cells, like dendritic cells (Tan et al.). In T1D, defective tTregs are unable to suppress the immune system while it attacks insulin-producing pancreatic beta cells. Thus, targeting Tregs has been a focus for many T1D therapies (Visperas and Dario).

Tregs in T1D

Tregs and their role in T1D have been studied in both humans and mice. Non-obese diabetic (NOD) mice develop diabetes similar to human T1D. In these mice, T Cells are reactive to multiple pancreatic antigens such as GAD65, IGRP, and insulin, leading to inflammation of the islets of Langerhans in the pancreas. However, Tregs specific to only one antigen were able to control autoimmunity caused by Teffs specific to multiple antigens according to a study. This

finding makes Treg therapies an even more viable option as a treatment for T1D, as a polyclonal Treg response is not required to control an autoimmune disease (Tarbell et al.).

Tregs also have multiple sub-phenotypes, but they can be further divided into three general types: resting Tregs, activated Tregs, and non-suppressive Tregs. One study found that activated Tregs are impaired in patients with T1D, but further research is needed to understand why. The number of activated Tregs is also higher in T1D patients, which is probably a response to counteract the reduced effectiveness of activated Tregs (Haseda et al.). Another study, this time in BDC2.5/NOD.Foxp3sf mice, found that Tregs don't inhibit Teff activation. They also don't suppress the proliferation of Teffs in the pancreatic lymph nodes. Rather, they slow down the rate of insulitis in the pancreas. The Tregs in the pancreas overexpress many genes compared to Tregs in the lymph nodes, which likely gives Tregs in the islets of the pancreas greater suppressive capability (Chen et al.). From these two studies, it can be concluded that targeting Tregs could help delay or completely prevent T1D in humans due to the role they play in the prevention of autoimmune diseases.

Modified Treg Therapies

Some Treg therapies aim to make up for defective Tregs by introducing altered Tregs, with the goal being to slow or reverse insulitis. One study enrolled 14 adult patients recently diagnosed with T1D. Tregs from peripheral blood were taken and expanded in a lab for 14 days. After 14 days, the expanded-Tregs expressed higher levels of markers associated with memory, development, trafficking to lymph nodes, and enhanced function, indicating that their suppressive capabilities were improved. This was further supported by the finding that they were more effectively able to suppress the proliferation of Teffs in the lab than non-expanded Tregs. When these expanded Tregs were injected back into patients, there was a spike in Treg numbers in peripheral blood between days 7 and 14 after injection. After the 14th day, numbers began to fall, but they were still found up to a year later. This therapy, however, did not prevent diabetes from progressing in patients (Bluestone et al.).

A shortcoming of this study was the fact that only Tregs in the periphery were analyzed. Tregs from the Islets of Langerhans and the inflammation site were not assessed, even though this subset of Tregs are the ones that prevent Teff proliferation. Even though this therapy did not prevent the progression of T1D, it was able to successfully and safely expand Tregs and inject them back into the patient without significant Teff contamination. It was also able to demonstrate that these expanded Tregs have a long life span, which future studies can build off of (Bluestone et al.).

IL2-Based Therapies

Other therapies, such as Interleukin-2 (IL-2) based therapies, affect Treg function indirectly. IL-2 is a growth factor essential to the development of Tregs. So, to improve Treg development, one lab injected NOD mice with mouse IL-2 (mIL-2). IL-2 has a short half-life, however, which is why it was combined with CD25, which increased the protein's half-life to 16

hours. At low doses, mIL-2/CD25 expanded Tregs in NOD mice without significantly expanding Teff populations. This resulted in a delay in diabetes progression. Also, younger mice were more responsive to treatment, suggesting that this therapy is more effective when diabetes develops slowly. When treatment stopped, however, some mice became diabetic while others didn't for another 10-15 weeks, meaning continuous dosing might be necessary (Ward et al.).

Another IL-2 based therapy utilized the antibody F5111.2 to stabilize IL-2 in such a way that it preferentially expanded Tregs. This resulted in selective expansion of Tregs in a humanized NOD mice model (Trotta et al.). This does not mean that this treatment can be translated into humans, however. An IL-2 treatment that combined Rapamycin and IL-2 proved to be effective in mice. But when injected into humans with T1D, it impaired the function of pancreatic beta cells. Other immune cells, such as natural killer (NK) cells and eosinophils, also temporarily spiked. This likely created an inflammatory environment that led to a reduction in beta cell function, worsening the disease. Furthermore, in this study, these negative effects most likely correlated to IL-2 and not rapamycin. However, it is important to note that even though the disease worsened, there was a temporary spike in Treg numbers (Long et al.).

Other Possibilities

Other findings about Tregs have also opened up new avenues for Treg-related therapies. For example, a study found that in the pancreas of NOD mice, a higher number of Tregs express the C-X-C Motif Chemokine Receptor 3 (CXCR3), a result of the T-box transcription factor TBX21 (T-BET). NOD mice with Tregs lacking this gene were reported to have more aggressive insulitis, suggesting that the T-BET gene helps prevent autoimmunity in NOD mice. As such, transferring CXCR3+ Tregs into autoimmune mice might be more effective at controlling autoimmunity than transferring non-specific Tregs. These findings can also be applied to humans because a subset of Tregs similar to the CXCR3 subset has been found in humans (Tan et al.).

Conclusion

Tregs play an important role in controlling autoimmunity, which is why modifying defective Tregs, either directly or indirectly, can potentially slow down or reverse insulitis. However, Tregs have multiple subsets, and variabilities in Treg subsets must be further researched in order to make Treg-related therapies more effective. This includes studying Tregs from inflamed and peripheral sites as well as during different stages of insulitis. As mentioned above, some therapies involving Treg modification and IL2 treatment show promises, but they need to be further tested in human subjects.

Works Cited

Bettini, Maria, and Matthew L Bettini. "Function, Failure, and the Future Potential of Tregs in Type 1 Diabetes." *Diabetes* vol. 70,6 (2021): 1211-1219. doi:10.2337/dbi18-0058

Bluestone, Jeffrey A et al. "Type 1 diabetes immunotherapy using polyclonal regulatory T cells." *Science translational medicine* vol. 7,315 (2015): 315ra189. doi:10.1126/scitranslmed.aad4134

Chen, Zhibin et al. "Where CD4+CD25+ T reg cells impinge on autoimmune diabetes." The Journal of experimental medicine vol. 202,10 (2005): 1387-97. doi:10.1084/jem.20051409

"Diabetes Basics." Center for Disease Control and Prevention, www.cdc.gov/diabetes/basics/index.html. Accessed 30 Sept. 2023.

DiMeglio, Linda A et al. "Type 1 diabetes." *Lancet (London, England)* vol. 391,10138 (2018): 2449-2462. doi:10.1016/S0140-6736(18)31320-5

Erdem, Neslihan et al. "Breaking and restoring immune tolerance to pancreatic beta-cells in type 1 diabetes." *Current opinion in endocrinology, diabetes, and obesity* vol. 28,4 (2021): 397-403. doi:10.1097/MED.0000000000000646

Haseda, F et al. "$CD4^{+}$ $CD45RA^{-}$ FoxP3high activated regulatory T cells are functionally impaired and related to residual insulin-secreting capacity in patients with type 1 diabetes." *Clinical and experimental immunology* vol. 173,2 (2013): 207-16. doi:10.1111/cei.12116

Long, S Alice et al. "Rapamycin/IL-2 combination therapy in patients with type 1 diabetes augments Tregs yet transiently impairs β-cell function." *Diabetes* vol. 61,9 (2012): 2340-8. doi:10.2337/db12-0049

Nordquist, Helen. and Radia T. Jamil. "Biochemistry, HLA Antigens." *StatPearls*, StatPearls Publishing, 24 April 2023.

Tan, Tze Guan et al. "Singular role for T-BET+CXCR3+ regulatory T cells in protection from autoimmune diabetes." *Proceedings of the National Academy of Sciences of the United States of America* vol. 113,49 (2016): 14103-14108. doi:10.1073/pnas.1616710113

Tarbell, Kristin V et al. "CD25+ CD4+ T cells, expanded with dendritic cells presenting a single autoantigenic peptide, suppress autoimmune diabetes." *The Journal of experimental medicine* vol. 199,11 (2004): 1467-77. doi:10.1084/jem.20040180

Trotta, Eleonora, et al. "A Human Anti-IL-2 Antibody That Potentiates Regulatory T Cells by a Structure-based Mechanism." *Nature Medicine*, vol. 24, no. 7, 25 June 2018, https://doi.org/10.1038/s41591-018-0070-2. Accessed 30 Sept. 2023.

Visperas, Anabelle, and Dario A A Vignali. "Are Regulatory T Cells Defective in Type 1 Diabetes and Can We Fix Them?." *Journal of immunology (Baltimore, Md. : 1950)* vol. 197,10 (2016): 3762-3770. doi:10.4049/jimmunol.1601118

Ward, Natasha C et al. "Persistent IL-2 Receptor Signaling by IL-2/CD25 Fusion Protein Controls Diabetes in NOD Mice by Multiple Mechanisms." *Diabetes* vol. 69,11 (2020): 2400-2413. doi:10.2337/db20-0186

Identifying Invasive Plant Species with Convolutional Neural Networks By Kwanhee Lee

Abstract

Invasive plants are species that do not originate in a certain area and dominate the particular ecosystem by directly and indirectly killing native plants and animals. Native plants are the keystone species that feed and shelter almost all the living organisms in its ecosystem. With the introduction of invasive plants, however, these keystone species are declining in number. Given this, it is important to advance our monitoring of known invasive species in ecosystems to formulate more directed interventions to support native plants and animals. An emerging way to monitor invasive species is through analysis of images using convolution neural networks, which help identify parts of an image and determine whether an image that someone is trying to identify shares features with invasive species images. The dataset used in this paper has images of invasive Hydrangeas in Brazilian National Forest that originated from Asia. The algorithm gave an accuracy of 71.73% in determining whether or not an image contains invasive plants. This shows how machine learning can have a positive impact on finding invasive plants to assist researchers in monitoring ecosystems, ultimately to support targeted interventions to help native species.

Introduction

Some people in particular may wonder why there seems to be fewer birds, fish, frogs, and other insects in general, and why the water quality of nearby creeks and streams seem to be getting worse. One major answer is invasive plants. According to the USDA, invasive plant species have contributed to the decline of 42% of U.S. endangered and threatened species, and for 18% of U.S. endangered or threatened species, invasive plant species are the main cause of their decline [1]. Invasive plants also contribute to altering the quality and timing of runoff, erosion, sedimentation, and may affect water availability. Some invasive species such as salt cedar, consume more water than their native counterparts, which lower the water table and dry up streams [2]. So, how can people combat this growing problem? We can solve this by eradicating the invasive plants with various methods such as pulling down, chopping off, or drying them out by cutting a part of their stem [3]. But, how will this be possible if people lack accurate and comprehensive monitoring of the whereabouts and features of invasive plants? One way to solve this problem is to devise a machine learning (ML) model, specifically with identifying features [4]. This method can be used to take an image of many different plants and determine whether there is an invasive plant to be monitored in that image.

Invasive Species and Neural Network Background

Now, what are invasive plants? Invasive plants are non-native plants whose introduction causes harm to the environment. Most invasive plants are introduced by human activities. However, unlike non-native insects or animals, which are usually introduced by accident, invasive plants are introduced deliberately. Many invasive plants are used as ornaments or

ground cover [5]. Because of that, most invasive vines tend to start their journey from someone's backyard. Once these plants spread to nearby forests, they reproduce. The seeds of the plants will then move into new areas by wind, running water, or animals such as birds. Once the invasive plants become present, they will outcompete native plants for resources such as water, sunlight and available space. Additionally, because the new area is often lacking predators or deadly disease that checks the population of the invasive plants, these plants can produce ample amounts of seeds and spread at an uncontrollable rate [6]. This uncontrollable spread is the reason why invasive plants are dangerous. As invasive plants take over, native plants die from mal-nutrients. Some vines can strangle and bring down trees that are many times larger than themselves. The invasive plant that was used for this research is called Invasive Hydrangea which is a shrub that has beautiful clumps of flowers [7].



Fig 1: Image of invasive Hydrangea photo by Gvozdev Antony [8]

Machine learning works by first acquiring the training data and sample data that is used to train the model. In a neural network (a type of ML algorithm), the model passes the images through layers. The first layer of the neural network receives the image and the last layer displays the output. The middle layers determine which category the image falls into by passing the image through convolutional filters and activation functions, which consist of weights and biases. Convolutional neural networks (CNN) as shown in Fig 2 use a matrix with numbers and zeros to find patterns which ultimately classifies the image into categories. If there is more than one middle layer, the neural network is called a deep neural network [9][10].

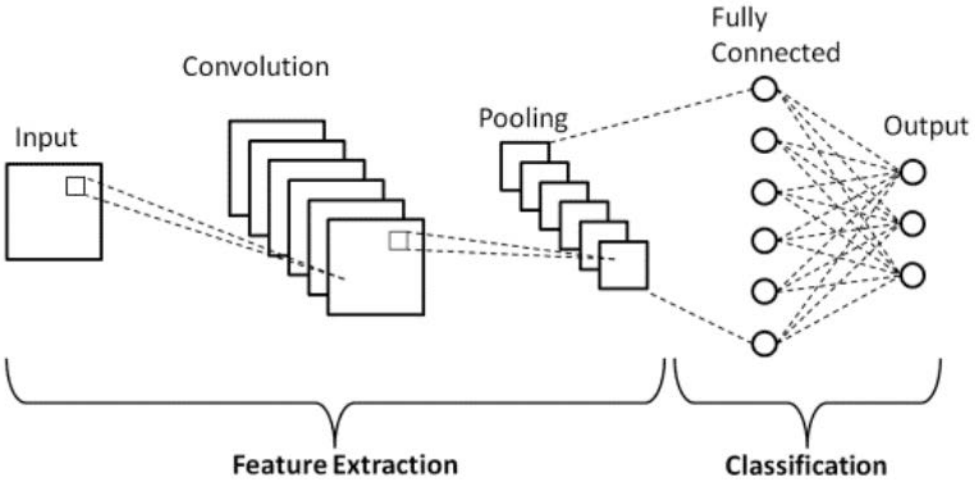


Fig 2: The Conceptual Architecture of Convolutional Neural Network [11]

Contextualizing Research

In Rockville, Maryland, there is a volunteer service program called "Weed Warrior" that works once every week to remove invasive plants in local parks. For years, the program co-leaders and the volunteer group including me have been encountering some issues. Oftentimes we would find one area consisting of far more invasive plants to remove than a different area. The work seemed like it would be more effective if there was something that could inform us that one area requires extra personnel support, so that the people involved in this work could be divided up accordingly and focus attention on the areas that need it. The ML algorithm was thought to be a useful tool in solving this problem. The advantage ML algorithms have in analyzing images to monitor areas is due to the difficulty of personally going around searching for invasive plants. The State of Georgia had similar concerns regarding the detection of invasive plants. Because of the overwhelming number of invasive Hydrangea in the state of Georgia, a competition in Kaggle was created with the goal being to identify invasive Hydrangea. The competition was based on solving one underlying problem that scientists could not possibly sample in a large quantity area, which was almost the same problem as ours. This is the basis of this research.

Experiment–1. Dataset

The data were gathered from the dataset published for the competition in Kaggle which is one of the data science community platforms [12]. It consists of real images with and without invasive Hydrangea, along with labels, either 1 or 0, indicating whether the image contains an invasive Hydrangea or not, respectively.

2. Procedure–Data Cleansing/ Preparation

With the given 2,295 images, some images were removed that could potentially harm the accuracy of the model: blurred images or images with bad lighting. This left 1,775 images;

roughly 900 images containing invasive Hydrangea and 800 without any Hydrangea plants present. The images were converted into 400 x 400 pixels and made into an array with RGB values. They were then matched with the labels on corresponding images.

A. Train Test Split

One crucial step is to split the data into training and testing sets. With this process, the machine is assigned some portion of the data for training and uses the rest of the data to test the accuracy. For this model, 750 out of 1,775 images were randomly chosen to be the test data, and the rest as training data.

B. Model Design

The final step is to get the accuracy and confusion matrix, which consists of rows and columns that represent actual class and predicted class, respectively [13]. In order to do this, the model must be trained using convolutional neural networks. By using keras API, tensorflow, and activation functions such as ReLU and sigmoid, the model is trained. Finally, the model gives accuracy based on how well it predicted the test set.

Results

The confusion matrix is as shown in Fig 3. Out of 330 test images without invasive Hydrangea, the model correctly identified 266 images as not containing Hydrangea plants in them and incorrectly classified 64 images as having Hydrangea in them. Out of 420 test images containing invasive Hydrangea, the model correctly identified 272 images as having Hydrangea in them and incorrectly classified 148 images as not having Hydrangea in them. The percentage of correctly predicted images, which is the accuracy of the model, comes out to 71.73%. The model runs 20 epochs which gets the most precise validation accuracy without overfitting the model as is shown in Fig 4.

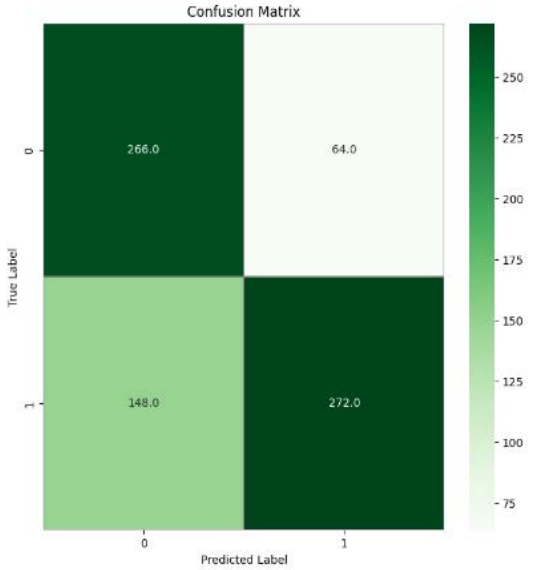


Fig 3: Evaluation of the identification of images with the confusion matrix (True label vs. Predicted label)

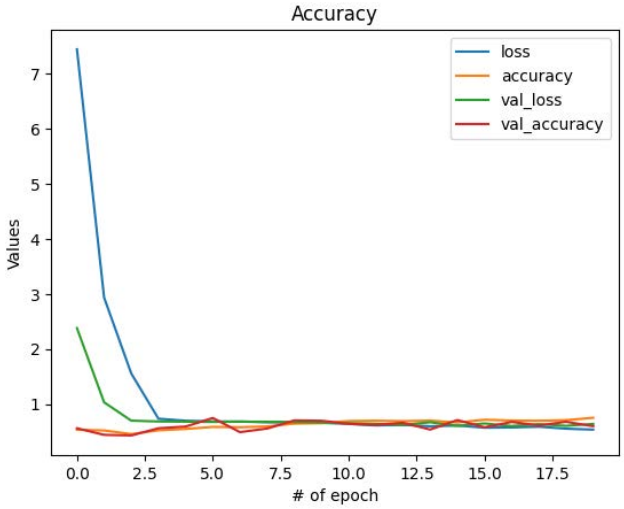


Fig 4: loss which dropped close to 0 and flattened and the accuracy showed increasing

Discussion and Conclusion

The model generated herein had a 71.73% accuracy, which is appropriate considering the data were real images taken by people. But why was the accuracy only 71.73% if the convolutional neural network can pick up patterns, and if images with blurs and bad lighting were removed previous to the training step? The answer still lies in the images. On some occasions, images labeled 1 (meaning an invasive Hydrangea is present in this image) had no distinguishing features except for leaves and branches which look very similar to an image labeled as 0 (meaning an invasive Hydrangea is not present in this image). In fact, some of the

images with only Hydrangea leaves were classified as if they contained no invasive Hydrangea. The same occurred for when the images labeled 1 had flowers visible but were too small such as having one flower in the image that looked like a dot in the background. One finding is that when gathering real images, it is helpful for the model if there are clear distinguishing features, such as flowers, or if the existing features are big and apparent. Further improvements that can be made to the model include adding extra images labeled 1 with only Hydrangea leaves so that the model can find patterns of the Hydrangea leaves that are distinguishable from other plant leaves, thus improving the models' accuracy. For the issue with images having small flowers, the image can be converted to have more pixels so that the model can detect it with this increased resolution. This model used a total of 5 layers, 1 input layer, 3 middle layer, and 1 output layer. For further improvements on the model as a whole, adding extra middle layers can bring promising results.

Acknowledgements

I acknowledge the co-leaders of the “Weed Warrior” program, Chuck Woolery and Monica Berger, for the motivation and interest to carry out this project. Also I acknowledge my mentor who assisted me when I had difficulties proceeding with this project.

Works Cited

US Forest Service. “Invasive Plants.” Wildflowers. Accessed September 05, 2023.https://www.fs.usda.gov/wildflowers/invasives/index.shtml

Heller, Laura. "Effects of Invasive Species on Water Quality in Freshwater Ecosystems." SOKA University. Accessed September 01, 2023.https://www.soka.edu/student-life/20th-anniversary-anthology/creative-coexistence-nature-humanity/effects-invasive

Mattrick, Christopher. “Managing Invasive Plants Methods of Control.” Conservation Notes of the New England Wild Flower Society 10, no.3 (2006): 20-23, Accessed September 04,. https://extension.unh.edu/sites/default/files/migrated_unmanaged_files/Resource000988_Rep1135.pdf

Molnar, Christoph. “Interpretable Machine Learning.” (2023), Accessed September 01, 2023.https://christophm.github.io/interpretable-ml-book/

The National Wildlife Federation. “Invasive Species.” Threats to Wildlife. Accessed September, 2023.https://www.nwf.org/Educational-Resources/Wildlife-Guide/Threats-to-Wildlife/Invasive-Species

Swearingen, Jil. “Plant Invasive of Mid-Atlantic Natural Areas, 5th ed.” (2014): 8-9, Accessed September 01, 2023.https://www.invasive.org/alien/pubs/midatlantic/midatlantic.pdf

US Forest Service. “Wild Hydrangea.” Wildflowers. Accessed September 05, 2023.https://www.fs.usda.gov/wildflowers/plant-of-the-week/hydrangea-arborescens.

EDD Maps. “Panicled Hydrangea.” Species Information. Accessed September 01, 2023.https://www.eddmaps.org/species/subject.cfm?sub=14077

Sarker, Md, K. ‘‘Explaining trained neural networks with semantic Web technologies: First steps.’’ Proc. CEUR Workshop, London, U.K. (2017): 1–10. Accessed September 01, 2023.https://arxiv.org/pdf/1710.04324.pdf

Sarker, Md, K.‘‘Efficient concept induction for description logics.’’ Proc. AAAI Conf. Artif. Intell., Honolulu, HI, USA (2019): 3036–3043, Accessed September 01, 2023. https://arxiv.org/pdf/1812.03243.pdf

Gurucharan, MK. “Basic CNN Architecture: Explaining 5 Layers of Convolutional Neural Network.” upGrad. Last modified July 27, 2022. Accessed September 03, 2023. https://www.upgrad.com/blog/basic-cnn-architecture/

Kaggle. “Invasive Species Monitoring.”, Competitions. Accessed June 01, 2023.https://www.kaggle.com/competitions/invasive-species-monitoring/data

Powers, David M. W. "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation." International Journal of Machine Learning Technologies 2, no.1 (2011): 37-63, Accessed September 01, 2011.https://arxiv.org/ftp/arxiv/papers/2010/2010.16061.pdf

Evolutionary Aviation Network Flow Optimization Using Topological and Demographic Features By Rahul Rao

Abstract

The rapid expansion of air transportation within the United States, particularly following the deregulation act of 1978, has triggered substantial growth in demand for air travel, a catalyst for global economic progress. This surge in demand is projected to continue its upward trajectory in the forthcoming years. However, this heightened demand has contributed to a surge in service degradation and flight delays across the National Airspace System (NAS). Addressing this, a novel research project introduces an innovative methodology for optimizing air network efficiency. Drawing from classical network theory and incorporating factors like population demographics, the project utilizes extensive big data resources. The study establishes an interconnected airport network enriched with demographic and topological data. The centrality of each airport, indicating its significance, is determined using Dijkstra's shortest path algorithm. The notion of removal importance is introduced to assess an airport's impact by contrasting centrality before and after node removal. The efficacy of this approach is evaluated through weighted and unweighted centralities, with the weighted centrality reducing RMS error by 15 percent compared to the unweighted. Visualization of results on a U.S. map highlights higher importance for popular airports, yet intriguing deviations, attributed to hub scale and distinction between removal and centrality-based importance.

Introduction

The expansion of air transportation is a driving force behind global economic growth, facilitating connections between people and businesses. As the demand for air travel continues to rise worldwide, ensuring the efficient operation of the National Airspace System (NAS) becomes increasingly vital. However, this rapid surge in aviation traffic has led to challenges such as delays and service degradation within the NAS. Addressing these issues requires the development of innovative tools for optimizing air network traffic and improving overall aviation system efficiency.

Problem Statement

The global aviation industry is witnessing an unprecedented surge in demand, projected to persist over the next few decades. The International Civil Aviation Organization (ICAO) predicts an average annual growth rate of 4.3% in air transportation demand over the next two decades (ICAO, 2023). The challenge lies in aligning the aviation infrastructure with this soaring demand. While research has been dedicated to understanding the complexities of air traffic, current projects often overlook essential air network characteristics.

Significance of Research

This study proposes an innovative approach that integrates classical network theory with demographic and topological features to address the intricacies and limitations of the contemporary

aviation network. By combining these diverse data components, this strategy aims to achieve a comprehensive understanding of airport significance and its impact on the aviation network. The research's goal is to enhance the effectiveness and efficiency of aviation network flow optimization through the application of evolutionary algorithms.

Literature Review

The rapid growth of the global aviation sector has attracted significant research attention toward mitigating the challenges posed by increased air traffic demand. For example, Pellegrini et al. (2020) endeavored to formalize the complex nature of air traffic and optimize the efficiency of the National Airspace System (NAS). Using airports as nodes and flights as edges, the study demonstrated the prevalence of classical network theory as a modeling approach for the air transportation network. However, previous research predominantly focused on route optimization and congestion control, without fully considering the broader factors influencing aviation network efficiency.

Recent studies by Zhang et al. (2019) and Al Sarrah et al. (2021) have highlighted the importance of integrating demographic and topological characteristics in air traffic modeling. Incorporating factors such as population density, geographic, socio-economic metrics, and spatial interaction allows for a more accurate representation of air travel demand and network dynamics. Demographic characteristics have been identified as crucial in understanding passenger demand trends, directly impacting airport significance and utilization (Lefevre et al., 2021).
Evolutionary algorithms have gained traction in transportation design due to their ability to address complex optimization problems. These algorithms emulate natural evolution principles to find optimal solutions in extensive solution spaces (Iyer, 2021). Moreover, big data and machine learning have opened new opportunities for aviation research. Big data analytics can analyze extensive aviation-related datasets, including past flight records, weather trends, and passenger demographics, enabling pattern recognition, demand prediction, and optimized scheduling (Miltiadou et al., 2020).

Furthermore, Nasrollahi and Abdi Kordani (2023) emphasize the importance of comprehending hub-and-spoke systems utilized by many airlines. Understanding the significance and impact of these nodes on network connectivity is vital for optimal air traffic flow. Al Sarrah et al. (2021) stress the urgent need for sustainable growth in the face of increasing air traffic demand. While past research has improved aviation network efficiency, there remain gaps, particularly in the incorporation of critical demographic information..

Methodology

This study proposes integrating demographic and topological characteristics into traditional network theory to enhance the comprehension of airport significance and influence within the network. To develop the framework, several materials were used, and various procedures were implemented.

Materials

This project utilised the following software and libraries:

- Homebrew: a macOS program manager.
- Python 3.9: a general-purpose, advanced programming dialect.
- FuzzyWuzzy: a Python module for pairing strings.
- NetworkX: a Python module for constructing, manipulating, and analysis of sophisticated networks.
- Java 19.0.2: a well-known programming language.
- Apache Commons: An Apache project concentrated on producing Java components that are reusable.
- IntelliJ: A software developer's Integrated Development Environment (IDE).
- Cytoscape: an open-source network visualization software engine.

Procedure

Following a systematic and structured methodology, Java and Python programming languages were utilized for data entry and analysis as part of the implementation of this research project. The initial phase involved collecting pertinent airport data from a publicly accessible dataset, "airport.csv," retrieved from OurAirports.com. This data set comprised essential details about airports around the globe and acted as the basis for building the air transportation system. The data was loaded and analyzed in the Java programming framework using a customized function called "loadiCities()." This function enabled the construction of a graph depicting the air network, with airports serving as nodes and flights as edges. Moreover, demographic characteristics were augmented by designating populace and city values to each airport depending on the most recent data. These demographic characteristics were deemed essential for comprehending the fundamental trends in air travel demand, thereby improving the precision of the subsequent analysis.

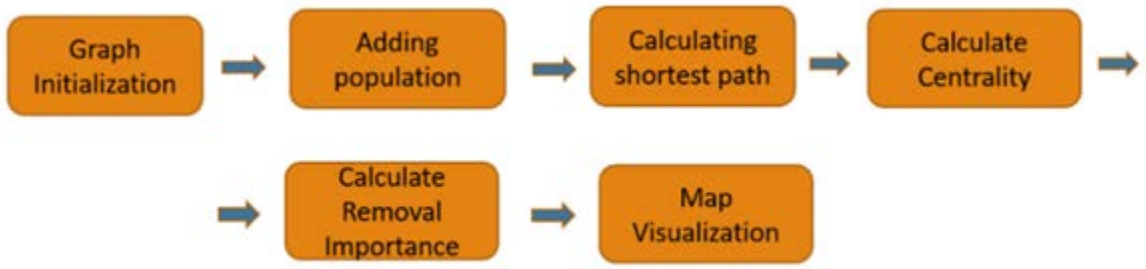


Figure 1: Implementation Strategy

Weighted centrality (WCentrality) and unweighted centrality (UWCentrality) were implemented as floating-point properties to the graph's borders and vertices to set up the data for subsequent calculations. These centrality measures are vital indicators of airport significance and impact on the aviation system (Yanjun et al., 2019). As a result, their incorporation was necessary to

attain an in-depth understanding of the network's operation. Next, the implementation transitioned to Python programming to assign each airport's shortest routes. The NetworkX library's implementation of Dijkstra's Shortest Path Algorithm was utilized for this specific reason. The algorithm computed the shortest routes between airports by receiving criteria such as origin, destination, weight, and technique. The nodes on the shortest paths were then indexed to modify the weighted centrality (WCentrality) by incorporating a "pathweight" figure and the unweighted centrality (UWCentrality) by adding 1 to each node on the shortest path.

A modified standard deviation procedure was fitted to the weighted and unweighted centralities to ascertain the removal importance of individual airports. By aggregating the square of each airport's distance from the mean, the algorithm determined the network value of each airport. Utilizing visualization software such as Cytoscape, the removal importance was displayed, providing helpful knowledge into network weaknesses and possibilities for targeted upgrades. To evaluate the precision of the centrality metrics, airport scores were computed for both weighted and unweighted centralities in relation to StratosJets.com's list of the 200 most significant airports (Stratos, 2021). Root mean square error calculations were performed to ascertain each centrality metric's precision. Lower error values accurately determined the network's most influential airports.

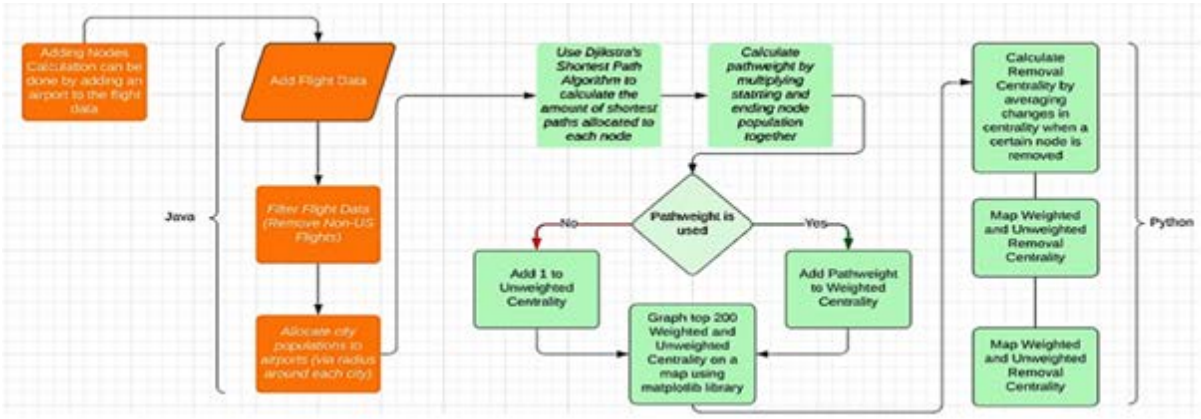


Figure 2: Implementation Flowchart

Findings

Sorting the airport network analysis by weighted centrality and unweighted centrality generated two separate city listings. The root-mean-square error (RMS) for both lists was calculated to assess the efficacy and precision of each centrality metric.

Altitude	City	connections	coord	Country	labelV	Latitude	Longitude	name	PIATA	PID	pop	selected	shared name	Type	UWCentralit	WCentrality
152	Anchorage	0	[-149.99600	United State	airport	61.1744003	-149.996	35740	ANC	3774	376602	FALSE	35740	airport	98441	1.60E+15
841	Minneapolis	0	[-93.221802,	United State	airport	44.882	-93.221802	36580	MSP	3858	5939236	FALSE	36580	airport	59861	3.44E+16
5431	Denver	0	[-104.67299	United State	airport	39.8616982	-104.673	35510	DEN	3751	5980542	FALSE	35510	airport	52679	9.01E+16
672	Chicago	0	[-87.9048,41	United State	airport	41.9786	-87.9048	36300	ORD	3830	1.53E+07	FALSE	36300	airport	37781	2.30E+17
433	Seattle	0	[-122.30899	United State	airport	47.449001	-122.309	33770	SEA	3577	7511791	FALSE	33770	airport	31059	4.01E+16
607	Dallas-Fort V	0	[-97.038002,	United State	airport	32.896801	-97.038002	34700	DFW	3670	1.18E+07	FALSE	34700	airport	28017	8.31E+16
1026	Atlanta	0	[-84.428101,	United State	airport	33.6367	-84.428101	34820	ATL	3682	7979477	FALSE	34820	airport	27237	1.06E+17
439	Fairbanks	0	[-147.85600	United State	airport	64.8151016	-147.856	36320	FAI	3832	113976	FALSE	36320	airport	24325	3.26E+14
126	Bethel	0	[-161.83799	United State	airport	60.7798004	-161.838	33990	BET	3599	6586	FALSE	33990	airport	20267	1.13E+14
645	Detroit	0	[-83.353401	United State	airport	42.2123985	-83.353401	34450	DTW	3645	7753419	FALSE	34450	airport	19865	1.79E+17
4227	Salt Lake Cit	0	[-111.97799	United State	airport	40.7883987	-111.978	33360	SLC	3536	4305134	FALSE	33360	airport	15645	3.00E+16
748	Charlotte	0	[-80.943099	United State	airport	35.2140007	-80.9431	36760	CLT	3876	3364685	FALSE	36760	airport	15211	9.38E+16
14	Kotzebue	0	[-162.59899	United State	airport	66.884697	-162.599	34930	OTZ	3693	0	FALSE	34930	airport	12817	0
312	Washington	0	[-77.455802	United State	airport	38.9445	-77.455803	35140	IAD	3714	1.41E+07	FALSE	35140	airport	11579	1.33E+17
20	Boston	0	[-71.005203,	United State	airport	42.3642998	-71.005203	32490	BOS	3448	1.37E+07	FALSE	32490	airport	11499	1.36E+17
125	Los Angeles	0	[-118.40799	United State	airport	33.9425011	-118.408	32850	LAX	3484	2.72E+07	FALSE	32850	airport	10841	1.57E+17
618	St. Louis	0	[-90.370003,	United State	airport	38.748697	-90.370003	34780	STL	3678	3705168	FALSE	34780	airport	10691	2.83E+16
21	Juneau	0	[-134.57600	United State	airport	58.3549995	-134.576	32930	JNU	3492	31974	FALSE	32930	airport	10477	9.24E+14
88	Aniak	0	[-159.54299	United State	airport	61.5816002	-159.543	46290	ANI	5967	0	FALSE	46290	airport	9641	0

Figure 3: Unweighted Removal Centrality Identifiers

Altitude	City	connections	coord	Country	labelV	Latitude	Longitude	name	PIATA	PID	pop	selected	shared name	Type	UWCentralit	WCentrality
36	Philadelphia	0	[-75.241096	United State	airport	39.8718987	-75.241096	35520	PHL	3752	1.51E+07	FALSE	35520	airport	9501	4.31E+17
672	Chicago	0	[-87.9048,41	United State	airport	41.9786	-87.9048	36300	ORD	3830	1.53E+07	FALSE	36300	airport	37781	2.30E+17
2181	Las Vegas	0	[-115.15200	United State	airport	36.080101	-115.152	36770	LAS	3877	3643294	FALSE	36770	airport	9049	2.03E+17
645	Detroit	0	[-83.353401	United State	airport	42.2123985	-83.353401	34450	DTW	3645	7753419	FALSE	34450	airport	19865	1.79E+17
125	Los Angeles	0	[-118.40799	United State	airport	33.9425011	-118.408	32850	LAX	3484	2.72E+07	FALSE	32850	airport	10841	1.57E+17
18	Newark	0	[-74.168701	United State	airport	40.6925011	-74.168701	32950	EWR	3494	4.36E+07	FALSE	32950	airport	1561	1.38E+17
20	Boston	0	[-71.005203	United State	airport	42.3642998	-71.005203	32490	BOS	3448	1.37E+07	FALSE	32490	airport	11499	1.36E+17
312	Washington	0	[-77.455802	United State	airport	38.9445	-77.455803	35140	IAD	3714	1.41E+07	FALSE	35140	airport	11579	1.33E+17
21	New York	0	[-73.872596	United State	airport	40.7771988	-73.872597	34970	LGA	3697	4.44E+07	FALSE	34970	airport	1619	1.33E+17
13	New York	0	[-73.778900	United State	airport	40.639801	-73.7789	35970	JFK	3797	4.35E+07	FALSE	35970	airport	1145	1.23E+17
896	Cincinnati	0	[-84.667800	United State	airport	39.0488014	-84.667801	32890	CVG	3488	3937511	FALSE	32890	airport	2155	1.22E+17
1026	Atlanta	0	[-84.428101,	United State	airport	33.6367	-84.428101	34820	ATL	3682	7979477	FALSE	34820	airport	27237	1.06E+17
620	Chicago	0	[-87.752403,	United State	airport	41.785999	-87.752403	35470	MDW	3747	1.50E+07	FALSE	35470	airport	4905	9.74E+16
748	Charlotte	0	[-80.943099	United State	airport	35.2140007	-80.9431	36760	CLT	3876	3364685	FALSE	36760	airport	15211	9.38E+16
5431	Denver	0	[-104.67299	United State	airport	39.8616982	-104.673	35510	DEN	3751	5980542	FALSE	35510	airport	52679	9.01E+16
607	Dallas-Fort V	0	[-97.038002,	United State	airport	32.896801	-97.038002	34700	DFW	3670	1.18E+07	FALSE	34700	airport	28017	8.31E+16
944	Ontario	0	[-117.60099	United State	airport	34.0559998	-117.601	35340	ONT	3734	2.91E+07	FALSE	35340	airport	1081	8.25E+16
791	Cleveland	0	[-81.849800	United State	airport	41.4117012	-81.8498	32870	CLE	3486	4831583	FALSE	32870	airport	8049	8.11E+16
778	Burbank	0	[-118.35900	United State	airport	34.2006989	-118.359	34440	BUR	3644	2.79E+07	FALSE	34440	airport	1395	8.05E+16

Figure 4: Weighted Removal Centrality Identifiers

RMS_Weighted	***0.1388***
RMS_Unweighted	***0.1633***

Figure 5: RMS Results

The results showed that the RMS for the city list ordered by weighted centrality was significantly lower than that for unweighted centrality. The lesser RMS value for the weighted index demonstrates a lower error rate, indicating greater precision in measuring airport centrality.

Unweighted Removal Centrality

Examining the unweighted removal centrality map reveals an inadequate and inaccurate representation of airport significance within the air transportation network.



Figure 6: Unweighted Removal Centrality Results

In this depiction, New York's airports, which have been viewed as critical nodes in the aviation system, seem less significant than anticipated. The map's markings range from dark blue to dark red, with dark blue representing the least important airports and dark red representing the most essential. Nonetheless, the unweighted removal centrality mapping method is insufficient for precisely assessing airport centrality and recognizing crucial regions requiring consideration in airport planning and layout. As demonstrated by underestimating the value of.

Weighted Removal Centrality

The weighted removal centrality mapping analysis depicts airport significance within the air transportation system plainly and accurately. Contrary to the underestimation observed in the unweighted removal centrality mapping, New York's airports, recognized as vital locations, appear essential in this representation. The weighted removal centrality map's labels range from dark blue to dark red, with dark red signifying the most significant airports and dark blue symbolizing the minor critical airports. This mapping method is an informative and valuable instrument for evaluating airport centrality, allowing transportation designers and researchers.



Figure 7: Weighted Removal Centrality Results

Weighted/Unweighted Airport Addition

Due to the inclusion of a Rhode Island airport, the markings in the weighted/unweighted airport inclusion images are different colors, indicating a hypothetical simulation of multiple links with other airports.

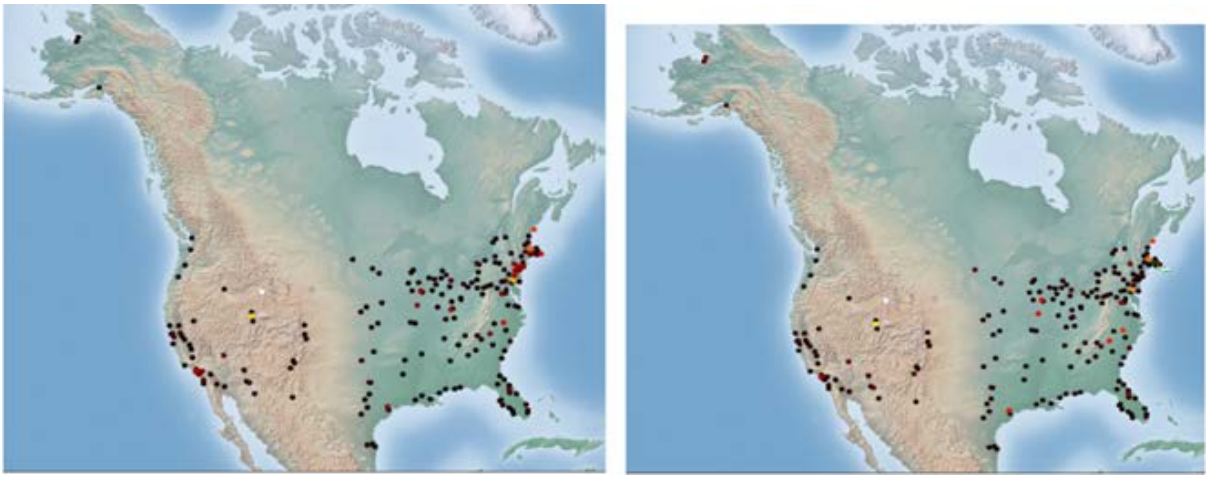


Figure 8: Weighted/Unweighted Airport Addition Results

This simulation offers a real-world use case to demonstrate the tool's effectiveness for assessing airport centrality. The lighter shadings across the complete network indicate an improvement in traffic flow. With the inclusion of a single airport, the impact of removing other airports is significantly mitigated, demonstrating the tool's considerable value to designers and strategists. In addition, the tool's potential stretches beyond airline transit, as it efficiently allows an assessment of the effects of adding nodes to any transportation system, showcasing its adaptability for a variety of transportation planning purposes. In use cases, airline planners are cautioned against counterintuitive outcomes, preventing issues relating to cost and time associated with the construction of suboptimal airports.

Discussion

The results of applying the proposed methodology to evaluating airport centrality and its implications for optimizing the air transportation network provide valuable insights. This section will elucidate the most important findings and their significance, the impact on transportation planning, and the broader implications of the research. The results demonstrated that weighted centrality sorting is superior to unweighted centrality sorting for precisely measuring airport importance within a network. The lower root-mean-square error (RMS) value obtained from the weighted list indicates greater accuracy and dependability. weighted centrality considers the actual path weights and distances between airports.

Study Implications

The more significant implications of this study stem from the prospective applications of the methodology outside of the aviation industry. Using evolutionary algorithms, incorporating

topological and demographic characteristics offers a comprehensive strategy adaptable to other transportation systems such as railways, roads, and maritime routes (Nasrollahi & Abdi Kordani, 2023). A precise evaluation of network weaknesses and optimization opportunities is essential for sustainable transportation planning and resource allocation amid rising demand and changing infrastructure requirements.

Conclusion

This study presented an extensive approach incorporating demographic and topological characteristics into classical network theory, providing a more precise measure of airport rankings than conventional unweighted centrality metrics. The implementation of weighted centrality yielded valuable insights into airport significance and its effect on the air transportation network. This method has the potential to significantly aid transport planners and researchers in maximising air traffic network flows, especially in catastrophic events such as airport closures triggered by weather.

Works Cited

Al Sarrah, M., Ajmal, M. M., & Mertzanis, C. (2021). Identification of sustainability indicators in the civil aviation sector in Dubai: a stakeholders' perspective. *Social Responsibility Journal*, *17*(5), 648-668. https://doi.org/10.1108/SRJ-06-2019-0203

ICAO. (2023). Future of Aviation. https://www.icao.int/Meetings/FutureOfAviation/Pages/default.aspx#:~:text=The%20most%20recent%20estimates%20suggest,GDP%20to%20the%20world%20economy.

Iyer, L. S. (2021). AI enabled applications towards intelligent transportation. *Transportation Engineering*, *5*, 100083. https://doi.org/10.1016/j.treng.2021.100083

Lefevre, J., Briand, Y., Pye, S., Tovilla, J., Li, F., Oshiro, K., ... & Zhang, R. (2021). A pathway design framework for sectoral deep decarbonization: the case of passenger transportation. *Climate Policy*, *21*(1), 93-106. https://doi.org/10.1080/14693062.2020.1804817

Miltiadou, D., Pitsios, S., Spyropoulos, D., Alexandrou, D., Lampathaki, F., Messina, D., & Perakis, K. (2020, December). A big data intelligence marketplace and secure analytics experimentation platform for the aviation industry. In *International Conference on Big Data Technologies and Applications* (pp. 48-62). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-72802-1_4

Nasrollahi, M., & Abdi Kordani, A. (2023). Designing Airline Hub-and-Spoke Network and Fleet Size by a Biobjective Model Based on Passenger Preferences and Value of Time. *Journal of Advanced Transportation*, *2023*. https://doi.org/10.1155/2023/2797613

Pellegrini, A., Di Sanzo, P., Bevilacqua, B., Duca, G., Pascarella, D., Palumbo, R., ... & Gigante, G. (2020). Simulation-based evolutionary optimization of air traffic management. *IEEE access*, *8*, 161551-161570. https://doi.org/10.1109/ACCESS.2020.3021192

Rademaker, M. E. (2020). *Composite-based structural equation modeling* (Doctoral dissertation, Universität Würzburg). https://opus.bibliothek.uni-wuerzburg.de/files/21593/Rademaker_Manuel_Dissertation.pdf

Stratos (2021). 200 Busiest Airports in the US. https://www.stratosjets.com/blog/busiest-us-airports/

Yanjun, W. A. N. G., Jianming, Z. H. A. N., Xinhua, X. U., Lishuai, L. I., Ping, C. H. E. N., & Hansen, M. (2019). Measuring the resilience of an airport network. *Chinese Journal of Aeronautics*, *32*(12), 2694-2705. https://doi.org/10.1016/j.cja.2019.08.023

Zhang, L., Du, H., Zhao, Y., De Maeyer, P., Dessein, B., & Zhang, X. (2019). Drawing topological properties from a multi-layered network: The case of an air transport network in "the Belt and Road" region. *Habitat International*, *93*, 102044. https://doi.org/10.1016/j.habitatint.2019.102

Type 1 Diabetes in Children: Affect on Cognition and Biotechnological Solutions By Shiksha Anand

Abstract

Type 1 diabetes in children is becoming increasingly prominent. Glucose is a crucial source of energy for the brain which enables the building of cerebral structures in a child's developing brain. As a result, the glycemic extremes of diabetes adversely impact children's brain structure, including decreasing gray matter and white matter volumes which are still under development in adolescence. Furthermore, earlier onset and frequent occurrences of hypoglycemia in children can increase the child's vulnerability to cognitive impairment, including psychomotor function, attention, and memory. However, there is a compensatory function in the brain's default mode network that attempts to prevent further damage in certain brain regions. As a result of these effects of type 1 diabetes on children, new biotechnological solutions are being researched to find safe and efficient ways to limit glucose fluctuations or prevent a longer duration of such fluctuations, consequently protecting children's cognitive function. This literature review focuses on cognitive impairment in children with type 1 diabetes and analyzes six major components, including glucose in the brain, fluctuations in glucose, and diabetes' impact on psychomotor function, attention, memory, and biotechnological solutions. In order to evaluate various previous papers, multiple databases were searched under the context of type 1 diabetes in children and cognitive impairments of this condition. The papers cited in this review consist of original research articles and literature reviews from PubMed, National Institutes of Health (NIH), and UCI databases.

Keywords: type 1 diabetes, hypoglycemia, hyperglycemia, cognitive impairments, biotechnological solutions, psychomotor function, attention, memory

Introduction

Type 1 diabetes is an autoimmune reaction in which the pancreas does not produce insulin, resulting in dysglycemia. Glucose is an essential resource required for brain maturation in children. Thus, extreme fluctuations in glucose levels could lead to impairment in children's cognitive function due to their brain vulnerability during the developmental stages. As a result, brain metabolism is critical in maintaining stability as glucose fuels the brain's neurons and cells. The prevalence of type 1 diabetes in children aged between 0 and 19 years has risen from 2001 to 2009 (Chiang et al., 2018). In particular, the SEARCH for Diabetes in Youth study recorded a 21.1% rise in diabetes in youth; therefore it is vital to conduct research on the nature of type 1 diabetes in children due to its increasing prominence. Although severe hypoglycemia and hyperglycemia can lead to persistent cognitive disorders in type 1 diabetic pediatric patients, there is a dearth of evidence of type 1 diabetes in children and brain function. Additionally, diabetes has a debilitating impact on children and there is a need for on-hand immediate devices to prevent such fluctuations. Therefore, this paper aims to provide an overview of these topics.

Specifically, it will delve into the role of glucose in brain functions and the impact of type 1 diabetes on psychomotor function, attention, and memory as a result of abnormal blood sugar levels with a focus on the current areas of development in the field of biotechnology for solutions. Ultimately, this review aims to emphasize the correlation between type 1 diabetes and brain activity in children while addressing the new biotechnology solutions used for treatment.

Methods

The PubMed, National Institutes of Health (NIH), and UCI databases were searched using the following combinations of keywords: type 1 diabetes in children, hypoglycemia, hyperglycemia, glucose in the developing brain, and cerebral function. To be included in the following literature review, papers focused on the role of glucose on brain functions, the impact of diabetes on young children with a focus on cognition as well as upcoming technologies to prevent cognitive impairment from diabetes. Additionally, the cited paper includes context to understand the various parts of the brain. The papers included in this study were published from the year 2005 to 2023; thus, including foundational knowledge on diabetes as well as updated research. Furthermore, the published papers also include literature reviews on type 1 diabetes. Excluded papers did not include diabetes's applications to children or provide complete access to the paper.

Glucose in the brain

Glucose

Glucose is a vital source of energy for the brain, thus impacting ATP production, oxidative stress management, and synthesis of neurotransmitters, neuromodulators, and structural components (Dienel, 2019). As a result, the intake and breaking down of glucose can affect cell signaling and brain states such as sleep or arousal (Nadim & Bucher, 2015). A study that tested the cognitive status of diabetic subjects by inducing hypoglycemia resulted in increased lethargy, stupor, and coma with extended periods of hypoglycemia (Dienel, 2019). Thus, glucose plays an important role in energizing the brain functions. Glucose is transported into cells and nervous tissues through the glucose transporters known as GLUT (Cacciatore et al., 2022). These transporters can be affected by the external factors of hyperglycemia, hypoglycemia, and an increased demand for glucose from the brain, therefore altering the expression of the GLUT transporters. Secondly, as the transportation of glucose in neurons is not affected by insulin regulation, neurons can increase their concentration of glucose based on the available concentration of glucose in their surrounding, thus leaving the brain more susceptible to damage. Essentially, if neurons have higher concentrations of glucose available in their extracellular environment, they can uptake more glucose which leaves the brain at risk for damage. Additionally, a child's brain requires more than 20-25% of the body's total glucose levels; thus, it can be safe to infer that a developing brain may be at higher risk of damage due to the neuron's independence from insulin and a higher presence of glucose in the brain. A child's demand for brain glucose is clear; however, due to the highly invasive methods of determining glucose levels

and glucose consumption in the brain, there is a lack of data on children's glucose activity in the brain. While children certainly have an increase in glucose consumption than adults, children at different age levels have varying usages of glucose in distinct regions of the brain. Lastly, as glucose levels rise, the blood-brain barrier becomes increasingly permeable, allowing the entry of substances that are harmful to not only the central nervous system but also the cellular structure involving the rearrangement of membranes. This phenomenon could even lead to the diagnosis of type 1 diabetes in individuals. Therefore, stable glucose levels are critical to maintaining adequate brain functions in children.

Aerobic glycolysis

Glycolysis, which is the breakdown of glucose for ATP, also occurs under sufficient oxygen conditions; this phenomenon is known as aerobic glycolysis (Cacciatore et al., 2022). Generally, oxidative metabolism pathways are preferred by the body to produce ATP unless there is a disruption in oxygen supply. However, occurrences of aerobic glycolysis indicate that glycolysis is not limited to low-oxygen levels. Aerobic glycolysis is also preferred as the primary pathway in multiple fast-growing single-cellular organisms and is seen in cancer cells, which have high growth rates. Thus, aerobic glycolysis is prevalent in rapidly growing organisms. Similarly, children's rapidly developing brains favor aerobic glycolysis, which explains the increased use of oxygen in children's brains. Another characteristic of aerobic glycolysis is that it increases glucose consumption to carry out the required reactions for the growth of cells; thus, children have an increased glucose consumption in order to build their cerebral structures. However, high aerobic glycolysis could negatively affect neural pathways and be associated with a higher risk of pathological hallmarks (Chen et al., 2021). While aerobic glycolysis carries out positive functions such as biosynthesis and rapid production of ATP, it also increases vulnerability to amyloid-β deposition which is a key feature in neurodegenerative diseases. Therefore, the greater prevalence of aerobic glycolysis in children could harm their neurological functions in the long run. However, there is an optimized wiring in high aerobic glycolysis regions that reduces this vulnerability in those brain regions. Further research must be conducted to understand the relationship between aerobic glycolysis and the brain's structural connections which is known as the connectome.

Fluctuations in glucose

Since children's brains are still developing, fluctuations in glucose levels including severe hypoglycemia and hyperglycemia in type 1 diabetic children can significantly alter their brain structure and neurocognitive function (Nevo-Shenker & Shalitin, 2021). Furthermore, longer durations of severe hypoglycemia can increase the potential for brain injury (Goyal & Raichle, 2019). In a developing brain, there is an increase in the total gray matter (GM) and white matter (WM) volumes as well as an increase in brain glucose consumption (Nevo-Shenker & Shalitin, 2021). Gray matter consists of the outermost layer of the brain (Mercandante & Tadi, 2023). This area has a high concentration of neurons present, which are primarily responsible for processing

and releasing information through the white matter's axon signaling. An increased density in gray matter enables high processing and increased mental development for an individual. As a result, damage in gray matter or a decrease in gray matter volume can result in cognitive impairment. Additionally, HbA1c levels and white matter express a strong correlation, signifying that higher blood glucose levels lead to lower cognitive function in children (Nevo-Shenker & Shalitin, 2021). Furthermore, there is a negative correlation between the gray and white matter volumes and glycemic levels in children (Cacciatore et al., 2022). Results indicated that consistent and severe hypoglycemic events lowered the gray matter and white matter volumes in the brain, thus impairing the overall cognition in children (Nevo-Shenker & Shalitin, 2021). However, as the brain matter volumes must be recorded during a severe hypoglycemic event, there is a lack of data clearly indicating a decrease in white matter volume from glycemic extremes. Similar to the impact of hypoglycemic events, hyperglycemia studies depicted an adverse effect on type 1 diabetic children's visual-spatial ability and memory tasks. Moreover, an earlier onset of type 1 diabetes in children led to poorer overall intellectual and informational processing as compared to children with a later onset of type 1 diabetes, indicating a relationship between cognitive function and glycemic extremes. This relationship between the earlier onset of type 1 diabetes and a more drastic change in cerebral function is also supported by Cacciatore et al. (2022). As a result, if a child had severe glycemic events before the age of five, then the child would be more likely to have lower memory and intelligence levels than of a child who did not experience such glycemic events at that age (Seaquist, 2017). Additional studies and research are required to determine the prolonged effects of hypoglycemia and hyperglycemia on cognition in children (Nevo-Shenker & Shalitin, 2021).

Diabetes' impact on psychomotor function

The psychomotor function involves cognitive tasks such as attention and problem-solving as well as physical movement such as motor responses. The white matter in the brain involves the cognitive domains of information processing, psychomotor function, visuoconstructive ability, and mental flexibility (Seaquist, 2017). As a result, a decrease in white matter function would most likely affect those cognitive functions. 25 adults were tasked to perform the ReyOsterreith Complex Figure Drawing test, a test that measures the visuoconstructive abilities of an individual. The results indicated a lower performance in the adults with diabetes while noting correlations with lower white matter microstructural integrity. Additionally, lower levels of white matter microstructural integrity indicated a decline in the cognitive domains mentioned above. While this study was performed on adults, it could be inferred that children with type 1 diabetes would likely have similar results from a decrease in white matter volumes. Furthermore, since children with type 1 diabetes have lower white matter volumes, it is possible that children would have more of a drastic effect on their cognition as compared to the adult participants in the previously mentioned study. In another study by the Epidemiology of Diabetes Intervention Complication (EDIC), participants with type 1 diabetes at a mean age of 27 performed several cognitive tests, and their HbA1c values were recorded 18 years later (Moheet et al., 2015). The

results indicated that higher HbA1c values (greater than 8.8%) meant a moderate decline in psychomotor abilities and motor speed, while lower HbA1c values (less than 7.4%) showed better psychomotor abilities. Therefore, since higher HbA1c values indicated high blood glucose in diabetic individuals, type 1 diabetes did impact adversely on psychomotor function. Adults with type 1 diabetes were more likely to have a higher decline in psychomotor function as they age as compared to adults without diabetes. Nevertheless, it was worth noting that there was a lack of data on the impact of type 1 diabetes on the psychomotor function of children as many studies combined analyzing psychomotor function with an aging population with or without type 1 diabetes.

Diabetes' Impact on Attention

Children with type 1 diabetes tended to have lower performance on sustained attention tasks than their diabetic peers; thus, attention was affected by type 1 diabetes (Nevo-Shenker & Shalitin, 2021). The default mode network (DMN) is linked to memory and executive functions and includes the anterior cingulate cortex and the posterior cingulate cortex which are both related to attention processes (Moheet et al., 2015). A DirecNet study involving type 1 diabetic participants and nondiabetic participants (control group) highlighted that there was greater activation in the executive control regions in the diabetic peers while they performed attention-related tasks (Nevo-Shenker & Shalitin, 2021). The executive control region was responsible for tasks involving attention and memory. The amount of activation in the executive control regions showed a correlation with deactivations in the posterior cingulate cortex. Thus, it could be analyzed that the higher activation of certain regions of the brain served as a compensatory function to counteract the impairments in the default mode network of type 1 diabetic brain. As a result, the DirecNet study concluded that the type 1 diabetic participants had equivalent cognitive performance as compared to the non-diabetic control group. Another study that used children's MRI confirmed the increased brain activation in certain control regions and concluded that executive control regions' increased activation could counteract the cognitive impairments produced by type 1 diabetes (Cacciatore et al., 2022). Therefore, this counteraction would allow for normal cognitive function in type 1 diabetic children.

Diabetes' Impact on Memory Function

The severity and onset of hypoglycemia were found to have an impact on long-term spatial memory performance (Hershey et al., 2005). A study that included participants aged from 6 years to 18 years, showed a correlation between hypoglycemia frequency and timing in type 1 diabetic children. In particular, if the first episode of hypoglycemia occurred before the age of 5 years with 3 or more severe episodes of hypoglycemia continuing to occur, then the child was more prone to worse memory function. Another study on children conducted by DirecNet analyzes memory function in relation to type 1 diabetes (Nevo-Shenker & Shalitin, 2021). The results produced that children with type 1 diabetes performed worse on the visuospatial working memory tasks while children with an earlier onset of type 1 diabetes displayed a more

pronounced effect on their memory skills. Secondly, similar to the activation during attention-related tasks, there is a higher activation in certain brain regions while performing memory-oriented tasks. However, since children with type 1 diabetes performed worse on memory tasks as compared to the control group, the DirecNet study concluded that the compensatory function was not effective in offsetting the memory-related impairments. As a result, type 1 diabetes and the severity of hypoglycemic events did impact memory function in children.

Biotechnological Solutions

Research on new technologies that prevent cognitive damage in children with type 1 diabetes has been underway. An earlier onset of type 1 diabetes in children suggested an increase in the chances of cognitive dysfunction that would affect the academic involvement of a child (Moheet et al., 2015). Thus, the development of new technologies has been necessary to protect a child's brain and cognition. An additional problem is that children are unable to express glycemic extremes until it may be too late, causing cognitive impairment (Nevo-Shenker & Shalitin, 2021). As a result, glycemic detection technology would also prove helpful in such situations. Technologies including, sensor-augmented pumps (SAP), continuous glucose monitoring (CGM), and continuous subcutaneous insulin delivery are useful solutions to improve glycemic outcomes. The sensor-augmented pumps include a continuous glucose monitor (CGM) and limit the severity as well as the duration of hypoglycemia (Steineck et al., 2017). As a result, if the severity and duration of a hypoglycemia event are restricted early on, then it can be assumed that the brain would experience less damage. Thus, SAP is useful in preventing cognitive damage by limiting severe hypoglycemic episodes. The SAP works by releasing insulin to the body upon a signal of low glucose levels from the CGM sensor to the insulin pump, allowing for automated suspension of insulin delivery (Steineck et al., 2017). CGM systems available include Dexcom G6, Senseonics Eversense, and Medtronic which gather real-time data such as the rate of change in glucose and hypoglycemia and hyperglycemia levels (Sherwood et al., 2020). This data is used to observe and analyze glycemia patterns in the patient. Additionally, the CGM's data can be shared with family members to promptly alert them of an alarming glucose level. Therefore, the CGM is beneficial for children as their parents can be notified when the child's glucose level requires attention, consequently limiting the severity of hypoglycemia and hyperglycemia. Next, insulin pumps, which are also referred to as continuous subcutaneous insulin delivery, track the insulin usage in the body and accordingly adjust the insulin doses. Insulin delivery can be automated to deliver insulin when the blood glucose levels drop below a certain threshold, therefore decreasing the amount of time a diabetic individual is in hypoglycemia. By limiting the duration of hypoglycemia, a type 1 diabetic child's neurocognitive damage is reduced. Thus, insulin pumps maintain and regulate glucose levels which reduce the frequency of severe hypoglycemia (Nevo-Shenker & Shalitin, 2021). Overall, these new technologies efficiently and safely monitor glucose levels and optimize glycemic

control in type 1 diabetic children; thus, lessening damage to a child's developing brain and its cognitive functions (Nevo-Shenker & Shalitin, 2021).

Conclusion

This literature review focused on the impact of type 1 diabetes on children's cognitive function. In particular, the review provided an overview on the (1) function of glucose, (2) fluctuations in glucose, (3) T1D's impact on psychomotor function, (4) attention, and (5) memory. Ultimately, it could be concluded that as a child's brain develops, it is more susceptible to cognitive impairments due to extreme and multiple glycemic events. Additionally, this vulnerability of the brain is due to the changes in a child's gray and white matter development. An earlier onset of such abnormal glucose levels indicated an overall lower cognitive function in children, including memory function. Type 1 diabetes negatively impacted psychomotor function in diabetic patients with higher HbA1c values. It is worth noting that this decline in psychomotor function has not been explicitly tested for children. Research showed a negative correlation between the severity and onset of hypoglycemia and performance on memory tasks. As a result, children with an earlier onset of hypoglycemia with multiple occurrences tended to perform worse on visuospatial memory tasks. Similarly, there is a negative effect of diabetes on children's attention function; however, research indicated that there is a compensatory function that prevents further impairments in the default neural network of type 1 diabetic individuals. These findings highlighted the importance of developing prevention and protection devices for type 1 diabetic children as diabetes can negatively impact children's academic performance. Sensor-augmented pumps, continuous glucose monitors, and insulin pumps are upcoming technologies to prevent cognitive damage following longer durations of glycemic extremes. These technologies are efficient in automatically regulating blood glucose levels in children who are often unaware of their glucose levels. Nevertheless, there is room for further research on the long-term benefits of these devices as well as other methods to prevent the debilitating consequences of diabetes.

Works Cited

Cacciatore, M., Grasso, E. A., Tripodi, R., & Chiarelli, F. (2022). Impact of glucose metabolism on the developing brain. Frontiers in endocrinology, 13, 1047545. https://doi.org/10.3389/fendo.2022.1047545

Chen, Y., Lin, Q., Liao, X., Zhou, C., & He, Y. (2021). Association of aerobic glycolysis with the structural connectome reveals a benefit-risk balancing mechanism in the human brain. Proceedings of the National Academy of Sciences of the United States of America, 118(1), e2013232118. https://doi.org/10.1073/pnas.2013232118

Chiang, J. L., Maahs, D. M., Garvey, K. C., Hood, K. K., Laffel, L. M., Weinzimer, S. A., Wolfsdorf, J. I., & Schatz, D. (2018). Type 1 Diabetes in Children and Adolescents: A Position Statement by the American Diabetes Association. Diabetes Care, 41(9), 2026–2044. https://doi.org/10.2337/dci18-0023

Dienel G. A. (2019). Brain Glucose Metabolism: Integration of Energetics with Function. Physiological reviews, 99(1), 949–1045. https://doi.org/10.1152/physrev.00062.2017

Goyal, M. S., & Raichle, M. E. (2018). Glucose Requirements of the Developing Human Brain. Journal of pediatric gastroenterology and nutrition, 66 Suppl 3(Suppl 3), S46–S49. https://doi.org/10.1097/MPG.0000000000001875

Hershey, Perantie, D. C., Warren, S. L., Zimmerman, E. C., Sadler, M., & White, N. H. (2005). Frequency and Timing of Severe Hypoglycemia Affects Spatial Memory in Children With Type 1 Diabetes. Diabetes Care, 28(10), 2372–2377.

Nadim, F., & Bucher, D. (2014). Neuromodulation of neurons and synapses. Current opinion in neurobiology, 29, 48–56. https://doi.org/10.1016/j.conb.2014.05.003

Nevo-Shenker, M., & Shalitin, S. (2021). The Impact of Hypo- and Hyperglycemia on Cognition and Brain Development in Young Children with Type 1 Diabetes. Hormone research in paediatrics, 94(3-4), 115–123. https://doi.org/10.1159/000517352

Mercadante, A. A., & Tadi, P. (2023). Neuroanatomy, Gray Matter. In StatPearls. StatPearls Publishing.

Moheet, A., Mangia, S., & Seaquist, E. R. (2015). Impact of diabetes on cognitive function and brain structure. Annals of the New York Academy of Sciences, 1353, 60–71. https://doi.org/10.1111/nyas.12807.

Seaquist E. R. (2015). The Impact of Diabetes on Cerebral Structure and Function. Psychosomatic medicine, 77(6), 616–621. https://doi.org/10.1097/PSY.0000000000000207

Sherwood, J. S., Russell, S. J., & Putman, M. S. (2020). New and Emerging Technologies in Type 1 Diabetes. Endocrinology and metabolism clinics of North America, 49(4), 667–678. https://doi.org/10.1016/j.ecl.2020.07.006

Steineck, I., Ranjan, A., Nørgaard, K., & Schmidt, S. (2017). Sensor-Augmented Insulin Pumps and Hypoglycemia Prevention in Type 1 Diabetes. Journal of diabetes science and technology, 11(1), 50–58. https://doi.org/10.1177/1932296816672689