

|CURIEUX|

ACADEMIC JOURNAL

January Issue

Issue 35: Part One

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: An Exploration of the Philosophy of Sufism Which Guides the Characters Within “The Forty Rules of Love” By Ahmad Adnan

Page 14: Analyzing the Correlation of fMRI Waves of Different Regions in the Brain to Determine Autism Using Deep Neural Networks By Ali Ahmad

Page 24: Roman Expansion and the Fall of The Roman Republic By Stella Anderson

Page 29: Fighting Time: Artificial Intelligence and the New Era of Musical Expression By Kali Anchlia

Page 46: UFT: A Review of Theoretical Physics By Ravi Shah

Page 73: Beyond Force: Korea, Nonviolent Coercion, and the Promotion of the Comfort Women System By Kate Lee

Page 81: Adolescent Substance Abuse: Role of Environmental and Psychological Factors by Hyowon Han

Page 89: The Impact of Airport Expansion on Local Communities By Inwoo Hwang

Page 94: Pharmaceutical and Non-Pharmaceutical Approaches to Managing Creutzfeldt-Jakob’s Disease Symptoms By Suhani Mehta

Page 105: Unraveling the Mysteries of the Chandipura Virus- A Comprehensive Literature Review By Naina Nagendra

Page 112: In What Ways Are People’s Private Information Protected in the Digital World, and How Can Related Issues Be Addressed? By Suahn Park

Page 119: Predicting the Size of Asteroids with Random Forest Classifiers By Alex Stoffel

Page 126: Three Traits to Rule Them All: A Recipe for Exceptional Business Leadership By Eric Azrak

Page 139: Z: The Key By Joseph Yanta

Page 152: Pathophysiology of Arthrofibrosis of the Knee after Ligamentous Injury and Repair: Synthesizing Past and Current Knowledge for Personalized Prevention By Grace Flury

Page 168: Who is A Genius? What is Beautiful? Who is a Terrorist?: Gender, Religious, and Racial Biases in Image Search Algorithms By Sema Akkaya

Page 177: Bitcoin's Carbon Footprint: Utilizing Artificial Intelligence to Analyze Future Energy Consumption and Environmental Impact By Varsha A, Eesha V

Page 187: Navigating the Landscape and Ethical Considerations of AI-Generated Visual Content By YunLe (Hanbin) Wang¹ and Brian Huh²

Page 193: Comparing Intrinsic and Extrinsic Motivation: An Overview of Experiments and Future Areas of Research By Juwon Hwang

Page 200: How to Become a Sports Agent By Daniel Contreras

Page 212: Exploring the Role of the Vagus Nerve in the Microbiota-Gut-Brain Axis By Kangyun Kim and Gudisa Tufa

Page 228: The Application of Enzymes in Industries By Preeti Vadlamani

Page 238: Standardizing Testing: Systematically unfair? By Justin Kang

Page 240: Harnessing AI for Cancer Therapeutics: The Predictive Power of Machine Learning in Anticancer Peptide Discovery By Cheuk Ling Choi

Page 257: The Primary Market of Concert Ticketing: Evaluating Live Nation Entertainment's Involvement with Recent Inefficiencies By Parker Upton

Page 273: Shattering the Glass Flask: A Women's Battle Against Sexism in STEM By Ashley Hurjak

Page 282: Investigating the Role of Epigenetics in Cancer By Anushree Choudhury

Page 308: Why Humans Dream: A Review of Sleep and Brain Biology, and Diseases in the Context of Dreams By Tyler Lovejoy

Page 323: Investigation of High-Performance Sodium-Ion Batteries' Electrode through the Optimization Materials and Manufacturing Processes. By Lam Anh Khoi (Kelvin Lam)

Page 333: Engineering a Deep Learning Model to Correlate Between Single-cell RNA Sequencing By Gene Datasets By Ameya Bhide

Page 337: Integrating Artificial Intelligence and Technology for Urban Environment Sustainability By Aarav Mittal

Page 345: Interdisciplinary Perspectives on Persuasion: A Comparative Analysis With Insights into Algorithmic Game Theory By Aarav Mittal

Page 358: Effect of Social Media on Neuroplasticity in Adolescents By Vanshi Garodia

Page 380: Overlooked Predispositions to Anorexia and Improving Treatment By Lainey Bradley

Page 399: The Efficacy of HPV Vaccines and a Healthy Microbiome on the Development of Cervical Cancer By Sarah Wang

Page 408: The Chronicle of Social Movements: How Museums Influence and Reflect Social Movements in the 21st Century By Laura Wylie

Page 417: Analysis of Common Injuries in Tennis By Sunay Iyer

Page 431: Panopticon Dilemma: Surveillance as Freedom and Captivity in Law By Zhenze Yu

Page 437: The Potential of Using Neural Networks in the Diagnosis of Cardiovascular Disease By Salil Belgal

Page 445: Art of Healing – A Comprehensive Review of Visual Art Interventions in Dementia Care and their Influence on Memory, Quality of Life and Well-Being By Yejie Choi

Page 454: Applications of Atom Manipulation in Scanning Tunneling Microscopy By Nikita Chanda

Page 465: Color Psychology and Pediatrician Attire By Katherine Pyasik

Page 482: Financing Modeling of Renewable Energy By Xavier Ma

Page 504: A Tale of Two Cities: A Comparative Study to Tackle the Issue of Affordable Housing by Joya Trivedi and Oliver Ong

Page 514: Biomass Conversion Methods for Green Energy Production in Rural Regions: A Review By Joseph Tso, Salma Baig, Jack Schimler, and Jordan Penland

Page 531: High School Transition: Hardships and Solutions By Anya Bhakta

Page 540: How did Intellectuals Influence Public Opinion and the Course of the Dreyfus Affair in France at the Turn of the 19th Century and in Subsequent French Politics and Society? By Nikhil Kothari

An Exploration of the Philosophy of Sufism Which Guides the Characters Within “The Forty Rules of Love” By Ahmad Adnan

Abstract

This research utilizes the hypothetical tenets of Sufism and the real-life effects through history and bridges the gap to analyze Ella and Rumi’s spiritual journey. The broader knowledge gained from the effects of spirituality outside the novel constructs a collective knowledge of societal norms. This leads to the capability to tackle usual gender or cultural stereotypes, such as in the case of Ella’s openness to spirituality and Rumi’s sabotage of his own reputation to grow spiritual awareness (Anjum, 2006). It would demonstrate multiple interactions between the tenets of Sufism and the progression of spiritual awareness.

Introduction

This essay will aim to explore “How does the implication of tenets of Sufism present in the parallel narrative transform the characters of “The Forty Rules of Love” by Elif Shafak?” The essay will seek to analyze the role of the primary tenets of Sufism, renouncing worldly matters, attributes, and desires and purifying oneself, as well as secondary tenets which are taking over God’s qualities to slowly achieve an annihilation of the personality, and gaining oneness with God in progressing the spiritual journey of Ella and Rumi. Sufism in the wider context of Islam is a specific mystic body that focuses on spirituality (Schimmel, *The Path*). This research will identify the characters’ current position along the path of enlightenment, and highlight their respective growth using literary devices and chronological position. Furthermore, to strengthen the credibility of the claim, this research will view the two simultaneously running plotlines together to discover a plethora of commonalities and strengthen the effect of the tenets of Sufism.

This essay will focus on the reoccurring motifs and symbols that exist between the simultaneously running plotlines to aid in effectively viewing the decomposition of significant events in the characters’ respective spiritual journeys. Utilizing this decomposition, a lens of a formalist archetype and assuming the individual stages of the tenets of Sufism to the individual spiritual level of Ella and Rumi. While these commonalities of acting upon the same ideal Sufistic path, can be used to identify the magnitude of effect of the tenets, it generally lacks the reference to the wider context of Sufism in Islamic beliefs. As a result, further research was done to effectively contextualize the real-life effects of the tenets. The research led to a consultation of Tanvir Anjum’s “Sufism in History and its Relationship with Power,” to illustrate how the tenets affect people in history. This allows the essay to utilize the formalist lens effectively, since the tenet’s effect can confer exactly with anecdotes in historical records, strengthening its legitimacy.

Primary Tenets of Sufism

Before this essay explores Shafak’s use of a parallel narrative to highlight the significance of the tenets of Sufism in inciting evolution in the characters’ spirituality such as in the case of modern-day wife and mother Ella and thirteenth-century mystical poet and scholar

Mawlana Jalal al-Din Muhammad Rumi as depicted in the *Sweet Blasphemy*, it is necessary to understand the path of Sufism which contextualizes the spiritual journey of the characters. It is generally regarded that Sufis undergo a standard course to achieve “fanā (“annihilation”), primarily an ethical concept of annihilating one’s own qualities,” (Schimmel, *The Path*) in order to exhibit the qualities of God. Abiding by the tenets of Sufism of abandoning one’s own material possessions, purifying the soul through means of worship, contemplation, and connection with God, and finally accepting God’s qualities over your own would allow the murid, student, to transcend past temporal worries and achieve a higher spiritual status. Schimmel suggests that “the ecstatic experience, called intoxication, is followed by the “second sobriety” - i.e., the return of the completely transformed mystic into this world where he acts as a living witness of God” (Schimmel, *The Path*) can be utilized to view the spiritual journey of the characters in the novel through transformative motifs and spiritual diction. This chapter will specifically discuss the primary tenets of Sufism, a renunciation of worldly matters and its effect on the purification of the soul, to represent the path taken by the murids, Ella and Rumi, through the lens of Sufism as a whole.

Shafak introduces the integration of the tenets of Sufism in Ella’s life by showing how the renunciation of worldly things incites changes in priorities. The extended metaphor of Ella’s monotonous routine represents constraints due to the typical demands of society, especially that of a maternal figure who must tend to the needs of her children, and husband, and work before herself, which allows Ella to question her usual conformation to these societal and gender norms. Consequently, it leads to an abandonment of the security of routine, in favor of uncertainty. As a result of this realization and multiple interactions with Aziz about the nature of spirituality and its growing role in her life, she monologued about if “she look[ed] younger? Prettier? Or perhaps more full of life? She could not see any difference. Nothing had changed, and yet nothing was the same anymore” (Shafak, 2009, p. 117). The creation of a semantic field of the words “younger, prettier, and full of life” portrays an overall reference to the primary preoccupation, that being worldly matters such as appearance and societal perception of oneself. It also references Ella’s cultural context as the semantic field highlights what her society deems to be most important for a woman to conform to. Her later statement that “nothing was the same anymore” juxtaposes the immense importance of gender norms that she previously felt, indicating a change in the change in priorities for her life by shifting the diction of the realization from appearance to spiritual uncertainty. Her differences in reaction to the positive changes in spirituality directly signify her changes in priorities by relinquishing attention to gender norms and accepting spirituality itself.

The parallel narrative nature of the story strengthens the hypothesis of the implication of the tenets of Sufism in spiritual evolution as Rumi, fictionally depicted in “*Sweet Blasphemy*,” undergoes a similar path to Fana as Ella. Before meeting Shams, Rumi was already an accomplished scholar and imam who conveyed the messages of God to many others. However, he realized his lack of spiritual awareness after his interaction with Shams, which caused a change in priorities similar to Ella's. Shams had asked Rumi, “Of the seven stages, which stage

are you at? And do you think you have the heart to go further, till the very end? Tell me, how big is your cup?” (Shafak, 2009, p. 111). The repetition of questions that Shams provides Rumi alludes to an overall lack of spiritual understanding, and simply being knowledgeable about Islam, may not be enough to truly be close to Him. It also creates the imagery of a path, alluding to the overarching path of the Sufistic tenets to achieve oneness with God and how first Rumi must be aware of his lack of spirituality in order to start following the tenets, with the first being to change priorities away from status and reputation. The symbol of the cup signifies this void of spiritual understanding, yet also the capability to achieve fulfillment.

This integration of the tenets of Sufism is further highlighted when Shams gives Rumi the task of obtaining alcohol to ruin his flawless reputation to gain a greater understanding of “spiritual growth [and how it] is about the totality of our consciousness, not about obsessing over particular aspects” (Shafak, 2009, p. 160). The synecdoche of “our” represents the position of spirituality in all planes of existence and its effect on all. While this is demonstrated through Shams’ interactions with Hasan the Beggar, Desert Rose the Harlot, and Suleiman the Drunk to guide them on their respective paths to enlightenment, it is especially prevalent for Rumi. Being linear in his approaches to obtaining higher levels of spirituality, Rumi experiences a trade-off that imbalances the priorities of closeness with God and with that of human judgment, reputation, customs, and traditions. Hu suggests that viewing the importance of one's path to spiritual enlightenment above material goods and attributes shifts priorities to embody the tenets of Sufism (Hu, n.d.). This becomes a clear indication of changing priorities as portrayed by Rumi’s acceptance of completing the task.

As Ella and Rumi further accept spirituality in their lives, the tenet of Sufism of disposing of materialism and societal pressure allows the characters to exhibit less control and a different ego. Ella’s decrease of control over her life stems from the acceptance of spirituality, and therefore the tenets of Sufism, which leads her to “break away from her staid and tranquil life. From a woman with lots of dull grays and browns on her life’s canvas, she was turning into a woman with a secret color—a bright, tantalizing red” (Shafak, 2009, p. 95). The chronological progression of painting imagery and color diction highlights the transformative nature of the integration of spirituality in Ella’s life. The “grays and browns” mentioned allude to the extended metaphor of routinely conforming to gender norms that inhibit individuality and limit spirituality. It highlights the draining disposition of the gender norms in modern society wherein the ideal physical and spiritual appearance of a woman must be kept indistinguishable from others. Furthermore, the bleak imagery transforms to become much brighter, symbolizing enhanced spiritual awareness, which also leads to improved self-esteem and happiness. The color imagery also alludes to another extended metaphor of light, often represented as auras in the novel, which is further evidenced by the role of light in Sufism as a whole. Ahmady suggests that “light is right or the divine being itself ... and colors are the variant form of the existence” (Ahmady, 2017). Connecting the use of color symbolism in Ella’s spiritual change signifies the growing acceptance of the tenets of Sufism. She has less control of her life due to the increasing

uncertainty that pairs with continuing on the path of the tenets of Sufism, which allows for divine enlightenment.

The parallel narrative between Ella and Rumi strengthens the evidence for the connections with the path of the tenets of Sufism as the characters interact with further stages of it. After giving up worldly matters to incite changes in priorities and control over one's life and ego, the purification of the soul is the next step of the characters' path toward enlightenment. "Once he made me beg across from the mosque where I used to preach, forcing me to put myself in the shoes of a leper beggar. He cut me off first from my admirers, then from the ruling elite, bringing me in touch with the common people" (Shafak, 2009, p. 188). The creation of the semantic field of "begging, forcing, cutting, and bringing" represents the forceful nature to purify the soul in order to advance the path of the tenets of Sufism. The semantic field also highlights the difficulty of Rumi's tasks which he had to undergo in order to purify himself and creates a tone of tribulations to achieve the fana (Saari, 2002). Therefore, the parallel narrative between Ella and Rumi demonstrates a collective development through the path of the tenets of Sufism.

Secondary Tenets of Sufism

The secondary tenets of Sufism regard the inheritance of God's qualities into oneself to invoke a journey to achieve Fana. It involves taking over divine attributes to annihilate a previous personality and give yourself fully to God to achieve oneness (Schimmel, *The Path*). The parallel narrative in the novel between the murid, students, Ella and Rumi, strengthens this hypothesis by demonstrating the path to spirituality in exact accordance with the tenets of Sufism as a whole. The subsequent tenets of Sufism that begin to configure, only occur after meticulous mystical contemplation and absolute adherence to the primary tenets. There is an important deviation between relinquishing worldly matters in a vacuum, than relinquishing them for God, or an attempt to become closer to Him. Without complete adherence to the primary tenets, the goal of true oneness with the divine becomes clouded with preoccupation about solely giving away material possessions. Following the primary tenets without the objective of fana restricts the potential to reach the secondary tenets of Sufism. Inheriting God's qualities into oneself and acting as a witness for His will is not possible without primarily building the fundamental principles of intoxication to worldly matters. Shafak illustrates Ella as one who overcame cultural obstacles regarding the openness of spirituality and the uniqueness of individuality. She begins to adopt the qualities of God such as forgiveness, as said in the next stage of the tenets of Sufism.

Firstly, Ella identifies the quality of forgiveness as one that is divine, and the absolute dissemination of it is only possible through the power of God. She says that there are more important things than "passion and love in a marriage, such as understanding, affection, compassion, and that most godlike act a person could perform, forgiveness" (Shafak, 2009, p. 10). The simile comparing the significance of the act of forgiveness itself with the attributes of God is important to acknowledge as it shows Ella's identification of God's qualities which is later used to evidence the significance of its adoption to her own personality. This identification

develops Ella's spiritual journey using the tenets of Sufism as she is able to adopt the qualities effectively. She further demonstrates this substitution of God's qualities for her own when she forgives herself for the first time. She tells Jeannette "I'm sorry for the things I do. I know I shouldn't complain when I'm so blessed. I'm so ... unhappy -" (Shafak, 2009, p. 35). The anaphora of "I" emphasizes the magnitude of forgiveness that she gives herself as this was the first time where she is able to be honest and open. It also reinforces the significance of forgiveness as being a quality of God which becomes the next step through the tenets of Sufism and spirituality. The development of the motif of forgiveness portrays the evolving progression of character which Ella demonstrates after continuous interactions with spirituality such as during the teachings of Aziz. Shafak's reoccurring mentions of the forgiveness she gives herself, and specifically the heightened involvement of the motif in Ella's spiritual journey as a whole, highlights the increasing progression towards higher stages of the tenets of Sufism.

Similarly, the fictional depiction of Rumi in "Sweet Blasphemy" demonstrates the insertion of God's qualities, a stage of the tenets of Sufism, when he forgives and loves all the outcasts of society. Progressing through the tenets of Sufism alongside Ella, a lens of a parallel narrative can be utilized to strengthen the claims of the implications of the tenets in the changes of characters' respective spiritual journeys. Aladdin mentions that "after Shams came into [Rumi's] life, his circle of love became so vast it included even the most fallen of society - prostitutes, drunks, and beggars, the scum of the scum. I believe he could even love Shams' killers" (Shafak, 2009, p. 218). The symbol of the circle represents a continuous love for all, equally stretching to infinitely many directions from his own single point as circles themselves often do. The significance of Rumi's vast love is that he could forgive anyone, even if they go against the teachings of Islam to which he devotes his life. For example, even though alcohol is forbidden, he comes to the tavern during one of his trials with Shams. He shares the wisdom of the negative nature of wine with courtesy and sincerity to the people at the tavern, despite it being prohibited in his religion. Suleiman the Drunk even states that "no matter what people say about what you did today, and I'm sure they are going to say plenty, I think as a preacher it was very brave of you to come to the tavern and talk with us without judgment" (Shafak, 2009, p. 157). Rumi had gently conversed with those ostracized and displayed his growing forgiveness even for people who were disobeying the Sharia law. Furthermore, this adoption of God's qualities as a tenet of Sufism increases substantially until others even notice the overwhelming vastness of it. Rumi attempts to adopt absolute forgiveness, potentially even to Shams' killers. This is only demonstrated by God's characteristics, and Rumi attempts to achieve spirituality through the tenets and gain a better understanding of the wider context of the path left to achieve fana.

The final stages of the tenets of Sufism are achieved by Ella and Rumi who go through the full path of relinquishing material and worldly obsessions, purifying themselves, and incorporating God's qualities into oneself to become closer to achieve the annihilation, fana, of the previous personality (Specia, 2017). Ella achieves this golden stage of the tenets of Sufism after she abandons her previous life of routine and security, and lives in the moment, in

accordance with the forty rules and Aziz's teachings. She reflects on her experiences and future uncertainty and states that a "life without love is of no account. Don't ask yourself what kind of love you should seek, spiritual or material, divine or mundane, Eastern or Western ... Divisions only lead to more divisions. Love has no labels, no definitions. It is what it is, pure and simple. Love is the water of life. And a lover is a soul of fire! The universe turns differently when fire loves water" (Shafak, 2009, pp. 224-225). The spiritual and mystical diction heavily utilized in this statement represents the enlightenment achieved by walking the path of the tenets of Sufism. It directly juxtaposes her initial personality and perception on life, symbolizing the annihilation of the past self, achieved through the path traversed throughout the novel. Furthermore, the imagery of fire and water, which is usually used to represent contrasting ideals, beliefs, or situations, now highlights the union between the past personality and the current one, signifying an enlightenment that transcends physical properties and interactions. The heavily contrasting nature of fire and water is now adjoined harmoniously to indicate the achievement of oneness with spirituality. The acceptance of all parts of the personality, therefore, allows the attainment of a higher level of spiritual understanding and awareness.

Rumi also experiences a similar fana, conversion of past personality, as the parallel narrative further demonstrates the implications of the tenets of Sufism through the lens of multiple characters with contrasting cultural and societal backgrounds. However, a stark difference between Ella and Rumi's spiritual journey is the initial starting point when adopting the tenets. Ella displayed naivety regarding spirituality as a whole and experienced the teachings fully all at once. Whereas Rumi had previous knowledge and capability for understanding spirituality. Yet, this previous knowledge which stemmed only from books did not allow Rumi to fully conceptualize true spirituality, especially in comparison to Shams. Regardless, Rumi achieved a form of fana after the death of Shams, which allowed him to convey his spiritual experiences and knowledge through poetry. He states that "we dervishes will dance our way through love and heartbreak even if no one understands what we are doing" (Shafak, 2009, p. 221). This is significant as it is an allusion to an overarching development of typical Sufism that uses dance and poetry to express spiritual understanding. Rumi abandons his old personality and transforms into someone with an oneness to God, similar to his teacher, Shams. He begins to view his connection with God beyond mystical contemplation and extensive studies, but rather by finding love and connection in all forms of expression. Furthermore, Fizza suggests that the whirling itself "has great significance in terms of teaching an individual to liberate oneself from being hedonistic" (Fizza, 2021). The specific references to tenets of Sufism, such as the relinquishing of worldly matters, further connect with the implication of the tenet and its use in the progression of Rumi in his spiritual journey. This evidences the transcendence that Rumi achieves by going through the tenets of Sufism and illustrates a transformation in his character.

Conclusion

Shafak uses extended metaphors and metaphysical diction to illustrate the spiritual progression of the characters, Ella and Rumi, through an overarching lens of Sufism and its

tenets as a whole. For Ella, spirituality is explored by dividing the evolutions of perspective for the subject, and each of its effects on the individual. Firstly, Ella harbored dismissal since cultural factors and tradition inhibited the exploration of the concept of spirituality. She later, after more interactions with spirituality, becomes more open to the idea and starts to ponder the possibility. She explores the positive impacts spirituality has on her happiness, relationships, and transformation from the mundane routine that was all she once knew. As a result, she slowly accepts and integrates it into her life until she fully embraces it and allows spirituality to change as a result. Her priorities change as a result of spiritual integration, and this evolves into Ella taking qualities of God, exhibiting the standard path to true oneness that extends beyond the novel and into Sufi beliefs. Her journey to the annihilation of a previous personality that solely conformed to gender and societal norms exists as a result of continuous, incrementing spiritual interactions, and its continued benefit in her life.

Shafak strengthens this claim of the significance of the Sufistic tenets' implications on one's spiritual journey by portraying a parallel narrative that exhibits it in different times and cultures. It allows for the elimination of potential factors such as cultural differences or specific people that may vary the effects of spirituality. By showing spirituality's equal effect in different time periods regardless of the cultural bias, the credibility of the tenets of Sufism stands. Rumi's role in developing this hypothesis is to demonstrate an unusual lack of spiritual awareness in forms of life other than books, prayer, and speech. Rumi's interactions with his teacher Shams, draw a connection to God in various ways that seem insignificant such as interactions with the ostracized and weeks of mystical contemplation and pondering. Rumi then traverses through the Sufistic path in order to heighten this enlightenment. (Schimmel, 1988)

However, a weakness of the essay stems from the fictional nature of Ella which differs from the fictional nature of Rumi. Ella is a character that is original to the novel and shows the effects of spirituality through a contemporary lens. However, a historical fictional novel, "Sweet Blasphemy," within this primary fictional novel holds Rumi. Ella is indirectly guided by this version of Rumi and brings into question the legitimacy of the effect of spirituality as they both exist in a fictional plane and may be altered at any time to align with the author's choices. It is possibly rooted in ambivalence, yet, the path the characters take with regard to the tenets of Sufism, is non-fictional to an extent, and as a result, it precisely conveys the spiritual interactions in the novel.

Works Cited

- Ahmady, Fariba. "The Symbolism Of Color In The Islamic Sufism & Mysticism." PhD Thesis. University of Religions and Denominations, 2017. 21 November 2022.
<<https://old.urd.ac.ir/en/research-archive/the-symbolism-of-color-in-the-islamic-sufism-mysticism/>>.
- Anjum, Tanvir. "Sufism in History and Its Relationship with Power." *Islamic Studies*, vol. 45, no. 2, 2006, pp. 221–68. JSTOR, 20 November 2022
<<http://www.jstor.org/stable/20839016>>
- Fizza. Exploring The Significance of Dance in Sufism. 29 October 2021. 16 November 2022.
<<https://houseofpakistan.com/dance-in-sufism/>>.
- Hu, Ya. Field Thirteen, Renunciation and Asceticism (Zuhd). n.d. 26 November 2022.
<<http://nurashkijerrahi.org/field-thirteen-renunciation-and-asceticism-zuhd/>>.
- Saari, Che Zarrina. A Purification Of Soul According To Sufis: A Study Of Al-Ghazali's Theory. University of Malaya. Malaya, 2002. 25 November 2022.
<https://www.researchgate.net/publication/313470158_A_Purification_Of_Soul_According_To_Sufis_A_Study_Of_Al-Ghazali's_Theory>.
- Schimmel, Annemarie. "Mystical Poetry in Islam: The Case of Maulana Jalaladdin Rumi." *Religion & Literature*, vol. 20, no. 1, 1988, pp. 67–80. JSTOR, 20 October 2022
<<http://www.jstor.org/stable/40059367>>
- Schimmel, Annemarie. *The Path*. Ed. Adam Zeidan. n.d. 16 November 2022.
<<https://www.britannica.com/topic/Sufism/The-path>>.
- Shafak, Elli. *The Forty Rules of Love*. The Penguin Group, 2009.
- Specia, Megan. "Who Are Sufi Muslims and Why Do Some Extremists Hate Them?" *The New York Times* 24 November 2017. 14 November 2022.
<<https://www.nytimes.com/2017/11/24/world/middleeast/sufi-muslim-explainer.html#:~:text=Sufi%20practice%20focuses%20on%20the%20spiritual%20learning%20known%20as%20tariqa.>>>.

Analyzing the Correlation of fMRI Waves of Different Regions in the Brain to Determine Autism Using Deep Neural Networks By Ali Ahmad

Abstract

Diagnosing autism is challenging because, in many cases, several of the outward-facing behavioral symptoms of autism are hard to quantify. In this study, we aim to overcome this challenge by innovating and analyzing Functional Magnetic Resonance Imaging (fMRI) brain scans from the Autism Brain Imaging Data Exchange (ABIDE) dataset to classify autistic subjects from neurotypical subjects using artificial intelligence. This dataset is a culmination of multiple studies originating from several universities in an attempt to consolidate all fMRI autism data available. From the fMRI data, connectivity between different areas of the brain was quantified using correlation coefficients. The connectivity data was used to train a machine learning model known as a multilayer perceptron to make the classification. To optimize prediction accuracy, the model's hyperparameters were tuned, and several different neural network architectures were tested and compared. In one set of trials, a network of two hidden layers was used, each with N nodes. In another set of trials, two hidden layers were also used, the first with N nodes and the second with $N/2$ nodes. When comparing all the models, architectures with hidden layer node sizes of 2500-1250 and 4500-2250 yielded the highest prediction accuracy of 69.92%. This study marks a notable progression in the exploration of advanced machine learning methods to understand the neural foundations of ASD, highlighting the capacity of deep neural networks to make valuable contributions to the realms of neuroimaging and clinical diagnosis.

I. Introduction

Autism, also known as autism spectrum disorder (ASD), is a neurodevelopmental disorder that affects social interaction, communication, and behavior. Though the exact cause of ASD is unknown, researchers have come up with several possible reasons related to brain development and autism, in particular genetics. Many studies show that certain genetic mutations can increase the risk of developing ASD [1]. These genetic factors can influence brain function, affecting the way neural circuits are formed and organized.

Another possible reason for autism is differing brain structure and connectivity. Brain imaging studies have shown differences in brain structure in individuals with autism. These differences take place in regions of the brain that determine social cognition, communication, and sensory processing. One other possible reason for autism is prenatal and early development factors. This includes maternal factors such as advanced parental age, certain infections during pregnancy, maternal use of certain medications, and exposure to environmental toxins. Additionally, disruptions in early brain development, including abnormal neuronal migration, altered synaptic pruning, and imbalances in neurotransmitters, may be responsible for the development of autism. It is important to understand this condition to foster inclusivity and acceptance for individuals with autism. By understanding and supporting individuals with

autism, we can work towards creating an inclusive society that values and respects the rights and contributions of all individuals, regardless of their neurodiversity.

Currently, the primary method of diagnosing autism practiced by medical professionals happens after several years into one's life. The parents/guardians of a child may notice certain features relating to their child's social capabilities and consult a doctor to see if the child has any condition. From there, the doctors undergo a series of tests to gauge the child's social abilities and determine whether or not they have ASD. This method is known as Developmental Screening or Observation and Behavioral Assessments [2]. Other methods practiced by doctors include Diagnostic Interviews with parents and Developmental and Psychological Assessments, both of which are similar to the previously mentioned methods. In addition to medically practiced methods, there are several proposed methods out there.

One of these methods is the use of machine learning/A.I. to detect autism. In this method, a neural network analyzes some input such as the MRI image of a brain, genetic data, or any phenotypical data, and determines if that person has ASD or not. This is performed by training the deep learning model with labeled data and making predictions. An example of a machine learning-based method of detecting autism is the use of A.I. to detect “abnormal” behavior in a person by analyzing a video of that person [3]. In the project, they developed deep learning models for the automatic identification of clinically relevant behaviors exhibited by infants in a one-on-one interaction setting with parents or expert clinicians. They achieved a 70% accuracy for smiles, 68% accuracy for look faces, 67% for look objects, and 53% accuracy for vocalization. Another proposed method can be seen through a study that utilized language and speaking patterns to determine if one had autism. The study measured expressive and receptive language on six different occasions between the ages of 2 and 19 in a cohort of 192 different children who were initially referred for autism. The results confirmed that those with specific language impairments typically had ASD.

Another study claims that C677T polymorphism plays a role in the classification of autism [4]. The study states “recent studies have revealed that the genes involved in the folate/homocysteine pathway, in particular C677T polymorphism, may be risk factors for autistic children.” A total of 98 children were diagnosed as autistic and 70 age and sex-matched children who did not have autism were tested for C677T polymorphism. This polymorphism was studied by using polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) methods. They concluded that MTHFR 677T-allele frequency was found to be higher in autistic children compared with children without autism (29% versus 24%), but it was not found statistically significant. Additionally, C677T polymorphism is a point mutation at position 677 on MTHFR gene with the substitution of cysteine to thymine nucleotide at that position. Essentially, the study concluded that C677T polymorphism can cause several mutations in the human body, one of which is autism.

In this paper, we use a Deep Neural Network (DNN) to analyze brain wave signals collected from MRI brain scans of subjects with and without ASD. With the rise of neuroimaging, researchers have turned to fMRI data analysis to explore Autism Spectrum

Disorder. Numerous studies propose that the impairments in social and communicative functions are linked to the functioning and connectivity of cortical networks. For the fMRI data analysis, machine learning classifiers are promising methods to investigate how a pattern of brain activity is related to different cognitive states. A growing number of studies have shown that machine learning classifiers can be used to extract useful information from neuroimaging data. Several of these studies utilize typical machine learning methods, but research shows that Deep Neural Networks are very promising in regards to voxel-based classification and functional connectivity-based classification [5]. The application of deep learning to fMRI data classification faces several challenges, including the limited availability of training samples. To address this, we utilized the Autism Brain Imaging Data Exchange (ABIDE) dataset, which provides over a thousand rs-fMRI samples from various sources worldwide [6]. The objective of this study is to create deep-learning models to classify ASD subjects and typically developing (TD) participants based on rs-fMRI data from ABIDE.

II. Methods

ABIDE Data Set

For our experiments, data from the Autism Brain Imaging Data Exchange (ABIDE) was used, consisting of fMRI scans of 1112 subjects with a roughly 50/50 split between people with typically developing brains and people with ASD. This dataset is a culmination of many studies originating from several universities in an attempt to consolidate any and all data available. This was mainly to provide researchers with an ample dataset to advance methods to diagnose ASD subjects.

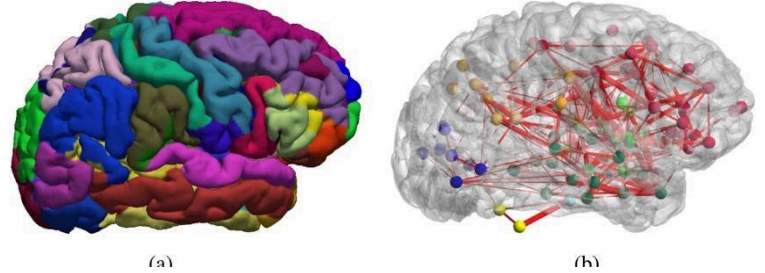
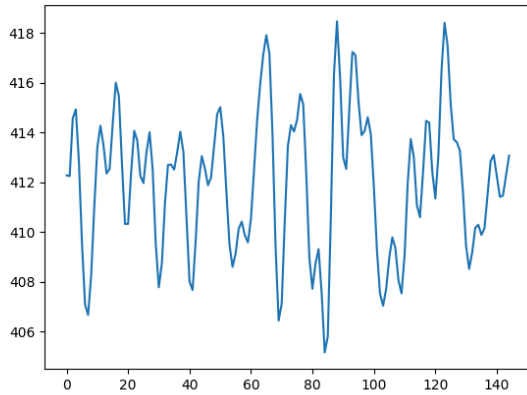
The raw fMRI data represents Blood Oxygen Level-Dependent (BOLD) signals as a function of time. For each location of the subject's brain, there is one such signal measured. Since hundreds of thousands of time series are measured per person, there is too much data to be analyzed for any sort of prediction. This is typically solved by additional processing steps to reduce the size of the dataset.

Preprocessing

Since the raw fMRI scans contain hundreds of thousands of time series per person, the ABIDE dataset offers several options for preprocessed fMRI scans which reduce the number of time series. It does this by splitting the brain up into regions and then averages the time series of each respective region. There are a number of different ways to split up the brain regions, and the ABIDE dataset provides those as options. We chose one that sectioned the brain into 200 regions for our experiments.

For this study, we compared data from different regions of interest (ROI) of the brain within subjects to see if there are similarities or differences between brain waves, or the fMRI signals, in subjects. In Figure 1, the fMRI time series for one of the 200 regions is illustrated on the left. The x-axis represents the progression of time during the acquisition of rs-fMRI data, and

the y-axis is indicative of the strength or magnitude of the blood oxygen level-dependent (BOLD) signal, which is a measure of brain activity in fMRI.



To compare the data, we take values from two of the fMRI waves and plug them into a correlation coefficient function (1). Where r is the correlation coefficient, x is the fMRI waveform of one part of the brain, and y is the fMRI waveform of another part of the brain. If the function outputs a 1, it signifies that both parts of the brain being compared have a positive correlation. An output of 1 does not necessarily mean that the waves are identical, rather it means that for both waves, as one “goes up” so does the other. An output of -1 signifies the opposite of 1: as one of the waves increases, the other decreases. If the function outputs a correlation coefficient of 0, the parts of the brain do not have any significant connection.

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (1)$$

To visualize the data, the values outputted from all the possible combinations of brain regions in the correlation coefficient function were displayed in a graph seen in Figure 2. At this point, the data is in the form of a 2-D dimensional matrix. For the data to be inputted into the Deep Neural Network (DNN) model, the data must be in the correct form. This is performed by taking data from half of the correlation coefficient graph (either below or above the diagonal line) and inserting the data into a one-dimensional array with 19,900 input parameters, each of which represents the correlation coefficient for one connection between two different regions of the brain. Only half of the data is used because, in the graph, the bottom left, and top right are identical in the correlation coefficient matrix.

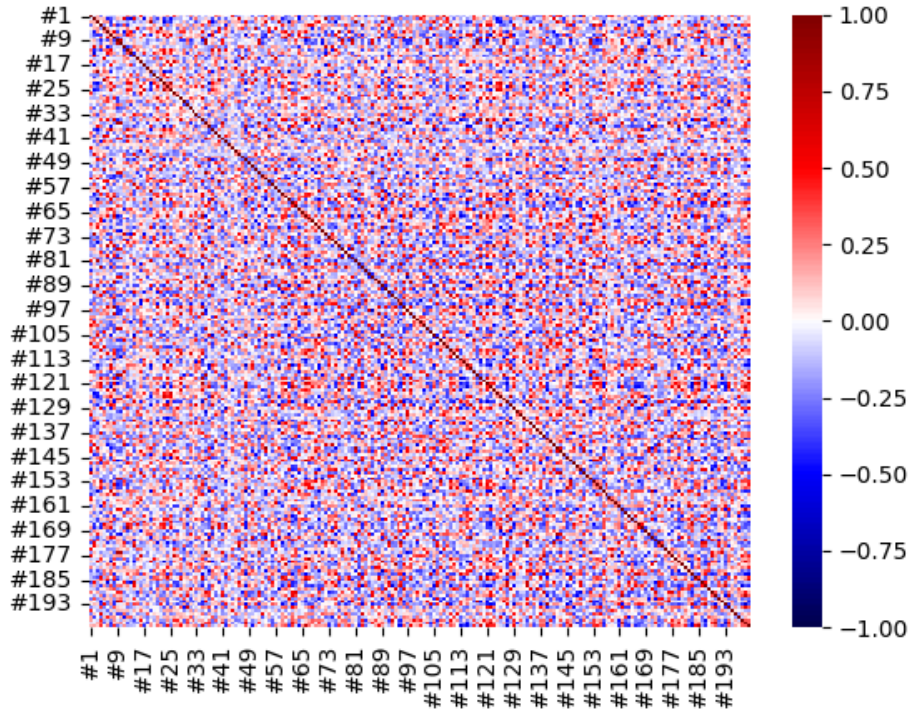


Figure 2

Classification

In the past, several studies have shown that the use of machine learning provides high efficiency in the analysis and prediction of fMRI data. Typically, for the classification of ASD and TD, studies applied the supervised learning method of support vector machines (SVM). These studies have shown to provide very high efficiency (outputting values as high as 97%) [8]; however, they all rely on data from small sample sizes. When the same methods were exposed to higher sample sizes, the values significantly dropped.

While many established machine learning studies predominantly employ traditional algorithms such as Support Vector Machines (SVMs), decision trees, and logistic regression for classification tasks, a noteworthy shift has emerged in the field. Deep Neural Networks (DNNs) have demonstrated superior effectiveness in predicting fMRI-related outcomes compared to conventional machine learning methods. In this study, we use a DNN to classify individuals with Autism Spectrum Disorder (ASD) and typically developing (TD) participants, utilizing functional connectivity features to achieve the highest achievable prediction accuracy.

Deep Neural Networks

Our DNN model is a multi-layer perceptron (MLP), consisting of an input layer, two hidden layers, and an output layer. The input layer is responsible for ingesting the provided fMRI data from the ABIDE data set, while the hidden layers play a crucial role in extracting relevant features from the input data. Within the hidden layers of the MLP, individual nodes receive input

from corresponding nodes in the preceding layer. These nodes applied non-linear transformations to the data by way of a ReLU activation function. Non-linear transformations introduce complexity and enable the DNN to understand intricate, non-linear relationships within the data.

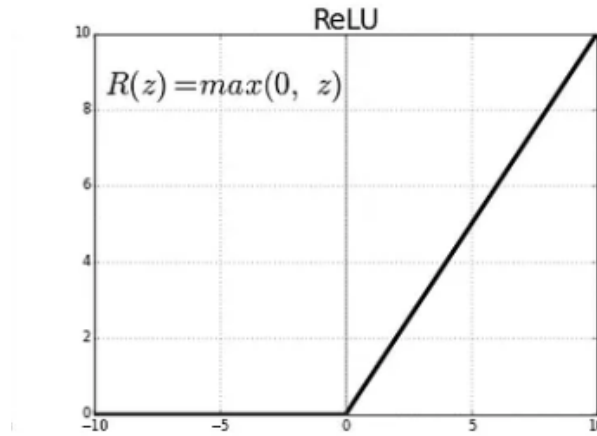


Figure 3

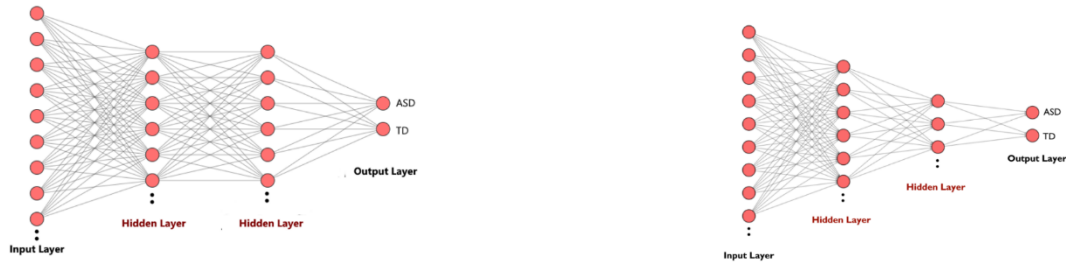
A neural network function with 4 main steps. First, the network assigns random weights to each of the channels (connections between layers). Secondly, the network performs what is referred to as forward propagation. Forward propagation pushes the input data through the neural network, applying the ReLU activation function at each node. The network's output is then compared to the actual output, and it performs stochastic gradient descent, which minimizes the loss function $J(w_i)$ (2). Finally, the network performs backward propagation, which updates all the weights in the network. This process is repeated thousands of times given the learning rate of 0.0001 until the network effectively minimizes the loss function.

$$w_i := w_i - \alpha \cdot \frac{\partial}{\partial w_i} J(w_0, w_1, \dots, w_n)$$

$$w_i := w_i - \alpha \cdot \frac{\partial}{\partial w_i} \cdot \frac{1}{n} \sum_{n=1}^m (y_i - w_0 - w_1 x_1 - w_2 x_2 - \dots - w_n x_n)^2$$

We utilized two main different categories of neural network architectures. The first of which included an input layer, two hidden layers (both of which had the same number of nodes), and an output layer. The second architecture of neural network was arranged in a different manner. The first hidden layer has N number nodes, while the second hidden layer has N/2 number of nodes as the first layer. The difference in nodes between the two hidden layers allows

the MLP to narrow down the model complexity to aid in approaching a more accurate final result consisting of only two nodes. The processed data is then transmitted to one of the two nodes in the output layer, either the node that determines that the subject is TD or the node that classifies the subject with ASD.



The MLP architectures we used can be seen in Figure 4 and Figure 5. These figures illustrate that the perception accepts an input of 19900, transmits the data to the first hidden layer with a node amount between 100-5000, moves the altered data to the second hidden layer with node amount between 50-5000, and then goes to the output layer, which has a binary outcome: positive or negative for autism.

Input Layer	Hidden Layer 1	Hidden Layer 2	Output Layer
19,900	100	100	2
19,900	500	500	2
19,900	1000	1000	2
19,900	1500	1500	2
19,900	2000	2000	2
19,900	2500	2500	2
19,900	3000	3000	2
19,900	3500	3500	2
19,900	4000	4000	2
19,900	4500	4500	2
19,900	5000	5000	2

Input Layer	Hidden Layer 1	Hidden Layer 2	Output Layer
19,900	100	50	2
19,900	500	250	2
19,900	1000	500	2
19,900	1500	750	2
19,900	2000	1000	2
19,900	2500	1250	2
19,900	3000	1500	2
19,900	3500	1750	2
19,900	4000	2000	2
19,900	4500	2250	2
19,900	5000	2500	2

The loss function is used to measure the prediction error that the model produced. The loss function we used is called cross entropy which is typically used for classification models. Cross entropy loss measures the difference between the discovered probability distribution of a machine learning classification model and the predicted distribution. In the function, p refers to the predicted possibility, y is the expected value, and l is the value of loss. The loss value can be between 0 and 1, with 0 being a perfect prediction.

$$l = -(y \log(p) + (1 - y) \log(1 - p))$$

(3)

While running the trials, we adjusted the learning rate and number of epochs from 0.00001 to 0.5 for the learning rate and 50 to 500 for the epochs. The learning rate controls the step size at which the model's parameters are updated during training. Additionally, the learning rate determines how much the model's weights and biases are adjusted during each iteration of the optimization algorithm. An epoch refers to one complete pass of the entire data set. In our experiment, the number of epochs was increased to the point where it was evident that the network was reaching its lowest possible training error, also known as early stopping. Additionally, the data had a 70/30 split between the training and testing sets and randomly assigned to each set. This means that the program used 70% of the data to learn and “train” on, and then evaluated the other 30% of the data, creating predictions.

III. Results

In this experiment, we applied a deep neural network, specifically a multi-layer perceptron to the resting state fMRI data extracted from the ABIDE repository. We utilized two main MLP architectures, both with differing numbers of nodes and transformations throughout the network. Tables 1 and 2 lay out the success of the network at each individual iteration. These values can be visualized in Figures 7 and 8, which show how the network potential was maximized. For architecture 1, hidden layers with node amount of 2500-2500 responded with 69.55% accuracy. For architecture 2, hidden layers with node sizes of 2500-1250 and 4500-2250 both accurately predicted 69.92% of the data set.

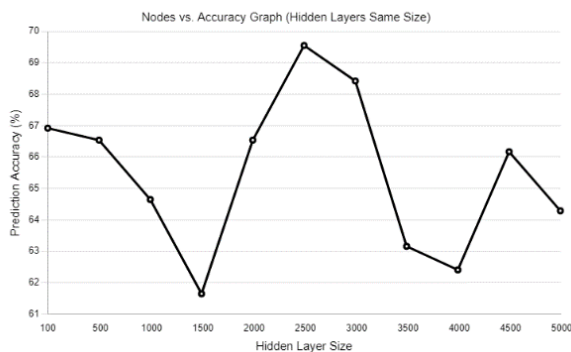


Figure 7

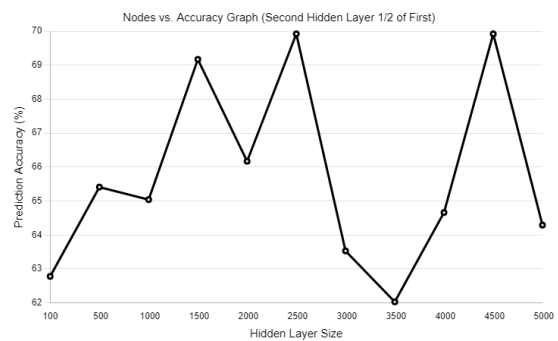


Figure 8

First Layer Size - Second Layer Size	Accuracy
100 - 100	66.92%
500 - 500	66.54%
1000 - 1000	64.65%
1500 - 1500	61.65%
2000 - 2000	66.54%
2500 - 2500	69.55%
3000 - 3000	68.42%
3500 - 3500	63.16%
4000 - 4000	62.41%
4500 - 4500	66.17%
5000 - 5000	64.29%

Table 1

First Layer Size - Second Layer Size	Accuracy
100 - 50	62.78%
500 - 250	65.41%
1000 - 500	65.04%
1500 - 750	69.17%
2000 - 1000	66.17%
2500 - 1250	69.92%
3000 - 1500	63.53%
3500 - 1750	62.03%
4000 - 2000	64.66%
4500 - 2250	69.92%
5000 - 2500	64.29%

Table 2

IV. Conclusion

In this study, we focused on the application of Deep Neural Networks to analyze MRI brain scans, specifically resting-state functional MRI data, of individuals with and without Autism Spectrum Disorder. The rise of neuroimaging has opened new avenues for understanding the neurological basis of ASD, with a particular emphasis on the functional and connectivity patterns of cortical networks. The primary objective of this study was to harness the potential of deep-learning models to classify individuals with ASD and typically developing participants based on rs-fMRI data from ABIDE.

The results indicated that a multi-layer perceptron with two layers, the second layer being half of the first layer, allows for the greatest accuracy in predicting the status of fMRI scans. When the network followed the architectures of 19900-4500-2250-2 or 19900-2500-1250-2, it accurately predicted 69.92% of the 1112 samples, more than any other iterations in this study.

This research represents a significant step forward in the exploration of advanced machine learning techniques for understanding the neural underpinnings of ASD, and it underscores the potential for deep neural networks to contribute to the field of neuroimaging and clinical diagnosis. As we continue to refine and expand our methodologies and datasets, we can look forward to further advancements in the early detection and characterization of ASD, potentially leading to more effective intervention and support for affected individuals. Possible future alterations in methodology include utilizing a different preprocessing pipeline, adding more hidden layers within the network, using different activation functions, and running the same network with access to more computing power to allow for greater precision.

Works Cited

- [1] "Autism." National Institute of Environmental Health Sciences, U.S. *Department of Health and Human Services*. Accessed 8 Nov. 2023.
- [2] "Screening and Diagnosis of Autism Spectrum Disorder." *Centers for Disease Control and Prevention*. 31 Mar. 2022
- [3] Wu, Chongruo. "Machine Learning Based Autism Spectrum Disorder Detection from Videos." *PubMed Central*. 14 Apr. 2021
- [4] Sener, Elif. "MTHFR Gene C677T Polymorphism in Autism Spectrum Disorders." *Hindawi Publishing Corporation*. 6 Nov. 2014
- [5] Pascanu, Razvan, et al. "How to Construct Deep Recurrent Neural Networks." *arXiv.Org*, 24 Apr. 2014, arxiv.org/abs/1312.6026.
- [6] Craddock C, Benhajali Y, Chu C, Chouinard F, Evans A, Jakab A, Khundrakpam BS, Lewis JD, Li Q, Milham M, Yan C and Bellec P (2013). *The Neuro Bureau Preprocessing Initiative: open sharing of preprocessed neuroimaging data and derivatives*
- [7] "Brain Network Image." *Nature Publishing Group*. Accessed 8 Nov. 2023.
- [8] M.A., Cherkassky, V.L., Buchweitz, A., Keller, T.A. and Mitchell, T.M., 2014. Identifying autism from neural representations of social interactions: neurocognitive markers of autism. *PloS one*, 9(12).
- [9] Dansbecker. "Rectified Linear Units (ReLU) in Deep Learning." *Kaggle*, Kaggle, 7 May 2018.
- [10] Maheshkar, Saurav. "What Is Cross Entropy Loss? A Tutorial with Code." *W&B*, 10 Sept. 2021

Roman Expansion and the Fall of The Roman Republic By Stella Anderson

The Roman Republic was one of Ancient Rome's most distinguished features and potentially the most influential to our modern world. They set a blueprint for how to create and maintain a republican government. The Roman Republic saw much political, social, and military success through territorial expansion, revolutionary innovations, and more. Nevertheless, eventually, the Republic collapsed and transformed into an empire. The causes of the fall of the Roman Republic have been widely debated by Classics scholars. Some argue that it was one event or person that initiated the collapse, while others theorize that certain inherent features of Rome's republic could not support an increasingly powerful Rome. However, at the heart of Rome's downfall was its very strength: expansion. Rome outgrew the republic both physically and politically. The Roman Republic fell gradually over the course of its last two hundred years because the people, particularly the aristocracy, no longer upheld the moral and ethical values binding the Republic together. This moral and political failure was due primarily to the unprecedented effects of Roman expansion: the luxuries that came with Roman expansion led to a corrupt aristocracy, ambitious individuals were able to acquire more *dignitas* through military conquests than ever conceived possible, and the institutions set in place were not equipped to manage such a vast territory. These consequences of Roman expansion intertwined with each other to create a violent, unstable Rome that eventually self-destructed.

At its very core, the Roman Republic was built upon the values of freedom, equality, virtue, and balance. During the Early Republic, citizens and aspiring politicians had a deep respect for these values and their laws: sacrosanctity was for the most part observed, no individual tried to seize absolute power, and each branch of government performed its assigned role. The Republic functioned with a great amount of discipline. There was an expected loyalty to the State, resulting in people putting the Republic first. However, sometime after 200 BC, this changed. The majority of historians have agreed that there was a turning point in which the respect for the rules and customs of the Republic faded away. There is some debate though around when this turning point was. Polybius and Livy, two of the oldest sources on this subject, argue that the fall of Carthage in 146 BC was the turning point. Other scholars such as Appian and Stockton point to the murder of Tiberius Gracchus. Further potential turning points include Sulla's first march on Rome, the emergence of the First Triumvirate, and Caesar crossing the Rubicon. Although satisfying, finding the precise date for the beginning of the fall of the Republic is not essential in finding the causes of this fall. Rather, it is essential to note that the turning point seems to have occurred as Roman expansion began to take off, sometime after 200 BC. Gradually, over this time period, as Rome expanded overseas, the principles of the *mos maiorum* that held the Republic together eroded away with the consequences of Roman expansion.

One of the first factors that contributed to the downfall of the Republic was the quantity of and indulgence in *luxuria* that came as a result of expansion. The Fall of Carthage in 146 BC, marked a major step in Roman expansion. Rome now controlled parts of North Africa, Spain,

Sardinia, and Corsica, while prior to the Punic wars, the extent of their expansion was contained to Italy. With this new territory, Rome acquired a great deal of resources including silver, iron, and gold from Spain. This influx of resources only added to the spoils the Roman elite obtained after Manlius Vulso's campaigns in the East in 187 BC. These new luxuries combined with the recent defeat of Rome's greatest enemy, led to an indulgent aristocracy. Polybius makes the argument that without the external threat of Carthage, Romans were now able to relax and indulge in ways that they couldn't before, thus making this the beginning of the collapse of the Republic. However, as Livy points out, this is not the first time that the Romans obtained this amount of luxury after a major defeat. In 187 BC, after Manlius Vulso's campaigns, Rome acquired great portions of wealth as Vulso's triumph in 187 demonstrated. The fall of Carthage was not the turning point, but rather an accelerated extension of the already impertinent, wealthy elite. Cato the Elder, who lived during the Punic Wars and witnessed the influx of *luxuria*, expresses his disgust for the elite's indulgent attitude. While he was fighting in Sicily with "great Scipio," as his quaestor, he denounced Scipio to his face for his outlandish, wasteful spending. (Plutarch, Life of Cato the Elder, 3). Cato argued that Scipio was corrupting his soldiers with his extravagant nature. In response, Scipio said he had no use for a grouchy, frugal quaestor, like Cato. Cato became so fed up with Scipio that he proceeded to leave Sicily mid-war. When Cato returned to Rome, he joined Fabius to condemn Scipio's "waste of enormous moneys" before the Senate. (Plutarch, Life of Cato the Elder, 3). It seems that excessively indulging in however much luxury one had was the norm in Roman culture at this point during the Punic wars, as Cato himself was considered an outlier for his frugal, virtuous actions. An incoherent and selfish upper class led to an ineffective and unequal Rome. The *luxuria* that Rome inherited from their expansion led to a corrupt, lenient, and greedy upper class that did not prioritize the *mos maiorum* values that fueled the Republic, leading to its demise. Furthermore, this greedy, power-hungry upper class spawned overly ambitious leaders such as Sulla, Marius, Pompey, and Caesar which further contributed to the collapse of the Republic.

The next consequence of expansion came out of the growingly powerful but corrupt upper class; individuals obtained an enormous amount of *dignitas* as a result of expansion that gave them too much power within the Republic. While fighting Carthage, Rome continued to expand East into the Hellenistic world. They were victorious in the Macedonian Wars, the Syrian War, the Numidian War, and more. The successes in these wars skyrocketed the careers of leading military generals and consuls in power during the time of the defeat, giving them an overwhelming amount of *dignitas*. *Dignitas*, translating to dignity, was acquired through noble, honorable, and virtuous acts, such as defeating an enemy or winning a war. Having *dignitas* boosted one's popularity and respect, helping them reach their next goal. Unfortunately for the Republic, the level of *dignitas* acquired through such expansive battles allowed certain generals or magistrates to defy laws and customs to fulfill their personal goals/ambitions. For example, during and after the Cimbric War in which Marius led Rome to victory, Marius was elected consul six years in a row. Considering consuls were only supposed to rule for one year at a time, six years is outlandish. Marius's *dignitas* due to his military successes in expanding and

defending Rome gave him an extreme amount of power that did not fit into the Republic as is shown in his six-year consulship term. Leaders like Marius, who broke the traditions of the Republic and tarnished the balance of power, set a precedent for future generations to do the same. Following Marius was Pompey the Great, who also had a level of *dignitas* capable of surpassing the laws of the Republic. Once Pompey the Great defeated the last of the Marius supporters for Sulla in North Africa, he demanded a triumph, a parade to honor victorious military leaders. Triumphs, however, were only granted to magistrates who held *imperium*. Not only did Pompey not have *imperium*, he wasn't even a magistrate. In fact, he wasn't even old enough to be the lowest-ranking magistrate, the quaestor. Despite all of these institutional restrictions, Sulla granted Pompey a triumph. The amount of *dignitas* and power that generals were able to obtain as a result of their Roman expansion efforts was too much for the Republic to handle.

Over time, the increased power and *dignitas* granted to individuals created a domino effect that tore apart the Republic. The actions of individuals such as Marius and Sulla set the precedent for the future generation. Slowly, from Marius to Sulla to Pompey to Caesar to Octavian, the amount of power one individual was able to retain within the Republic grew and snowballed until eventually, the power of one man got to such an extreme that there was no longer a Republic, but rather an empire. The Republic ended when one man, Augustus, retained power over the entirety of the Roman Empire. An overly ambitious atmosphere fostered a self-destructive society. This raises the question of how can a Republic exist in a world where ambitions are inevitable, yet destructive. What is the role of ambition in a Republic? Two of our world's most distinguished writers, Cicero and Machiavelli, both considered this topic. Cicero argued that an ambitious atmosphere was destructive and could not peacefully coexist with a Republican system. Machiavelli on the other hand critiqued Cicero for not understanding the complexities of the problem of ambition. He argued that ambition is necessary within a Republic and can be controlled with the right protections. The dual effects of ambition on a Republic, which Cicero and Machiavelli commented on, definitely impacted Ancient Rome.

Although an overly ambitious atmosphere contributed to the fall of the Roman Republic, it was also an essential quality that fueled different institutions essential to the Republic's success, such as the *cursus honorum*. Ambition was a quality that was inherent to Ancient Rome. The *cursus honorum*, for example, a ladder of magistrate positions, required ambitious men willing to work hard for Rome. The ambitions of the magistrates most likely stemmed from selfish goals of wanting more power and wealth, but in turn, their ambition helped Rome. An example of this mutually beneficial relationship between ambition and the Republic is the consular positions. The Roman Republic needed two consuls for the government to run smoothly, and every year there was an overwhelming number of men who wanted to be consul due to their personal ambitions. The Republic continues, while ambitious men accomplish their goals. The same idea applies to all the rankings on the *cursus honorum*. Ambitious men were eager to work hard for the Republic to achieve their personal goals. A Republic can coincide with the demands of individual ambition. The question then becomes: how much ambition can a

Republic handle? Ancient Rome provides an example of how too much ambition can lead to destruction.

The last of the three key consequences of Roman expansion was that the political institutions were not able to support an increasingly powerful Rome. The political structure of the Roman Republic relied primarily on three branches of government: the Citizen Assemblies, the Magistrates, and the Senate. This system stabilized the power in Rome, with each branch which weren't the most effective and led to extreme amounts of violence. The first way that Rome tried to compensate for the Republic's inability to control distant provinces was through the implementation of promagistrates. After the Punic wars, the Senate and Citizen Assemblies appointed promagistrates to rule parts of Spain and North Africa. Once one finished their term as either praetor or consul in Rome, they then had the opportunity to become a promagistrate in a farther province. Due to the geographical distance between Rome and promagistrate-run provinces, power amongst the promagistrates often went without restriction as they were far enough away from Rome for the Senate to intervene. Quintus Sertorius who was the governor of Spain is an example. He was sent out as governor during Marius' rule, so when Sulla took over Rome he refused to comply with Sulla's reforms. Because he was so distant from Rome he was able to run a sort of alternative Rome in Spain, basically running his own province without restriction of any other branch. After an initial victory against Sulla's army, Sertorius and his province were defeated by Pompey. The addition of the promagistrates added a new level of power that the Republic had to, but could not check. It became a breeding ground for overly ambitious, power-hungry rulers. In fact, Julius Caesar was the proconsul of Gaul before he became dictator of Rome. The implementation of promagistrates disrupted the power distribution in the Republic. Inherently, the institutions that the Republic was founded on did not support the level of expansion that Rome reached.

Expansion in the Ancient World was something the Romans did better than everyone else. Their brilliant military techniques combined with their unique ability to integrate diverse peoples into their society enabled them to become the dominant power in the Ancient World by 146 B.C. Their territory by the end of the Roman Republic spanned over three continents and contained thousands of different ethnic groups. However, Rome's greatest strength proved to be the Republic's biggest weakness. The consequences of expansion destroyed the Republic. The extreme amounts of resources and luxuries that came with the provinces Rome conquered, gave way to an indulgent, corrupt aristocracy that lost respect for the code of their ancestors, the morals of the Republic. In addition, the *dignitas* that military leaders and (pro)consuls were able to obtain as a result of expansion gave them an unprecedented amount of power that enabled them to defy the rules and customs of the Republic. Over time this fostered an overly ambitious society in which the power of the people was overcome by the power of the individual. The Republic was not built to support the increasingly powerful and expansive Rome. Ultimately, these three key factors led to the erosion of the core values of the Republic, which facilitated the collapse of Ancient Rome's Republic.

Works Cited

- Duff, Alexander S. "Republicanism and the Problem of Ambition: The Critique of Cicero in Machiavelli's Discourses." *The Journal of Politics*, 2011, <https://doi.org/41407594>. Accessed 19 Nov. 2023.
- Grainger, John D. "The Campaign of Cn. Manlius Vulso in Asia Minor." *Anatolian Studies*, vol. 45, 1995, pp. 23–42., doi:10.2307/3642912.
- HENDERSON, M. M. "TIBERIUS GRACCHUS AND THE FAILURE OF THE ROMAN REPUBLIC." *Theoria: A Journal of Social and Political Theory*, no. 31, 1968, pp. 51–64. *JSTOR*, <http://www.jstor.org/stable/41801828>. Accessed 19 Nov. 2023.
- Levick, Barbara. "Morals, Politics, and the Fall of the Roman Republic." *Greece & Rome*, vol. 29, no. 1, 1982, pp. 53–62. *JSTOR*, <http://www.jstor.org/stable/642930>. Accessed 19 Nov. 2023.
- Plutarch. "The Life of Cato the Elder." *The Parallel Lives*, Vol II ed., penelope.uchicago.edu/Thayer/E/Roman/Texts/Plutarch/Lives/Cato_Major*.html.
- "The Transformation of the Republic." *A Companion to the Roman Republic*, edited by Robert Morstein-Marx and Nathan Rosenstein, e-book ed., Blackwell, 2006.

Fighting Time: Artificial Intelligence and the New Era of Musical Expression By Kali Anchlia

Abstract

This paper explores the dynamics of and developing interaction between human creativity and artificial intelligence (AI) in the field of music. It mainly focuses on the emotional and psychological effects of music reception regarding AI-generated music. The study explores the emotional impact of music, concentrating on jazz as a special sub-genre characterized by improvisation, which provides insight into the cognitive abilities of AI. This analysis emphasizes lacking performance ability, along with the shortcomings of AI composition systems. However, the future potential of AI is also emphasized by the evolution of AI-technology in music. With the conversation situated within the context of western philosophical viewpoints, the hypothesis posits the growing importance for musicians to embrace AI, while still addressing the significant emotional gap that remains between listeners and AI-generated music.

1. Introduction

Imagine each individual life encircled within a sphere. Every person is the center of their own sphere. The things in this sphere are only that which we allow to enter. It is rare that we open the tiny door at the edge of our sphere and let something else in. Humans after all, at our very core, are survivors. Therefore, the dream of our world is carefully constructed in order for us to do exactly that: survive. When we listen to music however, the key that we have kept hidden finds its way into our hands, and we open the door before we realize what is happening. If music is our siren, then we are the sailors at sea. However, modern life is not ruled by myth, and our siren could be either human or technologically made, or even some combination of both. The question is then posed, can artificial intelligence produced music ever tempt us the way human touched music can?

This paper presents the relationship between AI and humans in the field of music, as well as the potential road to acceptance. It centers around the emotional and psychological aspects of the human mind and the place of music in this dynamic.

Our feelings no doubt play a large role in the way we receive the music that is communicated to us. Regardless of the genre, the overall effect of emotions is a powerful driving factor in how music is received by the listener. A sub-genre explored more in depth however is jazz music, as it includes an aspect not as apparent in other genres: improvisation. With an emphasis on the context of cognitive systems and the suitability of improvisation to AI, several perspectives are discussed. Along with this, the contrast of old and new AI systems and artificial intelligence's overall evolution, clarifies the hypothetical questions described above.

Based on the information that is presented, it will be clear that in order for AI-generated music to meet or surpass music created by humans, it must first evolve to develop cognitive thinking similar to humans, and meet emotional standards both in compositions and in live performances. Although it is currently lacking, AI has the technological potential to advance and

eventually generate music barely distinguishable from humans, at which point musicians must adapt to AI's presence in the field.

2. Emotional Connection, Human vs AI Generated Music

The strength of music is often underestimated. Music has the ability to close the gap between mind and body, providing an almost psychedelic experience. The purpose of music is to make others feel, and consequently, music is absolutely reliant on how well people are able to relate to and appreciate these combinations of sound. Music is an entity where other's opinions do matter, and there is always room for growth and development. Therefore in this section, we will touch upon to what extent AI versus human generated music can– or cannot– evoke emotion.

To lay the groundwork for further discussion, we must first consider the fact that the human brain is interested in changes in sound. We often cannot connect emotionally with only one constant frequency. When addressing the difficulty of 'modeling expressiveness' in music through AI, Ramon Lopez de Mantaras and Josep Lluís Arcos develop this necessity of variability by introducing and explaining the constantly firing auditory neurons in our brain (47). If auditory neurons repeatedly receive the same stimulus, their firing rate decreases over time: this process is known as *habituation*. In other words, without alterations in the music's pitch, rhythm, and dynamic, we perhaps unknowingly detach ourselves from the listening experience. It is simply no longer interesting enough for our mind to be actively following the music.

If avoiding such habituation is the case, then the basic assumption would be that only a distinct 'change in sound' is needed for artificial intelligence to match the emotional intelligence of human beings. However, we cannot forget to apply higher-level parameters based on music theory and machine learning, in order to prevent what is otherwise sure to be a cacophony of disruptant, horrifying notes. Although an evolution of AI systems and more specific parameters will be discussed later on, it is beneficial to introduce here the three blanketing techniques for incorporating AI algorithms into music as described by Robert Laidlow in his 2022 thesis (Laidlow). These techniques are not regarding the systems themselves, but rather what human musicians can do with already established assistance algorithms. However, it is important to note that these methods are more related in this case to classical and ensemble music than in a jazz-style approach. (The contrast between the two genres will be considered later)

1. Interlocking: acknowledges the limitations of AI, specifically its current inability to produce coherent music for lengthy durations, by alternating between AI-generated and human-created music.
2. Collaging: involves the combination of various layers of AI-generated music. Uses solid micro-level AI music, to piece together and formulate a final cohesive sound.
3. Hidden Layers: a blueprint of the music is formed solely by AI (allows the composer more creative possibilities). However, the individual elements of AI-generated music are not directly heard in the final piece.

The obvious interference here of humans in AI's own 'creative process' shows the limitations artificial intelligence has thus far when it comes to creating music, and expands the view from considering each note separately, to a more holistic perspective of the overall piece. What is more, the three techniques described above apply to ethical questions and the overall awareness of AI, and thus will be further explored. Still, Laidlow's assertion is narrow in the sense that it is written from the point of view of a composer, rather than a machine learning engineer. However, Laidlow's evaluation does convey the understanding human musicians have of feelings and emotions. The skill of placement (again referring to the three techniques) based on one's senses is something to consider as a powerful tool in music.

While one side of the relationship is the composer, the other is naturally the listener. After understanding not only the general process of our neural system when it comes to music, and techniques currently used by musicians to formulate music so that it fits these criteria, it is vital to consider the actual human to human connection and its relation to emotive responses. A feature by The Learning Network, prompted students from Ohio, Vancouver, Washington, Kentucky, and Washington, D.C., to submit writings on their thoughts on AI replacing artists in the music generation (Zvereff). In an almost unanimous agreement, they found it inconceivable that AI could replace the unique charisma and talent of live, human performers. The general consensus (although we cannot ignore that the sample size of the investigation was small) was that people like to know that they are listening to someone's music— someone being the key word here— who has gone through similar pains and obstacles as them. It is no surprise that a human finds comfort, hope, and excitement in relatability. Even when ignoring AI's impact on music, one can date as far back as the popularity of Kraftwerk in the 1970s, to represent this idea that new innovation can never really surpass the familiarity of the connection between the performer and the audience.

2.1. Jazz, AI Improvisational Systems

Jazz music is known for its expressive and authentic qualities, and here we must emphasize that the function of AI in jazz versus other genres must be different in order for the points addressed in the previous segment to be successful. There are several implications of accepting AI composed jazz music in the broader jazz community. Understanding cultural and societal ramifications, along with the preservation of jazz heritage is crucial. However, what must truly be perfected is the *art* of improvisation. Improvisation is more than vital to jazz: it defines the genre. Improvisation is both a reflection of individual artistry and a collective tradition passed down since the late nineteenth century.

Therefore, along with the history of music, there has been a subsequent history of what we classify as *Improvisation Systems*, which are related to the real time creation of jazz solos (44). The existing research synthesized by Mantaras and Arcos highlights chronologically some of these improvisation systems (which they state are one of three main subsystems of computer music in which AI techniques have been applied), along with their methods of functioning. For instance, the Flavors Band system attempted "random functions" paired with "musical

constraints” in order to generate variations in jazz (Mantaras and Arcos, 46). Its well known arrangement was an improvised solo of John Coltrane’s composition ‘Giant Steps.’ Noticeable was its use of procedural language to define pre-set rules for improvisation.

Another example is ~GenJam~ which interestingly uses something called a genetic algorithm to improvise (Mantaras and Arcos, 46). A genetic algorithm simulates the logic of Darwinian selection in the field of machine learning, where only the best attempts at improvisation are selected for replication and the rest are discarded. In the 1994 GenJam release, a human listener acts as a judge, and scores the improvisations of the system. Later iterations have also attempted the same method but with the use of automatic improvisation rating algorithms based on important melodic and harmonic qualities, rather than solely on the subjectivity of humans.

Employing neural networks was explored in 2001 with ~Franklin~, a system taught to improvise through transcriptions of human-solo improvisations. Most notably the saxophonist Sonny Rollins contributed to Franklin, both in solo recordings and with the overall style of the system after trial and error of imitation. In all of the three above examples, different algorithms in improvisation systems are used for the same ultimate goal: for AI to improvise on the spot, make something that sounds unique, and gives people the feels.

However, some criticize, “The lack of interactivity with a human improviser... on the grounds that they remove the musician from the physical and spontaneous creation of a melody” (Mantaras and Arcos, 46).

So we have a contradictory goal here. On one hand, we expect AI to fully improvise on its own, but on the other hand, we critique an AI system that does not include a human mind in the process. This duality brings to attention the varying degrees of acceptance prevalent in society (both in the past and in modern day) of AI in music. Thom notes how these expectations were met via the popular 2001 Band-Out-Of-A-Box (BoB), a “real-time” interactive system that uses an unsupervised probabilistic clustering algorithm (Mantaras and Arcos, 46). The BoB functions by learning all possible notes it could play at a given point in time, and then reacting to four-bar solos of human improvisers. The enhanced interaction between AI and humans did allow BoB to be a more successful machine than some previous attempts, especially because it adapted human musicians' intentions and artistic vision. These human-involved AI systems however, may defeat the point of improvisation altogether. If improvisation is defined as ‘creating and performing spontaneously and without preparation,’ then how could a system where a human composer is changing a generated solo until it becomes something that suits their style be classified as improvisation? We can also see a limit in authenticity within the jazz genre if AI is simply imitating the styles and chord progressions of human musicians it learns from.

2.1.a. Memories and Rules in Live Performance

One unanswered question is rooted in the challenge of live jazz performance, and the capability of AI in navigating between structure and improvisation in these scenarios. Ramalho and Ganascia claim that “Jazz musicians' activities are supported by two main knowledge

structures: memories and rules” (109). If we take this idea as the basis for everything that follows, more light will be shed on the specific mechanisms in which AI generates improvised solos. In one important way, artificial intelligence and humans are similar: both learn from past experiences. Let’s begin with ‘memory’ and the trigger point of memory, context. Musicians cannot always justify the actions they take during performances; several of them are actually the result of unconscious memories being activated mid-event. Most of the time these memories will appear as a result of the environment and piece one is performing, and alter the method of performance through “concepts such as tension, style, swing and contrast” (Ramalho and Ganascia, 109). As jazz is an intuitive and ever-developing genre, these concepts can only be mastered by observing, listening, and *imitating*. Whether this imitation is through watching videos of jazz legends performing, or playing in a band and bouncing ideas off of fellow members, storage of long term musical memory is important. AI seems to be learning in the same way as humans then, although it may be building up memory differently.

After looking at imitation from the perspective that it is simply a form of knowledge and learning, the benefits seem to outweigh the problem of authenticity posed before. Especially when we realize there are double standards when humans do the same thing. Now, this is not to say that imitation is equivalent to improvisation, and again, the ethics of taking one’s style as their own should still be debated. Regarding exactly how memory– imitation or not– is stored, we have a contrast between long term and short term memory systems. AI seems particularly adept at processing and communicating short term memory, but has difficulty with longer pieces of music. Typically its shortcomings include creating genuine emotional responses because of its limited interpretive abilities and the previously discussed repetitiveness. To combat these limits, Princeton computer science student Ji-Sung Kim developed his 2016 project ‘DeepJazz’, using a long-term short memory (LTSM) recurrent neural network (Weiner). Traditional RNNs (recurrent neural networks), work with short musical phrases, but LTSM’s do so with longer, more complex compositions. Ji-Sung created DeepJazz’s LTSM to be trained with only one piece, Pat Metheny’s “And Then I Knew.”, and subsequently generated piano solos. A chief technology officer and critic Ken Weiner, although acknowledging the use of the now popular LTSM as a good method, believes the reduction of the instrumentation to only piano is extremely simplistic (Weiner). He compares “fixed” notes of a piano with “malleable extended tones” of wind instruments such as trumpet, trombone, and saxophone, and believes that the complexity of vibrato and similar expressiveness is something AI cannot perform. David Borgo, a University of California, San Diego, music professor and saxophonist adds onto the concern: “Research in this area has tended to focus on getting computers to play the ‘right notes,’ but we are still a long way from designing systems capable of the micro and macro temporal, timbral and textural adjustments necessary to groove together and to develop high-level collective improvisation in an unscripted fashion with human musicians (rather than insisting that human musicians improvise with, or groove to, the computer)”. Clearly, progress still needs to be made in AI memory, especially when Jazz music is known for its *soul* and ability to create and control the mood in a setting (when played live).

Of course, feeling is only half of it; the other half is ‘rules’, which is much better suited to artificial intelligence in its current state. These rules consist of commonly agreed upon music theory, and some unavoidable incomplete thoughts (in which case it would rely upon memory to decide what exactly to do). Although jazz is seen as more of a free form of music, theory and understanding of the way notes act in relation to each other is extremely important. The most basic definitions of musical concepts integrated in AI would be on “pitch, duration, amplitude... temporal sequence of notes... set of simultaneous notes... Scales and rhythm... melodies and/or chords” (Ramalho and Ganascia, 108). The application of music theory can be ultimately divided into 4 categories; melodic analysis, harmonic analysis, rhythmic analysis, and structural analysis (Torrance and Schumann, 257). Shih-Lun Wu and Yi-Hsuan Yang present their ‘Jazz Transformer’ as a neural sequence model that analyzes these categories.

1. Melodic Analysis: melodic phrases and patterns (pitch range, note durations, intervals between notes)
2. Harmonic Analysis: chord progressions and harmonic structure (jazz chords, harmonic tension/release)
3. Rhythmic Analysis: rhythmic patterns and complexities (beat patterns, syncopation, swing)
4. Structural Analysis: overall form and organization (traditional jazz song structures like AABA, blues form, other common forms)

Although different machines integrate music theory into their algorithms in different ways, they are always thoroughly encoded, because without understanding of sound, AI would not be able to generate anything of substance— memory or not. The quantitative measures of the Jazz Transformer is a perfect example of this, and the conclusion Wu and Yang found correlates to previous knowledge of AI’s ability, as well as adding that it lacks new styles, techniques, and cultural influences. The challenge is giving machines intrinsic motivation and the ability to make creative decisions based on real-time feedback; this is called autonomous creativity (Baxt).

“If we produce sounds with the intent to make music, that is music,” said Shlomo Dubnov. “That means that, to become musical, the computer must have its own intent.”

French scientist and composer Francois Pachet promotes PACT’s as a generic framework for representing the potential actions (or intentions) that musicians may take within the context of performance (Ramalho and Ganascia, 109). PACT’s consist of a combination of low level and high level instructions. According to the authors of this idea, examples of low-level PACTs include "play loud," "play this rhythm," and "play an ascending arpeggio,". A low-level PACT on all three dimensions could be, “play this lick transposed one step higher” (Ramalho and Ganascia, 109). High-level PACT’s include "play syncopated" (on rhythm) and "use major scale" (on pitches) (Ramalho and Ganascia, 109). These can be activated either during a specific moment, or encapsulate the entire song; its flexibility is the reason it is considered by some, a basis of problem solving for musical creativity modeling. The AI relies on the playability of these PACT’s however, and because of its varying dimensions, the less abstract a PACT is, the more it is ‘playable’. For example, “play funk-rock” is less playable than “play an augmented

fifth on the third beat”. There are also restrictions based on combinability of PACT’s; a simple example is that “play loudly” and “play quietly” cannot be combined (Ramalho and Ganascia, 110). This potential actions structure doesn’t seem to utilize a musician’s (AI’s) intentions described earlier, however the perception, composing, and execution modules of PACT’s do lean towards it more than other systems. The three elements they rely on are short term memory contents, the context and mood, and a future chord grid segment (Ramalho and Ganascia, 110, 111). The simplicity is ingenious; first the perception module corresponds to the context and mood, and puts these events into the AI’s short-term memory (STM). Next, the composing module uses the context stored in the STM to combine playable PACT’s, and sends this information to the ‘future’ chord grid segment (similar to lead sheets). Finally, the execution module takes the information from the chord grid and executes the PACT’s at the correct time by sending information back to the perception module. In this way, the AI is able to develop some sort of musical intention, although it is still relying solely on the context fed to them. Rather than focusing on rules and memory separately, PACT’s combines both in an effective way. If this idea is perfected, AI already comes a lot closer to becoming the self-thinking and feeling improvisation systems we doubt they can be.

The Defense Advanced Research Projects Agency (DARPA) hopes to advance the progress of AI in music even faster, and are conducting research on jazz-playing robots that can interact and improvise with human musicians (Weiner). DARPA hopes to blur the lines between man and machine, by studying and implementing nuances of jazz music. If they succeed, we truly won’t be able to differentiate much between artificial intelligence and humans (at least in terms of live performance); however the physical aspect of this is very interesting. In 1970, a Japanese roboticist Masahiro Mori first proposed the *Uncanny Valley Theory* (Hillier and Isaacson). Based on collected data, Mori reached a conclusion that as robots appear more human-like, they become more appealing to humans— that is until this likeness reaches a figurative ‘uncanny valley’. At a certain point, the feelings of commonality and emotional connection devolve into a sense of unease. It’s when the robot becomes too unnaturally human-like, that we begin to feel frightened. The uncanny valley brings to attention the difficulty of accepting AI if it even reaches a point where it can look, act, and play live music exactly the same as humans.

2.1.b. Cognitive Systems and Philosophies

To add upon jazz and improvisation in live settings, we must explore cognitive systems and the psychology of the human mind, along with its application to artificial intelligence. Francisco Varela and Umberto Manturana propose an *autopoietic perspective*, which centers around the idea that living organisms are self-creating and self-sustaining (Torrance and Schumann, 253). In the previous segment, much of the ‘memory’ based discussion focused on external adaptation and remembrance depending on various contexts. However, the autopoietic perspective shifts the focus to internal coherence, and emphasizes the continuous self-renewal and unity of organisms (Torrance and Schumann, 253). With this shift in focus there is a whole

new dimension of thinking. Previously we could only look at the surface level impact of memory, rules, and context, but now we delve into mind-body connection and awareness. Steve Torrance and Frank Schumann expand upon the autopoietic perspective, and umbrella into the idea of Enactivism, which is defined by two key concepts. The first is "Laying down a path in walking", initially introduced by Varela in the conclusion of her 1991 work (Torrance and Schumann, 253). She proposes that organisms create their world as they interact with it, rather than passively mirroring it. To relate this concept to music, musicians generate music in the moment, and cannot take back the actions they have already performed. Varela says this is akin to laying down a path without the ability to correct; in other words, AI generates its own world with its perceptions and history of learnings, rather than just completely copying others. Therefore when AI makes a musical mistake, it is proof that it is attempting at being the 'self sustaining' entity that is the very definition of the autopoietic perspective. Varela quoted a poem in her work by Antonio Machado, a Spanish poet, and the last line according to Torrance and Schumann captures the essence of jazz improvisation and performance in general, "when turning around you see the road you'll never step on again". Following this, the second concept is "Sense-making", which was also first written in a paper by Varela in the early 1980s. "Sense-making" suggests that the nervous system creates meaning rather than processes information. She applies this to the interaction between an organism and its environment, relating it to musicians generating meaningful musical responses in real-time (Torrance and Schumann, 253). If the nervous system creates meaning, then AI has to be able to observe the environment it is performing in, rather than only focus on its own system and musical interests. Selfishness usually never makes a good performance; knowing what one wants to make the audience feel however, certainly gives intention to your instrument. Intention can also be communicated between fellow band members. Enactivism is a valuable theory, because we can attempt to look at things from the perspective of AI, and see how it may cross the line between technology and humans.

Since jazz involves temporal awareness, it often lends itself to be an extremely physical and emotionally charged performance. It's this energy and spontaneity that allows the audience to forget about effort required to get to a point where improvisation comes easily to the tips of their fingers. By definition, improvise actually originates from the latin 'improvisus' meaning 'unforeseen' (Torrance and Schumann, 256). Although the music should be unforeseen to the audience listening to the music for the first time, it is very rarely completely unforeseen to the performer. Based on evidence, we already know that experience of some sort is required for musicians and artificial intelligence alike, but we haven't yet truly discussed practice. Even a highly skilled musician— or rather *especially* a highly skilled musician— will tell you that practice is key; jazz or classical, it doesn't matter.

"... it is a mistake to link the "improvised" too closely with the "unprepared." In fact a considerable amount of improvised performance is extensively planned in advance—and much jazz performance includes, or is indeed mainly comprised of, playing from pre-written sheet music. Even in most contemporary jazz performance, where musicians may improvise singly or

collectively over extended stretches, there will usually be pre-arranged structural constraints” (Torrance and Schumann, 256).

In jazz, the line between preparation and performance is blurred, because the musician always learns something new about the way to approach a piece when performing. A dialogue between Herbie Hancock and Wayne Shorter (well known jazz- musicians) actually underscores the notion that improvisation in jazz requires “constant, unseen effort” to achieve true brilliance (Torrance and Schumann, 254). In this case, AI’s failures and attempts to improvise are only part of its “constant, unseen effort”, and will eventually (on its journey to development) lay the correct path and become a successful self-creating ‘organism’.

If, when applying the previously described ‘sense-making’ AI is creating meaning in real time, is it doing so consciously? If so, then to what extent of consciousness is required for AI to be on par with human musicians? We now discuss mindlessness versus mindfulness in order to understand AI’s (and human musicians) expertise development. Looking at American philosopher Hubert Dreyfus’s texts as a general starting point allows for an introduction to these questions, and acts as a basis for establishing complex views. Typically, there are two cognitive systems (with no designated names) which philosophers commonly debate. The first corresponds to Dreyfus’s notion of “immediate coping” where he argues that organisms operate automatically and pre-consciously (Torrance and Schumann, 261). In other words, he says that the expertise development of a musician is a state of unconscious thought, because in his words, “thinking disrupts smooth coping,” (Bergamin 406). He explains the idea with his famous interpretation of why Baseballer Chuck Knoblauch kept throwing errors (Bergamin 406). The second system emphasizes the early stages of development, where organisms act slower, more reflective, and rely on method-based thinking. With this, rational mindedness combats the supposed effortlessness of the actions described in System 1. Tensions arise because of the blatant contrast between Systems 1 and 2.

The definition of expertise to humans is subjective, because to someone who doesn’t play an instrument, expertise could be a few nice sounding chords. However, when musicians constantly judge and push themselves (and others) in order to reach the ‘next level’ and become ‘better’, only they know how high their goal exactly is. Dreyfus defines everyday expertise in this case, not as technique or skill, but as how well each individual meets their own needs. He mentions ‘expert flow in action’, or in other words, full engagement in one’s actions without overthinking it. This suggests that mindfulness is present in their actions, but detached contemplation is not involved. A common example would be making coffee everyday for years; at some point the experience– in Dreyfus’s argument– leads to expert motions without the requirement of explicit thought (Bergamin 405). Some people may not add an ideal amount of cream or sugar, but it suits their needs perfectly. Now, if ‘expertise’ truly is what Dreyfus argues, AI in music is nowhere near reaching the expertise level of humans. Since the model of the AI is its brain, depending on the kind of model the AI is being fed, it could involve thinking in the moment, or already have all the instructions ahead of time. Either way, this level of ‘thought’ in the action does not correspond to Dreyfus’s argument. Some similarities could be seen, because

if the AI has a genetic system (which was explored earlier on), only the good outcomes would sustain, and so the AI's system may have less qualms to reproduce those outcomes (which again could be seen as "less thought"). Still in the end, AI has no subconscious, so it's impossible for it to have 'mindful actions' without actually thinking. Philosopher John McDowell defends his claim that rational mindedness is 'pervasive in our lives' by arguing that Dreyfus's idea is a myth at its very core (Schear 2). McDowell specifically focuses on Dreyfus's case of a chess master absorbed in lightning chess. Joseph K. Schear explains McDowell's argument well in the introduction of his novel *Mind, Reason, and Being-In-The-World*:

"McDowell urges that the chess master's absorption does not prevent him from knowing what he is up to, and that, moreover, if the chess master really is a master, he will be able to give rational explanations of his moves as intelligible responses to the forces on the board. So understood, the chess master's expert play is a case of "cultivated rationality" in operation. This is precisely the "actualization" of conceptual capacities in experience and action that McDowell is keen to highlight".

In terms of humans, philosophers believe McDowell went too far by saying that there is no slow, reflective thinking (as described in *system 2*), only completely rational thought without actively using deep thinking. Although we previously disputed Dreyfus's argument, there is a truth to the connection he made between overthinking and disruption of an action. Stage fright is common for performers— however, experienced musicians can overcome this to some extent by becoming more absorbed in the music than focusing on their heightened awareness of their surroundings (which will no doubt be a crippling distraction). American Philosopher Taylor Carman takes a similar view, regarding McDowell's theory as 'scholastic fallacy' (Schear 12). He believes that McDowell is attempting to fit structured and rational thinking into experiences that are supposed to be immediate and not deeply thought out. In other words, he is trying to use complex, organized thinking where it doesn't belong, which again is similar to our previous thoughts. Having rational contemplation as McDowell argues, may prove disastrous to a human musician then. However, looking at this in terms of artificial intelligence, if McDowell's theory is correct, AI is (in contrast to Dreyfus's theory) already an expert. AI does think rationally after all. Still, the whole point of this section of the paper is to discuss AI's ability to express emotions and make listeners *feel* the same way they do when human musicians perform. Based on the hypothesis that emotions, and therefore understanding of both the context and the piece beyond the surface level of rationality is truly vital, we can discard McDowell's theory momentarily.

Barbara Montero, a professional ballet dancer turned philosopher takes another view on Dreyfus's theory in her "A dancer reflects," (Schear 16). First, she dismisses Dreyfus's principle of automaticity, by arguing that there is not enough psychological evidence to support the theory. She mentions that action-in-the-flow is not necessarily mindless, and that if it were it would conflict with scenarios where experts may need to reflect or monitor their actions. This is similar to classical musicians who often view mindless performance in-the-moment as a mistake. Musicians like cellist Inbal Segev believe that eliminating thought during performance eliminates the artist's ongoing conception and image of the music (Torrance and Schumann,

263). Of course, classical music in comparison to jazz music will be discussed further later on. Furthermore she diverts from Dreyfus's theory when she categorizes 'esoteric expertise' separate from 'everyday experience' (Bergamin 403). Esoteric expertise refers to specialized knowledge or experience in a certain thing: such as performance in a particular instrument. Joshua A. Bergamin acknowledges this in *Being-in-the-flow: expert coping as beyond both thought and automaticity*:

“But for Montero, esoteric expertise is marked by a continuous desire to improve and she insists that this highly-motivated attitude— called kaizen in Japanese— requires explicit cognitive attention if it is to succeed (Bergamin, 407)”.

Therefore once again the idea that thought and 'explicit cognitive attention' is important surfaces. However, there is also another idea here: for AI to be an esoteric expert, it must have the desire to improve. In revisiting both the autopoietic perspective and autonomous creativity (2.1a), for AI to be self-sustaining, it must also have these human-like attributes of wanting to evolve. This is fundamentally not possible for AI currently, as it is the person who has created the model who is thinking, and pushing the AI to improve. The AI is a clone of the engineer, not its own individual— at least not yet. Still, if we take Montero's theory and apply it here to AI musicians, there is a significant way to gain esoteric expertise. It also relates to the previous idea of a 'constant, unseen effort', because one would only put in the needed effort should they have the desire to (or in the terms of artificial intelligence are being programmed to). Comparing humans and AI in this manner doesn't help however, because we cannot expect a machine to replicate the emotions of humans. However, in terms of the acceptance of AI musicians in society, we can see that there will be a struggle to consider a machine as an expert when it cannot truly be classified as its own individual. Although there are several more philosophers such as Gallagher, Toner, and Moran, who contribute immensely to this discussion, the debate on expertise and mindfulness is helpful in understanding several perspectives on how human/AI performers (for jazz or otherwise), think and act in the moment.

2.1.c. Classical Music vs Jazz Music

Looking at the difference between classical and jazz music allows for a continuity in the philosophies explored in the last segment, and of course, brings new ideas about the relative success of both genres of AI in music. First, there is a clear distinction between jazz and classical music. Jazz thrives on imperfection, with musicians embracing and owning their "mistakes" to create artistic expression. In contrast, classical musicians are known for attention to technique in order to attain a performance as close to perfection as possible. Due to the expectation that classical musicians replicate a score to technical perfection, they are described by some as 'robots' (Pellegrin). The fact that classical musicians are even ranked by objective ability and seated as such in an orchestra, only adds to this conclusion. However, this is not to say that classical music has less feeling involved. Though the restrictions on how the music of the time period and written style by the specific composer, only make it more difficult to express emotions in phrases, a good classical musician is able to make every perfected note contain

beautiful emotion. However, the sound each genre produces is different— jazz is freeing in a way that classical music can't always be. It lies less so in the score, but more in the soul of the performance itself. Along with this, improvisation is rarely ever in the equation when it comes to classical performance. Most classical musicians in the 20th century cannot improvise at all. These boundaries between jazz and classical have resulted in segregated studies for both genres of music, and similarly require different types of AI systems to replicate/create (Pellegrin).

We see this difference with AIVA, an AI program which created a 2016 album “Genesis”, and drew from classical master composers such as Beethoven and Mozart to create its own music. The piece “Symphonic Fantasy in A minor, Op.21” specifically drew attention as an impressive creation, because of its use of dynamics and intensity (Hillier and Isaacson). The AIVA team gathered opinions through *Turing Tests*, in which humans try to distinguish whether the AI generated music is distinguishable from human compositions. As Laidlow (2.0) called it, a “musical guessing game” of sorts (Laidlow). They found that the majority were not able to tell there was a difference, and that the AI took the styles of the composers it had samples from and created something new and profound. However, upon listening to the recordings, they also found that it still lacked the ‘human touch’, and tended to be repetitive. Although motifs are often used by composers, the piece often seemed like it was meant to be leading somewhere, but still ended up back at ground zero. Classically trained musician Hoh Xi Ting, compared the Symphonic Fantasy and another composition “Scottish Fantasy” by Max Bruch (1800). She explains that they both begin as “a slow, mellow, low brass melody, later joined by a violin solo”, but the big difference is that “Humans can feel, and their emotions are reflected in the composition as themes, harmonies, and the occasional, but well-placed discordant chord (Ting). There is direction, tension building up to a climax, and a satisfying finale {in the human composed Scottish Fantasy}”. The imperfection shown in these Turing Tests and the comparison above illustrate— as James Vincent, a senior reporter who has covered AI, robotics, and more for eight years at The Verge writes— that AI may be better suited for less structured music. In this case, jazz would prove to be an extremely suitable genre for AI composition, as its lack of rigid structure aligns with the short-sightedness of deep learning systems (2.1a) (Vincent).

In order to test this conclusion and attempt to combine the best of both worlds, classical and jazz, together, Professor Geraint Wiggins of Queen Mary's University introduces “the first concert consisting almost entirely of music composed by artificial intelligence” (Wiggins) (Vincent). The first AI-written song was a jazz combo led by Mark d’Inverno— and it proved to be a remarkable success that “sounded just like the real thing”. However, the team still hadn’t yet experimented with a less structured classical piece. In order to make it more likable and easier to comprehend to the audience, the music was presented in the style of a familiar composer: more specifically, Mozart (with a jazz spin). The two pieces tested by Stefan Lattner were *Mozart Unchained* and *Mozart Constrained* (Vincent). In *Mozart Unchained*, the professor gave the deep learning system little guidance, and allowed the AI more freedom to explore and create. The result was music that sounded less recognizable as classical music, with a melody that was “all over the place”. *Mozart Constrained* on the other hand, limited the system’s ability to generate

melodies outside of the most statistically likely patterns. This attempt was certainly more coherent, but seemed to be nearly imitating whole 20 second phrases from Mozart's own work. With the direct comparison of live performances of jazz and two structures of classical music, we can draw the conclusion that AI makes better independent judgments in jazz music, but these judgments are out of place in less structured classical music. We can further connect the results of Mozart Constrained back to Montero's question of whether AI is able to function on its own and understand the image of a piece. In this case, the answer is no once again, because it was Lattner's thoughts and limitations that allowed the constrained version to perform well, not the AI's. However, we can see that AI is suitable in its own way to both genres, and as we will learn in the next segment, it has the potential to overcome its obstacles and reach a point where both its rational and irrational 'thinking' could be useful.

2.2. Micro and Macro Evolution of AI

The AIVA team contradicted their statement in the previous section that 'Genesis' lacked a human touch, by stating that times are changing, and that people have been getting used to AI generated music in the past 20 years (Hillier and Isaacson). AI 's evolution in the past few years is certainly characterized by its increased use to people in the macro-world, but can also be analyzed in the micro-world of each new individual musical piece.

A recent Pitchfork article by Marc Hogan is titled, "Musicians Are Already Using AI More Often Than We Think". The article details Grammy-winning engineer and producer Shawn Everett, experimenting with Open AI's song creation tool. Everett input a chord progression from an unreleased Killers song, and requested from the AI a piece in Pink Floyd's style. He found that the song was unique and a little strange, but overall he was surprised because he thought a human couldn't have done something as creative with the instructions as the AI did (Hogan). Another credited artist Danny Wolf, an Atlanta-based producer, actively integrates AI into his work. Wolf (as of May 2023) has been collaborating with ChatGPT to generate the concept for his entire solo album. He even mentions that he frequently uses AI to create symphonies, which he then breaks down using another AI program to remove vocals from beats, and then incorporates them in his work (Hogan).

Peter Martin, an Oscar nominated film producer "... recalls how it would have taken 11 hours of a voice for an AI to be able to mimic it just three years ago. "That is now down to less than two minutes," he says" (Hogan). Commenting on the recent Drake AI remake, Martin explains: "Generating a fake Drake song might involve four or five different AI tools now... But he's beginning to see one-stop shops like Uberduck" (Hogan).

The micro-implications however more so affect development within a single piece. We will be revisiting the ideas of Collaging and Hidden Layers, as well as the performance of additional AI-generative platforms (2.0) to see this. A 2019 composition for mezzo-soprano and ensemble called "Alter" uses several algorithms which serve different purposes: MuseNet as a primary symbolic-generative algorithm, WaveNet as an audio-generative algorithm, and a combination of a text-RNN and 'GPT-2' as a text generative algorithm (Laidlow). The

composition is divided into three narrative-driven sections, and is defined by early, middle, and end points in the AI's development. In the early AI development of "Alter", the AI algorithm was simpler and less refined, and the music/lyrics were naive. Here 'MuseNet' generations explore repetition as a hidden layer; starting from small-scale repeats, then evolving as the narrative develops, into larger, structural repetition. At the same time, 'WaveNet' is used to create evolving electronic soundscape elements of vocalizations, speech, and singing, that mirror the narrative development of "Alter". The AI becomes increasingly self aware in the middle section of the piece, and gains more maturity according to Laidlow (Laidlow). At the end of "Alter", each instrument or voice performs its own line, creating a collage effect. By this point, AI-generated material (in this case MuseNet), completely takes over the decision making process, and creates more intricate musical textures.

If AI has the ability to learn to go through the above process on its own, it may be able to reach the level of human emotional understanding we expect it to. Both in the macro and micro world of AI-generated music, AI is taking a firmer stance and each new system is becoming a stronger version of its previous self. The truth is that AI is evolving at a faster rate because humans are looking for faster ways to produce music. The ethics of AI use in music must be touched upon in the future in order to fully understand the risks and benefits of AI-generated music.

3. Conclusion

In conclusion, this research has illuminated the complex relationship between AI-generated music, and human created music. As we have reviewed, AI-generated music, although sometimes revolutionary for the time period, is insubstantial in comparison to the music human musicians produce.

AI's largest problem is regarding the low level of emotional capacity within its music. Its inability to create music that makes a significant emotional impact on listeners is no doubt a great hindrance to its advancement. Not only this, but AI is also unable to create a lasting connection with the audience, because it cannot interact with fellow humans the way living-performers can. As discussed before, AI is completely different from humans, and as such this comparison may not be appropriate. However, because music is almost wholly ruled by emotions, AI must be able to produce emotions in music to some extent.

We can see this need further in jazz music, where AI has a lot of trouble in creating the energy and atmosphere required of the genre. AI is unable to improvise without some level of imitation, and is repetitive to a fault due to limitations in its systems long term memory. In terms of the question of improvisation, we looked over arguments concerning how AI could be suitable– or unsuitable in terms of spontaneity. Here some positive light was shed on AI-generated music being 'unpredictable' in a surprisingly creative manner. However the counterargument to this is AI simply being 'out of touch' with music-theory in live circumstances. Then again it can be argued that making mistakes is important for AI to be considered an evolving 'organism'.

It is obvious that several questions are left unanswered: Is improvisation meant to be completely in-the-moment, or slightly practiced and structured? Is intrinsic motivation truly possible for AI? Is expertise described as detached or contemplative? Is the uncanny-valley theory pervasive or could AI's advancement lead to growing acceptance in the new generation? AI's definition in relation to music, and the way scientists and composers view its evolution will affect future research on this subject matter, and perhaps bring to attention new claims that could harm or help the case of AI- generated music.

As we contemplate the significance of AI in music, it becomes evident that understanding its role is not just a matter of technological advancement but also a profound exploration of human versus machine in terms of emotion and creativity. AI's advancement in music could lead to a future of increased collaboration with human musicians, or complete independence once it reaches appropriate cognitive ability (in terms of humans subjective critique). In essence, the acceptance of AI in music means redefining our spheres of creativity and embracing new, unconventional, artistic expression.

4. Acknowledgements

PHD Rob Shafer, dissertation mentor and guide

Works Cited

- Baxt, Josh. "Computers in a Jazz Ensemble? Inventing Improvisational AI." UC IT Blog, 23 February 2022, <https://cio.ucop.edu/computers-in-a-jazz-ensemble-inventing-improvisational-ai/>. Accessed 15 October 2023.
- Bergamin, Joshua A. 'Being-in-the-Flow: Expert Coping as beyond Both Thought and Automaticity'. *Phenomenology and the Cognitive Sciences*, vol. 16, no. 3, Springer Science and Business Media LLC, July 2017, pp. 403–424, <https://doi.org/10.1007/s11097-016-9463-1>.
- De Mantaras, R. L., and J. L. Arcos. "AI and Music: From Composition to Expressive Performance". *AI Magazine*, vol. 23, no. 3, Sept. 2002, p. 43, <https://doi.org/10.1609/aimag.v23i3.1656>.
- Hillier, Lejaren, and Leonard Isaacson. "could ai music affect jazz?" New York Jazz Workshop, n.d., <https://newyorkjazzworkshop.com/could-ai-generated-music-affect-jazz/>. Accessed 18 November 2023.
- Laidlow, Robert. "Artificial Intelligence Within the Creative Process of Contemporary Classical Music". Diss. The Royal Northern College of Music in collaboration with Manchester Metropolitan University, 2023.
- Pellegrin, Rich. "Why improvisation is the future in an AI-dominated world." *The Conversation*, 5 October 2021, <https://theconversation.com/why-improvisation-is-the-future-in-an-ai-dominated-world-166724>. Accessed 15 October 2023.
- Ramalho, Geber., and Ganascia, Jean-Gabriel. "Stimulating Creativity in Jazz Performance." *AAAI-94*, Aug. 1994, pp. 108-113, https://www.researchgate.net/profile/Jean-Gabriel-Ganascia/publication/2671507_Simulating_Creativity_in_Jazz_Performance/links/02e7e51867c15dc574000000/Simulating-Creativity-in-Jazz-Performance.pdf
- Schear, Joseph K. *Mind, Reason, and Being-in-the-World: The McDowell-Dreyfus Debate*. Routledge, 2013.
- Ting, Hoh Xi. "Scottish Fantasy' and AIVA: Machine Learning in Classical Music." *Medium*, 16 December 2022, <https://medium.com/@kyrinstitute/scottish-fantasy-and-aiva-machine-learning-in-classical-music-8a3b78936418>. Accessed 15 October 2023.
- Torrance, Steve, and Schumann, Frank., "The Spur of the Moment: What Jazz Improvisation Tells Cognitive Science". *AI & Society*, vol. 34, no. 2, Springer Science and Business Media LLC, June 2019, pp. 251–268, <https://doi.org/10.1007/s00146-018-0838-4>.
- Vincent, James. "A night at the AI jazz club." *The Verge*, 12 October 2016, <https://www.theverge.com/2016/10/12/13247686/ai-music-composition-jazz-club-london-deep-learning>. Accessed 15 October 2023.

Weiner, Ken. "Machines Can Create Art, but Can They Jam?" Scientific American Blog, 29 April 2019,
<https://blogs.scientificamerican.com/observations/machines-can-create-art-but-can-they-jam/>. Accessed 15 October 2023.

Zvereff, Daniel. "What Students Are Saying About A.I.-Generated Music." The New York Times, 11 May 2023,
<https://www.nytimes.com/2023/05/11/learning/what-students-are-saying-about-ai-generated-music.html>. Accessed 15 October 2023.

UFT: A Review of Theoretical Physics By Ravi Shah

Abstract

The greatest challenge facing the world of theoretical physics is unifying gravity with the other three fundamental forces. Three of the four fundamental forces of physics are described within the framework of quantum mechanics and quantum field theory. The current understanding of the fourth force, gravity, is based on Einstein's general theory of relativity, which is formulated within an entirely different framework. It states that objects at the largest scales bend and warp space-time. The problem with general relativity is that it breaks down in the quantum realm. This signals the need for a new theory. A theory which describes gravity at the smallest scales. Some worked-on-theories include string theory and loop quantum gravity. While not all of these theories try to reconcile every fundamental force, some like string theory try to unify every fundamental framework. This review paper's intention is to review important concepts, technologies, and how they affect society in theoretical physics such as relativity, quantum mechanics, quantum field theories, quantum gravity, and ultimately address the unified field theory (UFT).

1: Relativity

Maxwell's equations are a set of four differential equations that James Maxwell published in 1865. They describe how magnetic and electric fields are created by currents, charges, and changes in the fields. They also describe vector fields. In order to know everything about what is happening in a vector field, you must know its two properties: its divergence and its curl. Maxwell's four equations describe the divergence of the electric field, the divergence of the magnetic field, the curl of the electric field, and the curl of the magnetic field. When you combine these four equations, you end up with a wave equation. The wave equation for an electromagnetic wave, light. A wave equation gives you the speed of that wave. When you calculate this specific wave equation, you end up with the speed of light.

However, there is an issue. Let's say you were measuring the speed of an electromagnetic field on a train, and there was another guy measuring the speed of that same electromagnetic field outside the train. You and the guy outside the train should have measured the same speed for the electromagnetic wave you guys were measuring. This tells us that the speed of light should be the same in all reference frames. Which totally didn't make sense to the scientists at the time. At the time, we had a very intuitive relativity known as Galilean relativity. Which basically says that if I was in a train moving at 100 mph and there was another train moving at 90 mph, I would see my train move 10 mph faster than the other one. So you could see why it made no sense that the speed of light was constant in every reference frame. Scientists attempted to come up with solutions to this problem. One of these solutions was called the aether. The aether was said to be the medium through which light travels and is able to move and propagate through space. So, people thought that that was what caused the speed of light to be the same at all reference points. Then people went looking for this aether. But what came up

was considered the most famous null result of an experiment. It was called the Michelson-Morley experiment. It was designed to measure the extra time it took a light beam to travel there-and-back against the aether wind, compared to a light beam traveling sideways across the aether wind. They ended up being the same, which disproved the aether.

Now this is where our friend, Mr. Einstein, comes in. Einstein theorizes that something else is actually going on. He creates the theory of relativity. Relativity solves the contradiction between Galilean relativity and Maxwell's equations.

1.1: Special Relativity

Albert Einstein presented the theory of special relativity in 1905. This theory laid the foundation for a new understanding of spacetime and provided a resolution to the apparent contradictions in the behavior of light and the constancy of its speed.

Einstein proposed that space and time are interconnected as a four-dimensional continuum known as spacetime. Events in the universe are described by their coordinates in spacetime, which comprise three spatial dimensions (length, width, and height) and one temporal dimension (time).

Einstein's theory is based on two postulates: the principle of relativity and the constancy of the speed of light. The principle of relativity states that the laws of physics are the same for all inertial (non-accelerating) observers. The constancy of the speed of light means that the speed of light in a vacuum is the same for all observers, regardless of their relative motion.

According to special relativity, time is relative and can dilate or stretch depending on the relative motion between observers. When two observers are moving relative to each other at a significant fraction of the speed of light, they will measure different time intervals for the same event. The faster an object or observer moves, the more time appears to slow down for them compared to a stationary observer. As an object's speed approaches the speed of light, time dilation becomes more apparent.

This is shown in the famous Hafele-Keating experiment. Joseph Hafele, a physicist, and Richard Keating, an astronomer took four atomic clocks aboard commercial airlines. By flying twice around the globe, first eastward and then westward, they were able to compare their clocks to those still at the United States Naval Observatory. Ultimately, the clocks were discovered to disagree with each other when they were put back together. Although the disagreements in the clock readings were relatively small (around ten nanoseconds), they were significant in proving time dilation on our own planet.

Further evidence of time dilation can be demonstrated by the behavior of muons. Muons are particles with short lifetimes. When muons are emitted from the sun and travel at relativistic speeds towards the Earth's surface, they should decay relatively quickly according to their lifetime of 2.197 millionths of a second. However, due to time dilation, a significant number of muons reach the Earth's surface. The stationary reference frame in this phenomena is us, an observer at rest relative to the Earth's surface. They're usually the scientists or researchers that detect the muons. These muons moving at relativistic speeds appear to have longer lifetimes than

their natural lifetime at rest. The moving reference frame is the perspective of the muon itself. From the muon's viewpoint, it is moving through space and time. From the muon's point of view, it is at rest within its own frame of reference. In this frame, all the physical laws, including the behavior of its own internal clock, operate as they would in any inertial (non-accelerating) frame. Therefore, the muon perceives its clock to tick normally, and it experiences no time dilation.

Another consequence of special relativity is length contraction, which describes how objects moving at relativistic speeds appear shorter along the direction of their motion when observed by a stationary observer.

Time dilation and length contraction are both demonstrated in the well-known "Twin Paradox" experiment (Figure 1). Imagine two identical twins, Alice and Bob. Alice stays on Earth, while Bob embarks on a high-speed journey into space, traveling at a significant fraction of the speed of light to a distant star and then returns to Earth. The paradox arises when Bob returns to Earth. According to special relativity, during his journey, Bob's onboard clock ticks more slowly than Alice's clock on Earth. As a result, when Bob returns, he will be younger than Alice.

This paradox is solved in two ways:

From Bob's perspective, he remained at rest in his own reference frame, and it was Alice who was moving away from him and then back toward him. Therefore, according to Bob, it is Alice who ages more slowly. This might seem counterintuitive, but it's a fundamental consequence of the principle of relativity.

And the second way:

Consider that Bob experienced two phases of motion: accelerating to reach the high speed and then decelerating to return to Earth. These acceleration phases break the symmetry between the twins' frames. It is during these acceleration phases that Bob ages less than Alice. When he returns to Earth, he has experienced less time than Alice because of these periods of acceleration.

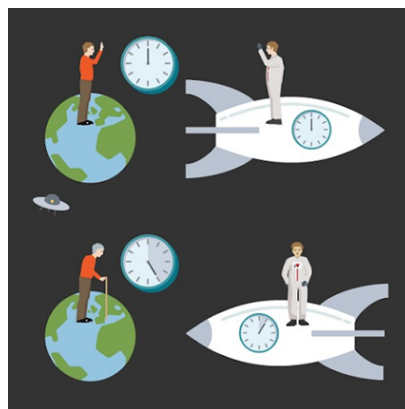


Figure 1: Twin-Paradox Experiment

Another thought experiment validating principles in relativity is the ladder paradox. Suppose a ladder travels horizontally at relativistic speeds and is experiencing length contraction. The ladder passes through an open barn door that is shorter than its rest length. If the ladder was not moving, it would not fit through the door. To a stationary observer, due to the contraction, the moving ladder is able to fit entirely inside the building as it passes through. From the point of view of an observer moving with the ladder, the ladder will not be contracted, and it is the building which will be contracted to an even smaller length and the ladder will not be able to fit through the door.

This contradiction rises from the mistaken assumption of absolute simultaneity. In relativity, simultaneity is relative to each observer, making the answer to whether the ladder fits inside the garage also relative to each of them.

1.2: General Relativity

Albert Einstein developed the theory of general relativity, presented in 1915, which revolutionized our understanding of gravity and extended the principles of special relativity to accelerated frames of reference and gravitational fields.

The Equivalence Principle in general relativity states that there is no local experiment that can distinguish between gravitational acceleration and acceleration due to other forces. This principle leads to the conclusion that gravity and acceleration are fundamentally indistinguishable.

Unlike Newton's idea of gravity as a force acting at a distance, Einstein proposed that gravity arises due to the curvature of spacetime caused by mass and energy (see figure 2). Objects with mass (like planets, stars, and galaxies) curve the spacetime around them, and other objects moving through this curved spacetime follow paths that we perceive as gravitational attraction. During a solar eclipse, Sir Arthur Eddington led an expedition to measure the apparent shift in the positions of stars near the Sun due to its gravitational field. According to general relativity, light rays passing close to a massive object like the Sun should experience gravitational deflection. The observations during the eclipse confirmed the predicted bending of starlight around the Sun, verifying Einstein's theory.

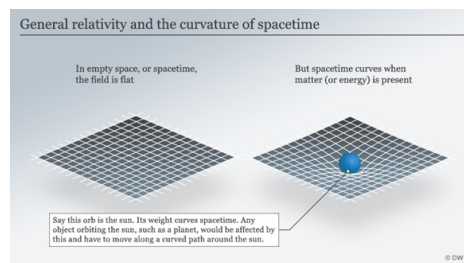


Figure 2: Curvature of Spacetime

Similar to the time dilation in special relativity due to relative motion, general relativity predicts gravitational time dilation. Clocks in stronger gravitational fields tick more slowly compared to clocks in weaker fields. Robert Pound and Glen Rebka measured the change in the frequency of gamma-ray photons emitted at the bottom of a tower and detected at the top. According to general relativity, light should lose energy when it climbs against the gravitational field of the Earth, leading to a redshift in its frequency. The experiment confirmed this effect, providing strong evidence for the gravitational time dilation predicted by general relativity.

Additionally, general relativity predicts the existence of gravitational waves, which are ripples in spacetime that propagate at the speed of light. These waves are generated by the acceleration of massive objects, such as colliding black holes or neutron stars. In 2015, the Laser Interferometer Gravitational-Wave Observatory (LIGO) detected gravitational waves for the first time, providing direct evidence for their existence.

2: Quantum Mechanics

One of classical mechanics' flaws was the inability to describe phenomena at the smallest scales. It couldn't explain what was happening with the behavior of particles or atoms. In the early 1900s, physicists made a shocking discovery. Consider a black body substance. A blackbody is a surface that absorbs all radiant energy falling on it. Visible light is absorbed rather than reflected on a black body, which produces the color black. It is possible to predict the intensity of the energy coming from a blackbody (blackbody radiation) based on its temperature. Physicists developed an equation to calculate the intensity, using the assumption that light is a wave. They found a problem with this theory. The equation was known as the Rayleigh-Jeans law, and it predicted that the higher the frequency of the radiation, the higher the intensity. It also says that the intensity of radiation emitted by a black body is directly proportional to the temperature and inversely proportional to the wavelength raised to the power of 4. This matched experimental conclusions, but only up to a certain point. The Rayleigh-Jeans law in terms of wavelength, λ , is denoted by:

$$B(\lambda) = \frac{2ck_B T}{\lambda^4}$$

Once the frequency of the light reached the ultraviolet range, the law broke down, as it didn't match the experimental results and it predicted an infinite intensity of the light. The law agrees with experimental results at large wavelengths, which have low frequencies, but disagrees at short wavelengths, which have high frequencies. This was known as the ultraviolet catastrophe and signaled the need for the development of something new, a different theory. This catastrophe was resolved by Max Planck, who created a different equation that successfully described what was actually happening.

Planck successfully resolved this catastrophe but unintentionally started a new one. According to Planck, electromagnetic energy takes the shape of tiny, discrete packets called quanta. After a certain point, you can't divide energy into anything smaller than these quanta packets. This was the beginning of the theory known as quantum mechanics, which describes energy and particles at the smallest scale.

2.1: Wave and Particle Theory of Light

Before Max Planck, energy was considered a continuous flow. Einstein argued that light energy traveled in packets called photons, which made light behave like a particle. But there were many experiments that showed that light behaved like a wave.

The double-slit experiment (see figure 3) was a famous experiment that demonstrated the wave behavior of light. It was conducted by Thomas Young in the early 19th century. Imagine a barrier with two closely spaced slits. On the other side of the barrier is a screen that acts as a detector. A light source emits a beam of light toward the barrier, and the light passes through the slits and reaches the screen. The goal of the experiment is to observe what pattern emerges on the screen.

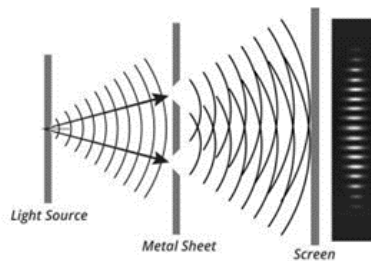


Figure 3: The Double-Slit Experiment

In classical wave theory, if light is treated as a continuous wave, the light waves passing through the two slits will interfere with each other when they overlap on the other side of the barrier. This is known as diffraction and is a fundamental law of waves. According to this prediction, you would expect to see an alternating pattern of bright and dark bands on the screen. These bands arise from constructive interference (where waves add up to create a bright spot) and destructive interference (where waves cancel out to create a dark spot). This is exactly what is observed in the experiment.

The interference pattern observed in the double-slit experiment is a clear indication that light is behaving as a wave. As I said, the pattern is a consequence of the constructive and destructive interference of light waves emanating from the two slits. Waves with troughs (low points) meeting troughs or crests (high points) meeting crests reinforce each other, creating bright fringes. Troughs meeting crests result in cancellation, forming dark fringes.

As I spoke about before, wave equations are a fundamental part of physics. Maxwell's equations produce a wave equation, which gives the speed of light. But it does so, assuming that

light is a wave. And yes, wave equations successfully describe waves. Those being electromagnetic waves or even mechanical waves. So light was officially deemed a wave for many years. Then Planck unintentionally provided evidence that energy could exist as discrete packets known as quanta.

Planck's discovery of quantized energy opened up many opportunities for physicists at the time to step into the world of quantum mechanics and make discoveries of their own. One of them was our good friend Einstein. Einstein is most known for his contributions to the theory of relativity, but he didn't actually win a Nobel Prize for them. He won the Nobel Prize for his work on the photoelectric effect.

The photoelectric effect (figure 4) was discovered by German physicist Heinrich Rudolf Hertz in 1887. Hertz discovered that the voltage at which sparking occurs changes when ultraviolet light is shined on two metal electrodes with a voltage connected across them. Einstein used the photoelectric effect to experimentally prove that light was actually a particle.

The photoelectric effect occurs when photons (light quanta) interact with atoms in a material. Each atom has a collection of electrons surrounding its nucleus, and these electrons are held in various energy levels or shells. When a photon with sufficient energy strikes an atom's electron, it can transfer its energy to the electron. If the energy of the photon is high enough to overcome the binding energy (work function) of the electron, the electron can be liberated from the atom and ejected from the material's surface.

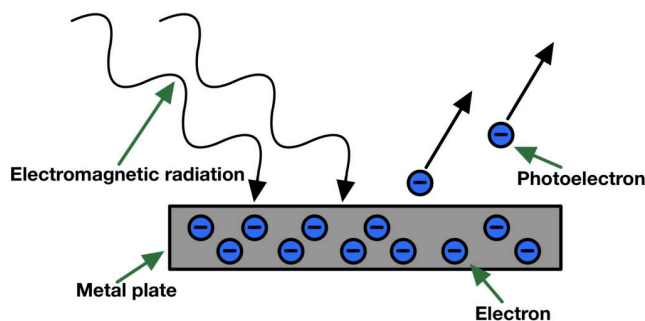


Figure 4: Photoelectric effect

Einstein also realized that you could learn a lot about the properties of light by studying the emission of electrons.

He discovered that the kinetic energy of the emitted electrons depends on the frequency of the electromagnetic radiation. There is also a threshold frequency below which no electrons are emitted for a given metal.

It is also important to know that particle theory and wave theory both agree that the electrons will be knocked out of the material, but they have different reasons why. According to wave theory, when an electron is struck by a light beam, the force generated causes the electron

to be ejected from the metal. The strength of the electric field acting as a force on the electrons will rise as the light intensity increases. As a result, more electrons are ejected from the metal with a higher speed and kinetic energy.

According to particle theory, when individual photons strike a metal, electrons are ejected from the metal. The photon is destroyed during the collision, and its energy is transferred to the electron, which is ejected from the metal. It takes a minimum of energy for this effect to occur.

So which one is correct, particle theory or wave theory? We can ask ourselves some questions to figure this out. Is there a specific frequency threshold of the photons below which electrons are not emitted? Indeed, there is. However, as we elevate the frequency of the photons, the maximum kinetic energy exhibited by the liberated electrons also increases. Now, does the intensity of light exert an influence on the electrons' maximum kinetic energy? Not quite. Intensity, instead, plays a role in the quantity of electrons ejected, leaving their energy unaffected.

Eventually, through calculation and consideration of these questions and answers, Einstein figured out that he was right. Light really does exist as tiny packets of energy called photons, and the photoelectric effect breaks down in terms of wave theory. The question many people were asking, though, was, Is light a wave or a particle?

Light is proven to be a particle, but there were many experiments before that showed that light behaved like a wave. You can't ignore experimental results that are valid.

So eventually, physicists had to admit that energy could exhibit wave and particle characteristics at the same time. This was called wave-particle duality, and we'll go in depth on it in the next chapter.

2.2: Wave Particle Duality

Wave-particle duality states that particles such as electrons or photons can exhibit wave-like behavior and particle-like behavior. Einstein said, "It seems as though we must sometimes use one theory and sometimes the other, while at times we may use either. We are faced with a new kind of difficulty. We have two contradictory pictures of reality; separately, neither of them fully explains the phenomena of light, but together they do."

Most physicists at the dawn of wave-particle duality thought that electrons were particles, but it turns out that if you emit electrons through a double slit, they diffract. This was proven to be true for other particles as well, proving that wave-particle duality applies to all matter.

The de Broglie wave equation is a wave equation that solves for the wavelength of a moving electron. It is denoted by:

$$\lambda = \frac{h}{mv}$$

Many physicists disagreed with each other on the concept of wave-particle duality. Niels Bohr thought that a particle would exhibit wave or particle properties depending on its physical setting. Including how it was observed, or where it existed in space.

Werner Heisenberg considered that wave-particle duality was present for all quantum objects. He coined the term wave-particle equivalence for his view on particle duality. Heisenberg's equivalence principle comes from two distinct approaches to wave-particle duality. One is the statistical interpretation of the wave function, and the second is that of quantized matter waves. Both of these approaches turned out to be mathematically equivalent and led to the same results. This convinced Heisenberg that the wave and particle depictions of electrons were different ways of describing the same thing.

The Schrodinger equation is a linear partial differential equation that governs the wave function of a quantum system. A wave function is a mathematical description of the probability of finding a particle in any given spot. It's denoted by the Greek symbol ψ (psi). Every wave needs a wave equation. That's a fundamental property of waves. So if particles are waves, then they need a wave equation. That's what the Schrodinger equation does for particles. It was named after Erwin Schrodinger, who postulated the equation in 1925 and published it in 1926. The Schrodinger equation can be thought of as Newton's second law of classical physics' quantum analogue. Newton's second law uses mathematics to predict the course that a particular physical system will take over time given a specific set of initial circumstances. The Schrodinger equation provides a wave function's evolution over time.

The equation looks like this:

$$i\hbar \frac{\partial \psi(x,t)}{\partial t} = -\frac{\hbar^2}{2m} \frac{\partial^2 \psi(x,t)}{\partial x^2} + U(x)\psi(x,t)$$

2.3: Uncertainty Principle

Heisenberg was one of the founding fathers of quantum mechanics. He accomplished many things in his lifetime. Most notably, the creation of the uncertainty principle. The uncertainty principle states that you can't know certain complementary properties of a quantum object with accurate precision at the same time. Some of these complementary properties include position and momentum, or energy and time.

It's basically a trade-off. If you wanted to know a particle's position, you would know less about its momentum, and vice versa. The same thing applies to every complementary property, including energy and time.

This problem arises due to the way scientists try to find a particle's position or momentum. Let's say you wanted to know the position of some particle propagating through space. In order to do that, we could use x-ray photons. X-rays have very short wavelengths, meaning they have very high energies. In order to find the position of this particle, we could knock this x-ray photon straight into it. I know it sounds a bit peculiar, but by tracking the

movement of our x-ray photon to the exact moment it hits our particle, we will know its position. Now, the reason why we can't know its momentum comes from classical physics. The conservation of momentum. Once we hit our particle with our x-ray photons, we completely throw off the momentum of our particle.

Now let's try to find the momentum of our particle. In order to do that, let's use radiation with a higher wavelength like a radio wave. If we knock the radio wave into our particle it won't disturb the particle much because it doesn't have as much energy as an x-ray. The flaw of using radiation with a higher wavelength is simply that the wavelength is longer, so we are less certain of its position.

Another uncertainty in quantum mechanics is the time-energy uncertainty. It states that there is a tradeoff (much like in position-momentum uncertainty) in which we can measure the energy of a quantum system and the precision with which we can measure the time at which that energy measurement occurs. This uncertainty comes from the conservation of energy, which is a huge fundamental law in classical mechanics. Quantum mechanics says that overall, energy is conserved, but as long as energy pops in and out of existence fast enough, it's not disobeying the laws of physics. Basically, energy can just exist then not exist. As long as energy pops out of existence in a small fraction of a second, it's all good. Particles do this all the time in space every fraction of a second. This way that energy in quantum systems can pop in and out of existence is a huge reason for the incoherence between two theories that we will come to later.

2.4: Spin

Angular momentum is a term used a lot in classical mechanics. It is the rotational analog of momentum, specifically linear momentum. It is a vector, meaning it has both a direction and a magnitude. The properties of bicycles, motorcycles, and even frisbees are described by the conservation of angular momentum.

Spin is a type of angular momentum that travels with all particles and even atoms. While classical angular momentum is associated with the rotation of an object around a point, spin is an inherent property of particles and is not tied to any physical rotation. It's more like a wave property. Spin is quantized and can only take certain discrete values, so spin is given a spin quantum number.

The Stern-Gerlach experiment is an experiment that demonstrates the quantization of angular momentum and the properties of spin. It was first performed by Otto Stern and Walther Gerlach in 1922.

The experiment involves a beam of particles being passed through an inhomogeneous magnetic field gradient. The magnetic field gradient varies along a specific direction, which we'll call the z-direction. When the particles (with spin) pass through the magnetic field gradient, they split into discrete, separate beams.

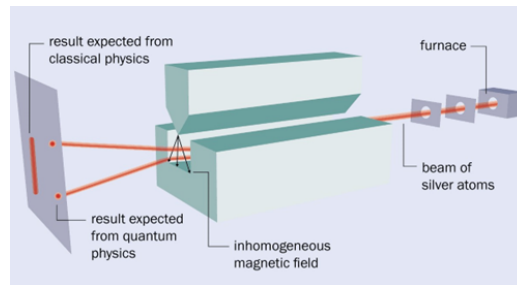


Figure 5: Stern-Gerlach Experiment

The quantization of angular momentum means that the component of the particle’s spin along the z-direction can only take certain discrete values. For a particle with spin 1/2, such as an electron, there are two possible spin states: spin-up (+1/2) and spin-down (-1/2) along the chosen direction. This leads to the splitting of the incoming beam, as mentioned earlier, into two separate beams—one deflected upwards and the other downwards—corresponding to the two spin states. Fortunately, this is exactly what was observed, demonstrating spin for the first time.

The first relativistic wave equation was the Dirac equation. Relativistic wave equations predict the behavior of particles at high velocities in terms of relativity and quantum mechanics. It describes all massive particles, called “Dirac particles”, such as electrons.

$$(i\partial - m)\psi = 0$$

This is the reduced Dirac equation. Seems simple but explains a lot. This one equation was the stepping stone for a new age of quantum mechanics. An age of reconciliation and quantization of fundamental interactions, quantum mechanics, relativity and other physical phenomena. These theories are known as quantum field theories (QFT).

3: Quantum Field Theory

There are four fundamental forces in the universe. Gravity, electromagnetism, and the weak and strong nuclear forces. The gravitational force is the force of attraction between masses. It keeps celestial bodies, like planets and stars, in orbit and governs the motion of objects on Earth. As I spoke about earlier in this paper, gravity is described as a curvature of spacetime in general relativity. Gravity is the weakest of the four fundamental forces. Its strength is proportional to the masses of the interacting objects and inversely proportional to the square of the distance between them, as described by Newton’s law of universal gravitation.

The electromagnetic force is responsible for interactions between electrically charged particles. The electromagnetic force is relatively strong and follows Maxwell’s equations, which describes the force between charged particles.

The weak nuclear force is responsible for the change of one type of subatomic particle into another. It also describes certain types of nuclear reactions, like beta decay. It plays a crucial role in processes like stellar nucleosynthesis (the process by which stars create chemical elements through nuclear fusion) and the behavior of neutrinos. Neutrinos are the most abundant particles that have mass in the universe. Every time nuclear fusion or fission occurs, it produces a neutrino. Even bananas emit neutrinos—they come from the radioactivity of the potassium in the berry. Additionally, the weak force operates over very short distances, typically within the size of atomic nuclei.

The strong nuclear force, also known as the strong interaction or color force, binds particles together to form nuclei. The strong force is exceptionally strong, but it is also short-ranged, rapidly decreasing with distance. If not for the strong force the positively charged protons in a nucleus would repel each other rather than bind together.

3.1: Quantum Electrodynamics (QED)

Quantum electrodynamics (QED) is a quantum field theory that describes the fundamental interactions between photons and charged particles, primarily electrons. Fundamentally, it describes the electromagnetic force in terms of quantum mechanics. Dirac's famous paper, titled "The Quantum Theory of the Emission and Absorption of Radiation," was published in 1927. The paper was the beginning of QED. The paper is notable for its introduction of the concept of "second quantization," which is a method for quantizing fields in a way that treats particles as excitations of those fields.

In his paper, Dirac derived expressions for the transition probabilities of electrons moving between different energy levels through the absorption or emission of photons. These transition probabilities were calculated using the newly developed formalism of second quantization.

Another idea that Dirac's paper introduced was antiparticles. Using his Dirac equation, Dirac proved the existence of antiparticles. Antiparticles are particles that have the same mass and characteristics as a standard particle, but with the opposite electric charge. Antielectrons, which are known as positrons, have a positive charge rather than a negative (like electrons do.) On the other hand, antiprotons have negative electric charges instead of positive (like regular protons do).

Although Dirac introduced a huge and successful quantum theory of electromagnetism, there was still much work to be done on QED. Richard Feynman first introduced his Feynman diagrams in 1948. Feynman diagrams are used by physicists for understanding and calculating particle interactions in QED. Prior to their introduction, physicists struggled with complex equations and lengthy calculations, making it challenging to gain insights into QED.

Feynman diagrams are composed of four fundamental elements: vertices, lines, arrows, and loops. Vertices in the diagrams represent interaction points where particles come together and interact. In QED, a common vertex involves one electron, one positron, and one photon. These points indicate the fundamental forces responsible for particle interactions.

Lines in Feynman diagrams represent particles, with different types of lines corresponding to different types of particles. For example, solid lines typically represent electrons, while wavy lines represent photons (see figure 6).

Arrows indicate the direction of particle flow. Some also point forward or backward in time, as antiparticles in Feynman diagrams move backward in time.

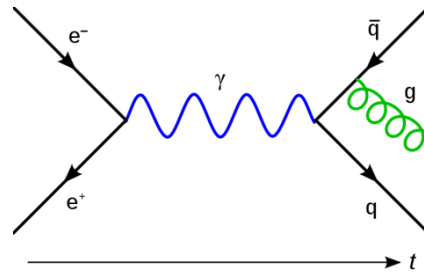


Figure 6: Feynman diagrams

This (figure 6) is an example of a Feynman diagram. It involves one electron, one positron, one photon, one quark, one antiquark, and one gluon. Quarks and gluons are fundamental particles. However, they have a little more to them that we'll get into in a bit.

This Feynman diagram depicts the path that a colliding electron and positron follow. First, the electron and positron collide. This collision produces a photon (represented by the blue wave pattern), which ultimately splits into a quark-antiquark pair. The antiquark also radiates a gluon which is represented by the green helix.

Additionally, Feynman diagrams must satisfy conservation laws, ensuring that charge, energy, momentum, and other fundamental quantities are conserved at each vertex and throughout the diagram. If a conservation law wasn't conserved throughout a diagram, then the diagram would be unusable.

Furthermore, the probability amplitude for a specific particle interaction is calculated by summing the contributions of all possible Feynman diagrams corresponding to that interaction. Each diagram represents a possible path, and the total probability of a process is obtained by summing over all contributing diagrams. This is known as the path integral formulation and is utilized in other areas of physics as well.

While Feynman diagrams are closely associated with QED, they have been extended to describe interactions involving other fundamental forces, such as the strong and weak nuclear forces. In this context, additional particles and vertices are introduced to accommodate the specific characteristics of those forces.

At the time, QED was a breaking point in quantum field theory, but there were still many steps to take in reconciling the other fundamental forces with quantum mechanics. The next force was the strong interaction.

The quantum field theory of the strong force is known as quantum chromodynamics (QCD).

3.2: Quantum Chromodynamics (QCD)

Quantum chromodynamics was formulated in the early 1970s as a result of the efforts of physicists Murray Gell-Mann and George Zweig. It was introduced as a theory to explain the strong nuclear force, which binds protons, neutrons, and other hadrons (any particle that is made up of two or more quarks) together within the atomic nucleus. Gell-Mann and Zweig's work introduced the concept of "quarks" as the fundamental building blocks of matter. Quarks are what make up hadron particles. They come in various flavors or types, such as up, down, strange, charm, bottom, and top. The different types of flavors make up particles, and the distinct combinations of these flavors make up different types of particles. For example, protons consist of two up quarks and one down quark. These quarks are held together by the exchange of gluons, which are massless vector bosons. Bosons are particles that mediate a fundamental force. Gluons are bosons, so they must carry a fundamental force. In fact, it's true that gluons are the force carriers of the strong force. The other fundamental class of particles is known as fermions. Quarks and leptons (particles that are not made up of quarks, like electrons), as well as most composite particles, like protons and neutrons, are fermions.

One of the most remarkable features of QCD is asymptotic freedom, a phenomenon that describes how quarks and gluons interact at very short distances or high energies. At these scales, the strong force weakens, allowing quarks and gluons to behave almost as free particles.

In contrast to asymptotic freedom, confinement describes how, at larger distances or lower energies, quarks and gluons are never found as isolated particles but are always confined within hadrons, like protons and neutrons. This phenomenon is still not fully understood and is one of the challenges facing QCD.

In QCD, there are three "colors" and their corresponding "anticolors", often denoted as red, green, and blue, and their antired, antigreen, and antiblue counterparts. These terms are mere labels, and they do not refer to the colors we perceive in everyday life. Instead, they are abstract properties assigned to quarks and gluons.

QCD is a non-abelian gauge theory, meaning that the order of operations (i.e., the sequence in which different color charges interact) is essential. Unlike QED, where the electromagnetic force is abelian (order independent), the non-abelian nature of QCD introduces unique complexities. Quarks and gluons interact with each other through the exchange of gluons, and these gluon-gluon interactions are influenced by the color charges involved.

Color confinement is deeply intertwined with color charge. When quarks attempt to move apart and increase the distance between them, the strong force attraction between them intensifies, not weakens, as in QED. The strong force attraction greatly strengthens as the distance between the two quarks approaches one femtometer. After the distance exceeds one femtometer, the strong force starts to weaken. But as the distance approaches one femtometer, the binding force is too strong for the quarks to overcome, and the quarks stay together. This behavior makes it impossible to separate quarks from one another, as the energy required to pull them apart becomes infinite. Instead, quarks are always found in color-neutral combinations, forming hadrons.

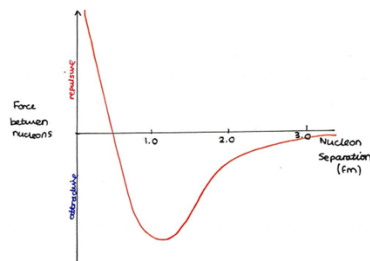


Figure 7: Color Confinement

The x axis shows the separation of nuclei and quarks in femtometers. The y axis describes the amount of force between the quarks. The higher y direction shows how the strong force is repulsive at short distances while the lower y direction shows how the strong force gets more attractive as it approaches one femtometer. It's important to note that we can ground these ideas in reality through experimental observations.

The European Muon Collaboration (EMC) experiment in the late 1970s and early 1980s provided crucial evidence for the existence of quarks within protons and neutrons. High-energy electrons and muons were used to probe the internal structure of nuclei, revealing the distribution of quarks within them.

Additionally, experiments on particle accelerators such as CERN's Large Hadron Collider (LHC) have produced strong evidence for QCD. In particle accelerators, high-energy collisions produce sprays of particles known as jets, which can be precisely predicted and explained by QCD calculations. The observation of jets provides experimental confirmation that QCD and building blocks of reality agree with each other.

Furthermore, numerical simulations of QCD on a discrete lattice have been instrumental in understanding the non-perturbative aspects of the theory, including confinement. These simulations provide insights and confirmation of the behavior of quarks and gluons at low energies and have been used to calculate the properties of hadrons from first principles.

3.3: Electroweak Interaction

The electroweak interaction unifies the electromagnetic and weak forces. This emerged during the mid 20th century as the three physicists Sheldon Glashow, Abdus Salam, and Steven Weinberg independently discovered that they could create a gauge-invariant theory of the weak force but they also included the electromagnetic force.

The unification of electromagnetism and weak interaction was achieved by introducing a new set of force carriers, known as the W^+ and W^- bosons and the neutral Z boson. The photon was also reconciled within the interaction as it's the force carrier of the electromagnetic force. Making the photon a boson as well.

The electroweak theory also incorporates the Higgs mechanism, proposed by Peter Higgs. This mechanism introduces a scalar field, the Higgs field, that gives mass to the W and Z

bosons while leaving the photon massless. It's also associated with a scalar boson called the Higgs boson, which is the quantum excitation of the Higgs field.

The most famous and significant experiments to discover the Higgs boson took place at the Large Hadron Collider (LHC) at CERN. They collide high-energy protons accelerated to nearly the speed of light. These collisions produce a wide variety of particles. The LHC is equipped with several detectors, with the two main ones being the ATLAS and CMS experiments. On July 4, 2012, CERN announced the discovery of a new particle that was consistent with the Higgs boson. This discovery was the result of extensive data analysis from both the ATLAS and CMS experiments. It successfully confirmed the existence of the Higgs field

The electroweak interaction actually became what is now known as the Standard Model of Particle Physics. This was done by reconciling QCD and QED with the electroweak interaction. Making the Standard Model actually a unification of electromagnetism, the strong force, and the electroweak interaction.

3.4: Standard Model of Particle Physics

The Standard Model of Particle Physics, as seen in figure 8, is the framework that describes and reconciles electromagnetism, the strong force, and the electroweak interaction. This chapter will serve as a review of this section while also introducing the next.

The Standard Model predicts a variety of elementary particles. These particles can be classified into two categories: fermions and bosons. Fermions are the building blocks of matter and comprise the particles we typically think of as "fundamental." There are two types of fermions: quarks and leptons.

Quarks come in six flavors: up, down, strange, charm, bottom, and top. They are never observed in isolation due to confinement (a property of quantum chromodynamics) and combine to form hadrons like protons and neutrons.

Leptons consist of the electron, muon, tau (like an electron but heavier), and their corresponding neutrinos. Leptons are electrically charged (except for neutrinos) and do not experience or react to the strong force.

Bosons are force-carrying particles responsible for transmitting the fundamental forces of nature. Photons, γ , mediate the electromagnetic force and are responsible for electromagnetic interactions like light and electromagnetic fields. W^+ and W^- bosons are the carriers of the weak nuclear force and are responsible for processes like beta decay in atomic nuclei. Z bosons (Z^0) are like the W bosons, but they mediate the weak nuclear force and are electrically neutral. Gluons mediate the strong nuclear force, acting as the exchange particle for the interaction. Through their strong interaction, gluons bind quarks into groups, forming hadrons such as protons and neutrons. The Higgs boson is associated with the Higgs field, which gives mass to particles via the Higgs mechanism. Additionally, the standard model also describes fermions' antiparticles. This (figure 8) is a diagram that represents the particles that the standard model

predicts. As you can see, it has every quark and antiquark. Every lepton and anti-pair and, of course, the bosons. It even labels each particle with its spin, mass, and charge.

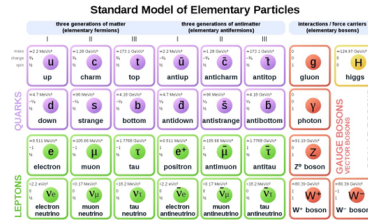


Figure 8: Standard Model of Particle Physics

While the Standard Model is highly successful, it has its limitations. For example, it doesn't incorporate gravity. Only three of the fundamental forces are reconciled within it. And so the search for what is known as the theory of quantum gravity is upon us.

4: The Unified Field Theory

Quantum gravity attempts to explain how gravity works on the smallest particles in the universe. In the beginning of this paper, I talked about how gravity is successfully described within general relativity. General relativity describes spacetime as a continuous, smooth fabric that can curve and warp in response to the presence of mass and energy. This makes quantum gravity an attempt at unifying general relativity with quantum mechanics.

Unfortunately, reconciling quantum mechanics with general relativity has some problems. Simply put, general relativity describes massive objects (like stars and galaxies), while quantum mechanics describes objects at the smallest scales (like particles and atoms). You can see the incoherence here. Additionally, general relativity predicts the existence of singularities at the centers of black holes, where spacetime curvature becomes infinitely steep. These singularities are considered problematic because they are very small but are points of infinite density. Basically, they are objects that are extremely tiny but are also extremely dense. General relativity and quantum mechanics provide different explanations for why this is, which ultimately signals even more incoherence between the two theories. Furthermore, quantum mechanics describes spacetime as fundamentally uncertain and unpredictable. It's constantly in a state of change and fluctuation.

These short bursts of energy are known as quantum fluctuations, and they are governed by the laws of the uncertainty principle (specifically the time-energy uncertainty). So spacetime almost looks like a collection of weird bumps. Almost like a pot of water boiling under a stove.

4.1: Quantum Gravity

In order to quantize gravity, we can look at other ways we quantized the other fundamental forces. The other fundamental forces are quantized by bosons that carry the fundamental forces. So maybe we can find a boson that carries the gravitational force. In the

context of gravity, the hypothetical particle associated with the gravitational force is referred to as the "graviton."

As done with electroweak interaction, we can treat gravity as a quantum field, and the graviton is the quantized excitation of this field. Much like other particles, gravitons would have quantum properties, including spin, energy, and momentum.

Unfortunately, when one attempts to quantize gravity in a straightforward manner, such as treating it as a perturbative quantum field theory, infinities arise that are not easily renormalized. Making the graviton quite a difficult task to even calculate. Additionally, unlike other fundamental forces, gravity is always attractive and universally so. This uniqueness presents challenges in constructing a quantum field theory that can quantize gravity and describe it consistently with the other forces.

Furthermore, gravitons remain theoretical constructs, and experimental evidence for their existence has not been found. Detecting gravitons directly is extremely challenging due to the extreme weakness of the gravitational force compared to other fundamental forces.

Another concept to note as we move further into quantum gravity is supersymmetry. Supersymmetry is a spacetime symmetry between two classes of particles: bosons and fermions. In supersymmetry, each particle from one class would have an associated particle in the other, known as its superpartner, the spin of which differs by a half-integer. For example, if the electron exists in a supersymmetric theory, then there would be a particle called a selectron (superpartner electron); a bosonic partner of the electron. In the simplest supersymmetry theories, with perfectly "unbroken" supersymmetry, each pair of superpartners would share the same mass and internal quantum numbers, besides spin. More complex supersymmetry theories have a spontaneously broken symmetry, allowing superpartners to differ in mass. Supersymmetry has no experimental evidence and has yet to be proven, but has become useful in later attempts at quantum theory of gravity.

Fortunately, the quest for a theory of quantum gravity is not over. There's much more to do. Many more theories of how space and time work to prove. One of these theories is loop quantum gravity.

4.2: Loop Quantum Gravity (LQG)

Loop Quantum Gravity (LQG) is a theory that seeks to unify two of the fundamental theories in modern physics: general relativity and quantum mechanics. It attempts to solve the problems of spacetime between the unification of general relativity and quantum mechanics and develop a quantum theory of gravity based directly on Einstein's geometric formulation rather than the treatment of gravity as a force. At its core, LQG departs from the traditional field-theoretic approach of particle physics and focuses on the quantization of space and time themselves. Other quantum theories, which usually assume spacetime to be continuous and unquantized, differ from LQG, as the central concept in LQG is the discretization of spacetime into what are known as "spin networks" or "spin foams."

In LQG, spacetime is viewed as a network of interconnected nodes and links, reminiscent of a complex web. These nodes represent discrete units of space, and the links between them represent discrete units of time. Mathematically, these networks of links and nodes are called spin networks, which are graph-like structures used to represent the quantum geometry of spacetime.

Holonomy Operators are operators that describe how objects, such as vectors or connections, evolve along a path in spacetime. In LQG, holonomy operators are associated with the links of the spin network. They encode how geometry changes as you move from one node to another along a link.

A key feature of LQG and spin networks are spin network states. Spin network states are mathematical representations of the quantum states of geometry associated with a given spin network. They are constructed by applying holonomy operators to a reference state, typically called the "vacuum state." This reference state is typically chosen to represent a featureless or "empty" spacetime. It serves as a starting point for building more complex quantum states.

Another feature of LQG is its quantization of geometry. Geometric quantities, like length, area, and volume, which are continuous in classical physics, are no longer treated as continuous variables but are discrete. This quantization is achieved through the assignment of half-integer spins (typically denoted as " j ") to the links or edges of the spin network. This leads to space and time becoming no longer continuous (like in relativity) but composed of tiny, discrete building blocks.

LQG also has implications for cosmology. Loop Quantum Cosmology (LQC), a branch of LQG, deals with the application of LQG principles to the universe as a whole. In ordinary cosmology, the universe's history is described by general relativity, but at extremely high energy densities, such as near the Big Bang or in a singularity, general relativity breaks down, and a quantum theory of gravity is required. LQC aims to provide a quantum framework for cosmology that explains these phenomena.

One of the greatest achievements of LQC is the resolution of the Big Bang singularity. In classical cosmology, the universe is thought to have originated from a singularity. This singularity exploded and propelled matter, among other things, to form the billions of galaxies in our universe. In LQC, the discrete nature of spacetime prevents such a singularity. Instead, the universe reaches a minimum volume, known as the "quantum bounce" (see Figure 9), at which point it begins to expand again. This bounce replaces the classical singularity, allowing for a continuous evolution of the universe.

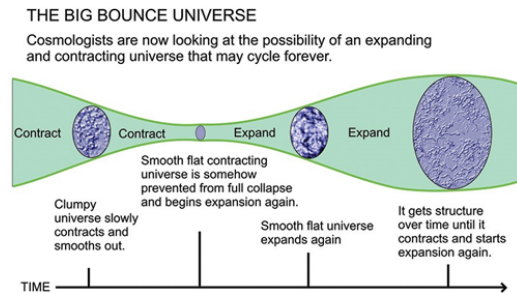


Figure 9: The Big Bounce

Unfortunately, while loop quantum gravity represents a promising approach to addressing the challenges of quantum gravity, it also faces its own set of formidable challenges and open questions. One of the most significant challenges is connecting LQG to the Standard Model of particle physics. LQG and the Standard Model are built on distinct mathematical formalisms. LQG relies on the quantization of spacetime geometry using techniques like spin networks and holonomy operators, whereas the Standard Model employs quantum field theory on a fixed spacetime background. The mathematical formalisms used in these theories are inherently different.

Additionally, LQG predicts modifications to the classical behavior of gravity, particularly in the strong gravitational field regimes near black holes and during the early universe. These quantum corrections need to be tested through precise observational data. Currently, there is no evidence proving LQGs predictions.

Fortunately, all is not lost, as there are actually two theories of quantum gravity. String theory is completely different from LQG and provides a distinct attempt at solving quantum gravity.

4.3: String Theory

Instead of only attempting to unify general relativity and quantum mechanics, string theory seeks to unify all four fundamental forces. It postulates that the fundamental building blocks of the universe are not made up of point-like particles but tiny, vibrating strings. These strings can vibrate at different frequencies, giving rise to different particles with distinct properties. The frequency of vibration of a string is given by the equation $f = \frac{n}{2L} \sqrt{\frac{T}{\mu}}$. In this equation, T is the tension in the string, L is the length, and n is the number of harmonics. A harmonic is a wave or signal whose frequency is an integral multiple of the frequency of the same reference signal or wave.

As I stated, the frequencies of the strings correspond to different particle types. For example, the lowest vibrational mode represents the theoretical graviton and higher vibrational modes give rise to particles like photons, electrons, quarks, etc.

String theory requires the existence of extra dimensions beyond the familiar three spatial dimensions and one time dimension. In some versions, there may be as many as 10 or 11 dimensions. These extra dimensions are typically compactified, meaning they are tiny and not directly observable. The existence of extra dimensions is crucial for the unification of the fundamental forces in the universe. In some string theory scenarios, the extra dimensions allow for the unification of gravity with the other forces at very high energy scales. String theory suggests that at these high energy scales, such as those near the Big Bang or within extreme environments like black holes, the forces of the Standard Model and gravity itself become indistinguishable due to the effects of the extra dimensions. This is sometimes referred to as "energy scale unification." Additionally, when all forces, including gravity, are considered in the higher-dimensional spacetime of string theory, they can be described by a unified set of equations. This unified description emerges because the geometry of the extra dimensions affects how forces propagate and interact. Furthermore, the concept of extra dimensions within string theory contributes to the notion of a landscape of possible string vacuum states. This landscape suggests a vast array of possible universes with varying physical constants and properties, giving rise to the multiverse hypothesis.

String theory also requires supersymmetry as it ensures the mathematical consistency of the theory. Without supersymmetry, certain mathematical infinities and inconsistencies would arise in the quantum calculations of string interactions. Supersymmetry helps regulate these infinities and ensures that the theory remains well-behaved at the quantum level.

Another reason for the necessity of supersymmetry in string theory is that during calculations, the contributions from supersymmetric partners tend to cancel out the contributions from their non-supersymmetric counterparts, effectively suppressing divergences that would otherwise make the theory inconsistent.

There are several distinct versions of string theory, such as Type I, Type IIA, Type IIB, heterotic SO(32), and heterotic E8 x E8. These variations initially seemed unrelated, but further developments, such as dualities, have shown that they are interconnected facets of a broader framework, known as M-Theory.

Type I string theory is characterized by the presence of open strings, which have endpoints that can move freely in spacetime. These open strings can couple to D-branes, which are extended objects in spacetime. Type I theory incorporates both open and closed strings (strings with no endpoints). It also exhibits an anomaly that necessitates the inclusion of supersymmetry.

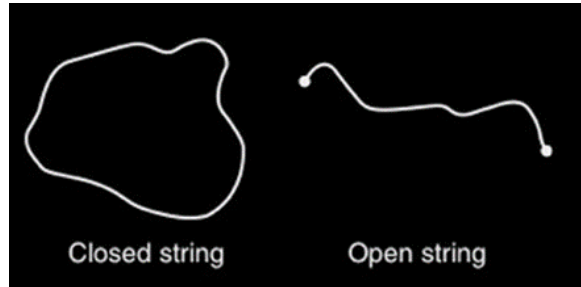


Figure 10: Open and Closed Strings

Type IIA and Type IIB string theories are closed string theories. Type IIA is noteworthy for being a non-chiral theory, meaning it doesn't distinguish between left- and right-handed particles. If a particle's spin and motion move in the same direction, it is said to be right-handed. If the spin and velocity directions are in opposition, the particle is left-handed. In contrast, Type IIB is a chiral theory, making such a distinction. These two theories are related by T-duality, a symmetry that interchanges small and large dimensions.

Heterotic string theories are hybrid theories that combine closed strings and open strings. These theories possess an intricate structure known as gauge symmetry, which allows for the unification of forces. Heterotic string theories come in two versions: $SO(32)$ and $E8 \times E8$.

M-Theory is a unifying framework that reconciles every string theory. It introduces an eleventh dimension and extends the notion of strings to include higher-dimensional objects called membranes.

Additionally, M-Theory is an extension of Type IIB string theory in which the spacetime contains varying, non-geometric backgrounds. It plays a crucial role in understanding certain aspects of compactification in string theory and has found applications in both particle physics and cosmology.

String Theory is not without its challenges. It requires the existence of extra dimensions, which have not been experimentally detected. Furthermore, the theory lacks a unique predictive framework, making it difficult to test empirically. Additionally, many versions of string theory incorporate supersymmetry and supersymmetric particles have not been observed, leading to concerns about whether supersymmetry is a valid symmetry in nature.

As LQG and string theory are the only theories of quantum gravity, there are no other frameworks that could provide insights into quantum gravity. We still need to work further into these theories to solve quantum gravity, but ultimately address the biggest problem facing the world of modern theoretical physics, the Unified Field Theory. We will go more in depth on the Unified Field Theory in the next chapter.

4.4: The Unified Field Theory (UFT)

A Unified Field Theory (UFT) is a theoretical framework that seeks to unify the fundamental forces of nature within a single, coherent, mathematical, description. It's often referred to as the holy grail of physics. Such a theory would provide profound insights into the nature of the universe, leading to a deeper understanding of how the fundamental building blocks of matter and the forces that govern them are interconnected and work. The Unified Field Theory could also just simplify the laws of physics and provide a more elegant and comprehensive description of the universe. The Unified Field Theory could also solve the inconsistencies presented by the attempts of quantum gravity.

Now, why is solving a UFT important? At its core, science is driven by the human desire and curiosity to understand the fundamental workings of the universe. A UFT would represent the pinnacle of this quest, as it could actually describe everything about the universe.

Additionally, A UFT could offer insights into the conditions of the early universe, moments after the Big Bang. It could potentially explain the origin of the universe and the initial conditions that set the stage for the formation of galaxies, stars, and all known matter and energy. Additionally, a UFT has the potential to provide profound insights into the inner workings of black holes.

Furthermore, advances in theoretical physics often lead to technological breakthroughs. The development of a UFT could potentially open up new avenues for technology, just as the understanding of quantum mechanics has given rise to new technologies as well. Such technologies could be advanced propulsion systems, as a UFT could provide insights into manipulating gravity or space-time; more efficient and sustainable energy sources, more powerful quantum technologies, advanced communication technologies, medical imaging, drug development, treatments for various diseases, and precise and less invasive surgical techniques; and even more advanced astronomical instruments and observatories.

A UFT could also explain the nature of dark matter and dark energy. Dark matter is a form of matter that does not emit, absorb, or interact with electromagnetic radiation. It is invisible to our current instruments. However, its presence is inferred through its gravitational effects on visible matter, such as galaxies and galaxy clusters. Dark energy, on the other hand, is a mysterious form of energy that is believed to permeate all of space and is responsible for the observed accelerated expansion of the universe. Unlike dark matter, which acts as a gravitational attractor, dark energy has a repulsive effect, causing the universe's expansion to accelerate.

Overall, physicists are still working on this framework and we've come very close, but we still have our ways to go.

5: Conclusion

Currently, in physics, we find ourselves in an exciting period characterized by ongoing research, discovery, and the pursuit of answers to some of the most profound questions about the universe.

The Large Hadron Collider (LHC) at CERN continues to operate, enabling scientists to explore the properties of fundamental particles and test various aspects of the Standard Model of particle physics. Researchers are still actively searching for new particles or phenomena beyond the Standard Model, such as dark matter particles or evidence of supersymmetry.

Advances in astrophysical observations, such as those made by space telescopes like Hubble and ground based observatories, have deepened our understanding of the universe's large-scale structure, the expansion rate of the cosmos, and the distribution of dark matter and dark energy. The study of exoplanets (planets outside our solar system) has expanded, with numerous discoveries of potentially habitable worlds and the search for signs of extraterrestrial life gaining momentum.

Quantum technologies are rapidly advancing, with developments in quantum computers, quantum cryptography, and quantum sensors. Companies and research institutions worldwide are investing in these technologies. Research in quantum mechanics continues to explore the foundations of quantum theory.

The detection of gravitational waves by LIGO and Virgo collaborations has opened a new era in the study of the universe. Gravitational wave observatories are becoming increasingly sensitive, allowing for the observation of more events involving compact objects like neutron stars and black holes.

The quest for a Unified Field Theory continues to be researched. Including string theory and LQG, with scientists exploring various formulations and consequences of these theories. However, they have not yet been experimentally validated and remain a topic of ongoing debate.

As you can see, ideas in modern physics are still being researched, tested, and debated on. We've come a long way since the beginnings of theoretical physics with Maxwell's equations.

Works Cited

- [1] Norton, J. “*Einstein's Special Theory of Relativity and the Problems in the Electrodynamics of Moving Bodies that Led him to it.*” 2003. University of Pittsburgh, Department of History and Philosophy of Science. 2023. <https://sites.pitt.edu/~jdnorton/papers/companion.pdf>.
- [2] Einstein, A. “*On the General Theory of Relativity.*” 1915. Princeton. University. 2023. <https://einsteinpapers.press.princeton.edu/vol6-trans/228>.
- [3] Einstein, A. “*Relativity: The Special and General Theory.*” 1920. HENRY HOLT AND COMPANY. 2023. https://www.f.waseda.jp/sidoli/Einstein_Relativity.pdf.
- [4] Einstein, A. “*On the Electrodynamics of Moving Bodies.*” 1905. University of Oxford: Physics Department. 2023. <https://users.physics.ox.ac.uk/~rtaylor/teaching/specrel.pdf>.
- [5] Original Papers by Einstein, A & Minkowski, H; Historical Introduction by Mahalanobis, Prasanta. “*The Principle of Relativity.*” 1920. UNIVERSITY OF CALCUTTA. 2023. https://www.rarebooksocietyofindia.org/book_archive/196174216674_10151069422921675.pdf.
- [6] 2021. “*Twin Paradox.*” MIT College of Railway Engineering and Research. 2023. <https://mitcorer-barshi.medium.com/twin-paradox-8ec93a729591>.
- [7] 2010. “*Light Clock.*” Wikipedia 2023. https://simple.wikipedia.org/wiki/Light_clock.
- [8] “*Einstein's Theory of Gravity.*” BC Campus; University Physics Volume 1. 2023. <https://pressbooks.bccampus.ca/universityphysicssandbox/chapter/einsteins-theory-of-gravity/>.
- [9] RhEvans. 2013. “*Derivation of the Rayleigh-Jeans law.*” The Curious Astronomer. 2023. <https://thecuriousastronomer.wordpress.com/2013/10/28/derivation-of-the-rayleigh-jeans-law-part1/>.
- [10] Binney, James & Skinner, David. 2008. “*The Physics of Quantum Mechanics.*” Cappella Archive. 2023. <https://www-thphys.physics.ox.ac.uk/people/JamesBinney/qb.pdf>.
- [11] Morin, David. “*Introduction to Quantum Mechanics.*” Harvard University. 2023. https://scholar.harvard.edu/files/david-morin/files/waves_quantum.pdf.
- [12] Hamer, Ashley. 2019. “*The Double-Slit Experiment Cracked Reality Wide Open.*” Discovery. 2023. <https://www.discovery.com/science/Double-Slit-Experiment>.
- [13] 2021. “*The Photoelectric Effect, Photons and Planck's Equation.*” Thescienceandmathszone 2023. <https://thescienceandmathszone.com/the-photoelectric-effect-photons-and-plancks-equation/>.
- [14] Einstein, A. 1905. “*On a Heuristic Point of View Concerning the Production and Transformation of Light.*” Princeton. 2023. <https://einsteinpapers.press.princeton.edu/vol2-trans/100>.
- [15] Franklin, Allan. 2023. “*Right Experiment, Wrong Theory: The Stern-Gerlach Experiment.*” Stanford Encyclopedia of Philosophy. 2023. <https://plato.stanford.edu/entries/physics-experiment/app5.html#:~:text=Stern%20and%20Gerlach%20had%20assumed,had%20not%20considered%20this%20possibility>.
- [16] “*Significance of de-Broglie Waves.*” QS Study. 2023. <https://qsstudy.com/significance-of-de-broglie-waves/>.

- [17] The Dialogues. “*What is Heisenberg’s Uncertainty Principle?*” 6 Nov 2016. Medium. 25 Sep 2023. <https://medium.com/thedialogues/what-is-heisenbergs-uncertainty-principle-7d8b44e39b07>.
- [18] Fermat’s Library. “*Dirac’s Equation (1928).*” 21 Nov 2020. Twitter. 25 Sep 2023. <https://twitter.com/fermatslibrary/status/1330155012752863238>
- [19] Dirac, Paul. “*The Quantum Theory of the Emission and Absorption of Radiation.*” 1927. St John’s College, Cambridge, and Institute for Theoretical Physics, Copenhagen. 27 Sep 2023. http://www.ffn.ub.es/luisnavarro/nuevo_maletin/Dirac_QED_1927.pdf.
- [20] Kaiser, David. “*Physics and Feynman’s Diagrams.*” 2005. Massachusetts Institute of Technology, The Scientific Research Society. 27 Sep 2023. <https://web.mit.edu/dikaiser/www/FdsAmSci.pdf>.
- [21] Wikipedia, “*Feynman diagram picture.*” 2021. Wikipedia. 27 Sep 2023. https://en.wikipedia.org/wiki/Feynman_diagram.
- [22] Mangano, Michelangelo. “*INTRODUCTION TO QCD.*” CERN <https://cds.cern.ch/record/454171/files/p41.pdf>.
- [23] G. Ecker, “*Quantum chromodynamics.*” CERN. <https://cds.cern.ch/record/943008/files/p55.pdf>.
- [24] Wilczek, Frank. “*QCD MADE SIMPLE.*” August 2000. PHYSICS TODAY; American Institute of Physics. 28 Sep 2023. https://frankwilczek.com/Wilczek_Easy_Pieces/298_QCD_Made_Simple.pdf.
- [25] Sophie, “*The Nuclear Atom.*” 28 May 2019. SCIENCE AND MATHS REVISION. 28 Sep 2023. <https://www.scienceandmathsrevision.co.uk/topic/the-nuclear-atom/>.
- [26] Weinberg, Steven. “*The Making of the Standard Model.*” 2003. CERN. 28 Sep 2023. <https://cds.cern.ch/record/799984/files/0401010.pdf>.
- [27] Kibble, Tom. “*The Standard Model of Particle Physics.*” December 2014. ResearchGate. 28 Sep 2023. https://www.researchgate.net/publication/269577741_The_Standard_Model_of_Particle_Physics.
- [28] Rovelli, Carlo. “*Loop quantum gravity.*” November 2003. PHYSICS WORLD; Hamilton College. 29 Sep 2023. <https://academics.hamilton.edu/physics/smajor/Papers/rovelli03.pdf>.
- [29] Rovelli, Carlo. “*Loop Quantum Gravity.*” 15 July 2008. Research Gate. 29 Sep 2023. https://www.researchgate.net/publication/260657560_Loop_Quantum_Gravity.
- [30] Ashtekar, Abhay. “*Loop quantum cosmology: an overview.*” 17 December 2008. Inspire HEP. 29 Sep 2023. <https://inspirehep.net/files/b1d179b8d40d3797091bbb4a7c320437>.
- [31] Bojowald, Martin. “*Loop Quantum Cosmology.*” 8 December 2005. Research Gate. 29 Sep 2023. https://www.researchgate.net/publication/277181408_Loop_Quantum_Cosmology.
- [32] Tong, David. “*String Theory.*” January 2009. University of Cambridge. 1 Oct 2023. <https://www.damtp.cam.ac.uk/user/tong/string/string.pdf>.

- [33] Becker, Katrin; Becker, Melanie; Schwarz, John. “*String Theory AND M-Theory.*” 2006. Cambridge University Press. 1 Oct 2023.
<http://www.nucleares.unam.mx/~alberto/apuntes/bbs.pdf>.
- [34] Hooft, Gerard 't. “*INTRODUCTION TO STRING THEORY.*” 14 May 2004. Utrecht University. 1 Oct 2023.
<https://webspacescience.uu.nl/~hooft101/lectures/stringnotes.pdf>.
- [35] “*Elementary Particle Physics.*” University of Catania, Department of Physics and Astronomy. University of Catania, Department of Physics and Astronomy. 2023.
<https://www.dfa.unict.it/en/content/elementary-particle-physics>.
- [36] Steber, George. “*DARK ENERGY AND THE EXPANDING UNIVERSE.*” Nuts and Volts. 2023.
<https://www.nutsvolts.com/magazine/article/dark-energy-and-the-expanding-universe>.
- [37] “*DIFFERENT THEORIES.*” String Theory. 2023.
<https://stringtheoryphysics.weebly.com/different-theories.html>.
- [38] Johnston, Hamish. 2022. “*How the Stern–Gerlach experiment made physicists believe in quantum mechanics.*” physicsworld. 2023.
[https://physicsworld.com/a/how-the-stern-gerlach-experiment-made-physicists-believe-in-quantum-mechanics/./](https://physicsworld.com/a/how-the-stern-gerlach-experiment-made-physicists-believe-in-quantum-mechanics/)

Beyond Force: Korea, Nonviolent Coercion, and the Promotion of the Comfort Women System By Kate Lee

“They had taken everything away from me: my youth, my self-esteem, my dignity, my freedom, my possessions, and my family.”¹ These haunting words from Lee Yong Soo, a former Korean “comfort woman,”² echo the profound suffering endured by countless other comfort women—their numbers estimated to range between 100,000 to 200,000—from 1932 to 1945.³ The comfort women system was organized and operated by the Imperial Japanese Army during World War II, and it forcibly recruited young women from occupied territories, primarily Korea, to serve in a system of nonconsensual sexual slavery for Japanese soldiers. Women were subjected to not only sexual exploitation, but also physical abuse and psychological torment in the network of brothels euphemistically named comfort stations. Japan's colonial rule over Korea, spanning 35 years, laid the foundation for the atrocities against Korean women during this period by impoverishing Korean families and intensifying their susceptibility to exploitation. Aided by a pre-existing culture of obedience, the Japanese government maintained strict command over both Japanese and Korean society, dictating people's beliefs, access to information, and capacity to protest. With this immense power, the Japanese state created a highly formalized organization: the comfort women system. Although Japan's reign ended in 1945, the legacy of the comfort women system is long lasting; it continues in the form of tense Japan-Korea relations, but more importantly, the legacy is visible in the scars of former comfort women who were subjected to abuse and terror. Given the stakes, it is critical that we understand all of the causes of this system. The use of violence and physical force are readily apparent in examining the creation of the comfort women system. But despite the prevalence of physical coercion in establishing and normalizing the comfort women system during World War II, nonviolent coercion in the form of socioeconomic conditions, state control of media and education, and legal deception were equally significant in promoting the comfort women system and in silencing women after the war.

It is reasonable that scholars have focused so intently on violent coercion in the establishment of the comfort women system in Japan.⁴ Violence played a crucial role in the recruitment, containment, and psychological control of comfort women. Japanese recruiters, government officials often aided by the police, pursued single girls under the official pretense of

¹ Yong Soo Lee, “Protecting the Human Rights Of Comfort Women” (Hearing of the Asia, the Pacific, and the Global Environment Subcommittee of the House Foreign Affairs Committee, 2007).

² The author acknowledges that the Japanese military-coined term “comfort women,” a euphemism for women who provided nonconsensual sexual services to the Japanese Imperial Army and lived under conditions of sexual slavery, masks the severity of the crime. Moving forward, the term will not be in quotation marks because in modern culture and literature, people recognize that the meaning behind comfort women is a system of sexual slavery.

³ Dai Sil Kim-Gibson, *Silence Broken: Korean Comfort Women* (Mid-Prairie Books, 1999), 40.

⁴ Carmen M. Argibay, “Sexual Slavery and the Comfort Women of World War II” (Berkeley Law Scholarship Repository, 2003); Yong-Shik Lee, Natsu T. Saito & Jonathon Todres, *The Fallacy of Contract in Sexual Slavery: A Response to Ramseyer's "Contracting for Sex in the Pacific War"* (Michigan Journal of International Law, 2021).

Kunro Jungshindae (Voluntarily Committing Body Corps for Labor), part of a human resource mobilization by Japan for military purposes.⁵ Kim Bok Sun, a former Korean comfort woman, recounted her experience: “They grabbed me by force and I was dragged out of our yard.”⁶ This forceful recruitment was utilized to meet quotas and was not uncommon, with Japanese soldiers often raiding Korean villages to kidnap women for service in comfort stations and, according to the summary findings of a 1993 Japanese report that collated documents from the Japanese state and former comfort women, “recruiters resorted in many cases to coaxing and intimidating these women to be recruited against their own will.”⁷ The violence extended to the comfort stations, where women underwent both physical and emotional suffering at the hands of Japanese soldiers and officials. Hwang Keum Ju, another former Korean comfort woman, detailed the extent of the brutality, recalling how many women were “beaten to death, sometimes shot to death or rotted to death with venereal disease, not to speak of suicides” and that they were “beaten every day.”⁸ Dehumanization coupled with the regularity of violence impacted comfort women psychologically; Hwang explained there was “no time for feelings.”⁹ Restraining the comfort women emotionally as well as physically through violence undoubtedly contributed to the system’s continuation and normalization. But because so much academic attention has been paid to the significance of violence in the comfort women system, this essay seeks to highlight other forces. Underlying trends, conditions, and attitudes in 1930s and 1940s Japan enabled this system as much as physical coercion did.

In comparison to the overt violence often associated with the comfort women system, nonviolent coercion can be overlooked due to its subtlety and indirect nature. Socioeconomic factors—a form of nonviolent coercion—like Korea’s colonial status and the patriarchal family structure allowed the comfort women system to proliferate with minimal opposition. Japan’s colonization of Korea from 1910 to 1945 granted Japan the unique ability to take advantage of Korea’s resources. In fact, Japan controlled all of Korea’s natural resources and its fishery and forestry, as well as most industries and commercial markets.¹⁰ More specifically, Japanese traders collected a large part of or even all of the Korean farmers’ crops at harvest time after routinely loaning them money to buy seed and other items.¹¹ This exploitation exacerbated the already extreme hardship of the farmers while simultaneously harming the Korean economy. Japanese recruiters capitalized on the prevailing socioeconomic difficulties by targeting young girls from families significantly in debt to Japanese-sanctioned financial institutions. The Korean women at Myitkyina, Burma, testified that “recruiters told them that volunteering to work at hospitals was a way to pay off their families’ debts.”¹² Additionally, the Korean patriarchal family structure

⁵ Dai Sil Kim-Gibson, *Silence Broken: Korean Comfort Women* (Mid-Prairie Books, 1999), 40-41.

⁶ *Ibid.*, 43.

⁷ Dai Sil Kim-Gibson, *Silence Broken: Korean Comfort Women* (Mid-Prairie Books, 1999), 41. Larry Niksch, *Japanese Military’s “Comfort Women” System* (Congressional Research Service, 2007), 10.

⁸ Dai Sil Kim-Gibson, *Silence Broken: Korean Comfort Women* (Mid-Prairie Books, 1999), 25-27.

⁹ *Ibid.*, 26.

¹⁰ *Ibid.*, 37.

¹¹ Dai Sil Kim-Gibson, *Silence Broken: Korean Comfort Women* (Mid-Prairie Books, 1999), 33.

¹² Larry Niksch, *Japanese Military’s “Comfort Women” System* (Congressional Research Service, 2007), 11.

hindered women's ability to challenge their families' decisions to send them away or inquire about the location or nature of their new work, let alone question the existence of the comfort women system. The societal pressure to fulfill their duty as daughters, support their families, or contribute to their brothers' education compelled many of these young women to be willing to work.¹³ According to Hwang Keum Ju, a former Korean comfort woman, "women at that time were not expressive."¹⁴ Furthermore, the Japanese state actively suppressed the local culture, banning newspapers published in Korean, requiring the study of the Japanese language in all public schools, and even mandating Koreans to change their family and personal names to Japanese under Ordinance No. 20 of 1939.¹⁵ By destroying Korean identity and weakening a sense of collective resistance, Korea was left vulnerable to coercion and manipulation. Thus, a prewar society with a culture of male prerogative combined with an impoverished Korean countryside contributed to the relative ease with which women were coerced into sexual slavery and the lack of public resistance to the comfort women system.

The Japanese state further employed nonviolent coercion through strict control of the media and education, inculcating ideas that aligned with the state's and discouraging intellectual curiosity; thus, the existence of the comfort women system was made obscure or was accepted as a necessary evil of war. To comprehend the feasibility of mass indoctrination within Japanese society, it is crucial to recognize the influence of extreme patriotism and the normalization of sacrifice. Fervent loyalty to the emperor was ingrained in the Japanese populace, with willingness to give one's life for one's nation honored as both an ideal and a norm.¹⁶ Such normalization of sacrifice not only fostered a prewar culture of obedience, but also contributed to the perception that the comfort women were simply fulfilling their duty during the war. The Japanese government also utilized this environment of control for both censorship and overt propaganda in the media during wartime. The Cabinet Information Bureau, a powerful agency staffed by military men on active duty, restricted newspapers and magazines from publishing articles with anti-war or anti-military opinions that appeared to reduce civilian support for the military, views of foreign newspapers, and any slandering of Japan's policies.¹⁷ By portraying their decisions positively in the media, a form of information that uniquely reaches a broad audience and molds cultural norms, the Japanese government was able to garner mass support from the public, regardless of the information's accuracy. Another powerful tool used by the Japanese government was the public education system. After 1904, the national governmental Ministry of Education took charge of compiling all elementary-school texts.¹⁸ For example, all elementary schools in Saitama Prefecture were informed to stress war and patriotism in every

¹³ Dai Sil Kim-Gibson, *Silence Broken: Korean Comfort Women* (Mid-Prairie Books, 1999), 41.

¹⁴ *Ibid*, 15.

¹⁵ *Ibid*, 36.

¹⁶ "Asia For Educators," *Japan's Quest for Power and World War II in Asia* (Columbia University).

¹⁷ Saburo Ienaga, *The Pacific War : World War II and the Japanese, 1931-1945* (New York : Pantheon Books, 1978), 100.

¹⁸ Saburo Ienaga, *The Pacific War : World War II and the Japanese, 1931-1945* (New York : Pantheon Books, 1978), 20.

subject.¹⁹ Japanese authorities even required emperor-centered patriotic rituals in school, like venerating the imperial photograph on the opening day each year.²⁰ With this glorification of the emperor and state, questioning the government's choices, let alone knowing of or recognizing the immorality of institutions like the comfort women system was unlikely. The lack of vocalized doubt, especially towards authority, was amplified because formal education for most individuals in prewar Japan ended with elementary school, meaning impressionable children would carry on these beliefs to adulthood.²¹ Thus, by stifling intellectual inquiry and instilling beliefs aligned with the state's interests, the Japanese government shaped public perception and compliance with their actions regarding the comfort women system. The combination of controlling culture in Japan and command of information channels, along with the flawed notion that sexual violence is inevitable during war, effectively minimized pushback and perpetuated the comfort women system.

In addition to controlling the media and education, the Japanese government meticulously organized legal stratagems like labor drafts and comfort station rules that legitimized the comfort women system. Mass rape in wartime requires significant planning, and Japan's comfort women system was not an exception.²² Comfort stations were set up by explicit order of the military, with the Imperial army and navy directing the procurement of women, some through private contractors.²³ Indeed, Hwang Keum Ju, a former Korean comfort woman, described how her family received a notice called *Kunro Jungshindae*, an official labor draft, to go work in a factory.²⁴ Numerous women who went under this voluntary work system were made sex slaves for Japanese soldiers at the comfort houses, contrary to what the notice promised.²⁵ The draft's sanctioning by the Japanese government instilled the order with a sense of authenticity, convincing girls like Hwang to believe they were contributing to the war effort. Hwang further detailed that "the military provided the accommodations and posted detailed regulations for soldiers, traders and women. 'Comfort House Regulations' specified who could visit, the schedule, fees, and rules regarding disease prevention."²⁶ Moreover, the Japanese military took identification of comfort women, selection of proper recruiters, norms for hygiene, and enforcement of temperance seriously.²⁷ These thorough regulations depicted the comfort women system as a well-organized establishment with responsible leaders, order, and structure. The official nature of the system prevented panic or suspicion amongst those who were aware of its existence, like soldiers, and instead increased credibility. Not only were the soldiers allowed to

¹⁹ Ibid, 23.

²⁰ Ibid, 21.

²¹ Ibid, 20.

²² Joanna Bourke, "A Global History of Sexual Violence" (Cambridge University Press, 2020), 159.

²³ Onozawa Akane, *Denying the Comfort Women* (Routledge, 2019), 75.

²⁴ Dai Sil Kim-Gibson, *Silence Broken: Korean Comfort Women* (Mid-Prairie Books, 1999), 18.

²⁵ Ibid, 38.

²⁶ Ibid, 50.

²⁷ Kyu-hyun Jo, "For the Sake of Providing Comfort to All Imperial Soldiers Progressing on Every Front": An Analysis of Regulations on the Establishment and Management of a Japanese Panopticon Over "Comfort Women" (International Journal of Korean History, 2023).

believe that their violent treatment of comfort women was permitted, but they were also able to subconsciously justify their actions, placing the responsibility on the higher-ups who managed and established the system. Consequently, with a veneer of legality and lack of retaliation, the comfort women system was maintained.

In the post-war era, silence surrounding the comfort women issue was perpetuated by nonviolent pressures, including the stigma attached to sexual violence, the Korean state's insistence that the nation move forward, and Japan's reluctance to acknowledge its historical responsibility. To understand all the factors contributing to the longevity of the silence, it is of course important to mention the role of psychological trauma—suffering that can remain for years or decades after an assault—in potentially deterring victims from coming forward. However, trauma is often used as an excuse by those responsible for hearing and acting upon survivors' testimonies. Leaders point to the difficulty of revisiting upsetting experiences to explain away women's silence after the war. Instead, the primary catalyst for the continuation of silence lies in the external obstacles that victims who were ready to share their experience faced. A barrier applicable to not only former comfort women, but victims to sexual abuse across the world in the past two centuries, is the hesitancy in reporting because of the likelihood of publicity, retaliation, and discrimination.²⁸ In fact, silence is encouraged in many cultures, where victims face humiliation due to assumptions that only shameless women talk publicly about sexual matters.²⁹ Unfortunately, this was true for many former comfort women in post-war Korea who found themselves discarded by their own communities as a result of conservative societal values and of the negative stereotypes tied to sexual subjects.³⁰ It took over four decades of silence for one former Korean comfort woman, Kim Hak Soon, to give a public testimony and demand an official apology from Japan in 1991.³¹ Was four decades of silence truly necessary for the stories of comfort women to become a public conversation? Long before Kim went public, Hwang Keum Ju, the woman who talked about Kunro Jungshindae, tried to report about the Korean comfort women. She told the then-First Lady her story and about the many women who served Japanese soldiers, and she asked “why the Korean government did not do something for us.” The First Lady responded, “Please don't ever repeat this story... Korea and Japan have already signed a treaty to take care of the matters concerning the two countries during the colonial period... Korea needs to move forward.”³² Pressure from the Korean state to move on stemmed from the unstable status of post-war Korea; Korea and Japan had signed the 1965 Treaty that established basic diplomatic relations between the two countries. Discussing the comfort women issue would bring back tension and disrupt the uncertain peace. Also, the unstable socioeconomic position of post-war Korea amplified the often already-deprived former

²⁸ Joanna Bourke, “A Global History of Sexual Violence” (Cambridge University Press, 2020), 150.

²⁹ Ibid, 151.

³⁰ Ibid, 153-154.

³¹ Rin Ushiyama, *‘Comfort women must fall’? Japanese governmental responses to ‘comfort women’ statues around the world* (Memory Studies, 2021).

³² Dai Sil Kim-Gibson, *Silence Broken: Korean Comfort Women* (Mid-Prairie Books, 1999), 20-21.

comfort women, further hindering them from speaking out.³³ Even after Kim Hak Soon and other victims' testimonies, the comfort women were not fully acknowledged. Instead, by the late twentieth century, the comfort women issue became not only deeply politicized in Korean public discourse, but also a source of diplomatic tension between Japan and Korea. The Japanese government made concerted efforts to stop the installation of, or demand the removal of, memorials dedicated to comfort women.³⁴ This diplomatic discord diverted focus away from survivors' main goal of receiving recognition from the Japanese government. Ultimately, the silencing of comfort women both before and after publicly speaking out can be attributed to coercion of a nonviolent nature: the societal shame associated with sexual violence, and the Korean and Japanese state's urge to leave the past in the past.

"I must stand up for myself and the others... I was robbed of my youth, and I want [Shinzo Abe] to apologize before I die."³⁵ Lee Yong Soo's wish for the former prime minister of Japan to acknowledge her suffering encapsulates not only the unyielding spirit of resilience, but also the long-lasting pain many of the women subjected to sexual slavery by the Japanese military during World War II experienced. Given the immense physical and emotional anguish the comfort women endured during and after the war, it is critical we gain a holistic view of the causes of the complex comfort women system. While physical coercion undoubtedly played a prominent role in advancing the system, the equally influential nonviolent coercion, like Korea's poor socioeconomic status, the Japanese government's strategic manipulation of media and education, and its legal tactics, sustained the system and suppressed women after the war. The Japanese government's role was substantial. Their colonization of Korea impoverished families and increased vulnerability to exploitation. Their authoritative control over both Japanese and Korean cultures shaped perspectives, controlled flows of information, and quelled dissent. Their purposeful organization of the comfort women system legitimized it and lessened backlash. But their apology remains inadequate, especially given the severity of the war crime. Victims continue to await proper reparations, such as a formal apology, legal compensation, a thorough investigation, and recognition of the atrocities they suffered in memorials and Japanese textbooks.³⁶ Many former comfort women like Lee Yong Soo have campaigned to expose the abuses, demanded Japanese atonement, and testified before commissions and legislatures.³⁷ Studying this issue can not only help bring justice to comfort women, but it might shed light on, or even help to prevent, similar violations of human rights in other countries and time periods. Wartime sexual violence—as well as the silencing of victims after—is a pattern. But it does not have to be.

³³ Rin Ushiyama, *'Comfort women must fall'? Japanese governmental responses to 'comfort women' statues around the world* (Memory Studies, 2021).

³⁴ Ibid.

³⁵ Pamela Constable, *70 years later, a Korean 'comfort woman' demands apology from Japan* (The Washington Post, 2015).

³⁶ Beverly Milner (Lee) Bisland, Jimin Kim, and Sunghee Shin, *Teaching about the Comfort Women during World War II and the Use of Personal Stories of the Victims* (Association for Asian Studies, 2019).

³⁷ Pamela Constable, *70 years later, a Korean 'comfort woman' demands apology from Japan* (The Washington Post, 2015).

Works Cited

- Akane, Onozawa. "Denying the Comfort Women." Routledge, 2019.
<https://www.routledge.com/Denying-the-Comfort-Women-The-Japanese-States-Assault-on-Historical-Truth/Nishino-Kim-Onozawa/p/book/9780367349660>.
- Argibay, Carmen M. "Sexual Slavery and the Comfort Women of World War II." *Berkeley Law Scholarship Repository*, vol. 21, no. 2, 2003.
<https://lawcat.berkeley.edu/record/1118669?ln=en>.
- Bisland, Beverly Milner (Lee), et. al. "Teaching about the Comfort Women during World War II and the Use of Personal Stories of the Victims." *Association for Asian Studies*, 2019.
<https://www.asianstudies.org/publications/ea/archives/teaching-about-the-comfort-women-during-world-war-ii-and-the-use-of-personal-stories-of-the-victims/>.
- Bourke, Joanna. "A Global History of Sexual Violence." *Cambridge University Press*, 2020.
<https://www.cambridge.org/core/books/abs/cambridge-world-history-of-violence/global-history-of-sexual-violence/0A071B08367803602D56840D6CFF1C63>.
- Constable, Pamela. "70 years later, a Korean 'comfort woman' demands apology from Japan." *The Washington Post*, 2015.
https://www.washingtonpost.com/local/70-years-later-a-korean-comfort-woman-demands-apology-from-japan/2015/04/22/d1cf8794-e7ab-11e4-9767-6276fc9b0ada_story.html.
- Ienaga, Saburo. *The Pacific War: World War II and the Japanese, 1931-1945*. Cambridge University Press, 2011.
<https://www.cambridge.org/core/journals/journal-of-asian-studies/article/abs/pacific-war-world-war-ii-and-the-japanese-19311945-by-saburo-ienaga-translated-by-frank-baldwin-new-york-pantheon-books-1978xvi-317-pp-notes-index-1000/5674400B33F2A92BE98FB4EEDE39ABA7>.
- "Japan's Quest for Power and World War II in Asia." *Asia for Educators, Columbia University*. Accessed June 15, 2023. http://afe.easia.columbia.edu/special/japan_1900_power.htm.
- Jo, Kyu-hyun. "'For the Sake of Providing Comfort to All Imperial Soldiers Progressing on Every Front': An Analysis of Regulations on the Establishment and Management of a Japanese Panopticon Over 'Comfort Women.'" *International Journal of Korean History*, 2023. <https://doi.org/10.22372/ijkh.2023.28.1.63>.
- Kim-Gibson, Dai Sil. "Silence Broken: Korean Comfort Women." Parkersburg, Iowa: Mid-Prairie Books, 1999.
<https://archive.org/details/silencebrokenkor0000kimg/mode/2up>.
- Lee, Yong Soo. "Protecting the Human Rights Of Comfort Women." *ETAN*. Hearing of the Asia, the Pacific, and the Global Environment Subcommittee of the House Foreign Affairs Committee, 2007.
<https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1026&context=humtraffdata>.
- Lee, Yong-Shik, et. al. "The Fallacy of Contract in Sexual Slavery: A Response to Ramseyer's 'Contracting for Sex in the Pacific War.'" *Michigan Journal of International Law*, vol. 42, no. 2, 2021. <https://doi.org/10.36642/mjil.42.2.fallacy>.

Niksch, Larry. "Japanese Military's 'Comfort Women' System." *Congressional Research Service*, 2007.

<http://japanfocus.org/data/CRS%20Comfort%20Women%203%20Apr%2007.pdf>.

Ushiyama, Rin. "'Comfort women must fall'? Japanese governmental responses to 'comfort women' statues around the world." *Memory Studies*, vol. 14, no. 6, 2021.

<https://doi.org/10.1177/17506980211054308>.

Adolescent Substance Abuse: Role of Environmental and Psychological Factors by Hyowon Han

Abstract

Substance abuse among adolescents remains a pressing issue in the United States, leading to severe physical, mental, and social consequences. Adolescence is a critical period marked by changes in brain structure, making it a vulnerable time for experimentation with psychoactive substances. This review explores environmental and psychological factors contributing to adolescent substance abuse and proposes intervention strategies targeting these risk factors. The identified factors include peer influence, family environment, urban vs. rural residence, physical inactivity, early-life stress, and personality traits. Findings suggest that while family dynamics and early-life stress increase the risk of substance abuse, peers, particularly best friends, exert more influence than parents. Disparities in parenting styles and substance accessibility between rural and urban areas contribute to variations in adolescent substance use. Physical inactivity also poses a risk, suggesting potential interventions involving exercise. Substance use often begins as a coping mechanism for adolescents experiencing early-life stress. Additionally, specific personality traits—high neuroticism, extraversion, openness to experience, low agreeableness, and low conscientiousness—impact the initiation of substance use. This review underscores the importance of identifying environmental factors in substance abuse and underscores the need for further research to tailor intervention methods accordingly. School-based programs leveraging peer influence, exercise interventions, and community-based approaches are recommended.

Introduction

The term "substance" in the context of substance abuse refers to any psychoactive compound capable of inducing physical, mental, and social problems, irrespective of its legal status. Nicotine, alcohol, cannabinoids, opioids, depressants, stimulants, and hallucinogens all fall under this category, with some being legal, illegal, or strictly available for medical purposes. The misuse of these substances presents a significant and pervasive issue in the United States, manifesting as binge drinking and the unauthorized consumption of drugs, including opioid pain relievers and sedatives, as reported by a staggering 61 million and 44 million people in the United States, respectively (McLellan, 2017).

These problems, rooted in persistent substance use, are categorized as substance use disorders, with some cases evolving into chronic or severe addiction. Addiction is characterized by a destructive three-step cycle: binge, withdrawal, and craving, progressively intensifying over time (Koob & Volkow, 2018). Diagnostic criteria for substance use disorders include parameters such as the quantity and duration of substance consumption, an inability to manage substance use, increased tolerance, withdrawal symptoms, and continued usage despite adverse social and dangerous consequences (McLellan, 2017). Recognizing and intervening during the mild or moderate stages of substance use disorders is critical, with several intervention methods available, including medications that target neurological circuits, behavioral therapies,

contingency management, and dual-diagnosis treatments, all tailored to the individual's lifestyle, underlying causes, and symptom severity.

However, substance use disorder is still a complex problem to solve, as repeated use fosters alterations in brain structure, which could undermine the individuals' capability to withstand the impulses of substance use, leading to frequent relapses. More specifically, drug effects and withdrawal symptoms are due to their neurobiological mechanism, which is associated with the dopaminergic reward system, prefrontal cortex, neurocircuitry in the amygdala, and several neurotransmitters such as dynorphin, serotonin, oxytocin, endocannabinoids, and acetylcholine. For instance, during withdrawal, the extended amygdala experiences a decrease in dopamine release in the reward system alongside increased corticotropin-releasing factor (CRF) and dynorphin levels, exacerbating withdrawal symptoms (Koob & Volkow, 2018).

This proves the adolescents' increased susceptibility to substance abuse, as their amygdala and prefrontal cortex are still developing, and their heightened emotional reactivity could lead to impulsive decision-making, such as experimentations with drugs. In fact, 90 percent of individuals who developed substance use disorder started using illegal substances, smoking, and drinking prior to reaching the age of 18. Furthermore, approximately 50 percent of high school seniors have experimented with illicit drugs, with 20 percent using prescription medications for nonmedical purposes. Shockingly, one in every eight high school students grapples with substance use disorder (Garofoli, 2020).

Given the persistently critical nature of adolescent substance abuse, it is imperative to identify vulnerability factors that contribute to this issue. While hereditary factors account for 40-60 percent of addiction, stemming from genetic variations in neurological mechanisms (Koob & Volkow, 2018), it is equally crucial to investigate the modifiable environmental and psychological factors that play a pivotal role in prevention and intervention for adolescent substance use. In this review, six environmental/psychological risk factors of adolescent substance abuse - peer influence, family environment, urban vs. rural residence, physical inactivity, early-life stress, and personality traits - and the associated methods of intervention are analyzed, emphasizing the importance of social interactions and education in adolescence substance abuse and further proposing an optimal method for intervention.

Methods

The study employed a comprehensive research methodology focused on gathering data from peer-reviewed articles available in scientific databases, specifically PubMed and ScienceDirect. The research process was divided into two primary sections. The first section centered on investigating various environmental factors contributing to substance abuse, while the second section delved into the examination of intervention methods.

To refine the search and ensure the relevance of the retrieved articles, a systematic keyword approach was adopted. In the initial section dedicated to environmental factors, the following keywords were utilized: environmental factors, substance abuse, drug addiction,

adolescence, peer influence, family environment, mental health, physical inactivity, and personality. These keywords were instrumental in narrowing down the search results and identifying pertinent articles.

In the subsequent section dedicated to intervention methods, specific keywords such as interventions, treatments, adolescent substance abuse, substance use disorders, and prevention were employed to pinpoint research papers, and studies focused on strategies for addressing substance abuse in adolescents.

Throughout the source selection process, paramount importance was given to both relevance and credibility. Articles that directly contributed to the understanding of the subject matter were prioritized, and reputable sources were relied upon. Additionally, a diverse range of source types, including reviews, meta-analyses, longitudinal studies, and surveys, were considered to ensure a comprehensive analysis of the topic.

Environmental Factors

Risk factors of substance abuse could be classified into genetic and environmental factors. Considering that adolescents are strongly influenced by social factors - such as peer pressure, media influence, and societal standards -, the 'nurture' aspects of risk factors could be particularly more dominant in the younger demographics. These environmental factors are usually the components that trigger individuals' start to substance use and also cause cravings or relapses. For instance, traumatic life events, such as child abuse, bereavement, and homelessness, could trigger the temptation to utilize substances to cope, potentially leading to substance use disorders or addiction.

Peer Influence

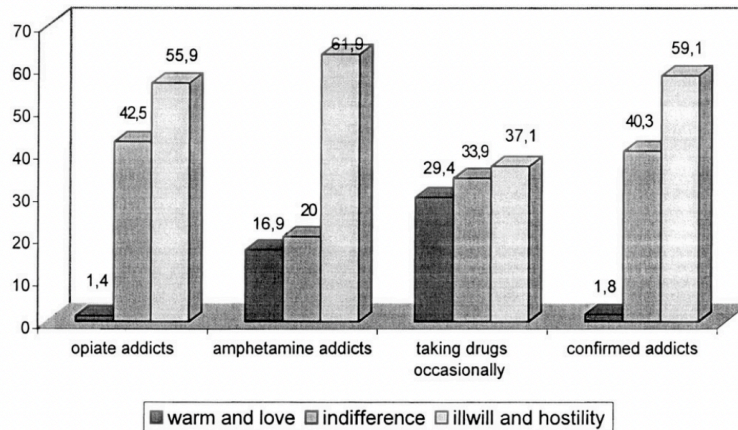
Substance use due to peer influence during adolescence could be attributed to both direct - such as being pressured into taking drugs - and indirect influences, which include substance use for social purposes and settings favorable to substance use. For instance, 84 percent of the individuals who called a cocaine hotline stated that the reason they started using drugs was active peer pressure, as they were forced to take illicit substances. There is also a result that peers, especially their closest friend, had more significant influence than parental wishes in one's substance use, supported by a statistic that 24 percent of individuals taking illicit drugs have a best friend that also takes drugs (Swadi 1999). The risk-taking behaviors and limited self-reliance of adolescents also contribute to the high impact of peers, as they are less likely to resist the temptation of substance use. On the other hand, according to a survey in Hong Kong secondary schools, adolescents with friends who object to smoking and drinking had significantly lower rates of substance use, highlighting the possibility of peers discouraging one another's drug use (Loke 2013). This stresses the importance of peer relationships in adolescents and poses a prospect of a prevention method that involves secondary school education and collaborations with peers.

Family Environment

Although whether children with alcoholic parents have an increased risk of alcohol use due to genetics is controversial, the consensus is that the general family structure and parenting style impact adolescent substance use, as parents serve as role models for a healthy lifestyle and behavior. Thus, when considering parental alcoholism as an aspect of parenting style, it could lead to increased substance use in the children due to the lack of parental monitoring and discipline, as well as the stress from the environment itself. Family dynamics such as parental divorce, family disruption, and lack of intimacy could also influence adolescent substance use, while the general personality traits of each parent also act as a factor. More specifically, the family environment with an overly involved mother and a distant father is known to cause a higher risk of substance abuse in their children (Swadi 1999). Moreover, adolescents in single-parent families were more likely to consume illicit drugs and smoke cigarettes, while individuals from intact families had a lower rate of becoming regular drinkers and smokers (Loke 2013). Finally, as shown in Figure 1, the family atmosphere of most substance addicts was classified as ‘ill will and hostility,’ in contrast to the significantly low rate of the ‘warm and love’ type, especially in opiate addicts (Jedrzejczak 2005).

FAMILY FACTORS SINGLED OUT IN THE GROUP OF RECRUITS CONSTANTLY TAKING DRUGS

Family Factors (%)				
Complete	Incomplete	Foster	Outside the Family	Own Family
21.9	42.4	0	0	35.7



*Figure 1: Family atmosphere and drug use by recruits. Source: Jedrzejczak 2005

Urban vs. Rural Residence

As cultures and beliefs differ by geographic location, the accessibility of various drugs - mainly depending on whether the location is rural or urban - affects adolescent substance use.

In general, urban locations, especially downtown areas, are more prevalent in marijuana, cocaine, and hallucinogens, while rural populations have increased adolescent alcohol and tobacco use due to more flexible attitudes regarding those issues. According to a survey of 513,909 middle and high school students in either rural or urban areas of Georgia, United States, rural students reported greater access to legal substances such as tobacco and steroids. In comparison, urban students had greater access to marijuana, prescription drugs, and inhalants. These differences in accessibility were primarily due to different parenting styles - attributed to socioeconomic status, ethnicity, and personal beliefs - which affected their attitude towards substance use to either be “restrictive” or “permissive” (Warren 2015).

Physical Inactivity

Physical activity is a significant indicator of individuals’ overall health as it is a risk factor for various diseases and symptoms, such as heart disease, obesity, diabetes, and osteoporosis. In a longitudinal experiment conducted based on 4240 individual twins in adolescence, the individuals were classified into three groups - persistently inactive, occasional exercisers, and persistent exercisers - and were observed throughout a time period of 6 to 7 years with a questionnaire regarding alcohol and illicit substance use. Only twins were used to rule out the possibility of genetic factors and to observe the contrast solely in the environmental factors - physical activity. As a result, the correlation between physical inactivity and drug use was found to be stronger than that between physical inactivity and alcohol use. This may be due to physically active individuals having health-related goals and recognizing the detrimental impact of substance use, while alcohol use is also connected to social activities and is not perceived as harmful as illicit substances. Additionally, as young women tended to be less physically active than young men, they had a higher risk of experiencing problems due to alcohol and drug use (Korhonen 2009).

Early-Life Stress

Early-life stress is an umbrella term for child abuse, stressful life events such as parental loss and divorce, and emotional neglect. When the start age and motives in drinking were observed in US drinkers, sixty-six percent of the respondents had one or more adverse events in childhood. Such events were parental divorce, family members with drinking problems or mental illness, and sexual abuse. Most individuals with early-life stress were shown to use drinking as a coping mechanism for stress, starting alcohol use early in adolescence (Rothman 2008).

Psychological Factors

Personality and Behavior

The five personality traits that increase an individual's vulnerability to substance abuse are commonly addressed as the “big five personality traits”: high neuroticism, extraversion, openness to experience, low agreeableness, and low conscientiousness. More specifically, according to a survey based on 980 twin pairs, high neuroticism is related to cannabis and

sedative use, and high extraversion is connected to cocaine and prescription stimulant misuse. Meanwhile, openness was associated with all forms of drug use throughout life (Dash 2023). However, those correlations might also be due to other environmental factors that are common to both twins, such as family environment and early-life stress, rather than each twin's personality. It is still noteworthy that some of these traits are prevalent in adolescence due to the neurological changes that take place during puberty.

Methods of Intervention

As it is essential to prevent or treat substance abuse as early as possible, intervention methods specifically targeted for adolescents are needed. These generally include school-based, family-based, community-based, digital, and policy-related interventions. Among these, digital advertisements are known to influence youth tobacco use when they include strong images, sound, and narrative; however, there is insufficient evidence, and further research is required. Community-based methods had success in increasing adolescent knowledge regarding substance abuse, although the evidence is not strong as well (Das 2016). In this paper, the significance of different environmental factors is recognized, and the intervention methods based on those risk factors - specifically peer influences and physical inactivity - are evaluated.

School-based interventions are successful in preventing the initiation of tobacco use when using the aspects of peer influence and social competencies of adolescents. One of those methods is cooperative learning, which reduces deviant peer clustering, where socially marginalized adolescents aggregate and influence each other to perform certain behaviors such as illicit drug consumption and premature alcohol use. This procedure significantly decreased the rates of alcohol and tobacco use in students as they were encouraged to interact with various friends - chosen by the school - rather than self-selected groups of peers, which has the potential to cause deviant peer clustering (Ryzin 2018).

As physical inactivity was one of the risk factors, regular exercise-related activities could serve as an intervention method. Studies regarding exercise interventions reveal that exercising is especially beneficial to individuals who are in the process of substance use disorder treatment. Individuals who maintained an active lifestyle throughout the treatments had a more extended abstinence period than the ones who did not (Weinstock 2009). Moreover, exercising could be helpful for the reduction of not only feelings of "craving" illicit substances but also stress and anxiety, which is a significant vulnerability factor in adolescent substance abuse.

Discussion

Adolescent substance abuse is influenced by numerous risk factors, with key contributors including peer influence, family environment, urban or rural residence, physical inactivity, early-life stress, and personality traits. However, it is important to acknowledge that the interplay between these factors is as significant as each factor individually, making it challenging to isolate any single determinant. For instance, adolescents experiencing mood dysregulation are at a heightened risk of social marginalization in school, potentially leading to deviant peer clustering

(Swadi, 1999). Moreover, adolescent behavior and peer influence are closely intertwined, as risky behavior and sensation seeking, common traits in adolescents, tend to manifest more prominently in group settings. Additionally, the family environment plays a substantial role in shaping the personality and behavior of adolescents, as evidenced by the increased likelihood of delinquent behavior among children of divorced parents (Loke, 2013).

Furthermore, it is crucial to recognize the significant impact of genetic factors. Genetic influences on substance use escalate during the transition from adolescence to adulthood, accounting for one-third of the variance at age 16 and increasing to one-half at age 18, while the role of environmental factors diminishes from 70 percent to 15 percent between ages 14 and 18 (Meyers, 2013). This underscores the importance of environmental factors, particularly during early adolescence. Importantly, environmental factors are more amenable to modification than genetic factors, which are largely hereditary.

Conclusion

In conclusion, alterations in environmental aspects of life, such as social interactions and education, have the potential to mitigate substance abuse among adolescents. Despite the complexities arising from the interaction of various risk factors, targeted interventions hold promise in addressing this critical issue. The strategies discussed in this review, including exercise-related activities, collaborative learning, and family and community-based education, serve as primary avenues for enhancing both physical and mental well-being while concurrently preventing substance abuse. These interventions are instrumental in guiding adolescents toward a healthy and substance-free adulthood.

Works Cited

- Das, J. K., Salam, R. A., Arshad, A., Finkelstein, Y., & Bhutta, Z. A. (2016, October). Interventions for adolescent substance abuse: An overview of Systematic Reviews. *The Journal of adolescent health : official publication of the Society for Adolescent Medicine*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5026681/>
- Garofoli, M., Yudko, E., Esser, M. B., Griswold, K., Johnston, L. D., Dennis, M., Robertson, E. B., & Knight, J. R. (2020, February 21). Adolescent substance abuse. *Primary Care: Clinics in Office Practice*. <https://www.sciencedirect.com/science/article/pii/S0095454320300154?via=ihub>
- H;, S. (1999, July 1). Individual risk factors for adolescent substance use. *Drug and alcohol dependence*. <https://pubmed.ncbi.nlm.nih.gov/10428362/>
- Koob, G. F., & Volkow, N. D. (2016, August). Neurobiology of addiction: A neurocircuitry analysis. *The lancet. Psychiatry*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6135092/>
- Korhonen, T., Kujala, U. M., Rose, R. J., & Kaprio, J. (2009, June). Physical activity in adolescence as a predictor of alcohol and illicit drug use in early adulthood: A longitudinal population-based Twin Study. *Twin research and human genetics : the official journal of the International Society for Twin Studies*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2723718/>
- Loke, A. Y., & Mak, Y.-W. (2013, August 27). Family process and peer influences on substance use by adolescents. *International journal of environmental research and public health*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3799532/>
- M;, J. (2005, August 1). Family and environmental factors of drug addiction among young recruits. *Military medicine*. <https://pubmed.ncbi.nlm.nih.gov/16173210/>
- McLellan, A. T. (2017). Substance misuse and substance use disorders: Why do they matter in healthcare?. *Transactions of the American Clinical and Climatological Association*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5525418/>
- Meyers, J. L., & Dick, D. M. (2010, July). Genetic and environmental risk factors for adolescent-onset substance use disorders. *Child and adolescent psychiatric clinics of North America*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3661209/>
- RW;, R. E. E. T. (2008, August). Adverse childhood experiences predict earlier age of drinking onset: Results from a representative us sample of current or former drinkers. *Pediatrics*. <https://pubmed.ncbi.nlm.nih.gov/18676515/#full-view-affiliatio>

The Impact of Airport Expansion on Local Communities By Inwoo Hwang

Abstract

The number of commercial flights has been steadily increasing every year for the past two decades and is expected to continue as international travel recovers from COVID-19 restrictions. To meet the demand for flights, many governments have expanded, or will expand, their airports. Statistics show that the majority of growth in the commercial aviation industry has already happened in North America and Europe, but in the coming years, commercial aviation is expected to grow the most on the continents of Africa and Asia. In light of this inevitable shift, it is important to understand the benefits and drawbacks of airport expansion. This review paper examines literature about airports in Tanzania and Hong Kong to demonstrate that, when poorly planned, airport expansion can have devastating effects on local communities. Residents who are forced to move suffer psychological as well as financial setbacks despite financial compensation packages from the government. Those in the areas surrounding the newly expanded airport are impacted by noise and air pollution, which takes a toll on their health and wellbeing. These negative effects are an indication that government policies must be improved when it comes to airport expansion.

Introduction

In our increasingly interconnected and global society, air travel is more important now than ever. In fact, since the early 2000s, the number of commercial flights has increased every year, peaking at 38.9 million flights in 2019 (Statista Research Department). The increased demand for flights over the past few decades means that the demand for aircrafts and airport space has also been increasing. When the number of aircrafts exceeds the limit the airport can manage, airport expansion is required to ensure efficiency of airport operations and the safety of passengers. In order to expand the terminal and the runway, which is normally 2,700 meters long, it is sometimes necessary to acquire more land in the surrounding area. To do this, the government must obtain consent from the residents and provide acceptable compensation. However, recent studies about airport expansion and displacement show that across the world, policies regarding financial aid for displaced peoples are insufficient. Moreover, airport expansion comes with additional issues such as noise and air pollution, which also threatens the community around the newly-expanded airport. Thus, airport expansion comes with a myriad of ethical issues. Against this backdrop, the purpose of this review paper is to evaluate the pros and cons of airport expansion and explore how governments around the world navigate the complexities to meet travel demands while managing community development.

The Problem with Financial Compensation

One of the most prevalent issues with airport expansion is the displacement of peoples in the surrounding residential areas. Governments usually provide financial assistance to individuals being displaced due to airport expansion, but studies show that financial assistance

alone is not sufficient compensation. First, financial assistance offered is not equal to the financial burden of displacement. Because the financial assistance is usually based on the value of the resident's land and the home built on it, the money is not enough to cover the costs of moving or loss of opportunities after displacement. For example, in a case study about long-term impacts of airport expansion on displaced communities in Tanzania, Mtaki et al. (2016) found that affected communities felt they were being underpaid because the government is not transparent about how they calculate compensation rates. Moreover, "compensation is based on the market value of the lost asset," meaning that compensation is often less than the amount required to replace the lost asset, or the house in this case (Mteki et al.). Additionally, people who were displaced were forced to move locations that were farther away from their jobs or other opportunities that supported their financial livelihood. Because the financial compensation does not account for such possibilities, families are forced to change jobs or incur additional costs for transportation to work, both of which adds to their financial burden. In the case of Tanzania, there was a sharp increase in the unemployment rate after communities were forced to relocate, showing that there was no concern for residents' financial situation after displacement.

Second, financial compensation is also insufficient because it does not solve psychological damage that comes from displacement. Through in-depth interviews, Mtaki et al. (2016) also found that residents were generally very dissatisfied with their new environments compared to their original homes. The top reasons for dissatisfaction include family separation and a loss of community. Because there is no guarantee that the displaced community can find another area to live in together, members of the community inevitably become separated due to airport expansion. For tight knit communities who rely on each other for trade and bartering, as well as emotional support, the separation can have dire consequences for individual members of the community who are now without essential support systems (Magembe-Mushi). In light of these challenges, it is evident that the impacts of displacement extend beyond finances, impacting the psychological well-being of those displaced from their homes. Therefore, financial assistance is both insufficient and neglects to address all the costs of displacement.

Health and Environmental Effects

Another major issue with airport expansion is the impact on both the local ecosystems and the health of nearby communities. Airport expansion is a massive project and certainly brings physical impact at first. For creating just a new runway at Seattle-Tacoma International Airport, it required about 13 million cubic meters, whereas in total of 307 million cubic meters of land were consumed to construct the Chek Lap Kok Airport in Hong Kong (Upham). Thus, environmental issues associated with airport expansion, such as deforestation and water and air pollution, are significant and only scale with the size of the airport expansion project. Research conducted in Hong Kong reveals that the construction and operation of airports are often synonymous with noise pollution, lessening the quality of life for those living nearby. Some may believe that it is only the construction which causes air/noise pollution, however just operating the airport on a daily basis certainly does create a noticeable amount of air/noise pollution in the

nearby area. In fact, according to Corpuz's study in 2012, he claimed that 7 percent of the public housing population with 24 percent of the private housing population suffers from noise pollution by the airport on a daily basis (Corpuz). An increase in noise pollution often decreases the quality and quantity of sleep of family members, which causes a ripple effect of other negative consequences in their everyday lives. Moreover, the increased air pollution brought on by airport expansion can have a more direct impact on the health of nearby residents. In fact, 41 percent of respondents in Hong Kong attribute recent respiratory problems with the airport, and 44 percent of respondents expected their health to worsen as a direct consequence of the airport's expansion (Corpuz). As demonstrated in case studies from around the world, including Hong Kong, Boston, the Netherlands, and Manchester, the expansion of airports often leads to a shift in the balance between safety and convenience. Homes near the airport compound the environmental issues, leading to decreased air quality and increased noise levels that can have long-term physical and psychological effects. According to the article from Medical News Today, people may experience difficulty falling asleep or waking too early (Millar), which is a problem particularly prevalent when the airport is located near the green belts area, surrounding major urban areas where the ecological balance is easily disrupted by the airport's construction and operations.

Solutions

Therefore, when deciding if and how to pursue airport expansion projects, socioeconomic and environmental issues should be carefully weighed against the economic benefits of airport expansion. Generally speaking, significant advantages of airport development include increased economic output, more job possibilities, a rise in tourism, and improved operational effectiveness. When the Singaporean government announced its plans to expand Changi airport in 2018, for example, government messaging highlighted that airport expansion would make "Singapore attractive to global investments," and bring in more business and trade opportunities, therefore more jobs to the Singaporean market ("Three reasons why we need to expand our airport infrastructure"). However, upon closer inspection, it is possible that these benefits mostly benefit the government instead of the general population. As shown with the studies above, displaced peoples have socioeconomic challenges as a result of the airport expansion, and displacement of people results in an influx of new residents in other residential areas, which would affect the jobs and housing markets. While the government benefits from increased revenue and an improved national image, residents must suffer from overcrowding, social disintegration, and a steep increase in living costs long term. These problems show that the government must prepare and provide better alternatives that take into account how people's lives and communities are affected due to airport expansion.

Existing policies and compensation for the community displacement and land use change is not enough to outweigh the disadvantages to displaced and fractured communities. However, some airports around the world, such as Incheon Airport in South Korea and Heathrow Airport in England can stand as examples of possible solutions. During the initial construction of Incheon

Airport, the government marked out a large area and cleared land and water with future expansions in mind. Because this was done in the early stages, the government could expand the airport as necessary without the need to relocate communities or cause additional light and noise pollution for residents. If this is not possible because the airport was already built on a smaller plot of land, the government should make the expansion benefits more apparent to the general public. For example, when Heathrow Airport was adding a Northwest runway in 2016, the government did not stop at claiming that it would provide more local jobs; the airport made a pledge to create 5,000 more apprenticeships by 2030 and use these opportunities to decrease youth unemployment in the neighborhoods closest to the airport (Department for Transport). This plan demonstrates how the government has taken long-term repercussions of airport expansion into account. While it is not perfect because it does not address issues outside of unemployment, a plan that goes beyond basic financial compensation is an improvement.

Conclusion

In conclusion, airport construction and expansion can have devastating drawbacks on local communities that overshadow the benefits of larger, more functional airports. In Tanzania, airport expansion fissures communities and can add to the family's financial burden because compensation packages are insufficient. In places like Hong Kong and the Netherlands, studies show that airport expansion adds to environmental pollution and disrupts the lives of people in nearby residential areas. Because of the construction and noise from airplanes, residents report a decrease in the quality of sleep and their overall health. These effects are compounded by other issues such as overcrowding, limited housing, and steep increase of cost of living as neighborhood demographics in surrounding areas shift. Therefore, airport expansion is not always the boon that the government portrays it as. To reduce the negative impact on the communities, governments can improve the financial compensation package to account for loss of employment or other financial opportunities and make a greater effort to consider the psychological burden of displacement. When building new airports, it would be ideal to set aside a larger amount of land than initially necessary to account for later airport expansion. This would reduce the need to disrupt communities again later. With these changes, it is possible for the government to create airport expansion plans that allow all members of the community to benefit from increased air traffic in the area.

Works Cited

- Corpuz, L., Hines, J., Hogan, G., & Paredes, M. (2012). *Effects of Hong Kong International Airport: An Evaluation of the Perceived Impacts in Tung Chung*. WORCESTER POLYTECHNIC INSTITUTE in partial fulfillment of the requirements for the Degree of Bachelor of Science.
- Department for Transport. "Heathrow Northwest Runway: economic benefits." *GOV.UK*, 2016, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/562567/heathrow-north-west-runway-economic-benefits.pdf.
- Magembe-Mushi, D. L. (2018, November 8). *Impoverishment risks in DIDR in Dar es Salaam City: The case of airport expansion project*. *Current Urban Studies*.
https://www.scirp.org/html/1-1150321_88348.htm
- Millar, H. (2020). *Noise pollution health effects: Impact on mental and Physical Health*. *Medical News Today*.
<https://www.medicalnewstoday.com/articles/noise-pollution-health-effects#mental-health>
- Mteki, N. (n.d.). *Social Impacts induced by a development project in Tanzania: a case of airport expansion*. Taylor & Francis Online.
<https://www.tandfonline.com/doi/full/10.1080/14615517.2017.1322806>
- Statista Research Department. "Airline industry worldwide - number of flights 2023." *Statista*, 29 August 2023,
<https://www.statista.com/statistics/564769/airline-industry-number-of-flights/>.
- Three reasons why we need to expand our airport infrastructure." *Singapore Government*, 1 March 2018,
<https://www.gov.sg/article/three-reasons-why-we-need-to-expand-our-airport-infrastructure>
- Upham, P., Thomas, C., Gillingwater, D., & Raper, D. (2003). *Environmental capacity and airport operations: current issues and future prospects*. Academia.edu.
https://www.academia.edu/3357108/Environmental_capacity_and_airport_operations_current_issues_and_future_prospect

Pharmaceutical and Non-Pharmaceutical Approaches to Managing Creutzfeldt-Jakob's Disease Symptoms By Suhani Mehta

Abstract

Human prion diseases are part of a family of rare neurodegenerative disorders, affecting 1-2 out of every million people globally. The most prevalent type of prion disease is Creutzfeldt-Jakob disease, also known as CJD. Due to the misfolding of a prion protein, there is rapid degeneration of the brain. These misfolded prions lead to the formation of colonies, destroying surrounding brain tissue, leading to inevitable fatality. As the disease progresses, symptoms continue to worsen, ultimately causing complete mobility loss, dementia, slurring of words, abnormal jerking movements, and more. The central issue with CJD is a lack of treatment options available. As of now, there is no cure, instead, therapeutic methods focus on providing relief to the patient. Exploring potential therapies for CJD patients is crucial, as a patient's physical and mental state is unpredictable throughout the journey with the disease. Furthermore, the rapidly progressing nature of the disease leaves little time for the trial of therapeutics, as it's extremely difficult to predict the timeline of the symptoms, and the symptoms require constant care and attention. Thus, it is necessary to find a manner in which symptoms can be managed as efficiently as possible, without disrupting the patient's comfort. This review summarizes a wide variety of pharmaceutical and non-pharmaceutical therapies for CJD patients, with the goal of highlighting therapies that best relieve patients of their symptoms. Ranging from the potential use of antibodies to acute rehabilitation, these studies explore methods that can help CJD patients live out the rest of their lives in the most satisfactory manner.

Introduction

Prion diseases, also known as transmissible spongiform encephalopathies, belong to a large family of devastating neurodegenerative conditions. Creutzfeldt's Jakob's disease (CJD) is the most prevalent type of prion disease, impacting people all over the world. In this disease, proteins in neurons are spontaneously converted into prions, a type of protein that triggers the misfolding of normal proteins, leading to their dysfunction. The formation of prions is exponential, causing significant damage to the patient's neurons (Cleveland Clinic, 2022).

Diagnosis of CJD is a tricky process, and oftentimes diagnosis occurs only after symptoms begin to appear. In fact, CJD can be contracted for many years with the patient being asymptomatic. However, the rapid degeneration only starts to occur after symptoms begin to show (Cleveland Clinic, 2022). There are three key ways to diagnose CJD.

Electroencephalography (EEG), a type of test that records the electrical patterns of the brain to find any irregularities, allows for the discovery of which stage of CJD the patient is currently in (Wieser et al., 2023). Findings such as diffuse slowing, a specific type of brain activity pattern, appear in EEG tests during early stages of CJD, as these are the first signs of cerebral dysfunction (Emmady and Anilkumar, 2023). Furthermore, magnetic resonance imaging (MRI), a non-invasive technique used to obtain detailed anatomical images, is a helpful strategy in the

diagnosis of CJD, proving to be more than 90% accurate in diagnosis of CJD (Muacevic et al., 2020). Typically, abnormalities that display in CJD are found within the cerebral cortex or deep gray matter of the brain (Ranchod, 2023). Performing a lumbar puncture is also effective, as cerebrospinal fluid (CSF) is extracted from the spinal column and analyzed for prions (*Prnp* protein) (Johns Hopkins Medicine). Specifically, lab specialists are looking for the 14-3-3 protein, a marker for prion diseases such as CJD (Zerr et al., 1998). However, the only way to definitively confirm a diagnosis of CJD on a living patient is by performing a brain biopsy. In this procedure, a small portion of brain tissue is extracted and examined, but this process is considered a major risk to the patient (McPherson and DiNapoli, 2018).

There are four types of CJD: sporadic, familial, variant, and iatrogenic (NHS, 2021). Starting off with sporadic CJD, this is the most prevalent variant, comprising about 85% of all cases (CDC, 2022). There is no known cause for sporadic CJD, simply the fact that regular proteins in the brain spontaneously convert into prions, and this conversion persists indefinitely (NHS, 2022). Ninety percent of patients die within a year of symptoms onset, with the average duration of the disease being between 4-6 months (UCSF). Familial CJD is the result of a mutation in the *Prnp* gene, the gene responsible for making PrP proteins, which are known to be involved in cell signaling and neuronal homeostasis (True, 2017). This mutation displays an autosomal dominant inheritance pattern, meaning the closest relatives of an affected individual have a 50% chance of inheriting the pathogen. This variant of CJD accounts for about 10-15% of patients (UCSF, 2023). fCJD differs from sporadic CJD, as the disease course can vary from either a few months to over 5 years. Variant CJD is rooted from consuming infected cattle meat. The cow would have to have bovine spongiform encephalopathy (BSE - also known as mad cow's disease), which is similar to CJD, as it is also a prion disease. Prions convert into a pathogenic form that eventually damages the central nervous system of the cattle (CDC, 2021). However, there is an unclear timeline from infection to symptoms showing in cases of variant CJD. Finally, iatrogenic CJD spreads through the contamination of surgical equipment after the treatment of a CJD patient. This spread occurs due to lack of proper sterilization on surgical instruments before performing a procedure on the next patient (NHS, 2021). Though there are four variants to CJD, they all display a range of similar symptoms, including: memory loss, confusion, changes in behavior and personality, hallucinations, delusions, ataxia (problems with muscle coordination), balance issues, dystonia (muscle spasms and uncontrollable jerking movements), seizures, muscle atrophy (rapid loss of muscle mass and weight), and eventually, paralysis (Sitammagari, 2022).

Over the last decade, the presence of CJD in countries all over the world has increased, in some places more than in others. Studies have shown that there has been a rise in CJD cases and CJD related deaths in countries such as Japan, the Czech Republic, Slovakia, the UK (Nishimura et al., 2020), and South Korea. In fact, researchers noticed a unique change in the increase of cases in South Korea, specifically that there was an “unusually high percentage” of CJD patients part of the 30-39 years age group, rounding to about 1% of total cases in South Korea (Kim and Jeong, 2022). Though many of the listed countries noticed an increase in sporadic CJD cases, the

UK had a dramatic rise in the number of variant CJD cases, a rarer type of the disease. As of spring 2022, the UK has become the country with the current most active variant CJD cases, topping at 175 cases. France and Ireland follow the UK, but are nowhere near the number of active vCJD cases that the UK contains (The University of Edinburgh, 2022). It is unclear the cause of this global increase in cases, and whether it indicates increased prevalence of CJD or an increase in detection.

Pharmaceutical approaches to treating CJD

Antibodies

Antibodies have been proposed as a potential therapy to help a patient's body fight foreign prion proteins. Typically, antibodies are proteins that are naturally produced and serve to help the immune system fight foreign substances, known as antigens, such as infections, allergens, and toxins. They are produced by white blood cells (WBC) and the structure of each antibody slightly differs. It is a Y-shaped molecule, however each tip has a unique amino acid sequence required to fight the antigen. When an antigen comes in contact with a WBC, the WBC divides and multiplies to release millions of antibodies into the bloodstream. The treatment for CJD differs though, as it is a clinically created antibody (in a lab setting), known as a monoclonal antibody, rather than a naturally produced antibody from the human immune system. Monoclonal antibodies are created to target a specific antigen, classifying them as a type of immunotherapy for diagnoses such as cancer, rheumatoid arthritis, heart disease, lupus, multiple sclerosis, and more (Cleveland Clinic, 2022). Prior knowledge about antibodies has led to the creation and clinical trial of PRN100, a monoclonal antibody designed to prevent prions (PrP^C protein) from continuing to multiply by destroying them. In a study conducted by University College London (UCL) researchers, PRN100 was offered to 6 CJD patients, who were not in the terminal stage of the disease. The PRN100 was given intravenously in small increments every 2 days, and eventually larger increments every 2 weeks, if negative effects were not recorded. According to the results, this monoclonal antibody method was safe to use and successfully reached encouraging CSF and brain tissue concentrations that were noted in patient autopsies. In fact, the use of PRN100 led to the longest clinical duration of iatrogenic CJD ever recorded. Though death still occurred, there were no clinically adverse effects seen, giving hope for the use of PRN100 in the future, if its success continues to be seen with more extensive testing (Mead et al., 2022).

Antisense Oligonucleotides

The use of antisense oligonucleotides (ASOs) is a potential therapy for CJD utilized prior to the formation of prions, targeting these proteins at the level of their transcripts. ASOs bind to specific RNA sequences to control the production and expression of their respective proteins. These sequences, generally 15-20 nucleotides in length, target complementary RNA, with the Watson and Crick base pairing rules. Though ASOs have been in practice for the last few decades, new modifications have been explored for increased stability, efficiency, and response.

Typically, ASOs function either through RNA cleavage or RNA blockage. Within RNA cleavage, there are two main methods. To begin, there is RNase mediated degradation, in which a complex is formed between the ASO and its complementary RNA. This complex acts as a substrate for the enzyme RNase, which has the key role of degrading RNA from the ASO-RNA complex, essentially blocking the formation of the protein. The second method is known as RNA interference, where small interfering RNAs (siRNAs - 22 nucleotide double stranded RNA sequences) interact with an enzyme known as the Argonaute 2 enzyme, to form an RNA induced silencing complex (RISC). RISC then degrades the passenger strand of the siRNA, and the remaining strand guides RISC towards the complementary mRNA. The enzyme Argonaute 2 will then cleave this mRNA and silence its expression. Moving onto RNA blockage techniques, there is steric hindrance, which is the slowing of a molecule's efficiency due to its bulk. These ASOs bind to the target RNA sequence and prevent their assembly with a ribosomal subunit, inhibiting translation. These ASOs have a very high binding affinity for their target RNA strand. Finally, there is splice modulation, a technique used to address alternative splicing and frameshift mutations that alter pre-mRNA splicing patterns. ASOs can either bind to the pre-mRNA transcripts to fix the reading frame and produce a functional protein, or they can bind to the pre-mRNA to prevent the transcript sites from being accessed (Dhuri et al., 2020).

ASOs can play a role in the formation and inhibition of proteins, which leads to the discussion of how they can be used to treat prion diseases. In a 2019 study, researchers worked on the use of ASOs and how they can extend the survival of mice infected with prions. ASOs that were complementary to the endogenous prion mRNA were screened and the researchers came up with ASO1 and ASO2 as two potential therapeutics. The primary difference between ASO1 and ASO2 is how they target different regions of the PrnP gene (the 3' UTR and intron 2, respectively). For preliminary results, both ASOs were dosed in the brain and spinal cord of uninfected wild-type (WT) mice and there was a significant reduction of *Prnp* mRNA noted in the ipsilateral entorhinal cortex, hippocampus, and thoracic spinal cord, in comparison to mice treated with saline. Next, mice were infected with an established prion disease similar to that of CJD. These mice were injected with ASO1 about 15 days before the expected onset of symptoms. The results show that the use of active ASO1 delayed the onset of symptoms by 33% and the clinical phase of the disease lasted 3x longer than the mice treated with saline. Overall, the survival time of treated mice increased by 55%. This study highlights the successful delivery of ASOs to the brain through the spinal cord in mice, achieving similar brain distribution as spinal injections in primates. This sheds possible light into dosing humans with ASOs through a lumbar puncture, which is already utilized as a CJD diagnosis technique (Raymond et al., 2019).

Drugs

Though researchers have not found drugs to target prion proteins specifically yet, drugs have been implemented to manage common CJD symptoms. Benzodiazepines are a family of depressant drugs used to slow down signals between the brain and body. Though benzodiazepines can be used to relieve anxiety, insomnia, and even alcohol withdrawals, they

can also be used to treat brain related issues such as seizures (ADF, 2023). Specifically, clobazam, a long-acting benzodiazepine, has been used in treating CJD related seizures. Clobazam increases the permeability of chloride ions through neuronal cells, by increasing the frequency of ligand gated ion channel openings. The increased cell potential as a result activates GABA, an inhibitory neurotransmitter. As a result, there is a decrease in neuronal activity and excitability as a whole in the cell, stopping the specific signals that cause seizures (Maille et al., 2023). A study found that initiation of clobazam along with levetiracetam and lacosamide (Liu et al., 2023), two supporting anti-epileptic drugs, led to a significant decrease in seizures in CJD patients, stopping them all together within 72 hours. Researchers suggest that benzodiazepine receptors remain intact in CJD patients, allowing them to be effective in helping patients (Maille et al., 2023).

Furthermore, sodium valproate has also been used to reduce seizures, by controlling excessive electrical activity in the brain. Sodium valproate can achieve this by blocking certain voltage gated ion channels to reduce the high frequency transmission of neurons (Rahman and Nguyen, 2022). This is a type of slow release medicine, typically taken in the form of a tablet (NHS, 2021).

Finally, flupirtine maleate (FLU) is a non-opioid pain relieving drug, which functions by blocking the glutamate N-methyl-D-aspartate receptor (Harish et al., 2012), a primary excitatory neurotransmitter receptor in humans (Jewett, B and Thapa, B., 2022). In previous cell culture experiments, FLU was able to prevent apoptotic cell death of neurons that were previously treated with parts of prion proteins. Because of this, FLU was implemented in a study with CJD patients, and the results seemed promising. CJD patients treated with FLU showed significantly less deterioration when periodically tested for dementia under the Alzheimer's Disease Assessment Scale. Though further studies are necessary to determine whether FLU could be part of a stable CJD treatment plan, it has had positive effects on cognitive function of these CJD patients; there is great hope for its potential (Otto et al., 2004).

Non-pharmaceutical methods

Pharmaceutical methods have the capability to help CJD patients with their symptoms, however in some cases, the progression of the disease can be too quick for the implementation of these methods. For this reason, non-pharmaceutical methods can be key in managing the less complex symptoms, providing quick comfort to patients.

Visual art therapy

Visual art therapy consists of the use of creative techniques to allow patients to express themselves, in an artist's manner. Through art, patients' psychological and emotional undertones can be analyzed, making it a great way for therapists to dig deeper into a client's life. Therapists are trained to draw connections between the patients' artistic choices and what is occurring within their life in order to help them. Scientifically, art work has proven to reawaken memories and reveal stories, and display unconscious thoughts (Psychology Today, 2022). In relation to

CJD, art therapy can be useful in increasing patient awareness of what is going on, along with evoking spontaneous expression of emotion. As the disease progresses, there are obvious physical challenges, but there are a great deal of mental challenges as well. The effects of art therapy on CJD was tested in a case study, where the key subject was Ms. A, a middle-aged woman with disease progression in its early stages. For 23 weeks, she participated in weekly art therapy sessions, with the goal being to increase her comfort with the circumstances of her disease. The researchers noticed that art therapy provided her with short windows of time where she was able to deeply express her feelings about what was happening to her. This showed that despite rapid changes in her brain and the continuous decline in her physical capabilities, she was still aware of her surroundings. In fact, art therapy created a safe environment for Ms. A, where she could feel less overwhelmed and anxious by the reality of the disease (Shrestha et al., 2016). As shown in this case study, visual art therapy has the potential to calm the minds of CJD patients, creating a calmer and safer environment for them as the disease progresses.

Acute Rehabilitation

Acute rehabilitation is a type of care where a patient receives whatever medical attention they may require, whether its constant medication or the monitoring of the progress of a disease, at the same time as some sort of physical therapy (Santé, 2016). This can prove to be quite beneficial to patients with CJD, as analyzed in a case study from 2022. A 62-year old man, after a recent diagnosis of CJD entered a 14 day acute rehabilitation program, in hopes to improve his mobility. During this two week study, the patient underwent a series of treatments, including therapeutic exercise, gait training, neuromuscular reeducation, cognitive behavioral therapy, and voice therapy (Copeland et al., 2022). The main purpose of therapeutic exercise is to correct any impairments to the anatomy and restore skeletal/muscular function (Bielecki and Tadi, 2023), whereas gait training focuses on improving the ability to walk (Eng and Tang, 2007). Neuromuscular reeducation worked to restore normal function of his nerves and as a result, muscular function (AmeriHealth, 2018). Throughout the process, he was also treated with midodrine to manage orthostatic hypotension, to keep his standing blood pressure levels normal (Prescr, 2021). The results were quite promising, as by the time of his discharge, the patient was able to complete daily activities with minor assistance. At the beginning of the 2 weeks, he required moderate assistance, around 50% caregiver effort and 50% of his own effort. However after 2 weeks, it shifted to 75% patient effort and only 25% caregiver effort for day-to-day tasks. Though this was just a case study, acute rehab could be a promising therapy to improve the mobility of CJD patients (Copeland et al., 2022).

Discussion

CJD is the most prevalent type of prion disease, which is truly a tragedy considering its rapidly progressing nature and the lack of consistent treatment options available (John Medicine Hopkins). However, there are therapies, both pharmaceutical and non-pharmaceutical that can be implemented to help alleviate symptoms, to hopefully lead to a more comfortable journey with

the disease. From the trial of monoclonal antibodies that destroy prions in the body, to the successful implementation of ASOs to target prions before they are even formed, and finally, the use of clinically-approved drugs such as clobazam, sodium valproate, and flupirtine maleate to control seizures and dementia, these pharmaceutical methods demonstrate heavy potential for attacking CJD from a molecular perspective. Through further experimentation with these proposed solutions, there is much hope for solving the mystery behind CJD.

On the other hand, non-pharmaceutical approaches can be an easier method of providing relief to patients, without partaking in any complicated processes. However, there is a key issue with the direction that CJD research has fallen under. Rather than targeting the root of the issue, which is the spontaneous multiplication of prions, research has been focused on addressing symptoms of the disease. This can prove to be useful in the short term, but as the prevalence of the disease increases, there is a greater need for finding a reliable solution to CJD, in order to save lives. As mentioned above, there has been a rise in CJD related death rates in the following countries: Japan, the Czech Republic, Slovakia, the UK (Nishimura et al., 2020), and South Korea (Kim and Jeong, 2022). Again, the reason for this growth in cases is unknown, however it's hypothesized to partially be due to heightened awareness of CJD and the advancement of diagnostic technology over time.

Overall, there is a desperate need for more research in the prion-disease field of neuroscience, as the lack of knowledge about the disease is a problem in itself. There are still many unanswered questions, making it extremely difficult to design a solution to control the disease. Without concrete answers on the basics of CJD, such as how prions initially form, there is no way to combat the disease once and for all. However, CJD is similar in nature to other neurodegenerative diseases, in terms of being caused by the misfolding of proteins that ultimately lead to neuron loss (Lamprey et al., 2022). For example, Alzheimer's disease is caused by the build-up of the misfolded amyloid-B protein (due to mutation) in neuronal cells (Ashraf, 2014). A common Alzheimer's treatment utilizes cholinesterase inhibitors, which are a group of drugs that block the breakdown of acetylcholine into acetate and choline. As a result, there's an accumulation of acetylcholine, and difficulty with balance, coordination, and even walking and talking as the disease progresses. Accumulation of a misfolded form of the alpha-synuclein protein leads to the formation of toxic clumps, which ultimately lead to cell death (Cure Parkinson's). Deep brain stimulation is a treatment recommended to Parkinson's patients that don't respond well to medication, where electrodes stimulate specific areas of the brain that are involved in movement, to stop symptoms such as tremors, slowness in movement, and rigidity. Again, DBS has never been tested on CJD patients, and there's no research to demonstrate the reactions of prions to it, however, the similarities in causes and symptoms of CJD and Parkinson's make the future use of DBS in CJD an appealing option (NIA, 2022).

To summarize, there are still ways to go in order to find a cure that is reliable across a wide variety of CJD cases and that targets the root cause of the disease. The lack of research in the field is the biggest obstacle in the journey to discovering how to end CJD once and for all.

Works Cited

- “What Is Acute Rehabilitation?” *Santé Cares*, 21 Dec. 2016, santecares.com/2016/12/21/what-is-acute-rehabilitation/. Accessed 4 Sept. 2023.
- Alcohol and Drug Foundation. “Benzodiazepines - ADF - Alcohol & Drug Foundation.” *ADF - Alcohol & Drug Foundation*, 10 Jan. 2023, adf.org.au/drug-facts/benzodiazepines/.
- “Art Therapy | Psychology Today.” *Www.psychologytoday.com*, www.psychologytoday.com/us/therapy-types/art-therapy#:~:text=Art%20therapy%20involves%20. Accessed 4 Sept. 2023.
- Ashraf, Ghulam Md, et al. “Protein Misfolding and Aggregation in Alzheimer’s Disease and Type 2 Diabetes Mellitus.” *CNS & Neurological Disorders Drug Targets*, vol. 13, no. 7, 2014, pp. 1280–1293, www.ncbi.nlm.nih.gov/pmc/articles/PMC5193501/.
- Basit, Hajira, and Chadi I. Kahwaji. “Clonazepam.” *PubMed*, StatPearls Publishing, 2020, www.ncbi.nlm.nih.gov/books/NBK556010/.
- Bielecki, Joseph E., and Prasanna Tadi. “Therapeutic Exercise.” *PubMed*, StatPearls Publishing, 2021, www.ncbi.nlm.nih.gov/books/NBK555914/.
- CDC. “Occurrence and Transmission | Creutzfeldt-Jakob Disease, Classic (CJD) | Prion Disease | CDC.” *Www.cdc.gov*, 14 Nov. 2022, www.cdc.gov/prions/cjd/occurrence-transmission.html#:~:text=The%20majority%20of%20cases%20of.
- Centers for Disease Control and Prevention. “Bovine Spongiform Encephalopathy (BSE).” *Centers for Disease Control and Prevention*, 2019, www.cdc.gov/prions/bse/index.html.
- Cleveland Clinic. “Antibodies: Definition, Types & Function.” *Cleveland Clinic*, 5 June 2022, my.clevelandclinic.org/health/body/22971-antibodies.
- Clinical Policy Title: Neuromuscular Reeducation Clinical.*
- Copeland, Royce, et al. “The Management of Newly Diagnosed Probable Creutzfeldt-Jakob Disease in Acute the Management of Newly Diagnosed Probable Creutzfeldt-Jakob Disease in Acute Rehabilitation Setting: A Case Report Rehabilitation Setting: A Case Report.” *Clinical Medical Research and Healthcare Delivery*, vol. 2, no. 3, 2022, scholar.rochesterregional.org/cgi/viewcontent.cgi?article=1116&context=advances, <https://doi.org/10.53785/2769-2779.1116>. Accessed 3 Mar. 2023.
- “Creutzfeldt-Jakob Disease (CJD): Types, Causes, Symptoms & Outlook.” *Cleveland Clinic*, my.clevelandclinic.org/health/diseases/6001-creutzfeldt-jakob-disease.
- “Creutzfeldt-Jakob Disease Symptoms.” *Ucsfhealth.org*, www.ucsfhealth.org/conditions/creutzfeldt-jakob-disease/symptoms.
- Dhuri, Karishma, et al. “Antisense Oligonucleotides: An Emerging Area in Drug Discovery and Development.” *Journal of Clinical Medicine*, vol. 9, no. 6, 26 June 2020, p. 2004, <https://doi.org/10.3390/jcm9062004>.
- “Diagnostic Criteria | Creutzfeldt-Jakob Disease, Classic (CJD) | Prion Disease | CDC.” *Www.cdc.gov*, 25 Jan. 2019, www.cdc.gov/prions/cjd/diagnostic-criteria.html.

- Edinoff, Amber N., et al. "Benzodiazepines: Uses, Dangers, and Clinical Considerations." *Neurology International*, vol. 13, no. 4, 10 Nov. 2021, pp. 594–607, www.ncbi.nlm.nih.gov/pmc/articles/PMC8629021/, <https://doi.org/10.3390/neurolint13040059>.
- Emmady, Prabhu D., and Arayamparambil C. Anilkumar. "EEG Abnormal Waveforms." *PubMed*, StatPearls Publishing, 2023, www.ncbi.nlm.nih.gov/books/NBK557655/#:~:text=Diffuse%20slowing%3A%20Diffuse%20slowing%20indicates. Accessed 4 Sept. 2023.
- Eng, Janice J, and Pei-Fang Tang. "Gait Training Strategies to Optimize Walking Ability in People with Stroke: A Synthesis of the Evidence." *Expert Review of Neurotherapeutics*, vol. 7, no. 10, Oct. 2007, pp. 1417–1436, www.ncbi.nlm.nih.gov/pmc/articles/PMC3196659/, <https://doi.org/10.1586/14737175.7.10.1417>.
- European Centre for Disease Prevention and Control. "Facts about Variant Creutzfeldt-Jakob Disease." *European Centre for Disease Prevention and Control*, 26 June 2017, www.ecdc.europa.eu/en/vcjd/facts.
- "Familial Prion Disease." *Memory and Aging Center*, memory.ucsf.edu/genetics/familial-prion-disease#:~:text=Genetic%20Creutzfeldt%2DJakob%20disease%20. Accessed 4 Sept. 2023.
- Harish, S, et al. "Flupirtine : Clinical Pharmacology." *Journal of Anaesthesiology Clinical Pharmacology*, vol. 28, no. 2, 2012, p. 172, www.ncbi.nlm.nih.gov/pmc/articles/PMC3339720/, <https://doi.org/10.4103/0970-9185.94833>.
- Ho, Mai-Lan. "Creutzfeldt-Jakob Disease | Radiology Reference Article | Radiopaedia.org." *Radiopaedia*, radiopaedia.org/articles/creutzfeldt-jakob-disease?lang=us.
- Jewett, Benjamin E., and Bicky Thapa. "Physiology, NMDA Receptor." *PubMed*, StatPearls Publishing, 2022, www.ncbi.nlm.nih.gov/books/NBK519495/#:~:text=The%20N%2Dmethyl%2DD%2D.
- John Hopkins Medicine. "Prion Diseases." *Johns Hopkins Medicine*, 2019, www.hopkinsmedicine.org/health/conditions-and-diseases/prion-diseases.
- Kim, Yong-Chan, and Byung-Hoon Jeong. "Creutzfeldt-Jakob Disease Incidence, South Korea, 2001–2019." *Emerging Infectious Diseases*, vol. 28, no. 9, 1 Sept. 2022, pp. 1863–1866, www.ncbi.nlm.nih.gov/pmc/articles/PMC9423913/, <https://doi.org/10.3201/eid2809.212050>. Accessed 4 Sept. 2023.
- Legname, Giuseppe. "Elucidating the Function of the Prion Protein." *PLOS Pathogens*, vol. 13, no. 8, 31 Aug. 2017, p. e1006458, <https://doi.org/10.1371/journal.ppat.1006458>.
- "Lumbar Puncture." *Www.hopkinsmedicine.org*, [www.hopkinsmedicine.org/health/treatment-tests-and-therapies/lumbar-puncture#:~:text=A%20lumbar%20puncture%20\(LP\)%20or](http://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/lumbar-puncture#:~:text=A%20lumbar%20puncture%20(LP)%20or).

- McPherson, Christopher, and Vincent DiNapoli. "Brain Biopsy | Mayfield Brain & Spine." *Mayfieldclinic.com*, [mayfieldclinic.com/pe-brainbiopsy.htm#:~:text=In%20an%20open%20biopsy%2C%20a](https://www.mayfieldclinic.com/pe-brainbiopsy.htm#:~:text=In%20an%20open%20biopsy%2C%20a)n. Accessed 4 Sept. 2023.
- Mead, Simon, et al. "Prion Protein Monoclonal Antibody (PRN100) Therapy for Creutzfeldt–Jakob Disease: Evaluation of a First-In-Human Treatment Programme." *The Lancet Neurology*, vol. 21, no. 4, Apr. 2022, pp. 342–354, [https://doi.org/10.1016/s1474-4422\(22\)00082-5](https://doi.org/10.1016/s1474-4422(22)00082-5).
- Miller, Colton , and Edward Harris. "Antisense Oligonucleotides: Treatment Strategies and Cellular Internalization." *RNA & DISEASE*, 7 Nov. 2016, <https://doi.org/10.14800/rd.1393>. Accessed 6 May 2019.
- National Institute on Aging. "Parkinson’s Disease." *National Institute on Aging*, 14 Apr. 2022, www.nia.nih.gov/health/parkinsons-disease#:~:text=Parkinson.
- "New Drug: Midodrine for Orthostatic Hypotension." *Australian Prescriber*, 17 Dec. 2020, <https://doi.org/10.18773/austprescr.2020.081>. Accessed 9 Jan. 2021.
- NHS. "Creutzfeldt-Jakob Disease." *NHS*, 2019, www.nhs.uk/conditions/creutzfeldt-jakob-disease-cjd/.
- NHS Choices. "Diagnosis - Creutzfeldt-Jakob Disease." *NHS*, 2020, www.nhs.uk/conditions/creutzfeldt-jakob-disease-cjd/diagnosis/.
- Nishimura, Yoshito, et al. "A Nationwide Trend Analysis in the Incidence and Mortality of Creutzfeldt–Jakob Disease in Japan between 2005 and 2014." *Scientific Reports*, vol. 10, no. 1, 23 Sept. 2020, p. 15509, www.nature.com/articles/s41598-020-72519-0, <https://doi.org/10.1038/s41598-020-72519-0>.
- "PRNP Gene: MedlinePlus Genetics." *Medlineplus.gov*, medlineplus.gov/genetics/gene/prnp/.
- Qamar, Muhammad Sohaib, et al. "Creutzfeldt-Jakob Disease with Atypical Magnetic Resonance Imaging Features." *Cureus*, 2 Nov. 2020, <https://doi.org/10.7759/cureus.11294>. Accessed 19 Nov. 2020.
- Raymond, Gregory J., et al. "Antisense Oligonucleotides Extend Survival of Prion-Infected Mice." *JCI Insight*, vol. 4, no. 16, 22 Aug. 2019, insight.jci.org/articles/view/131175#SEC5, <https://doi.org/10.1172/jci.insight.131175>.
- Reimhult, Erik . "A Review of the Common Neurodegenerative Disorders: Current Therapeutic Approaches and the Potential Role of Nanotherapeutics." *ProQuest*, vol. 23, no. 3, 2022, p. 1851, www.proquest.com/docview/2627713382/7A5A08B4E0934401PQ/3?accountid=8042, <https://doi.org/10.3390/ijms23031851>.
- Singh, Ravneet, and Nazia M Sadiq. "Cholinesterase Inhibitors." *Nih.gov*, StatPearls Publishing, 19 Jan. 2022, www.ncbi.nlm.nih.gov/books/NBK544336/#:~:text=Cholinesterase%20inhibitors%20function%20to%20decrease.

- “Sodium Valproate: Medicine to Treat Epilepsy and Bipolar Disorder.” *Nhs.uk*, 27 Sept. 2018, www.nhs.uk/medicines/sodium-valproate
- “The Science behind Parkinson’s.” *Cure Parkinson’s*, cureparkinsons.org.uk/what-is-parkinsons/the-science-behind-parkinsons/.
- Valqui, Melissa. “Steric Hindrance.” *ChemTalk*, 9 June 2021, chemistrytalk.org/steric-hindrance/#:~:text=Definition%20of%20Steric%20Hindrance&ext=Whe. Accessed 4 Sept. 2023.
- WIESER, H, et al. “EEG in Creutzfeldt–Jakob Disease.” *Clinical Neurophysiology*, vol. 117, no. 5, May 2006, pp. 935–951, <https://doi.org/10.1016/j.clinph.2005.12.007>.
- Zerr, I., et al. “Detection of 14-3-3 Protein in the Cerebrospinal Fluid Supports the Diagnosis of Creutzfeldt-Jakob Disease.” *Annals of Neurology*, vol. 43, no. 1, 1 Jan. 1998, pp. 32–40, pubmed.ncbi.nlm.nih.gov/9450766/, <https://doi.org/10.1002/ana.410430109>. Accessed 27 July 2022.

Unraveling the Mysteries of the Chandipura Virus- A Comprehensive Literature Review

By Naina Nagendra

Abstract

In recent years, knowledge of emerging infectious diseases, or “EID,” has been of great significance to researchers and the general public. Chandipura virus, shortened as CHPV, is an EID belonging to the Rhabdoviridae family of viruses. First occurring sporadically, the virus transcended into a genuine epidemic in 2003. This epidemic occurred in Andhra Pradesh and resulted in a fatality rate of 56%, killing 183 children. CHPV is a zoonose, specifically borne by vectors such as mosquitoes, ticks, and sand flies. CHPV raises concern among the Indian population due to its differentiating nature from similar viruses of this type, namely VSV (Vesicular Stomatitis Virus), as the mortality rate for CHPV is significantly higher than that of VSV. Alongside this, CHPV primarily targets children, symptomized by acute encephalitis.

Furthermore, CHPV produces influenza-like symptoms as well as neurological dysfunction. Virologists are still searching for an accurate CHPV test but have recently used ELISA and PRNT testing. With this, Chandipura currently does not have an effective homogenized treatment despite its contributions to premature deaths in India. However, certain drugs are showing promise, such as the antiviral drug Favipiravir. This drug was tested for its efficiency on inoculated SCID mice and proved to be used as a possible preemptive measure against the fatal virus.

Unraveling the mysteries of the Chandipura Virus- a comprehensive literature review

Emerging infectious disease, referred to as EID, has been a topic of particular interest within the scientific community over the years. An emerging disease is described as a disease that is newly identified but is rapidly increasing in either incidence or geographic range¹. EID falls into two categories that can accurately define a *disease* as emerging: known and unknown. The illness can be an entirely new disease forming from numerous evolutionary or changes of organisms or an old infection reappearing due to changes in antimicrobial resistance in organisms². Alongside these factors, particular EIDs are zoonoses, diseases spread primarily by animals or insects. New zoonotic diseases have appeared to be prevalent in society, namely Zika Virus and SARS-CoV-2, the latter being the more relevant due to its sudden emergence and mortality, which was discovered to be linked to bats³. One emerging disease that has continuously plagued the Indian subcontinent for decades is the zoonotic disease known as the Chandipura virus.

Chandipura virus, commonly known as CHPV, is an emerging disease borne by vectors such as sandflies, ticks, and mosquitoes and is a leading cause of pediatric encephalitis in India. In 1965, near the village of Chandipur in Nagpur district, Maharashtra, India, Indian scientists discovered a new virus isolated from the blood of two patients with fever of uncertain cause⁴. Other febrile vector-borne illnesses, such as dengue and chikungunya virus, coincided in Chandipura, but this febrile illness seemed different. Based upon the discovery location, the virus

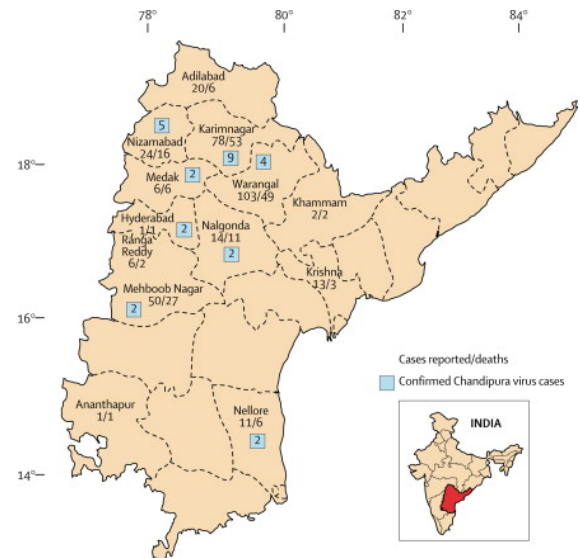
was named Chandipura virus⁵. Retrospective serological studies showed that the virus may have been present in the 1950s⁴. In 1954, there was an outbreak of fatal fever of uncertain cause with neurological complications; this illness was harmful for all known agents. This illness, Jamshedpur fever, was later considered CHPV⁶. The infection was proven to be zoonotic in 1967 by scientists at the National Institute of Virology in Pune from investigating the vectorial capacity and found that sand flies, in particular, had a high susceptibility to CHPV and could carry a significant amount of virus, proving that the disease was vector-borne⁵. CHPV was first isolated from the *Phlebotomus* species of sandflies inside human houses and cowsheds in Aurangabad, Maharashtra, in the 1960s⁷. *Phlebotomine* and *Sergentomyia*, genera of adult female sandflies, were also noted to carry the virus by reverse-transcriptase-polymerase chain reaction⁸⁻¹⁰. Although the Chandipura virus was primarily isolated in 1965, the cases were dismissed, and the virus was considered a lesser threat due to its low pathogenicity at the time to cause infections to either animals or human¹¹. Over time, CHPV cases occurred sporadically; for example.

Between June and September 2003, 329 children between 9 and 14 were stricken with encephalitis in the southern Indian state of Andhra Pradesh (Rao). One hundred eighty-three children consequently died - a fatality rate of 56%¹³. Initially, Japanese encephalitis was suspected. However, the patients progressed rapidly from an influenza-like illness to coma and death with accompanying abdominal pains and vomiting¹⁴. The fatal nature of the disease and the fleeting progression raised suspicion among virologists and led them to believe that Chandipura was indeed the cause. However, with constant sporadic cases littering India, the severity of the virus relating to Indian children came into question.

After the 2003 outbreak, Chandipura viral outbreaks remained sporadic. Small outbreaks occurred in Gujarat in 2010, 2014, and 2019, with a 50% mortality. The latest outbreak occurred in 2019 when a five-year-old girl died two days after rapid onset encephalitis¹⁵. Alongside these findings, CHPV was isolated from *Phlebotomus* sandflies in Senegal¹⁶ and four-toed hedgehogs in Nigeria¹⁷ in Sri Lanka from Macaques¹⁸, establishing the virus' transmission is not only to humans. These cases were suspected to be transmitted by phlebotomine sandflies, as they showed a higher susceptibility to the virus in specific geographic regions¹¹. Despite transmission of CHPV to animals in other countries, no documented cases of Chandipura virus have been identified outside of India.

Virology

2CHPV is a 180-nanometer organism belonging to the *Rhabdoviridae* family of negative-stranded RNA viruses, with its genus *Vesiculovirus* and order *Mononegavirales*.



Rhabdoviruses are typically characterized by their complex bacilliform or unique bullet-shaped particles. CHPV is approximately 180 nanometers in length¹⁹ The Rhabdoviridae family comprises over a hundred differentiating viruses that transmit the infection to various hosts, including non-vertebrates, vertebrates, autotrophs, fungi, and protozoans (Rupprecht). It holds four genera of animal viruses: *Lyssavirus*, *Ephemerovirus*, *Novirhabdovirus*, and *Vesiculovirus*²⁰. The transmission routes for these viruses vary in some ways and are similar in others. The most commonly known disease Rhabdovirus produces is rabies, which is spread by the genera *Lyssavirus*. This prevalent disease spreads directly from mammals to other mammals. The *Novirhabdovirus* (NV) is a strictly aquatic genus transmitted through contaminated eggs or water-borne viruses²¹. NV infects most marine life. Hematophagous insects, such as sandflies, ticks, and mosquitoes, bear *Ephemerovirus* and *Vesiculovirus*²⁰. The difference between them lies in their structure from isolating the strands of the viruses. CHPV lies within the subgenera of *Vesiculovirus*, but it differs from other viruses of this type, namely VSV (*Vesicular Stomatitis Virus*), due to its high fatality rate. CHPV and VSV furthermore differ in geographic origin, as the former was isolated throughout India, whereas the latter is almost exclusively found in the Americas²²

Symptoms and Diagnosis

The Chandipura virus is an integral contributor to Acute Encephalitis Syndrome, becoming one of India's leading causes of premature deaths. The virus is characterized by its influenza-like symptoms, such as vomiting and fever. Unlike influenza, CHPV is also distinguished by neurological dysfunction, producing an altered sensorium and decerebrate posture, leading to death within the first two days of hospitalization²³. However, virologists have not explained how the virus developed quickly and deteriorated rapidly. Furthermore, CHPV does not have a rapid test similar to the Coronavirus. Instead, the general diagnosis is made by examining the symptoms and differentiating CHPV from established encephalitis through a CT scan²⁴. Another way to diagnose Chandipura is the Plaque reduction neutralization test (PRNT). This is considered a "gold standard" type of test due to its accurate detection of antibodies against the virus. The serological PRNT test also neutralizes and prevents virions from infecting specific cells. Despite this, the test is highly cumbersome, taking a long time to complete¹³. The most recent type of test to detect the presence of CHPV is an ELISA test, which stands for enzyme-linked immunoassay. Because the test takes less time to be fulfilled, it can be substituted for other serosurveillance tests²⁵.

Treatment and Prevention

Despite the disease's tropism towards youth and its immense mortality rate, CHPV remains vastly neglected. This is due to the virtually nonexistent treatment to relieve emerging virus symptoms. Currently, the most obtainable way to avoid infection in local areas where CHPV is a threat is to prevent vectorial infection entirely by using rudimentary means of protection, such as insect repellent or mosquito netting²⁶. In truth, the idiosyncratic mutating

nature of the Chandipura virus is the obstacle to attempting to create a treatment for the disease. This mutating nature prevents virologists from developing long-lasting vaccines or therapeutic drugs. This is because viral inhibitors cannot be used as an agent to slow virus replication, unlike similar viruses as CHPV, which can quickly develop resistance to these inhibitors over time, rendering them useless. Several *in vitro* experiments have been conducted using drugs such as ribavirin and a synthetic peptide to slow CHPV growth. However, a homogenized therapeutic option is yet to be created²⁷. However, certain drugs are showing promise. As of 2023, the antiviral drug Favipiravir has garnered attention among virologists due to its distinctive clinical productivity and, therefore, its potential to be utilized in a possible treatment for CHPV. A team of virologists tested the theory of Favipiravir efficiency against CHPV through a functional *in vivo* model. The experiment consisted of female SCID mice divided into pre-symptomatic and post-symptomatic treatment and a standard control group. All groups were inoculated with an isolated strain of Chandipura, and both the pre and post-symptomatic groups sustained an oral dose of Favirapir twice a day. The results claimed that Favirapir was relatively successful in prolonging survival with the inoculated SCID mice for both the pre and post-symptomatic groups. Furthermore, Favirapir lessened the viral loads in specific organs, including blood, the brain, and kidneys. Researchers speculated that Favirapir might be used as a preemptive measure therapy for CHPV or to prevent the early infection from developing into something more lethal with further research²⁸.

Ultimately, the Chandipura Virus is severely under-reported, as its increase in lethality is worrisome due to its severe mutation rate through a series of genetic changes. Although the disease only affects the Indian sub-continent, it can become a more significant global threat if not adequately handled.

Works Cited

- Morse, S. S. (1995). Factors in the Emergence of Infectious Diseases. *Emerging Infectious Diseases*, 1(1), 7-15. <https://doi.org/10.3201/eid0101.950102>.
- McArthur D. B. (2019). Emerging Infectious Diseases. *The Nursing clinics of North America*, 54(2), 297–311. <https://doi.org/10.1016/j.cnur.2019.02.006>
- Ye, Z. W., Yuan, S., Yuen, K. S., Fung, S. Y., Chan, C. P., & Jin, D. Y. (2020). Zoonotic origins of human coronaviruses. *International journal of biological sciences*, 16(10), 1686–1697. <https://doi.org/10.7150/ijbs.45472>
- Bhatt, P. N., & Rodrigues, F. M. (1967). Chandipura: a new Arbovirus isolated in India from patients with febrile illness. *Indian Journal of Medical Research*, 55(12).
- Menghani, S., Chikhale, R., Raval, A., Wadibhasme, P., & Khedekar, P. (2012). Chandipura Virus: an emerging tropical pathogen. *Acta tropica*, 124(1), 1–14. <https://doi.org/10.1016/j.actatropica.2012.06.001>
- Chari , M.V and Swamy, (1955). Jamshedpur fever. *British Medical Journal*, 2(4951), 1298–1303. <https://doi.org/10.1136/bmj.2.4951.1298>
- Dhanda, V., Rodrigues, F. M., & Ghosh, S. N. (1970). Isolation of Chandipura virus from sandflies in Aurangabad. *The Indian journal of medical research*, 58(2), 179–180.
- Depaquit J, Grandadam M, Fouque F, Andry PE, Peyrefitte C. Arthropod-borne viruses transmitted by Phlebotomine sandflies in Europe: a review. *Euro Surveill*. 2010;15(10):pii=19507. <https://doi.org/10.2807/ese.15.10.19507-en>
- G. Geevarghese and others, Detection of Chandipura Virus from Sand Flies in the Genus *Sergentomyia* (Diptera: Phlebotomidae) at Karimnagar District, Andhra Pradesh, India
- Kamath S. S. (2015). Monsoon maladies. *Indian pediatrics*, 52(8), 655–656. <https://doi.org/10.1007/s13312-015-0691-1>
- Sudeep, A. B., Gurav, Y. K., & Bondre, V. P. (2016). Changing clinical scenario in Chandipura virus infection. *The Indian journal of medical research*, 143(6), 712–721. <https://doi.org/10.4103/0971-5916.191929>
- Rodrigues, J. J., Singh, P. B., Dave, D. S., Prasan, R., Ayachit, V., Shaikh, B. H., & Pavri, K. M. (1983). Isolation of Chandipura virus from the blood in acute encephalopathy syndrome. *The Indian journal of medical research*, 77, 303–307.
- Sapkal, G. N., Sawant, P. M., & Mourya, D. T. (2018). Chandipura Viral Encephalitis: A Brief Review. *The open virology journal*, 12, 44–51. <https://doi.org/10.2174/1874357901812010044>
- Van Ranst M. (2004). Chandipura virus: an emerging human pathogen?. *Lancet (London, England)*, 364(9437), 821–822. [https://doi.org/10.1016/S0140-6736\(04\)16995-X](https://doi.org/10.1016/S0140-6736(04)16995-X)
- Van Ranst M. (2004). Chandipura virus: an emerging human pathogen?. *Lancet (London, England)*, 364(9437), 821–822. [https://doi.org/10.1016/S0140-6736\(04\)16995-X](https://doi.org/10.1016/S0140-6736(04)16995-X)
- One case of Chandipura virus infection detected in Gujarat: All you need to know about the virus that affects children. (2019, July 17). The Indian Express.*

- <https://indianexpress.com/article/lifestyle/health/chandipura-virus-infection-gujarat-maharashtra-all-you-need-to-know-about-virus-affects-children-5833334/>
- Ba, Y., Trouillet, J., Thonnon, J., & Fontenille, D. (1999). Phlébotomes du Sénégal: inventaire de la faune de la région de Kédougou. Isolements d'arbovirus [Phlebotomus of Senegal: survey of the fauna in the region of Kedougou. Isolation of arbovirus]. *Bulletin de la Societe de pathologie exotique* (1990), 92(2), 131–135.16. Ba Y, Trouillet J, Thonnon J, Fontenille D. [Phlebotomus of Senegal: survey of the fauna in the region of Kedougou. Isolation of arbovirus]. Bulletin de la Societe de Pathologie Exotique (1990). 1999 May;92(2):131-135. PMID: 10399605.
- Kemp G. E. (1975). Viruses other than arenaviruses from West African wild mammals. Factors affecting transmission to man and domestic animals. *Bulletin of the World Health Organization*, 52(4-6), 615–620.
- Peiris JS, Dittus WP, Ratnayake CB. Seroepidemiology of dengue and other arboviruses in a natural population of toque macaques (*Macaca sinica*) at Polonnaruwa, Sri Lanka. *J Med Primatol*. 1993;22:240–5.
- Louten J. (2016). Virus Structure and Classification. *Essential Human Virology*, 19–29. <https://doi.org/10.1016/B978-0-12-800947-5.00002-8>
- Dubovi, E. J. (2011). *Fenner's Veterinary Virology* (6th ed., pp. 327-341). Academic Press. <https://www.sciencedirect.com/science/article/abs/pii/B9780123751584000183>
- Thoulouze, M. I., Bouguyon, E., Carpentier, C., & Brémont, M. (2004). Essential role of the NV protein of Novirhabdovirus for pathogenicity in rainbow trout. *Journal of virology*, 78(8), 4098–4107. <https://doi.org/10.1128/jvi.78.8.4098-4107.2004>
- Granoff, A. (1999). *Encyclopedia of Virology* (R. G. Webster, Ed.) (1st ed.). Academic Press. <https://www.sciencedirect.com/referencework/9780122270307/encyclopedia-of-virology#book-info>
- Pavitrakar, D. V., Atre, N. M., Tripathy, A. S., & Shil, P. (2021). Cyclophilin A: a possible host modulator in Chandipura virus infection. *Archives of virology*, 166(11), 3143–3150. <https://doi.org/10.1007/s00705-021-05237-1>
- Menghani, S., Chikhale, R., Raval, A., Wadibhasme, P., & Khedekar, P. (2012). Chandipura Virus: an emerging tropical pathogen. *Acta tropica*, 124(1), 1–14. <https://doi.org/10.1016/j.actatropica.2012.06.001>
- Damle, R. G., Patil, A. A., Bhide, V. S., Pawar, S. D., Sapkal, G. N., & Bondre, V. P. (2017). Development of a novel rapid micro-neutralization ELISA for the detection of neutralizing antibodies against Chandipura virus. *Journal of virological methods*, 240, 1–6. <https://doi.org/10.1016/j.jviromet.2016.11.007>
- Singh, B. (2023, June 17). *BHU Research On Chandipura virus spells hope for diagnosis*. The Times of India. Retrieved August 19, 2023, from <https://timesofindia.indiatimes.com/city/varanasi/bhu-research-on-chandipura-virus-spells-hope-for-diagnosis/articleshow/101058122.cms?from=mdr>

- Balakrishnan, A., & Mun, A. B. (2020). Ribavirin inhibits Chandipura virus replication in Vero cells. *Journal of medical virology*, 10.1002/jmv.26184. Advance online publication. <https://doi.org/10.1002/jmv.26184>
- Kitaura, S., Tobiume, M., Kawahara, M., Satoh, M., Kato, H., Nakayama, N., Nakajima, N., Komeno, T., Furuta, Y., Suzuki, T., Moriya, K., Saijo, M., Ebihara, H., & Takayama-Ito, M. (2023). Evaluation of a novel severe combined immunodeficiency mouse model for antiviral drug evaluation against Chandipura virus infection. *Antiviral research*, 213, 105582. <https://doi.org/10.1016/j.antiviral.2023.105582>

Photo Credits

Colored map of India indicating the amount of confirmed Chandipura virus cases and deaths. (2004). *The Lancet*. Retrieved October 31, 2023, from <https://www.thelancet.com/journals/lancet/article/PIIS0140673604169821/fulltext>.

In What Ways Are People's Private Information Protected in the Digital World, and How Can Related Issues Be Addressed? By Suahn Park

Abstract

This article examines people's rights to protection of private information online, which has been constantly violated, to address possible solutions to alleviate issues on privacy. In fact, over 422 million people in the US were victims of data breaches in 2022 due to inadequate protection policies and cyber attacks. There are countless illegal, unethical, and immoral uses of personal information such as using them for phishing, fraud, or identity theft. This paper analyzes the policies of companies and governments that already exist and cases of personal information leakage. Data breach usually happens through insufficient security measures, social engineering attacks, and data selling and trading even though there have been movements like the General Data Protection Regulation (GDPR) in the European Union (EU). It reviews how possible solutions in the scope of companies and government can prevent further harassment and alleviate current problems that can be observed in various cases, especially in social media.

Key Terms

For further understanding, there are some essential vocabularies. First of all, digital right refers to human rights encompassing freedom and protection all people deserve in the context of the internet, technology, and digital communication. Major rights include the right to privacy, which ensures the protection of personal information; freedom of expression, which allows one to express one's opinion freely; right to data protection, which gives responsibilities to organizations that have one's personal information to secure it; and right to encryption, which safeguards individuals from unauthorized access by coding the data with a key that only authorized people can access to. Also, privacy refers to one of the basic rights to keep certain information from being disclosed to the public or uncertified access. It can keep one's life and information confidential and autonomous. There are several fields in privacy: information, physical, communication, and personal opinions. These are defended under a legal framework through various conventions, laws, and constitutions in the interconnected world. Furthermore, data leakage is when privacy is violated through unauthorized access, system vulnerabilities, insecurity in data storage, etc. It includes sharing private information with unauthorized groups and disclosing it to the public.

Introduction

Everyone has experienced an uncanny situation online where it becomes clear that your data is being used to influence you, whether it be targeted marketing of products, or at its worst polarizing political views through perpetuation of misinformation (Selman). For instance, if one discusses headphones, one's social media pages will be filled with headphone advertisements the next day. Events like this have caused many people to question the rights they have in the digital world. One of the major issues involving online rights violations was the protection of personal

information, which includes sensitive information such as name, signature, address, phone number, and date of birth. As much as electronic devices are being used and developed, it is essential to examine what kinds of protection people have online and how to improve them.

Even though there are several existing policies in different countries to prevent harm (“What Is GDPR, the EU’s New Data Protection Law?”), violation is unavoidable with current policies. Considering the fact that significant policies apply to only a few countries and are not strongly implemented since there are no global internet security organizations, resulting in country by country making efforts to establish security, leakage is a very prevalent issue. Common misuses of personal information happen through data breaches, which are when unauthorized individuals or groups gain access to databases that contain personal information. This can lead to complex problems like identity theft and phishing, an intended social engineering attack to steal personal information, which can be very harmful to the victims. Thus, to prevent current and further problems with personal information leakage, there needs to be a feasible and clear universal policy to protect people.

Violations in History

As mentioned before there were data breaches resulting in illegal uses of them. One of the most significant ones was the Facebook-Cambridge Analytica data breach in 2015. Cambridge Analytica was a political consulting firm that had the personal data of approximately 87 million Facebook users collected by a third-party app “This Is Your Digital Life,” engineered by a researcher at Cambridge University. This app gathered data through surveys via Facebook that extracted information not only about personal data but also about users’ interests and preferences. It exploited the data for political advertisements during several election campaigns, including the 2016 US presidential election, to influence the voters. The data usage, however, violated Facebook’s policies. The users lost their trust and raised concerns about the protection of personal information on social media platforms, prompting contentious debates about how to enhance user privacy (Wong). Afterward, Facebook encountered countless criticism from the public and regulators, resulting in the implementation of stricter data access policies (Confessore).

Users of Yahoo, an internet company, also experienced data breaches. In 2013, about 1 billion users’ information was disclosed to a few hackers who perpetrated the system. This information was taken and used for spamming, identity theft, and other malevolent activities (Perlroth). Consequently, Yahoo made efforts to improve its security practices to protect user data by strengthening encryption methods, implementing two-factor authentication (2FA), and enhancing security infrastructure (Redfern).

Harms Due to Personal Information Leakage

There are several activities that companies can do with the personal information they stole. First of all, identity theft, one of the most significant risks, is when cybercriminals use stolen personal information like names, Social Security numbers, and financial details to

impersonate individuals and commit criminal activities in their name. This can lead to financial loss including damaging the victim's credit scores and damaged reputation. For instance, in the US, 40 million people were affected, resulting in \$43 million in total. Also, damaged credit scores can be critical for getting loans and even jobs. This indicates that it is not just money that is lost, but an entire professional life or plan to buy a house could go awry. In addition, phishing and scams are also widely seen. Using personal information, the criminals craft compelling phishing emails and messages to make users provide further information or click malicious links that will hack the device to get personal information. This usually leads to financial loss and further privacy breaches, which can link to another problem like an account takeover. Account takeover is when cybercriminals gain access to individuals' online accounts, which will lead to unauthorized activities, loss of control over personal information, and reputational damage.

For instance, Lt. Mike Gibson's case shows how identity theft led to economic loss and reputational damage. There was unauthorized access to his online trading account, filing a loan of \$10,000 from a university under his name. It was difficult to vindicate himself and to prevent further financial withdrawal. He could not get his money back and was framed as a thief by the university, causing unnecessary damage (Aaronson).

Current Policies

Even though countless people experience harm due to data breaches, some policies protect them. For instance, the European Union (EU) and the European Economic Area (EEA), which also includes Iceland, Liechtenstein, and Norway, have a regulation called the General Data Protection Regulation (GDPR). It contains strict regulations for commercial entities when they are collecting, progressing, and storing personal data and imposes financial penalties for non-compliance ("What Is GDPR, the EU's New Data Protection Law?"). Similarly, California has the California Consumer Privacy Act (CCPA), which grants people certain rights regarding personal information such as the right to know that data is being collected and the right to opt out of data sharing and obliges companies to have data protection and transparency (Korolov). On the other hand, even though some countries have tried to implement strict policies, other countries often do not have any general laws protecting people from data leakages. They would not be able to penalize or regulate cybercriminals adequately, indicating a need for the implementation of general international laws (Pirvan).

Causes of Private Information Leakage

Even though there are efforts to protect people from data leakage, data can and is being collected and misused without individual awareness. For instance, the Facebook-Cambridge Analytica data breach (Wong), which is mentioned above, shows the vulnerabilities and strong necessity to strengthen the policies to secure the data and ensure transparency. Weak security policies, lack of adequate encryption, and insufficient network monitoring can expose the data to external attacks (Kaspersky). In fact, 47 percent of data breaches occur due to malicious attacks, and these can be easily prevented by reinforcing security measures ("Top 3 Causes of Data

Breaches”). The inadequate implementation of these policies is also an issue (Aaronson). Even though there are policies regarding privacy, it is complicated to be checked and applied. Many companies do not have enough security measures, and there are very few worldwide policies regarding the protection of private and sensitive information. Consequently, the safety protocols are often violated, and many people attempt to use the data maliciously (Aaronson). Especially with weak policies, data safety is a very severe issue worldwide, as seen in the exponential growth of its cases: 157 cases in 2005 and 1802 cases in 2022, affecting approximately 422.14 million people. Moreover, data can leak due to the complex data ecosystem, leading to a leakage from third parties. It is very difficult for a company to monitor all the third parties involved in the data processing, showing the necessity for more robust regulations.

Possible Solutions

As digital crimes are increasing these days, there is a need to enhance digital rights. First of all, more robust compliance with the privacy policies is needed. Companies should enhance accountability through clear descriptions about their responsibility, reasons, and authentication when they collect personal data. This transparency will ensure users a safer and more protected digital life. Considering that plenty of companies generate revenue by selling the personal information of their customers, governmental policy and incentives should ensure that each and every personal information is kept safe. Additionally, companies themselves have to make efforts to implement stronger policies. This can be done by restricting access to sensitive data through authentication methods, role-based access controls, and least privilege principles. It will ensure that only authenticated people access personal data. To make the protection even stronger, companies can encrypt the data and perform regular security assessments. Encryption allows the data to be unusable without appropriate decryption keys even if it gets stolen. This diminishes the risk of malevolent misusing of the data and allows another layer of protection. Conducting regular security audits can help companies identify potential weaknesses in the system and address them before any unauthorized access happens. Moreover, third-party risk management is very important. When companies work with several different partners, they should measure other companies’ security practices to ensure that they abide by apt data protection standards and policies. Moreover, punitive damages indemnification, which is compensation for those who experienced unpleasant situations, should be reinforced. Many countries have very weak compensation policies, reimbursing very small amounts of money. By improving safety measures, the companies would try to prevent data breaches to avoid financial damage. Adopting these measures can significantly reduce the risk of data breaches and enhance data security. Companies should take proactive actions to protect data.

At the governmental level, there are various actions a government can take to prevent data breaches. Government action is very crucial since it has the power to impose a policy that can protect all citizens from potential threats. For instance, implementing and enforcing data protection laws can help enhance the digital rights of the people. By listing clear guidelines for what personal data organizations can collect and how they subsequently handle it, governments

can help companies form appropriate and feasible policies. It can include requiring understandable consent agreements for data collection and encryption of data and penalties for non-compliance. For example, confining the degree of data that can be collected by each type of business, also known as data minimization, can minimize the damages by gaining only information that is needed. Also, the government can support research and development of cybersecurity systems that can prevent data breaches effectively by providing financial support. Conducting frequent security audits and assessments at the governmental level can bring larger impacts, too. It allows the government to keep companies under surveillance. It would make organizations and companies follow rules more strictly, which ensures adequate implementation of policies. To assist small and medium-sized entrepreneurs who have difficulties making strong cybersecurity measures due to a paucity of resources, the government can supply them with incentives and resources to strengthen data protection. These efforts to safeguard the sensitive information stored in the digital world would certainly reduce unwanted incidents.

Conclusion

To sum up, due to the dearth of adequate implementation of privacy policies online, millions of people have experienced unauthorized access to and detrimental consequences of their data being misused. To address this issue, companies and governments should make efforts to create clearer policies with regular audits, strengthen existing policies with strict penalties, and create a general law that applies to more people and regions. Since data misuse can cause reputational damages, identity thefts, and financial loss, which are very crucial, adopting policies to cope with rising problems is essential and urgent to protect individual rights online.

Works Cited

- Aaronson, Susan Ariel. "Inadequate Data Protection: A Threat to Economic and National Security." *CEPR*, 5 Feb. 2020, cepr.org/voxeu/columns/inadequate-data-protection-threat-economic-and-national-security.
- Aaronson, Trevor. "Hacking Team Data Breach Provides Links to Florida Law Enforcement." *Tampa Bay Times*, 25 Aug. 2019, www.tampabay.com/news/publicsafety/hacking-team-data-breach-provides-links-to-florida-law-enforcement/2237006/.
- Confessore, Nicholas. "Cambridge Analytica and Facebook: The Scandal and the Fallout so Far." *The New York Times*, 4 Apr. 2018, www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html.
- "Cyber Security Solutions for Local Government." *TitanHQ*, www.titanhq.com/cyber-security-solutions-for-local-government/?utm_campaign=THQ-PM-T1&acctid=THQ&utm_source=Adwords&utm_medium=PPC&keyword=&matchtype=&campaignid=14035790152&adgroupid=&gclid=Cj0KCQjw5f2lBhCkARIsAHeTvlgZkb2FxfPtYCtaRPT6EdjvKilSoZXfJU1WJdEPNWl_Kxh5hLYg45IaAuMjEALw_wcB&network=x&device=c. Accessed 25 July 2023.
- "Data Breach: Examples, Causes, and How to Prevent the next Breach." *HackerOne*, www.hackerone.com/knowledge-center/data-breach-examples-causes-and-how-prevent-next-breach. Accessed 25 July 2023.
- "A Guide to the Data Protection Principles." *ICO*, ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-protection-principles/a-guide-to-the-data-protection-principles/. Accessed 25 July 2023.
- Heiligenstein, Michael X. "Facebook Data Breaches: Full Timeline through 2023." *Firewall Times*, 31 May 2023, firewalltimes.com/facebook-data-breach-timeline/.
- "How to Effectively Manage a Data Breach." *SecurityMetrics*, 19 July 2023, www.securitymetrics.com/learn/how-to-effectively-manage-a-data-breach.
- Kaspersky. "How Data Breaches Happen." *Www.Kaspersky.Com*, 18 May 2023, www.kaspersky.com/resource-center/definitions/data-breach.
- Korolov, Maria. "California Consumer Privacy Act (CCPA): What You Need to Know to Be Compliant." *CSO Online*, 7 July 2020, www.csoonline.com/article/565923/california-consumer-privacy-act-what-you-need-to-know-to-be-compliant.html.
- Perlroth, Nicole. "All 3 Billion Yahoo Accounts Were Affected by 2013 Attack." *The New York Times*, 3 Oct. 2017, www.nytimes.com/2017/10/03/technology/yahoo-hack-3-billion-users.html.
- Petrosyan, Ani, and Apr 1. "Number of Data Breaches and Victims U.S. 2022." *Statista*, 1 Apr. 2023,

- www.statista.com/statistics/273550/data-breaches-recorded-in-the-united-states-by-number-of-breaches-and-records-exposed/.
- Pirvan, Petruta. "EU GDPR Applicability to International Organizations." EU GDPR Applicability to International Organizations, 12 Mar. 2021, iapp.org/news/a/eu-gdpr-applicability-to-international-organizations/#:~:text=The%20statement%20that%20an%20international,and%20rigors%20of%20the%20GDPR.
- Redfern, Elizabeth. "The Yahoo Cyber Attack & What Should You Learn from It?" Cashfloat, 1 Mar. 2023, www.cashfloat.co.uk/blog/technology-innovation/yahoo-cyber-attack/.
- Selman, Hanna. "Why We See Digital Ads after Talking about Something." *McNutt & Partners*, 25 Jan. 2021, www.mcnuttpartners.com/why-we-see-digital-ads-after-talking-about-something/.
- "Top 3 Causes of Data Breach Are Expensive." Calyptix Security Top 3 Causes of Data Breach Are Expensive Comments, 29 June 2017, www.calyptix.com/technical-insights/top-3-causes-data-breach-expensive/.
- "What Is GDPR, the EU's New Data Protection Law?" *GDPR.Eu*, 26 May 2022, gdpr.eu/what-is-gdpr/.
- Wong, Julia Carrie. "The Cambridge Analytica Scandal Changed the World – but It Didn't Change Facebook." *The Guardian*, 18 Mar. 2019, www.theguardian.com/technology/2019/mar/17/the-cambridge-analytica-scandal-changed-the-world-but-it-didnt-change-facebook.

Predicting the Size of Asteroids with Random Forest Classifiers By Alex Stoffel

Abstract

Asteroids are remnants from early solar systems, often caused by large bodies of mass colliding with each other. To make predictions about these asteroids, we can use machine learning. In this project I used a random forest classification model, a set of decision trees that a computer compiles to classify these asteroids according to a particular heuristic. A decision tree is a series of yes-no questions the computer answers, eventually resulting in an object classification. After using a total of 50 decision trees, our program could classify with 99.86% accuracy whether an asteroid was large or small using the Open Asteroid Dataset. We consider minor asteroids smaller than the mean of those in our data set and large asteroids to be larger than the mean of our data set. Classifying asteroids is crucial because sometimes we cannot measure the size of an asteroid directly. Still, by using this program, we can predict its size by measuring other aspects. It is essential to know the size because it will help us predict asteroid collisions with a high degree of accuracy.

Introduction

Did you know that on average 17,000 asteroids hit Earth yearly [11]? Although many of these asteroids are not large enough to cause any real damage to Earth, it is essential to know the size of these asteroids so that appropriate measures can be taken, in the case of destructive asteroids; these measures would be used to evacuate populations endangered by the colliding asteroid. Our program will be able to predict the dimensions of these asteroids without directly being given the information of how big it is. The program that we code will need to be able to classify an asteroid as either large or small by analyzing several different statistics about the asteroid and taking that information to predict its size.

Asteroids are rocky, metallic objects that orbit the sun and can be considered minor planets. They range in size from pebbles to large bodies like Ceres [9]. Most asteroids are in the central asteroid belt between Mars and Jupiter [9]. They are remnants from the early solar system and can provide insights into planetary formation. Meteorites, remnants of meteoroids that survive atmospheric entry, help us/scientists/meteoriticists study asteroid composition [10]. Spacecraft missions like Galileo and NEAR have provided us/scientists/meteoriticists with close-up asteroid encounters [5].

Here, we explore how we can use machine learning to predict the size of asteroids by using several different characteristics. Table 1 shows definitions of some of the key features used to describe asteroids in the dataset used. Are these all the variables considered by your decision tree? If so, state that, if not, you need to go into more detail about how the decision tree works

Table 1: Key features present in the asteroids' dataset and their definitions.

semi-major axis (au)	An asteroid's average orbital distance from the sun. [4]
eccentricity	How far away from a perfect circular orbit an asteroid is. If the eccentricity is lower, the asteroid's orbit is more circular, and vice versa. [15]
magnitude slope parameter	How big the asteroid would look if placed exactly one AU away from the observer. An AU is an astronomical unit. [8]
inclination with respect to x-y ecliptic plane (deg)	The tilt of the asteroid's orbit around the planet. [12]
Longitude of the ascending node	The angle in which the orbit is rotated from the reference plane's vernal point. The vernal point is the point on the elliptic when the sun crosses from north to south. [6]
argument of perihelion	The angle between the ascending node and the point where the orbit is closest to the Sun. The ascending node is where the asteroid's orbit intersects with the regular orbit and passes from north to south. [2]
perihelion distance (au)	The point where an asteroid is nearest to the Sun. [14]
aphelion distance (au)	An asteroid's furthest distance from the sun. [1]
orbital period	How long it takes for an asteroid to make one orbit around the Sun. [13]
geometric albedo	The ratio of a body's brightness at zero phase angle to the brightness of a perfectly diffusing disk with the same position and apparent size as the body. [7]

rotational period	The time it takes for an asteroid to rotate once.
-------------------	---

Table 2 shows the machine learning definitions required to complete this project:

Table 2: Definitions of important machine learning terms.

random forest classifier	A random forest classification model is a set of decision trees that a computer will use to classify asteroids according to a particular heuristic. [19]
decision tree	A decision tree is a series of yes-no questions the computer can answer, eventually resulting in a classification of the object being tested. [16]
supervised learning	A type of machine learning in which the data contains features and labels, and an algorithm is used to predict the labels from the features. [17]
unsupervised learning	The data does not have labels. The goal is to model the underlying structure of the data to learn more about it and understand the input. [18]
reinforcement learning	The machine decides what action to take and is given positive and negative rewards. The program must find the best possible path to achieve the reward. [3]

Results & Data Set:

For this project, we imported a data set from Kaggle titled “Open Asteroid Data Set.” It includes a list of the characteristics of over 800 asteroids in the solar system. To be able to make predictions about the data, we needed to use a machine learning algorithm, and we decided to use the random forest classifier method. A random forest classifier groups the data into a classification we chose. We decided that our machine-learning algorithm would classify the asteroids as large or small. Large asteroids are those whose size is larger than the mean of all of the sizes of the asteroids in our data set, and the smaller asteroids are those whose size is smaller.

After running the algorithm, the accuracy with which our program could correctly classify the asteroids is 99.86%, which is an excellent result. Here is a look at some of the trees our program used:

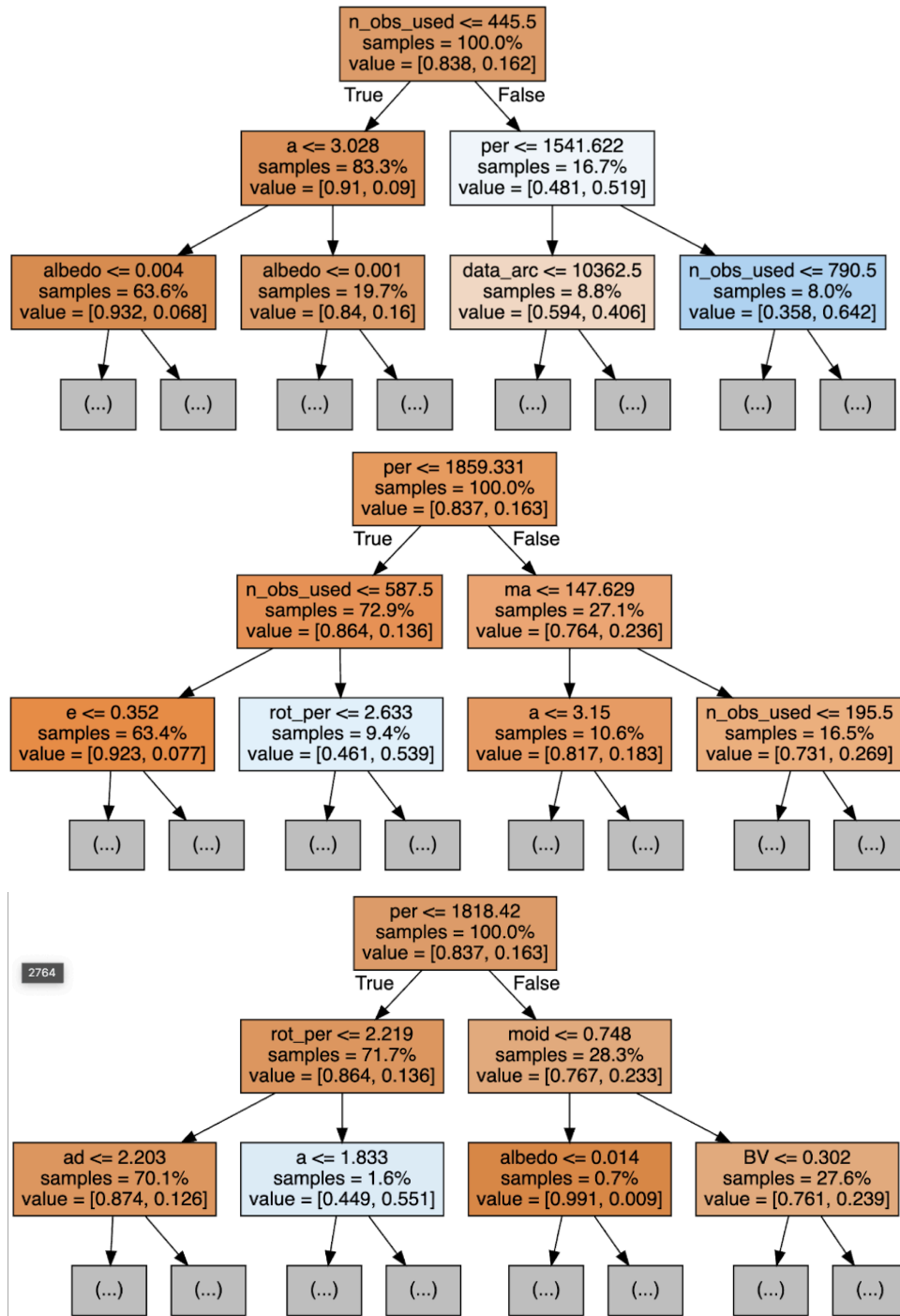


Figure 1: Sample decision trees from the random forest classifier.

After these trees, the program gave us a list of all of the different feature importances (how important each feature of the data set was to make the classifications). These are shown in Table 3.

albedo	7.45e-01
H	8.72e-02
n_obs_used	7.94e-02
data_arc	2.67e-02
i	6.98e-03
per_y	6.20e-03
per	5.84e-03
rot_per	5.78e-03

Table 3: Relative feature importances of different features in the dataset.

Discussion

Albedo was the most crucial feature because it can help tell us the surface area of the asteroid due to how much light is being reflected off the surface. The second most important was H, which is how large the asteroid would appear if it were placed one astronomical unit away from the observer. The feature that leads to the greatest split is the number of data pieces used (n_obs_used), which splits the data that separates large and small asteroids quite well (closest split to 50/50). Another feature that led to a significant data split was Earth's Minimum Orbit Intersection Distance (MOID), the distance between the closest point of the asteroid and Earth's orbits around the sun. Because of our findings on what characteristics of asteroids best classify them as large or small/

Conclusion

In the future, to make our algorithm more accurate, I would use more decision trees to allow for a more in-depth analysis of the data, which would ultimately lead to higher accuracy. Some other things to increase the accuracy of the program are to input more data into the algorithm, use more epochs, use a neural network instead of a decision tree, or use a different data set altogether.

Works Cited

- “Aphelion | Definition & Facts.” *Britannica*, Accessed 25 August 2023
- “Argument Of Perihelion | COSMOS.” *Centre for Astrophysics and Supercomputing*, Accessed 25 August 2023.
- Bhatt, Shweta. “Reinforcement Learning 101. Learn the essentials of Reinforcement... | by Shweta Bhatt.” *Towards Data Science*, 19 March 2018, Accessed 25 August 2023.
- “Data Table for Planets & Dwarf Planets: AU, Inclination, Eccentricity.” *Windows to the Universe*,
https://www.windows2universe.org/our_solar_system/planets_orbits_table.html. Accessed 25 August 2023.
- “Exploration | Asteroids – NASA Solar System Exploration.” *NASA Solar System Exploration*,
https://solarsystem.nasa.gov/asteroids-comets-and-meteors/asteroids/exploration/?page=0&per_page=10&order=launch_date+desc%2Ctitle+asc&search=&tags=Asteroids&category=33. Accessed 25 August 2023.
- “Glossary.” *Glossary*, <https://cneos.jpl.nasa.gov/glossary/node.html>. Accessed 25 August 2023.
- “The H and G magnitude system for asteroids.” *NASA/ADS*,
<https://adsabs.harvard.edu/full/2007JBAA..117..342D>. Accessed 25 August 2023.
- “In Depth | Asteroids – NASA Solar System Exploration.” *NASA Solar System Exploration*,
<https://solarsystem.nasa.gov/asteroids-comets-and-meteors/asteroids/in-depth/>. Accessed 25 August 2023.
- “In Depth | Meteors & Meteorites – NASA Solar System Exploration.” *NASA Solar System Exploration*,
<https://solarsystem.nasa.gov/asteroids-comets-and-meteors/meteors-and-meteorites/in-depth/>. Accessed 25 August 2023.
- “Meteorites on Earth: how many fall per year and why don't we see them?” *Iberdrola*,
<https://www.iberdrola.com/innovation/meteorites-earth>. Accessed 25 August 2023.
- “Orbital Inclination | COSMOS.” *Centre for Astrophysics and Supercomputing*,
<https://astronomy.swin.edu.au/cosmos/O/orbital+inclination>. Accessed 25 August 2023.
- “Orbital period | astronomy.” *Britannica*, 15 August 2023,
<https://www.britannica.com/science/orbital-period>. Accessed 25 August 2023.
- Pultarova, Tereza. “Perihelion: What is it and when does it occur?” *Space.com*, 4 January 2023,
<https://www.space.com/what-is-perihelion>. Accessed 25 August 2023.
- Suter, Paul M. “Astronomy Jargon 101: Eccentricity.” *Universe Today*, 8 February 2022,
<https://www.universetoday.com/154323/astronomy-jargon-101-eccentricity/>. Accessed 25 August 2023.
- “What is Decision Tree Analysis? Definition and FAQs.” *HEAVY.AI*,
<https://www.heavy.ai/technical-glossary/decision-tree-analysis>. Accessed 25 August 2023.
- “What is Supervised Learning?” *IBM*, <https://www.ibm.com/topics/supervised-learning>.

Accessed 25 August 2023.

“What is Unsupervised Learning?” *IBM*, <https://www.ibm.com/topics/unsupervised-learning>.

Accessed 25 August 2023.

Yiu, Tony. “Understanding Random Forest. How the Algorithm Works and Why it Is... | by Tony Yiu.” *Towards Data Science*, 12 June 2019,

<https://towardsdatascience.com/understanding-random-forest-58381e0602d2>. Accessed 25 August 2023.

Three Traits to Rule Them All: A Recipe for Exceptional Business Leadership By Eric Azrak

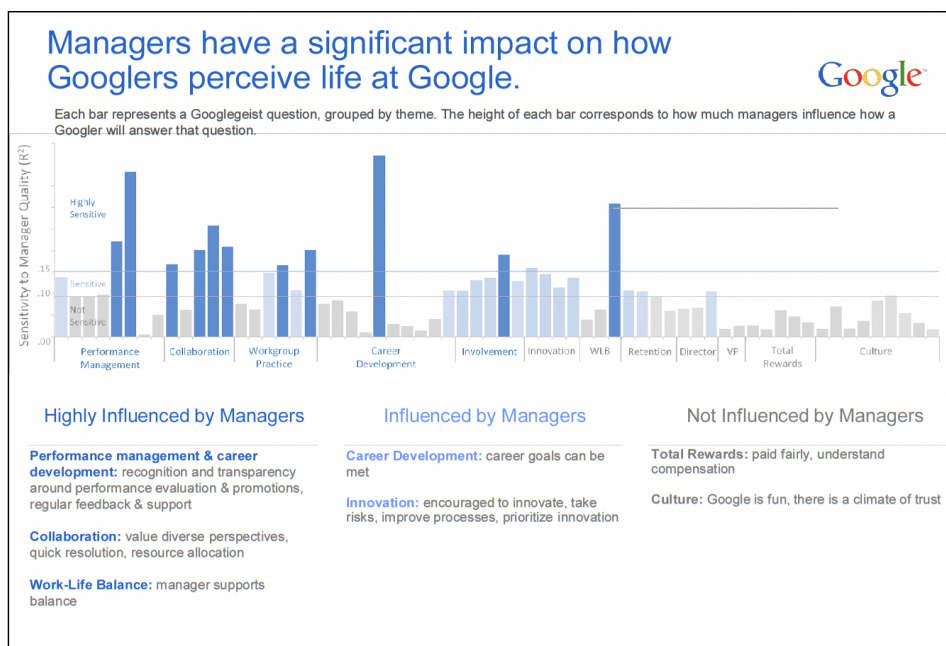
Introduction

How can you extract the very best from your team? It is the function of those who lead to enhance the led, not only as individuals but as part of something greater. While each leader employs their own style inspired by their strengths and weaknesses, a handful of managerial qualities tend to overlap among the finest directors. Understanding those specific traits and their effect on others is the key to taking your organization to the next level.

Section 1: Introducing Google’s Project Oxygen

In the early 2000s, Google LLC asked themselves a simple question: Do managers matter? As Google’s labor force was mainly composed of engineers, the introduction of subordinate and superordinate positions was not welcomed with open arms. Nadav Eiron, Google LLC engineering director, expounded: “Engineers generally want to spend their time coding and debugging. Many think that talking to direct reports gets in the way of getting that work done. And without training, some engineering managers have a hard time striking a balance between providing direction and micromanaging.” Thus, Google searched for hard evidence that would persuade their employees – and perhaps themselves – to believe in the benefits of proper management. Under the leadership of Google’s People and Innovation Lab (PiLab), as well as a handful of research experts, Project Oxygen was born to answer Google’s question.

Neal Patel and Michelle Donovan, Project Oxygen co-founders and directors, began their quest to prove managers *did* matter by doing the exact opposite, trying to prove managers *didn’t* matter. “Luckily, we failed,” commented Patel. The team had turned to performance reviews and



Googlegeist scores, an annual employee satisfaction survey regarding leadership, for data on existing management at Google. By comparing high-scoring (top quartile) and low-scoring (bottom quartile) managers, Patel and Donovan hoped to spot significant differences that distinguished the two groups. The discrepancies were incremental but monumental; the slightest variation in manager quality had notable effects on job satisfaction, employee retention, work-life balance, career development, and performance (see **Exhibit 1** above for more). Now that managers definitively mattered, naturally, the following question was, “What do our best managers do?”

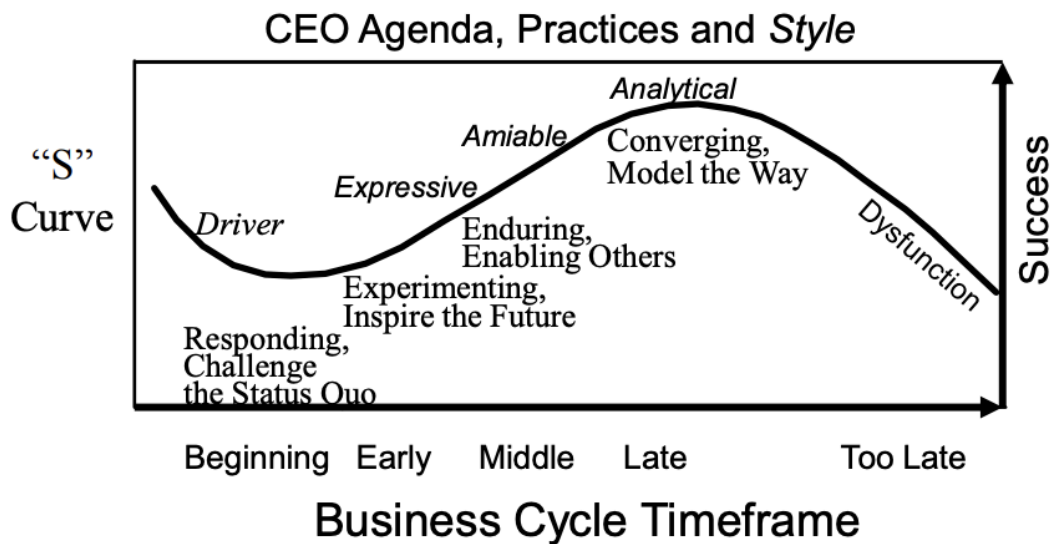
Utilizing double-blind qualitative interviews, the Project Oxygen team had collected a vast amount of data that pointed them to eight characteristics of great managers. Interviewers were given a script of questions regarding managerial behavior/tendencies, and polled high-scoring and low-scoring managers across all departments, levels, and geographic locations. After combing over the survey responses, thousands of Googlegeist survey comments, and more performance reviews, PiLab identified eight behaviors that were often shared by the high-scoring managers. The eight attributes were soon nicknamed the ‘Oxygen 8.’

At the top of Google’s Oxygen 8 list came **“being a good coach,”** understanding the niche of employees and pushing them to grow in a healthy manner. At the number two spot, a great manager **“empowers the team and does not micromanage,”** instilling a strong sense of trust within their team yet still available to answer questions. The following trait seems basic, but it’s essential to **“express interest/concern for team members’ success and personal well being;”** care about those who you work with. The fourth quality fits the more traditional role of a manager, as **“productive and results-orientated”** are the cornerstones of efficiency. Next up, trait five emphasizes that communication is key: **“is a good communicator - listens and shares information.”** An excellent manager shouldn’t only care about your personal well-being but also your career’s well-being, and hence, the sixth characteristic, **“helps with career development.”** The seventh trait is an essential function of leadership, underscoring direction, **“has a clear vision/strategy for the team.”** Last but certainly not least, a distinguishing characteristic of any team leader is that a great manager **“has key technical skills that help him/her advise the team.”** On their own, these eight qualities are influential, but as a group, the Oxygen 8 are exceptionally powerful. The aim of this paper, though, is to determine which *three* of the *eight* Google Oxygen traits are the *most* crucial for managerial success, later substantiated by the experience of two notable business leaders.

While Google – and this paper – use the terms interchangeably, “manager” and “leader” aren’t always synonymous. John P. Kotter of Harvard Business Review (HBR) explained: “Management is about coping with complexity; it brings order and predictability to a situation. But that’s no longer enough—to succeed, companies must be able to adapt to change. Leadership, then, is about learning how to cope with rapid change.” Kotter’s distinction offers a novel perspective to those in charge. While it’s *important* to adhere to Oxygen 8 traits, it’s *vital* to evaluate the need for leadership versus management during times of stress.

Section 2: Essential Context

Any analysis of business leadership would be incomplete without acknowledging the pivotal research that precedes it. Especially relevant to the discussion of Google’s Project Oxygen, *Why Should Anyone Be Led by You?: What It Takes To Be An Authentic Leader* by Robert Goffee and Gareth Jones explores the notion of authentic vs. quality-based leadership. The pair explain that the greatest leaders must leverage their own qualities to develop their own management style, and “In our view, there are no universal leadership characteristics. What works for one leader will not work for another.” A stark departure from an “Oxygen 8” trait-centric recipe, Goffee and Jones emphasize that leaders come in all shapes and forms. Some are technical experts in their fields, such as Bill Gates, who they dub the “ultimate computer geek,” and others are magnetic personalities like Steve Jobs or Richard Branson. Goffee and Jones also highlight that leadership is fundamentally situational, and success oftentimes depends on how well one’s style fits the current niche. Likewise, another essential piece of leadership literature, The Leadership S-Curve (see **Exhibit 2** below) supports the fickle nature of management. Each section of the cycle calls upon an alternate leadership style, ultimately requiring those at the top to shift gears in response. It’s crucial to remain dynamic and avoid staticity.



Section 3: Yvon Chouniard vs. Tim Cook in the Context of the Oxygen 8

In order to effectively sift through the “Oxygen 8” and extract the three foremost important management traits, let’s take a look at two monumental yet entirely different business leaders: Yvon Chouniard and Tim Cook.

Yvon Chouinard

Born in 1938 to a French Canadian farmer in Lisbon, Maine, Yvon Chouinard instinctively inherited the family values, hard work, and appreciation of quality tools. His mother, Yvone, exuded fearlessness and responsibility. She made many of the family decisions, and took that responsibility very seriously. Yvonne also took responsibility in another way, which Chouinard would later note “was probably my first lesson in philanthropy.” Seeking an arid climate better suited for her husband’s asthma, Yvonne had auctioned all of their family’s possessions and landed on Burbank, California. At some point along the 3,000-mile journey across the country, she spotted an Indian hogan, and on it, a Hopi woman with several hungry children. Yvonne gave her all of their preserved corn for the trip - mother to mother.

Even in high school, Chouinard could never be put in a box. He was a straight D-student, and had little interest in the subjects except, of course, for the shop classes. A self-labeled troublemaker, Chouinard would frequently find himself in detention writing “I will not...” lines hundreds of times. A great example of his entrepreneurial spirit, though, Chouinard boasted he would “take three pencils and line them up with sticks and rubber bands so I could do three lines at once.” Chouinard attended community college after his 1956 high school graduation, but taught himself Blacksmithing 101 as multi-day Yosemite climbs demanded. At the time, climbing gear was dominated by the European “conquer” culture; move on and don’t look back. All the gear was made single-use, and once inserted into the mountain/wall, the iron piton heads tended to break off. This didn’t sit well with American climbers. “You [Americans] climb the mountains or visit the wilderness but leave no trace of having been there,” explained Chouinard, who attributed the perspective to a swirling transcendentalist thought.

Recognizing the need for reusable state-of-the-art pitons, Chouinard got to work. He sold out of his garage to friends and then friends of friends for \$1.50 each, pricy next to the European twenty-cent standard, but you had to use Chouinard’s gear for the cutting-edge climbs they were doing. The margins remained slim, yet once he netted enough to fund his own climbs, Chouinard would head back to what he loved: the natural world. He slept two-thirds of the year in an army-surplus sleeping bag and didn’t buy a tent until he was forty.

After the Vietnam War came knocking in the fall of 1962, Chouinard reopened the shop closer to the waves in Ventura, California, in 1966. Demand soon outgrew supply, however, forcing Chouinard to scale up with advanced machinery and new recruits. Chouinard Equipment mostly hired friends, and Chouinard added that “none saw the business as an end itself. It was just a way to pay the bills so we could go off on climbing trips.” Fast forward four years, and Chouinard Equip. became the largest supplier of climbing hardware in the United States; it had also become an environmental menace. The continued hammering of steel pitons, which made up most of the business, in and out of the fragile cracks permanently disfigured the very rocks Chouinard adored. Luckily, there existed an alternative in the form of aluminum chocks that wedged in between cracks, and Chouinard Equipment shortly phased out steel pitons.

As the years piled up, so did Chouinard Equipment’s profits, and Yvon Chouinard saw a new opportunity within the climbing market: apparel. In the late sixties, active sportswear wasn’t

fashionable. From the cut to the colors, it was a bland trainwreck. So, when Chouinard returned from a Scotland climbing trip in a regulation team rugby shirt, people noticed. Patagonia was born three years later in 1973, and it is now an office and household name fifty years later.

In the wake of Chouinard Equipment and Patagonia's success, Yvon Chouinard could no longer duck the title of "businessman." Yet, he was adamant about "distance[ing] myself as far as possible from those pasty-faced corpses in suits I saw in airline magazine ads. If I had to be a businessman, I was going to do it on my own terms." He further declared that work *must* remain enjoyable on a daily basis, and it should be surrounded with friends who can dress however they choose, even barefoot.



Chouinard envisioned a radical realm that blurred the line between work and family: the office. He read every book on business he could get his hands on and took a special interest in Japanese plus Scandinavian styles of management. Chouinard prided himself on his MBA theory of management, management by absence. He advocated that a company needs an explorer to go out and get the "temperature of the world," meanwhile Patagonia would continue on. His attitudes and efforts all culminated in Patagonia's unique environment. Many employees surfed at lunch or played volleyball in the sandpit, and most frequented the slopes for ski and climbing trips - often sponsored by Patagonia! During Chouinard's tenure, he would regularly bring top managers to Argentina, nurturing their creativity and moral codes in tandem. He would also lead weekly employee seminars to Yosemite National Park or San Francisco's Marin headlands for the same purpose at a more granular scale. Chouinard, a Zen-Buddhist at heart, gleaned from wherever he could, notably the Iroquois tribe. As a people with a deep connection to the Earth, the Iroquois couldn't afford to be near-sighted. Therefore, to ensure the needs of the future were considered, they designated one of their own as a representative of seven generations ahead; Chouinard realized the custom's value and instituted it into Patagonia. The administrative building is devoid of private offices, even for executives. Perhaps the most remarkable feature of the office, though, is the on-site child care center, Great Pacific Child Care Development Center, Inc. Contemporaneously, it was one of only a hundred-twenty nationwide; today, there are over 8,000. Chouinard commented, "The presence of children playing in the yard with their mothers and fathers in the cafeteria helped keep the company atmosphere more familial than corporate." Patagonia also offers flexible work hours, maternity/paternity leave, and job sharing in addition to child care.



Consistency is key, especially with messages. Patagonia was one of the first companies to dedicate itself to impact-conscious business in the seventies with recycled catalog paper, and decades later, their branding or company values haven't budged. Each year, the company donates ten percent of profit or one percent of sales (whichever is greater) to environmental causes, principally grassroots efforts. Since 1985, the business has given away north of \$140 million. Patagonia funds and facilitates employee environmental action, as well as sponsors small groups of peaceful, civil disobedient protesters. In 2007, Patagonia was one of seventy-two original enterprises to incorporate as a B-corporation, a number which has grown exponentially since. When asked why he's in business, Chouinard explained that he wanted to inspire other enterprises to environmental action just as their top-of-the-line gear has inspired other manufacturers. Patagonia's mission statement captures its values quite nicely: "Make the best product, cause no unnecessary harm, and use business to inspire and implement solutions to the environmental crisis."

Tim Cook

Timothy Donald Cook was born in Mobile, Alabama, on the first of November, 1960. The Cooks lived the quintessential small-town, rural life of the mid-20th-century South. His parents, Donald and Geraldine, had grown up not more than 30 miles apart, and farms dotted the horizon as far as the eyes could see. The two tended to keep to themselves, a trait later adopted by their children, and Donald worked long hours at low-level jobs to support their family. Tim knew his father didn't enjoy it and grew up dreaming of a different – more fulfilling – life, one where he worked at what he loved.

In school, Tim was a shy intellectual who never failed to submit homework on time. He was named "Most Studious" by his classmates for several years, and set his eyes on Auburn University. During downtime, Cook could usually be found with fellow members of the school band or socializing with the athletes and drama students. He was the school yearbook business manager, and facilitated quarter-page ad sales from local businesses. Notably, Cook was the first in the position to collect payments periodically instead of scrambling to grab the lump sum at the deadline, according to his teacher, Ms. Davis.

Early on in his life, Cook was forced to reconcile the racially divided nature of his surroundings. One day in middle school, as he was riding home on his bicycle from school, Cook stumbled upon a crowd of hooded men standing in front of a burning cross driven into the front yard of a local Black family. He shouted, "Stop!" and was shocked when one of the men shifted his gaze toward him, lifted his hood, and was revealed to be a deacon from a nearby church; Cook pedaled away as fast as he could. The scene stayed with him for decades.

In 1977, Cook tore open his long-awaited letter of acceptance to Auburn. He had started on his goal to be an engineer and enrolled in the industrial and systems engineering department. He had learned to question everything and challenge those who argue, "It's just the way it's done." Not surprisingly, he was offered a position at Anderson Consulting and General Electric in 1982 but ended up at International Business Machines. At that time, IBM's Personal Computer

(or PC) had just flown off shelves and netted the company nearly a billion in sales. The environment at IBM was perfect for Cook, riddled with bureaucracy, structure, and engineering jargon. It was the soulless corporate empire to its inspiring rebel rival: Apple.

Cook was a natural. He quickly made his mark in materials management, and was classified as a HiPo (high potential) by senior management. Recognizing his immense promise, IBM soon paid for his MBA at Duke University's Fuqua School of Business, and Cook took on a new appreciation for marketing, finance, and other foreign business disciplines. In 1988, he graduated in the top ten percent of his class and headed back to IBM until 1994, when Louis Gerstner became CEO. At that moment in time, IBM was transitioning from primarily product-based to service-based, a shift that would diminish the company's relevance in Cook's eyes. Fortunately, Intelligent Electronics poached Cook shortly beforehand for an operations position, which he accepted as a way out of IBM.



Cook's timing couldn't be worse. Cook quickly learned that Intelligent Electronics (IE) was an operations disaster. Wasted warehouse space and convoluted software systems as a result of sloppily acquiring rivals left IE's margins paper thin. Even after Cook came to the rescue, total profits still sank, leading the company to ultimately agree upon selling the business. Competitor Ingram Micro offered a price that Cook thought to be too low, and while some executives pushed to sell, Cook convinced them to patiently wait for a higher bid. He eventually persuaded Ingram Micro to raise their price by several millions of dollars, striking a \$78 million dollar acquisition.

In 1997, Cook headed to, arguably, the number one tech company in the world at the time: Compaq. There, he was confronted by another inventory headache with the ratio of warehouse-to-manufacturing space at two-to-one. In response, Cook worked to shorten the amount of time between product manufacturing and delivery; it did the trick. Only a year later, however, Cook decided on exiting Compaq after picking up a call from Rick Devine, executive recruiter at what would become the number one tech company of all time, Apple.

Early on at Apple, Cook called an operations meeting to familiarize himself with every detail of the company's infamously terrible supply chain management. Acting head of operations when Cook arrived, Joe O'Sullivan, commented, "I saw grown men cry," referencing Cook's myriad of questions to the operations department. "He went into a level of detail that was phenomenal." The meeting would foreshadow Cook's leadership style, whipping staff into shape with rigorous interrogation at an extremely precise level; there was no room for error. Tripp Mickle, author of *After Steve: How Apple Became a Trillion-Dollar Company and Lost Its Soul*, put it best, "His Socratic style created a tense atmosphere that caused staff to squirm... The operations team learned to scour every aspect of production and prepare to answer any question

Cook could imagine. They drilled down into the performance of specific parts and each assembly line’s production results. Their boss’s appetite for specifics led everyone to become ‘almost Cook-like,’ O’Sullivan said.” Incredibly, Cook soaked up all the new information and began to work at the company’s operation troubles immediately.

Cook wanted more. After eliminating all excess inventory (which Cook called “fundamentally evil”), he traveled to Singapore to hear firsthand about operations (ops.) overseas. The Singapore ops. team described their success throughout the prior year and, specifically, how they bumped inventory turns (a measure of the frequency in which inventory was used and replaced) from eight times to twenty-five times per year. Regardless of the rate being second in the industry, Cook had pushed, “How would you get it to a hundred turns?” O’Sullivan answered, “I knew you’d say that. We’re nearly there.” Stone-faced, Cook asked him another question, “How would you get to a thousand?” Some laughed, but those who



knew Cook understood he was serious. In a few years, Apple would soon mark no inventory on their books through a process that would be called Line Stripping or Floor Marking. The ops. team painted a yellow line down the center of factory floors; one side belonged to suppliers and was written on supplier’s books until the components were moved across the line, only when used in on-demand assembly. Thus, on paper, there hadn't been any inventory. The practice soon became standard across “Big-Tech.” Cook had instilled in them a hunger for pure excellence (see **Exhibit 3** below).

Company politics were anathema to Cook, another one of his leadership quirks. At Apple, known to have its fair share of hotheads and drama, Operations was the odd one out. Toward the end of each fiscal quarter, Cook called for an Operations meeting to discuss any shortcomings. About a dozen higher-ups would write what they felt had been the root cause of an issue on a Posit-It and proceeded to stick it to a whiteboard. The notes were then grouped,

Exhibit 5 PC Manufacturers’ Key Operating Measures, 1997–2014

	1997	2000	2003	2006	2008	2010	2012	2014
Gross margins (%)								
Apple	21%	27%	29%	29%	35%	39%	44%	39%
Dell	23%	21%	19%	17%	18%	19%	21%	--
Hewlett-Packard	38%	31%	29%	24%	24%	22%	22%	23%
Lenovo	--	13%	15%	14%	15%	11%	11%	13%
R&D/Sales (%)								
Apple	12%	5%	8%	4%	3%	3%	2%	3%
Dell	1%	2%	1%	1%	1%	1%	2%	--
Hewlett-Packard	7%	5%	5%	4%	3%	2%	3%	3%

Source: Compiled from Capital IQ and ThomsonOne, accessed April 2012, March 2015.

Note: All information is on a fiscal-year basis. Apple’s fiscal year ends in September, HP’s in October, Dell’s in January, and Lenovo’s in March.

ranked, and discussed. The practice helped Cook weed out top performers from the rest of the team, usually those who were the most “Cook-like”: MBAs with a relentless work ethic. Yet, Cook went out of his way to hire from all areas of Apple in order to have diverse perspectives on problems.

Some describe Tim Cook as robotic - emotionless, calculated, and all-knowing - and it’s not hard to see why. Over his career, Cook’s track record for rooting out all operational inefficiencies is almost inhuman. He has continuously redefined the impossible, and his results speak for themselves. After Steve Jobs’ death on October 5, 2011, many were skeptical about the prospect of Cook at the helm; the two were complete opposites. Thus, Jobs directed Cook to lead Apple and never look back, never question what “Jobs had wanted.” And he did just that. He took the 1984 “Think Different” branding and flipped it on his head, turning Apple into one of the most recognizable faces of corporate America. Under his leadership, Apple would soon be the first company to be valued at a market cap of a trillion dollars (three trillion as of October 2023), a monumental feat. The small-town kid from Mobile, Alabama, who dreamed of a life different from his father’s now lives in a \$2.3 million condo in Palo Alto, California.

Section 4: Oxygen 8’s Best Traits

When considering which three of Google’s Oxygen Eight Traits are the most critical, the backgrounds and experiences of Yvon Chouinard and Tim Cook are pivotal. Against the backdrop of these two mega-successful business leaders and managers, picking and choosing from Oxygen 8 becomes much more manageable. Some qualities are simply much more pervasive throughout the lives of both leaders.

Trait #1

The first quality of importance is **“empowers the teams and does not micromanage”** (quoted from Google’s Oxygen 8 list directly). Take Yvon Chouinard first. As leader of Patagonia, he took pride in developing his own style of leadership, his Management By Absence (MBA) theory of management. Taking off to foreign countries and leaving your company in the hands of your executive team is the ultimate form of empowerment, trust, and not micromanaging. Now, the applicability of this theory, specifically to other companies, is up for debate, but the idea is there: sometimes, it’s best to take a step back. This allows employees to problem solve on their own in order to reach their deliverables, building grit and the ability to adapt. Additionally, a measure of distance oftentimes allows subordinates to develop their own ideas, contributing to a team’s collective intelligence. (Collective intelligence is the wealth of knowledge contained by all members of your group, from the various degrees to the small, random tidbits of information that each person calls their own.) Leveraging collective intelligence, as well as a diversity of perspectives, promotes a variety of potential solutions; natural selection will then isolate which proposal is best.

Looking at Tim Cook, it may be harder to see how he empowers those he manages. It seemed as if he was in complete control of his operations team at Apple, but taking a closer look,

Cook empowered his team with his high expectations and demand for excellence. Because he will eventually nitpick all information given to him, those in ops. were pushed to smooth out any bumps before presenting to Cook. He didn't instruct them HOW to get things done, but more TO get it done. A prime example, Cook didn't lay out a path for the Singaporean ops. team to reach a thousand inventory turns, rather expanded their horizons to even consider the goal. He genuinely asked O'Sullivan "How" they could reach the mark, signaling his trust in the team's knowledge of operations. And they came through in a big way, developing their ingenious yellow-line-inventory strategy. The growth of Apple and Patagonia would simply be impossible without Chouinard and Cook's trust in their respective teams.

Trait #2

The second quality that appears over and over again between master operator Tim Cook and eccentric entrepreneur Yvon Chouinard is "**productive, and results orientated.**" While Cook measured results quantitatively via metrics such as inventory turnovers and gross margin, Chouinard saw things qualitatively, choosing to compare himself against competitors' piton quality. The two took different approaches in "measuring what matters" (an executive strategy developed by John Doerr that focuses on establishing and determining key performance indicators), but regardless, both managers zeroed in on results. At the end of the day, if Chouinard's original pitons weren't cutting edge and essential to *the* climbs of the time, his efforts to offset mountain damage would most likely have been futile. If Cook's numbers and track record weren't in the right spot, his passion for detail alone wouldn't have landed him esteemed openings at Big Tech firms. Without a focus on constant improvement and a limitless dedication to polishing a product, even great leaders can fall short. The point is simple: deliver. Ancillary benefits or passions simply don't stack up to results.

In terms of highlighting productivity, once again, both leaders pass with flying colors. A true trailblazer, Yvon Chouinard revolutionized the workplace by allowing employees a degree of freedom (paid maternal and paternal leave, flexible dress code, Patagonia Outings, etc.), which he believed would be a net gain. However, there was one caveat: quality and timeliness of work wouldn't drop. As soon as responsibility fell, so did the privileges. Cook, on the other hand, intimidated the Apple Operations team in order to mold them all in his own image. By asking the right, nuanced questions to presenters, he earned his fledgling employees' trust and lit a fire under them to become increasingly "Cook-like." This was Cook's version of mentally aligning his team to prioritize the small cogs of the larger operational machine and eliminate all room for error, or in other words, prioritize productivity.

Trait #3

The final and perhaps most critical quality of exceptional business leadership is "**has a clear vision/strategy for the team.**" Initially, the characteristic seems advantageous but not necessary, and it's easy to see why under the traditional definitions of vision and strategy. Vision is not solely about setting a goal or looking ahead toward the future; it's also about noticing new

opportunities in the periphery that many others deliberately ignore. In the early '60s, Chouinard saw the opportunity and so did Tim Cook in the late '90s. Whether it be specialized, reusable pitons or a thousand inventory turns, many of the pair's ideas were thought to be preposterous upon conception. Strategy, alternatively, describes the method in which vision is carried out, but also the path in which mental alignment is achieved by your team. As discussed previously, Cook's strategy for team synergy chiefly consisted of bombarding employees with questions; Chouinard's strategy was much more experiential than Cook's, leading groups of employees on Patagonia-sponsored seminars.

Route and direction. A great leader can provide one, but an extraordinary leader wields both functions of leadership to capture their employees. When employees understand the bigger picture - their role in the company reaching new heights -, suddenly, their actions take on a much larger weight. Oftentimes, this novel sense of purpose can reinvigorate employees and spark them to deliver their best work. Vision (direction) and strategy (route) are pivotal to success, and are leveraged by those at the top to instill meaning to those subordinate.

Conclusion

"The whole is greater than the sum of its parts." In the context of their respective teams, leaders are responsible for the veracity of this statement. Of course, that's easier said than done, which is why Google's Project Oxygen provides such valuable insight to any aspiring leaders. For those struggling to elevate their employees, the Oxygen Eight characteristics of the top managers at Google are a great starting point. Yet, it's difficult to focus on the application of all eight traits at once. Thus, there is a need for a boiled-down version, highlighting only the top three traits of exceptional leaders and managers.

"Why Tim Cook and Yvon Chouinard?" you may ask. The choices seem random, as neither is connected to Google or Project Oxygen. Yet, when considering the diversity of perspectives, experience, fields, personalities, and more between Cook and Chouinard, it becomes a bit more clear that their names weren't pulled out of a hat.

Tim Cook is the face of corporate America. Cultivated in Alabama upon the values of hard work and passion, Cook came from a family with a rock-solid foundation. His scholastic aptitude and natural methodical inner-workings quickly set him aside from his peers at an early age. He was accepted into his dream school, Auburn University, studied industrial engineering, and headed straight to white-collar world: IBM. Cook, eventually, became known to obsess over every last detail in order to minimize risk as much as possible and was poached by an Apple recruiter in 1997. He has made big waves at the company and gained notoriety for his deal with ChinaMobile in 2013 that allowed Apple to tap the billion potential customers in China. Currently, Cook leads Apple (the largest company in the world by market cap) as CEO, but will always maintain his reputation as the master operator.

Yvon Chouinard's story is almost completely opposite Cook's. Chouinard is an entrepreneurial spirit who never fit the education system or corporate ladder; the only place he felt he *truly* belonged was among the Tetons. His love of nature was contagious and would play a

massive role in his life and the founding of Chouinard Equipment (CE). As a leader, Chouinard preached his “MBA” style of management and openly disliked “corporate corpses in suits.” He was adamant about improving employment practices and the work environment. Patagonia Inc., which sprung from CE, is now a model for people management and treatment.

The contrast between Cook and Chouinard is impossible to miss. One an entrepreneur at heart, the other an operator through and through. Yet, the beauty of the two together is that Cook and Chouinard’s differences allow for the essential three traits to really shine. Regardless of company size, type, or reputation, the three traits mentioned above stay constant. At any stage, it's important not to micromanage and lose the trust of your team. At any stage, it’s important to remain product and results-oriented in order to complete deliverables. And at any stage, it’s **vital** the leader has a clear vision and strategy to guide subordinates. Armed with this knowledge, elevating your organization and enhancing those who you lead is much more manageable. Apply the three golden traits, continuously improve your delivery, and polish your unique management style. That’s the recipe for exceptional business leadership.

Works Cited

- Chng, Daniel Han Ming, et al. "Why People Believe in Their Leaders - Or Not." *MIT Sloan Management Review*, 17 Aug. 2018, sloanreview.mit.edu/article/why-people-believe-in-their-leaders-or-not/.
- Momin, Weena Yancey M, and Dr. Kushendra Mishra. "Managing People Strategically with People Analytics: A Case Study of Google Inc." *Research Gate*, www.researchgate.net/profile/Weena-Yancey-Momin/publication/335320161_Impact_Factor_52_IJAR/links/5d5e3e1f92851c37637152e0/Impact-Factor-52-IJAR.pdf. Accessed 30 July 2023.
- McGuffie, Adam. "Yvon Chouinard's Leadership of Patagonia." *Adam McGuffie*, 2 Apr. 2016, adammcguffie.wordpress.com/2015/04/20/essay-yvon-chouinards-leadership-of-patagonia/.
- Kim, Hyeonjoo. "Comparison of Strategic Leadership: Steve Jobs and Tim Cook." *Research Gate*, www.researchgate.net/profile/Hyeonjoo-Kim/publication/344193721_Comparison_of_Strategic_Leadership_Steve_Jobs_and_Tim_Cook/links/5f5a6366299bf1d43cf9741e/Comparison-of-Strategic-Leadership-Steve-Jobs-and-Tim-Cook.pdf?origin=publication_detail. Accessed 30 July 2023.
- Chouinard, Yvon, and Vincent Stanley. "What We Do For a Living." *The International Journal of Ethical Leadership*, vol. 2, no. 1/7, 2013, <https://scholarlycommons.law.case.edu/cgi/viewcontent.cgi?article=1025&context=ijel>.
- Mickle, Tripp. *After Steve: How Apple Became a Trillion-Dollar Company and Lost Its Soul*. HarperCollins Publishers, 2023.
- Goffee, Robert, and Gareth Jones. *Why Should Anyone Be Led by You?: What It Takes to Be an Authentic Leader*. Harvard Business Review Press, 2019.

Z: The Key By Joseph Yanta

A wealth of research on the political habits of zoomers, as well as the steps necessary to engage them in the political system they will inherit.

Research Question: Which rhetorical devices, tactics and platforms of communication are best at mobilizing Gen Z in support of a candidate or cause?

The twenty-first century has seen a revolution in the means of political communication. The popularization of sites such as Twitter have made political socialization occur earlier and also likely contributed to intense polarization. An example of a politician whose social media usage defined their campaign would be Donald Trump, whose Twitter garnered 88 million followers and granted Trump massive media attention. In addition, vitriolic presidential debates have shaped the political landscape and defined it as one where insults, ad hominem tactics and buzzwords are common. The purpose of this project is to discover which rhetorical tactics garner politicians support, and which may be outdated or ineffective. This project will include the use of polling data, which can be used to discover the issues that are important to Gen Z and reveal trends amongst Gen Z voters that can be utilized by candidates and political organizations to increase youth engagement and support.

Introduction

The internet has brought profound change to nearly every aspect of our lives. It has created an entirely new reality through the introduction of social media and new debates over whether platforms such as Instagram, Twitter and Tiktok serve to unite and divide. Born from the early-to-mid 1990s through the early 2010s, Gen Z, referred to commonly as Zoomers, is the first generation to grow up with access to the internet and social media since birth. Despite growing up with easy access to information and data via the internet, Gen Z voters are not as politically informed as previous generations, and it is important that candidates bring attention to the issues in order to educate the new generation on political issues. If Gen Z does not become more politically informed, democracy may become less effective as democracy requires an educated populace in order to function properly (Wood 2013).

Research on Gen Z's voting patterns has defined clear patterns of voting preference, generally leaning left and supporting minority candidates, but existing studies regarding Gen Z are lacking due to the fact that many members of Gen Z are unable to vote yet. This study will investigate both what Gen Z seeks in a candidate and how candidates can tailor their messaging to appeal to young people.

This paper seeks to discover which rhetorical tactics and platforms can be utilized in order to engage Gen Z in the political process in support of a candidate or cause. In addition, this paper looks both backward and forward to see how Gen Z is currently voting and predict how they will vote in future elections. This paper analyzes data from multiple sources, including previous studies regarding Gen Z's political leanings as well as voting data from national

elections. This paper also analyzes campaigns that performed well amongst Gen Z voters. It analyzes the rhetoric and strategies of these campaigns to see which aspects of them appealed ‘]generally to Gen Z voters.

This paper discusses ways that politicians and issue based accounts already do, and can better inform people on important issues that drive potential voters to the polls, donors to their wallets, and volunteers to doors and phones. Ultimately, this paper seeks to discover how Gen Z thinks, why they think that way, and how they can be convinced to support a cause or candidate. This data can be used by campaigns, movements and advertising agencies in order to create strategies of reaching and engaging young people that will work for years to come as more of Gen Z becomes able to vote. This paper considers the methods of political advertising currently in use, whether or not these methods are effective at engaging young people, and which methods are most effective and should be leveraged to benefit campaigns and movements. Effective methods of engaging Gen Z include leveraging the use of infographics and social media platforms as well as social movements. Gen Z cares deeply about social issues and is critical of party affiliation.

Literature Review

Knowing how Gen Z thinks and votes is an important litmus test for the future of America, as Gen Z will create and support policies they desire now and in the future. The voting and thinking patterns of Gen Z have been researched, but currently that research is limited by the fact that many members of Gen Z are too young to vote or participate in studies, seeing as Gen Z currently ranges from 9 to 24 years of age. Gen Z cares deeply about social justice, and this passion for equity may bring about further democratization. Knowing what issues and messaging resonate with younger generations will benefit politicians who care to run youth-backed campaigns, as well as issue-based organizations and movements that seek to engage reform-minded youth. This section will cover what Gen Z seeks in a candidate, Gen Z’s motivations for voting, how and where Gen Z receives information, and who this information can benefit.

It is important when attempting to engage Gen Z in the voting process to understand what Gen Z looks for in a candidate. Gen Z is more likely to vote for an activist candidate who takes strong positions on issues than a candidate they see as upholding the status quo (Kelly 2023). This may explain Donald Trump and Bernie Sanders’s popularity amongst Gen Z. Bernie Sanders’s messages including student loan debt relief resonated with Gen Z and millennials to the extent that those two groups comprised 67% of Sanders voters (Altamura, Oliver 2022). Gen Z also tends to prefer female and minority candidates (Kelly 2023). Understanding the kind of candidate Gen Z prefers could lead political parties to back candidates who will have youth support and a growing share of the vote as more of Gen Z reaches voting age.

Understanding why Gen Z votes is key when creating rhetoric targeting young voters. Gen Z does not hold the same motivations for voting as previous generations. While baby boomers were driven to vote largely by a sense of civic duty, the voting motivations of Gen Z are

unclear (Alexander 2022). This can likely be explained by the fact that it is an extremely diverse generation in an ethnic, religious, political, and psychological sense. In fact, nearly half of Gen Z are racial or ethnic minorities (Fry, Parker 2018). This data may be incomplete simply due to the fact that not all of Gen Z has reached voting age. It is important that political parties and candidates understand what rhetoric will motivate Gen Z to take the time to vote. Gen Z also cares deeply about convenience (Wood 2013). This makes them unlikely to expend a large amount of effort in favor of a candidate or issue. This is unfortunate for any politician seeking Gen Z volunteers but has an upside. Social media posting is convenient for Gen Z, and can reach a large number of people. Gen Z can be effective politically by leveraging digital reach.

Gen Z consumes media in ways different than previous generations, in part because of the development and proliferation of social media. For instance, TikTok, an app which features short videos, has become wildly popular amongst members of Gen Z. Conversely, traditional media such as television news networks, newspapers, and radio shows have fallen in popularity amongst young people. Gen Z does not frequently read, watch, or listen to traditional news media, and this leaves younger people with a lack of convenient, informative and unbiased political media. Because the traditional news media networks are hesitant to alter their methods of delivering information, many members of Gen Z are left uninformed on political matters (Harder 2020).

Gen Z is able to multitask well, but has a very low attention span (Rothman 2016). This means that boring political advertisements and ones that do not target the issues that engage Gen Z will be ineffective at engaging them politically. Gen Z's ability to multitask makes them effective and efficient at spreading messages, especially through social media, when they care about a certain candidate or cause. Gen Z's proficiency in social media and internet communication makes them effective at spreading information rapidly and reaching a wide range of people. Campaigns can benefit by hiring Gen Z communications staff to more effectively reach out to young people.

Understanding Gen Z's sense of humor and style of communication can also help campaigns resonate with young people. Gen Z's humor is more absurdist than that of previous generations (Partlow, Talarczyk 2021). Knowing what makes Gen Z laugh is important as that is one way to engage a group with a low attention span. Politicians that make jokes that resonate with Gen Z may be able to reach them better and earn their vote by means of identifying with them. One example of a politician effectively engaging with Gen Z through comedy is Joe Biden's embrace of the "Dark Brandon" meme. Gen Z is also thoughtful, determined, and open-minded (Seemiller, Grace 2016). Gen Z's open-mindedness is important because they do not have many preconceived notions that might prevent them from supporting a certain candidate or issue. The largest obstacle to them supporting a given candidate or issue is their relative lack of political education.

Gen Z is an incredibly diverse generation. There is a gender divide between Gen Z men and women in terms of political views. Gen Z women tend to be more liberal than men (McDonald, Rouse, Kromer 2020). Knowing this, candidates could attempt to advertise to young

men as a conservative candidate or young women as a liberal candidate in order to find voters more likely to believe in their causes. Alternatively, candidates could attempt to push their messaging towards the gender that aligns differently from their political views, in order to attempt to reach out to people who might be easily convinced to vote for a different party or candidate.

Gen Z, along with every other generation, underwent political socialization through family members, friends, media, and landmark events. Some landmark events that shaped the political conscious of Gen Z include the 2008 recession, the wars in the Middle East, mass shootings, the Covid-19 pandemic and the legalization of gay marriage (Real 2022). These events have made Gen Z conscious of their safety and supportive of higher government regulation of weapons. Understanding the events which shaped the consciousness of Gen Z is incredibly important as these events often shape the values of people, such as how the 9/11 attacks spurred greater support for the Patriot Act and increased military funding amongst older generations.

Social media has become the main mode of communication for Gen Z. 73% of American teens use some form of social media, which has risen from 55% in 2006 (Lenhart, Purcell, Smith, Zickuhr 2010). Social media's increasing commonality amongst teenagers has led to advertisement companies utilizing social media advertising in order to appeal to younger groups as well as in order to target their advertising to people with specific interests. Activist causes on social media spread quickly amongst young people, with an example being the Black Lives Matter movement. The Black Lives Matter movement gained wide traction amongst Gen Z due to the fact that social media was commonly used to promote it, and it was a galvanizing social issue. While candidate promotion has faltered amongst young people on social media, issue-based accounts have engaged millions of people and created large following bases for their causes.

Activist social media accounts have commonly used infographics to inform potential voters about issues they are passionate about. Infographics are collections of imagery, visualized data, and text that often inform Gen Z social media users about political and social issues. These infographics are extremely effective and garner engagement at astounding rates due to their ability to allow a viewer to effectively process the information presented to them (Siricharoen 2013). Infographics have been found to be three times as effective as regular image posts on Instagram (Martinez 2021). One reason infographics are incredibly effective at mobilizing Gen Z is that it grants them more information on a subject. Gen Z is lacking in political literacy and infographics serve as a method to educate young people on political and social issues (Robin, Alvin, Hasugian 2022).

Twitter is an extremely important platform to study because it is a platform that many political figures have utilized to disseminate messaging to supporters, media and critics en masse. Twitter's text-based posts seem tailor-made for political messaging. In addition, the average Twitter user is able to view up to 1000 Twitter posts per day, which can convey a significant amount of messaging (Pancer, Poole 2016). Twitter became arguably the predominant

political social media site during the Obama presidency, but it became an object of controversy during Donald Trump's presidency (Minot, Arnold, Alshaabi, Danforth, Dodds 2021). Twitter is also simple to obtain data from, as likes and retweets are easily analyzable.

Another tactic that is likely to engage Gen Z is the strategic use of hashtags (Sipocz, Freeman, Elton 2020). Hashtags allow social media users to view more posts about certain topics. Some hashtags also become popular, and become similar to a slogan, benefitting a candidate or issue. Hashtags can be leveraged effectively in political campaigns. #MeToo and #BlackLivesMatter both became movements and spread quickly throughout the use of hashtags.

Knowing what sorts of issues the youth care about can help PR firms, political parties and candidates tailor messaging to increase youth support, which will become more significant as more of Gen Z reaches voting age. The youth presently deeply care about issues of race, gender, and class (Gordon 2009). Gen Z's political activism is largely social media based, and several issues that saw large Gen Z engagement include Black Lives Matter and human trafficking (Buzetto-Hollywood, Hill, Banks 2021). It is possible that they may care more about economic issues once more of them become financially independent, as has been a trend in the past. Knowing which issues Gen Z cares about is important because politicians will need to tailor their messaging accordingly in elections, as Gen Z currently makes up 40% of consumers and soon a large percentage of the voting base (Martinez 2021). Much research has been compiled about the psychological traits associated with right and left wing beliefs, and these traits can be targeted in political advertising in order to get potential voters involved with a candidate (García-Díaz, Colomo-Palacios, Valencia-García 2021). One method that has been used in order to reach potential voters is "political micro-targeting" (PMT) (García-Díaz, Colomo-Palacios, Valencia-García 2021). Political microtargeting allows campaigns to use voter behavior data and tailor ad campaigns to specific types of voters. PMT can be used to target Gen Z voters with different motivations for voting as well as patterns of thinking, which vary greatly amongst Gen Z as it is an extremely diverse generation.

Methodology

This project analyzes the methods of political messaging that effectively engage Gen Z in an effort to determine what drives Gen Z to the polls. To do so, this paper analyzes voting data as well as information from various opinion polls regarding Gen Z's voting, thinking and learning tactics. Both information found in studies and quantitative polling data has been collected and analyzed in order to paint a comprehensive portrait of the specific methods of messaging that effectively engage Gen Z in support of a candidate or cause. In addition to analyzing rhetorical choices, the effectiveness of modes of communication such as infographics were analyzed as they pertain to mobilizing young voters.

This project analyzes polling regarding national candidates. There are several reasons why we should focus on national candidates. First, there is more data on national candidates. Second, focusing on national candidates allows us to create a more holistic perception of Gen Z's political preferences. Studying state level candidates would not factor for regional differences

in youth perspective and culture. Data has been collected from reputable polls such as those performed by Pew Research, and no polls with obvious bias have been used. Obvious bias includes polls from partisan organizations or polls that utilize language that is preferential towards one option. No polls conducted unprofessionally (i.e. through Twitter) have been consulted, either. Polls that differentiate based on age groups have been used, as Gen Z may not think the same about a candidate or issue as other age groups. This project analyzes candidate support amongst Gen Z in both primary and general elections.

Politically or rhetorically significant communications are defined as those which reached voters either negatively or positively beyond the realm of basic political campaigning. A positive reach is defined as one that leads a potential voter or supporter to have a more positive view of a candidate or cause, more likely to vote for a candidate or more likely to donate or volunteer for a certain candidate or cause. These communications may include slogans, such as “make America great again” and “build back better,” which are tailored in order to communicate specific messages and themes to voters. Important indicators of a candidate’s performance with Gen Z are polling and voting data, and this paper analyzes which national candidates perform well with Gen Z and why. A limitation of polling data is that often people may not be interested in answering polls and may skew data. Additionally, distrust in media may skew polling. However, polling data is the best method to discover the sentiments of a group because it allows mass data collection and the factoring for different age groups as well as other demographics that might be important such as race and gender.

When creating this paper, I also considered using Google search and Twitter data. Several related research papers used this form of data effectively (Stefanov, Darwish, Atanasov, Nakov 2020). However, I decided not to for several reasons, including the fact that that data cannot be separated by age of author/searcher. Differences between generations is the largest factor in this research and that data would not be helpful for discovering which tactics influence Gen Z but rather people in general. Because of the lack of specificity of these data collection methods, I decided to not use them and rather use polling data.

This paper also considers the factors that cause Gen Z to think the way that they do, including the important events and milestones that have politically socialized young people in recent years. Political socialization depends on many factors, including friends, family and socioeconomic class. Events that have politically socialized Gen Z include the 2008 recession, the wars in the Middle East, mass shootings, and the legalization of gay marriage.

Findings

The research conducted points to several hypotheses regarding how Gen Z thinks and regards the political landscape surrounding them. We have proved these hypotheses using polling and voting data.

I hypothesized that Gen Z cares about the political issues that directly affect them and their generation and will be motivated to vote for candidates because of these issues. Issues that directly impact Gen Z include student debt relief and school safety. Both of these issues will

directly affect Gen Z financially and personally, and are issues that I hypothesize will motivate Gen Z to support a candidate. I found studies that rank which issues Gen Z cares about the most, and will use that to prove my hypothesis. In addition, I found studies regarding how much of Gen Z was concerned by inflation and rising gas prices. Gas prices are especially impactful to younger generations with lower incomes and less money in savings, and if Gen Z is deeply politically motivated by rising gas prices, then they are voting in favor of their economic well-being.

I also hypothesized that Gen Z is motivated by social justice. In order to evaluate my hypothesis, I found studies which evaluate how much Gen Z values and supports various social movements. A social movement is defined by the Encyclopaedia Britannica as “a loosely organized but sustained campaign in support of a social goal, typically either the implementation or the prevention of a change in society’s structure or values”. Some recent and relevant causes I would define as social movements under that definition include Black Lives Matter, #MeToo, Schools Strike for Climate, and the school walkouts against gun violence. These causes often received mass media attention, with social media serving as a means for increasing awareness of perceived issues. These causes also attempted to change society’s structure or values through the implementation of either political or social pressure. If Gen Z largely supports these causes, then Gen Z is motivated by social justice.

I found several studies that prove that Gen Z highly supports the causes that fit the definition of social movements. According to a poll by Morning Consult, 53% of Gen Z has a positive view of social justice, which is 9% more than Millennials, 14% more than Gen X, and 13% more than Boomers (Laughlin, 2020). It is important, however, to understand Gen Z’s support of individual social movements. Black Lives Matter is an influential social movement that seeks to end police brutality and perceived unequal treatment of black individuals by police. 81% of Gen Z is supportive of Black Lives Matter protests. Gen Z supports these protests more than Millennials, at 70%, Gen X, at 62%, and Boomers, at 56%. In addition, 39% of Gen Z participated in a Black Lives Matter-related social media trend, #BlackOutTuesday (Hurst, 2022).

#MeToo is another influential social movement with the goal of ending sexual abuse and harassment by means of increasing awareness and applying social pressure. According to a study by the Pew Research Center, 64% of 18-29 year olds who had heard of Me Too were supportive of it (Savage 2023). The issue of climate change is another issue that has spurred various social movements, especially through the medium of social media. Polling by Pew Research finds that 56% of Gen Z has seen content addressing climate change in the past several weeks, and 45% say that they engage with climate change content on social media (Thigpen, Tyson 2021). Gen Z’s overwhelming support for social movements shows that Gen Z is motivated by social issues.

The polling referenced above shows that Gen Z is supportive of various social movements, both organized and unorganized, and often interacts with these movements through the medium of social media. This data proves the hypothesis that Gen Z is motivated by social justice. Candidates, political parties and consulting firms can leverage this information in order

to increase Gen Z support for candidates. Politicians with large amounts of Gen Z support have referenced and lent support to social movements, and benefit by doing so. One such politician is Bernie Sanders, a Senator from Vermont. During the 2015 Democratic primary debate, Bernie Sanders lent support to the Black Lives Matter movement. Sanders' youth support is large, and he earned 57% of votes from voters 18-44 in the Michigan Democratic Primary in 2020. While Gen Z largely supports social movements and supports politicians who interact with them, few politicians have tapped this valuable method of garnering Gen Z support. Politicians can gain Gen Z support by interacting with social movements by attending protests, reposting movement-related content on social media or simply making statements in support of causes.

I hypothesized that Gen Z can be effectively reached through the medium of infographics. In order to evaluate my hypothesis, I found data from companies that have used infographics for marketing purposes and saw if they found infographics as helpful and effective at driving sales. I also sought data on how the use of infographics impacts web traffic and engagement with posts and advertisements.

I found data to evaluate my hypothesis through Infographic World, a communications company who surveyed 1,100 consumers and 100 businesses on their use and familiarity with infographics. They found that 84% of companies who had used infographics before found them effective (Woodward 2023). Infographics were the most effective medium for learning and retaining information, with 61% effectiveness. For comparison, commercials were 55% effective, PowerPoints 48% and blog posts 36%. In addition, the researchers found that infographics made 57% of individuals surveyed think deeply about a subject. Part of the reasons infographics are so effective is likely the use of eye-catching visuals as well as the combination of image and text. Overall, infographics are an effective method of communication, and especially effective at spurring deep contemplation about a subject.

Infographics are commonly used on social media, which has a young user base. If candidates and political organizations infographics effectively by, those groups can gain youth support and engagement while effectively politically informing young people. While social movements such as Black Lives Matter have effectively leveraged infographics to expand their reach and gain new supporters, politicians have lagged behind and not harnessed the potential of infographics. This paper should serve as motivation for politicians and political parties to find new and innovative strategies to engage and reach Gen Z voters, one of those strategies being the use of infographics in communications, especially through the medium of social media.

According to polling by SocialSphere, 50% of Gen Z voters surveyed were influenced by the issue of school safety and security (Volpe, 2023). Millennials, Gen X and Boomers all were influenced within 5% of that figure. Though Gen Z was not substantially influenced by school safety and security more than non-college age adults, the fact that half of Gen Z voters were influenced by that issue shows that Gen Z is often motivated by issues that affect their generation specifically.

With regards to economic issues, CIRCLE at Tufts released a poll showing that 39% of Gen Z listed inflation and gas prices as one of their top three political priorities (Booth and

Suzuki, 2023). Respondents chose this response the most, with abortion in second place with 30% and jobs that pay a living wage in third with 26%. When it comes to student debt relief, an issue that deeply affects Gen Z as a highly college-educated generation, Gen Z politically supports its own economic interest. According to a USA Today/Ipsos poll, Gen Z is the least likely generation to support the Supreme Court's overturning of the student loan forgiveness proposal (Newell et al., 2023).

In several ways, Gen Z has been shown to prioritize its economic and personal interest politically. Gen Z is motivated to pursue political action such as voting, protesting, or donating to candidates or causes by the issues they prioritize. Candidates can utilize this information and create political messaging that speaks to Gen Z on the issues that directly affect them, including student loans, school safety and inflation. The effect on Gen Z support created by this messaging can be maximized by posting it on Gen Z-dominated media such as Instagram. One example of a politician leveraging Gen Z issues to gain support is when Joe Biden posted a message in support of student debt relief on June 30th on Instagram. The post received 78,000 likes and performed better than most of Biden's other Instagram posts. This post could have perhaps been even more effective if it were made in infographic format. Gen Z can be politically motivated to vote, donate or support candidates through the use of messaging on issues that specifically affect Gen Z, and Gen Z can be reached effectively through the platforms they commonly use.

I originally hypothesized that Gen Z would be critical of party affiliation. I made this hypothesis because I felt as though recent protests such as those at the 2016 Democratic National Convention showed that young people were critical of party structure and thus party affiliation. Bernie Sanders supporters protested the nomination of Hillary Clinton at the DNC in 2016, and many of Bernie's supporters were young progressives. I also thought that young people's low approval of the electoral college system would translate to disagreement with the two party system and other elements of the American electoral system. However, my research seems to negate my hypothesis and I believe I have disproved it.

Polling by Gallup on party identification among different age groups found that 31% of Gen Z identifies as a Democrat, 17% of Gen Z identifies as a Republican and 52% of Gen Z identifies as an Independent (Jones 2023). While the amount of Gen Z who identifies as Independent may seem large, the same percentage of Millennials identify as Independent. 44% of Gen X identifies as Independent, which is not significantly different from the amount of Millennials and Gen Z who identify as Independent.

While I did not find evidence that supported my hypothesis, I am still searching for reasons that Gen Z and Millennials may identify as Independent more frequently than Gen X and Boomers. Some of the reasons behind this generational difference may lie in political literacy, as Gen Z is under-informed on political matters (Robin et al., 2022). Despite Gen Z perhaps not shunning party labels significantly more than other generations, the level of independent voters is still high. Identifying strongly with a political party may not benefit candidates' Gen Z support as much as it may benefit support among older generations with stronger party identification. Seeing as the trend of democratization continues, party labels may see a trend of decrease. There

is still much future research that needs to be conducted as to the party identification of Gen Z and their feelings on party structure and the system of the electoral college.

Political parties can earn Gen Z support by supporting young candidates who are in touch with youth culture. Alexandra Ocasio-Cortez (commonly referred to as AOC), a 33 year-old Congresswoman from New York, gained 435,000 viewers streaming the hit videogame “Among Us”. While some may regard streaming video games as not a proper form of political outreach, New York Times columnist Charlie Wazel argues that AOC “represents the future of how politicians will build support and then use that support to accomplish their political and policy goals”.

Conclusion

The question of this research paper was how Gen Z could be motivated to engage in politics by politicians, movements and political parties. We posed this question due to its timeliness and importance in a new digital era in which information is transferred through different mediums and at a faster pace. With the introduction of social media, Gen Z is learning and becoming politically socialized by different sources and it is important that politicians reach Gen Z through social media in order to ensure that young people feel represented in electoral politics.

I answered this question using polling data on Gen Z as well as statistical data concerning efficacy of mediums such as infographics. In addition, I used polling data of candidates who appealed to Gen Z effectively, as well as news concerning candidate communications.

My proven hypotheses are that Gen Z is motivated by social issues and can be engaged through social movements, that Gen Z voters can be motivated by candidates who prioritize young people’s wellbeing, both physical and economic, and that Gen Z can be informed and motivated through the use of fast-spreading infographics. My initial hypothesis that Gen Z is critical of party affiliation was unfounded, and is not a proven conclusion.

These conclusions contribute knowledge about Gen Z which will become increasingly important as the Gen Z voting population increases. These conclusions tell how politicians can effectively shift their campaigns so that they can succeed in the digital era. Additionally, the findings of this paper can shed light on how the proliferation of social media and the internet have affected politics, and how they will affect politics in the future.

There are several next steps researchers might take in order to determine how to effectively reach Gen Z. These steps include polling Gen Z on the specific candidates they like and dislike, and attempting to find the reasoning behind the choices they make. Researchers could also utilize social media tools to uncover which posts on social media are effective at garnering engagement from young people, and discovering if that engagement is mostly positive or negative. With Gen Z becoming an increasingly large voting bloc, there is much research to be done as to how politicians can reach young people. We count on the next generation to build a better future, but they cannot do that if they are not heard.

Works Cited

- Kelly, D. (2023) Inspiring Gen Z Voters to Participate in Voting and Volunteering. *Advances in Applied Sociology*, 13, 43-46. [10.4236/aasoci.2023.131004](https://doi.org/10.4236/aasoci.2023.131004).
- Martinez, M., Holt, J., Lamm, A., & Borron, A (2021). Generation Z and CRISPR: Measuring information processing using animated infographics. *Journal of Applied Communications*: Vol. 105: Iss. 3. <https://doi.org/10.4148/1051-0834.2394>
- García-Díaz, J., Colomo-Palacios, R., & Valencia-García, R. (2021). Psychographic traits identification based on political ideology: An author analysis study on Spanish politicians' tweets posted in 2020. *Future Generation Computer Systems*, 130, 59-74
- Pancer, E., & Poole, M. (2016). The popularity and virality of political social media: Hashtags, mentions, and links predict likes and retweets of 2016 U.S. presidential nominees' tweets. *Social Influence*, 11(4), 259–270. <https://doi.org/10.1080/15534510.2016.1265582>
- Sipocz, D., Freeman, J. D., & Elton, J. (2021). A Toxic Trend?": Generational Conflict and Connectivity in Twitter Discourse Under the #BoomerRemover Hashtag. *The Gerontologist*, 61(2), 166–175. <https://doi.org/10.1093/geront/gnaa177>
- Partlow, C., & Talarczyk, P. (2021). Absurdism and Generation Z Humor: the Effects of Absurdist Content on Perceived Humor Levels in Generation Z Students. *Journal of Student Research*, 10(4). <https://doi.org/10.47611/jsrhs.v10i4.2011>
- Deckman, M., McDonald, J., Rouse, S., & Kromer, M. (2020). Gen Z, Gender and Covid-19. *Politics and Gender*, 16, 1-16 <https://doi.org/10.1017/S1743923X20000434>
- Seemiller, G. (2016). *Generation Z Goes to College*. Jossey Bass Publisher. <https://www.apuaf.org/wp-content/uploads/2017/03/Book-review-Generation-Z-goes-to-college.pdf>
- Robin, P., Alvin, S., Hasugian, T. (2022). Gen-Z Perspective on Politics: High Interest, Uninformed, and Urging Political Education. *Jurnal Ilmu Sosial dan Ilmu Politik*. 11. 183-189. <https://doi.org/10.33366/jisip.v11i3.2550>
- Gordon, H. (2010). *We fight to win: Inequality and the politics of youth activism. We Fight to Win: Inequality and The Politics of Youth Activism*. Rutgers University Press.
- Lenhart, A., Purcell, K., Smith, A., & Zickuhr, K. (2010). Social Media & Mobile Internet Use among Teens and Young Adults. Millennials. Pew Internet & American Life Project. https://www.researchgate.net/publication/286376855_Social_Media_Mobile_Internet_Use_Among_Teens_and_Young_Adults
- Siricharoen, W.V. (2013). Infographics: The New Communication Tools in Digital Age. https://www.researchgate.net/publication/256504128_Infographics_The_New_Communication_Tools_in_Digital_Age
- Buzzetto-Hollywood, N., Hill, A., & Banks, T. (2021). Early Findings of a Study Exploring the Social Media, Political and Cultural Awareness, and Civic Activism of Gen Z Students in the Mid-Atlantic United States [Abstract]. *Proceedings of the 2021 InSITE Conference*. <https://doi.org/10.28945/4762>

- Harder, D. (2020). New voters, new platforms: How first-time Generation Z voters used social media in the 2019 Canadian federal election.
<https://doi.org/10.7939/r3-5cm5-ag47>
- Alexander, A. (2022). Voting Behavior and Motivations across Generations: Evidence from a Nationally Representative U.S. Survey. *Honors Theses*, 2684.
https://egrove.olemiss.edu/hon_thesis/2684
- Rothman, D. (2016). "A Tsunami of Learners Called Generation Z."
http://ce.wvu.edu/media/15624/needs-different_learning_styles.pdf
- Minot, J. R., Arnold, M. V., Alshaabi, T., Danforth, C. M., & Dodds, P. S. (2021). Ratioing the President: An exploration of public engagement with Obama and Trump on Twitter. *PLoS one*, 16(4), e0248880. <https://doi.org/10.1371/journal.pone.0248880>
- Real, O. (2022). "In the context of a pandemic and social development, how do critical events shape political socialization, and influence the political attitudes of generational cohorts?"
- Altamura, C., & Oliver, B. (2022). Who Feels the Bern? An Analysis of Support for Bernie Sanders in the 2020 Democratic Primary. *American Politics Research*, 50(5), 643-656.
<https://doi.org/10.1177/1532673X221112390>
- Wood, S. (2013). "Generation Z as Consumers: Trends and Innovation."
- Stefanov, P., Darwish, K., Atanasov, A., & Nakov, P. (2020). Predicting the Topical Stance and Political Leaning of Media using Tweets. 527-537.
<https://doi.org/10.48550/arXiv.1907.01260>
- Fry, R., & Parker, K. (2018). "Early Benchmarks Show 'Post-Millennials' on Track to Be Most Diverse, Best-Educated Generation Yet."
https://www.researchgate.net/publication/329428685_Early_Benchmarks_Show_%27Post-Millennials%27_on_Track_to_Be_Most_Diverse_Best-Educated_Generation_Yet
- Laughlin, N. (2020). *Gen Z's worldview after 2020*. Morning Consult Pro.
<https://pro.morningconsult.com/trackers/gen-z-worldview-tracker>
- Hurst, K. (2022). *U.S. teens are more likely than adults to support the black lives matter movement*. Pew Research Center.
<https://www.pewresearch.org/short-reads/2022/06/15/u-s-teens-are-more-likely-than-adults-to-support-the-black-lives-matter-movement/>
- Savage, M. (2023). *Did me too change the workplace for gen Z?*. BBC Worklife.
<https://www.bbc.com/worklife/article/20230307-did-me-too-change-the-workplace-for-gen-z#:~:text=The%20%20Pew%20Research%20Center%20%20study,movement%20than%20any%20other%20group.>
- Thigpen, C. L., & Tyson, A. (2021). *On social media, gen Z and millennial adults interact more with climate change content than older generations*. Pew Research Center.
<https://www.pewresearch.org/short-reads/2021/06/21/on-social-media-gen-z-and-millennial-adults-interact-more-with-climate-change-content-than-older-generations/>
- Woodward, M. (2023). *Infographic statistics: How effective are infographics today?*. SearchLogistics.

- <https://www.searchlogistics.com/learn/statistics/infographic-statistics/#:~:text=Key%20Infographic%20Statistics,-Tons%20of%20businesses&text=Content%20posts%20that%20include%20images,persuasive%20than%20presentations%20without%20them>
- Conner, C. (2017). *The data is in: Infographics are growing and thriving in 2017 (and beyond)*. Forbes.
<https://www.forbes.com/sites/cherylsnappconner/2017/10/19/the-data-is-in-infographics-are-growing-and-thriving-in-2017-and-beyond/?sh=7ca63017137c>
- Newell, M., Jackson, C., & Sawyer, J. (2023). *Support for student loan forgiveness varies across different ...* - ipsos. Ipsos.
<https://www.ipsos.com/en-us/support-student-loan-forgiveness-varies-across-different-amounts>
- Volpe, J. D. (2023). *New report reveals what issues motivate gen Z and their future priorities*. Walton Family Foundation.
<https://www.waltonfamilyfoundation.org/new-report-reveals-what-issues-motivate-gen-z-and-their-future-priorities>
- Blad, E. (2023). *What do gen Z voters care about most? A survey offers insights*. Education Week.
<https://www.edweek.org/teaching-learning/what-do-gen-z-voters-care-about-most-a-survey-offers-insights/2023/02>
- Booth, R. B., Medina, A., & Suzuki, S. (2023). *Gen Z, aware of its power, wants to have impact on a wide range of issues*. Circle at Tufts.
<https://circle.tufts.edu/latest-research/gen-z-aware-its-power-wants-have-impact-wide-range-issues>
- Jones, J. M. (2023). *Millennials, gen X clinging to independent party ID*. Gallup.com.
<https://news.gallup.com/poll/397241/millennials-gen-clinging-independent-party.aspx>

Pathophysiology of Arthrofibrosis of the Knee after Ligamentous Injury and Repair: Synthesizing Past and Current Knowledge for Personalized Prevention By Grace Flury

Abstract

Arthrofibrosis of the knee is one of the most common complications following ligamentous injury, affecting substantial amounts of patients following trauma and surgical procedures. Known symptoms of arthrofibrosis include limited range of motion, severe pain, joint stiffness, and joint effusion.¹ Patients who develop arthrofibrosis often have a longer recovery time and an impaired healing process compared to non arthrofibrosis patients post ligamentous injury, due to long-lasting effects and stressors such as pain and instability. Such stressors can cause mechanical and psychological impairments that commonly prevent patients from doing normal day-to-day activities. In severe cases, arthrofibrosis can lead to unemployment due to the inability of patients to perform basic activities.¹ Additionally, very few non-invasive treatment methods exist. A customary method of treatment for arthrofibrosis is arthroscopy to remove the deep tissue. Having to undergo a second surgery not only causes patients to be out of work for longer than planned, but also leads to higher hospital fees and bills. Reviewing the current state of literature, I identified diverse risk factors associated with arthrofibrosis, including age, sex, prior knee surgery, mass index, mechanism of injury, anatomical variations, autoimmune diseases, hormonal disorders, and more. Although these risk factors showed up in a wide range of the literature, there is a poor understanding of how these diverse risk factors affect the mechanism of deep tissue scarring. Additionally, many successful methods are being developed for treating arthrofibrosis in its late stages, such as surgical intervention and physio-therapeutic techniques. Still, no methods exist to evaluate the individualized prospective risk of patients. My paper is a narrative review of the literature, treatment strategies, and future directions, in which I included the epidemiology of arthrofibrosis, how multifaceted risk factors contribute to scarring, and a proposed protocol for a personalized risk score. While discussing scar-prevention strategy methods for arthrofibrosis, individualizing patient care must be considered in the treatment. The conclusions highlight the need for a multimodal approach, including a novel point-of-care system for high-risk individuals and a scar prevention strategy.

1. INTRODUCTION

Arthrofibrosis is abnormal formation of scar tissue in response to inflammation after an injury or surgery. It can occur in most joints, but is most common in the knee. There is currently no available cure for arthrofibrosis, so treatments (surgical, pharmacologic, etc.) focus on treating symptoms.² There are an estimated 85,000 patients who develop postoperative knee arthrofibrosis in the United States per year.³ The rate of reported incidence of arthrofibrosis following ACL reconstruction ranges from 4% to 38%.⁴ The types of surgeries and the percentage of patients that have continued severe motion loss beyond 1 year after that surgery (in

parentheses) include anterior cruciate ligament (ACL) reconstruction (2–35%), multiple ligament reconstruction (13–22.4%), total knee arthroplasty (1.3–12%), tibial plateau fracture repair (14.5%), and femur fracture repair (29.5%).³ Mild, moderate, and severe arthrofibrosis were classified as having a flexion range of 90°–100°, 70°–89°, and less than 70°, respectively, and/or a loss of extension of 5°–10°, 11°–20° and more than 20°, respectively.⁵ Arthrofibrosis is also characterized by an excessive production of scar tissue. This deep scar tissue builds up, restricting joint motion and causing loss of flexion and extension within the joint.⁵ The tissue is mostly made up of collagen and is known as extracellular matrix (ECM). As ECM builds up, adhesions form, contracting tendons and bursa around the joint, leading to severe pain and loss of motion.⁵ In summary, people with arthrofibrosis experience a wide range of symptoms including significant loss of joint range of motion, instability, pain, and discomfort are all known issues that affect people with arthrofibrosis.

2. SYMPTOMATOLOGY OF ARTHROFIBROSIS

Arthrofibrosis is often diagnosed based on clinical assessment and histopathological analysis. The diagnosis can be further confirmed by magnetic resonance imaging (MRI), which allows physicians to assess the soft tissue within the periarticular region with high resolution.^{6,5} Categorized by a significant loss of knee extension and/or flexion, a physical examination can expose the debilitating effects and loss of range of motion associated with arthrofibrosis. When evaluating the range of motion of prospective patients, quantitative thresholds of flexion and/or extension loss are taken into consideration. This commonly includes a flexion contracture of >15 degrees and/or <75 degrees of flexion, >10 degrees of extension deficit and/or <95 degrees of flexion, or a total knee arc range of motion <70 degrees.⁷ The severity of arthrofibrosis was classified by flexion range, with a range of 90°–100° classified as mild, 70°–89° as moderate, and less than 70° as severe. Similarly, a loss of range of extension by 5°–10° is considered mild, 11°–20° is moderate, and more than 20° is severe.⁵ Additional symptoms along with loss of range of motion include anterior knee pain, flexed-knee gait, quadriceps weakness, and patellofemoral painful crepitation.⁵ Arthrofibrosis of the knee is particularly difficult because symptoms often become intensified with movement such as walking and standing, making the condition even more debilitating than the original injury.⁵

2.1. Range of Motion/ Functional Limitations

Along with an increased cell proliferation in the knee, there is also increased synthesis and deposition of matrix proteins (collagen I, III, and VI) which accumulate, forming bundles and greatly affecting the range of motion of patients.⁸ Fibrous nodules of granulation tissue, which are known as localized anterior arthrofibrosis or cyclops lesions, are extremely common post ACL surgery.⁹ Occurring intra articularly, periarticularly, or both, bundles of tissue/ECM (Extracellular matrix) commonly form in various localized areas.⁸ Localized ECM can form in many places such as the suprapatellar pouch, anterior interval, intercondylar notch, medial and

lateral gutters, and posterior capsule and infrapatellar fat pad (IFP or Hoffa's fat pad). Symptoms vary depending on the location and extent of the ECM and adhesions.⁵ In the suprapatellar pouch, typically, the adhesions pull the walls of the pouch together, while the ECM contracts in the space, preventing normal movement and limiting range of motion.⁵ Similarly, in the posterior capsule, scar tissue often contracts the folds of the capsule, tightening them and preventing full extension of the leg, which causes abnormal gait.⁵

Frequently, excessive ECM around the Infrapatellar Pouch (IFP) and/or shortening of the patellar tendon can cause patella infera.⁵ This may cause marked patellofemoral pain and even osteoarthritis in later stages.⁵ Cyclops lesions are a common form of localized anterior arthrofibrosis occurring in the infrapatellar fat pad.⁹ The incidence rate of Cyclops Lesions ranges from 25% to 47% and most commonly occurs in patients who have undergone ACL reconstruction.⁹ Symptoms of Cyclops Lesions include pain, painful cracking, grinding with attempted extension, locking of the knee, joint line pain, and stiffness.

2.2. Pain

Often overlooked as a cause of arthrofibrosis, pain plays a crucial role post ligamentous injuries in the buildup of scar tissue. 3-10% of Total Knee Replacement (TKR) patients develop arthrofibrosis and report pain with activities from < 30% for light manual work to 78% for jumping and pivoting.¹⁰ The cause of pain in arthrofibrosis is often attributed to femoropatellar joint impairment from increased peripatellar scar tissues, which leads to joint overload and structural changes, such as cartilage destruction.¹⁰ Pain in arthrofibrosis can also be attributed to reflex sympathetic dystrophy, which is a spontaneous regional pain that happens in 15.2% of patients.¹⁰ In addition, 79% of patients who have arthrofibrosis later develop osteoarthritis, creating more pain and long term obstacles.¹⁰ For this reason, it is important to treat patients' pain levels at early stages with appropriate dosing so that the proper physical therapy goals are met in the least painful way possible. Trying to recover post ligamentous surgery without proper pain control can not only cause patients to not reach their physical goals during recovery such as gaining full range of motion, but it can also create a much longer and slower recovery. Not only is poorly managed pain after a ligamentous injury repair a cause of arthrofibrosis, but pain as a result of developing arthrofibrosis can worsen symptoms and conditions. The formation of painful adhesions (abnormal tissue connections) within the knee joint can create difficulties in recovery.¹⁰ In conclusion, strict pain management should be implemented in patients' recovery in order to fully allow patients to recover without discomfort or longer recovery time.

3. PATHOPHYSIOLOGY

The etiology of arthrofibrosis is complex and poorly misunderstood. arthrofibrosis most commonly occurs postoperatively, due to a stressor such as a ligamentous injury, but may also occur post-injury and prior to surgery.⁵ Additionally, in rare cases, arthrofibrosis can occur idiopathically.¹¹ Approximately 85,000 cases of arthrofibrosis occur following knee surgery in

the United States per annum.⁶ Other factors include poor surgical technique, non-compliance of patients in rehabilitation, and poor management of inflammation.⁵

Although the cause of arthrofibrosis is poorly understood, there are various risk factors associated with arthrofibrosis that contribute to the understanding and etiology. Some of these risks are considered modifiable, while others are non-modifiable. Non-modifiable risk factors include sex, the severity of trauma, systemic disease, pre-existing stiffness/inflammation, muscle imbalances, inflammatory disorders including but not limited to: crohn's disease, endometriosis, asthma, diabetes, inflammatory bowel disease, and other disorders such as osteomalacia, sex chromosome disorders, Turners, severe eczema, celiac disease, nutritional deficits, rickets, varus/valgus deformities, and bone/connective tissue disorders, endocrine disorders, and genetic predisposition are all considered risk factors of arthrofibrosis.^{5,12}

Modifiable risks include: poor post-operative rehabilitation, surgical technique, pain management, BMI, prolonged immobilization, smoking, and tourniquet time, injury pattern, cartilage injury classified as Outerbridge grade ≥ 2 , meniscus injury, time from injury to surgery, staged surgery, and preoperative range of motion (ROM).^{12,13} A history of previous surgery has been identified as a possible risk factor of arthrofibrosis, as the rate has seen to double or triple in patients who have undergone previous surgeries for multiple ligamentous injuries.⁷ Other surgical factors include timing of surgery, preoperative range of motion, malposition of graft, excessive graft tension, associated extra-articular procedure and meniscal repair. Post-operative factors can greatly contribute to the formation of arthrofibrosis such as prolonged immobilization, poor rehabilitation, complicated reflex sympathetic dystrophy, infections and synovitis.⁴ Adverse childhood experiences such as neglect or abuse are associated with disease and disability later in life, causing higher Th17 cell numbers, an elevated IL-6 response to stress, and autoimmune and inflammatory diseases.⁵ Additionally, the possibility of early onset osteoarthritis has been predicted to have a positive association with the development of arthrofibrosis. In a study conducted by Remst et al. over half of patients with OA were found to have fibrosis of the synovium.⁵ It is important to note that there are currently no established methods to determine the risk of developing arthrofibrosis following surgery.⁵ As such, understanding and acknowledgment of these multifaceted risks is crucial in order to take the next steps in arthrofibrosis treatment.

Gender, a factor that is often overlooked by many studies, plays a major role in the rates of arthrofibrosis. Females are far more likely to develop arthrofibrosis than men post ligamentous injury. According to a study by Malays Orthop, females who underwent ACL reconstruction faced a higher risk of developing arthrofibrosis at 16% higher than male patients at 4%.⁴ It has been reported that females are 2.5-2.8 times more likely to develop arthrofibrosis after ACL reconstruction.⁵ Additionally, female patients have been identified to report worse outcome scores than men on the self-reported knee function.¹⁴ Both psychological and anatomical differences support the fact that women have a higher propensity to develop arthrofibrosis.

Asaeda et al. suggested that gender differences in knee kinematics during functional activities, such as a naturally-higher knee adduction movement seen in females as well as an increased tibial rotation, may indicate a delay in the recovery of knee kinematics in ACL reconstructed females, suggesting an increased risk of developing arthrofibrosis. Usher et al.'s 2019 review suggested several psychological etiologies for higher rates in females include a female propensity to be less active postoperatively, not performing rehabilitation as well as men, seeking more medical interventions, and having “different” pain tolerance than men.⁵ Although this information is relatively new and not supported by all, gender differences in immunological responses are well established.⁵ 80% of autoimmune diseases occur in females, creating a much higher risk of developing arthrofibrosis.⁵ Additionally, females on average have a stronger immune response than men do, leading to increased rates of inflammatory diseases, another risk factor of arthrofibrosis.⁵

RISK FACTORS	
MODIFIABLE	NON-MODIFIABLE
<ul style="list-style-type: none"> • Surgical technique • Tourniquete time • Pain management • Poor post-operative rehab • Meniscus Injury • Pre-operative ROM 	<ul style="list-style-type: none"> • Sex • Connective tissue disorders • Inflammatory disorders • Systemic disease • Muscle imbalances • Severity of trauma

Figure 1: Modifiable and Non-Modifiable Risk Factors

The figure depicts modifiable and non-modifiable risk factors for arthrofibrosis.

4. INFLAMMATION AND MICROSCOPIC MECHANISM OF SCARRING

4.1. Pathology

The pathology of arthrofibrosis is often misunderstood. It is best characterized as the dysregulation of physiological wound healing responses and the increased production of

myofibroblast production and reduced apoptosis rate, causing a high concentration of myofibroblasts in the synovial fluid.² Typically, in wound healing, a stressor initiates the coagulation cascade and the immune system is activated to repair the damage by a complex process, beginning with the formation of a provisional fibrin clot via the enzyme thrombin.¹¹ Next, cytokines are released such as transforming growth factor β 1 (TGF- β 1) or fibroblast growth factor (FGF) through platelets, macrophages, neutrophils and helper T cells. This promotes tissue remodeling via ECM protein synthesis and fibroblast migration.¹¹ Although ECM deposition and tissue remodeling are inevitable during wound healing, they are prone to dysregulation.⁶ This dysregulation is often due to abnormal cytokine levels and ultimately defines pathologic fibrosis.^{5,11} The presence of TGF- β 1 found in arthrofibrosis proves that arthrofibrosis is caused by an inflammatory response. This inflammatory response is induced from the presence of IL-1 α and IL-1 β .¹⁰

Not only is there a lack of understanding of the pathology of arthrofibrosis, but the inflammatory phase of wound healing and tissue formation is largely misunderstood. The inflammatory response can be separated into two phases: type 1 and type 2. Type 1 inflammatory phase is characterized by an influx of neutrophils and monocytes, as well as coronial proinflammatory cytokines such as TNF and IL-1, and lasts up to 3–4 days.¹⁵ Type 1 initiates the formation of tissue to “clean up” the wound site and begin the subsequent layers of tissue that form as part of the healing process.¹⁵ The tissue can be categorized as fibrovascular granulation tissue and ECM and is a result of concomitant activation of mesenchymal cells and fibroblasts, myofibroblast induction, and angiogenesis.¹⁵ This formation occurs 3-10 days post injury and is followed by a type 2 immune response mediated by macrophages and T-cell subsets: These cells release anti-inflammatory molecules such as IL-10 and angiogenic molecules such as vascular endothelial growth factor and placental growth factor. They also release growth factors like insulin-like growth factor 1 platelet-derived growth factor, Wnt family member 3A, fibroblast growth factor, and activators of myofibroblasts and ECM production such as TGF- β .¹⁶ Arthrofibrosis, on the other hand, is categorized as an excessive type 2 response, with increased levels of IL-4, IL-13, or TGF- β expression.¹⁵

Furthermore, there is a clear association between inflammatory cytokines and collagen production. Many studies have also shown that certain chemokines, cytokines and proteins, can be potent mediators in the pathology of fibrosis.¹² Transforming Growth Factor Beta or (TGF- β) is a necessary cytokine that exists in most cells to regulate immunity and wound healing.¹² TGF- β is a major contributor to fibrosis due to its role in sustaining myofibroblast activation, increased production of disorganized ECM, and the inhibition of myofibroblast apoptosis and collagen degradation.^{12,17} Also, “TGF- β 1 has been known to increase expression of α -smooth muscle actin which causes the activation of fibroblasts as well as other fibrotic associated proteins, e.g. collagen I/III, fibronectin.”¹⁰ Tumor necrosis factor-alpha (TNF- α), platelet-derived growth factor (PDGF), and interleukin IL-1, IL-6, and IL-17, all other proinflammatory cytokines, are believed to be crucial mediators of arthrofibrosis. Additionally, high levels of

TGFBR1 (TGF Beta Type I Receptor) have been found to be associated with arthrofibrosis. According to a study done on rabbits, the level of TGFBR1 was significantly elevated in the rabbit synovial tissue after 4-weeks of fixation ($p<0.05$), displayed a positive correlation with loss of range of motion, and high hydroxyproline contents in the animal model.¹⁸ Increased levels of TGF β 1, TGF β receptors, and enhanced TGF β signaling have been reported in many studies, supporting the role that cytokine are present in the myofibroblast activation and in the pathogenesis of the fibrosis observed.¹⁷

5. LIVELIHOODS AND JOBS

Arthrofibrosis can last for up to 2–3 years, often restricting patients from their normal day-to-day activities and causing pain with movements such as flexion and extension. There are over 85,000 per year major knee surgery patients affected by arthrofibrosis in the US can be conservatively estimated to be.¹⁹ Pain, delayed recovery, and motion limitations are all symptoms of arthrofibrosis post-ligamentous surgery that can significantly affect the livelihoods of patients and not only prevent them from work, but also create financial stress.³ Even a small loss of knee extension, such as 5°, can debilitate a patient from walking normally and loss of flexion can create issues with climbing stairs, sleeping, driving, and getting in and out of chairs.⁵ The inability to do these simple everyday tasks can significantly affect patients' careers, resulting in job loss in severe cases.

Additionally, for patients undergoing Total Knee Arthroplasty (TKA), arthrofibrosis is estimated to be responsible for 28% of 90-day hospital admissions and 10% of revision surgeries within the first 5 years, placing a significant burden on societal costs.⁶ Revision arthroplasties have been seen to cost more than a primary TKA (1.6 times less cost-effective for the same increase in Western Ontario and McMaster Universities Arthritis Index score, and an increased hospitalization cost of \$7,000 over a primary.⁷ Patients who receive a revision surgery are two times more likely to require readmission due to higher rates of complications such as infection, leading to the possibility of worse outcomes due to continuous stiffness after the procedure.⁷ Rehabilitation costs are approximately \$1,500 per knee arthroplasty, adding to the full cost of recovery.⁷ This shows the high costs of revision surgeries, one of the most common forms of treatment for arthrofibrosis, and the possibility of complications, which can be a burden to patients.

The rates of both primary and revision TKA are expected to continue to increase exponentially during the next several decades.²⁰ Additionally, public health insurance has been associated with higher rates of hospital readmissions.²⁰ Although treatment is not always successful, one study found that all patients with arthrofibrosis underwent manipulation under anesthesia within 3 months of the original procedure, but that half of patients required a revision procedure for continued stiffness.²⁰ This shows that in addition to hospital readmissions, patients with arthrofibrosis are at risk for a continued high cost of treatment.²⁰ One study that tested the effectiveness of mechanical therapy devices on knees attributed medical costs and risk of surgery

among patients with arthrofibrosis found that in patients with no device in the 6-month post-index period, medical costs related to their knee were \$9,345± 14,120.²⁰

Another factor that should be noted is the lack of workers' compensation given to patients with arthrofibrosis. We found that very few patients who develop arthrofibrosis and undergo knee surgery receive worker's compensation. According to a study done to test the efficiency of a high-intensity home mechanical stretch therapy device in patients who have arthrofibrosis, only 12 out of 966 patients received HIS device on workers' compensation.³ This is key to note because most patients were not compensated for their recovery time off of work. This lack of compensation could mean patients will feel rushed to return to work and can create a much longer recovery with the possibility of more issues, such as an increased risk of capsular contracture via aberrant ECM deposition, which can manifest secondary to periarticular trauma, surgical insult, postinfectious arthritis, or hemarthrosis.¹²

Additionally, the effects of arthrofibrosis can create psychological issues for patients, such as major depressive disorder. Many studies point to the fact that “depressed patients have been found to have higher levels of proinflammatory cytokines, acute phase proteins, chemokines and cellular adhesion molecules.”²¹ Additionally, therapeutic administration of the cytokine interferon- α leads to depression in up to 50% of patients. “A variety of studies have reported increased levels of inflammatory cytokines and their soluble receptors in peripheral blood and cerebrospinal fluid (CSF) of patients with major depression.”²² Furthermore, it is known that the elevation of inflammatory cytokines can be reflected in the brain itself such as including IL-1 α , IL-2, IL-3, IL-5, IL-8, IL-9, IL-10, IL-12A, IL-13, IL-15, IL-18, interferon-gamma (IFN- γ), and lymphotoxin-alpha (Shelton et al., 2011).²² Not only that, but proinflammatory cytokines have been found to interact with neurotransmitter metabolism, neuroendocrine function, synaptic plasticity, and behavior, which are all related to depression.²¹

Stress can also play a major role in the mental health of patients with arthrofibrosis. Typically, precipitating depression, stress can promote inflammatory responses through sympathetic and parasympathetic nervous system pathways. Although, it is important to note though that both phenotypes, depression and increased immune activation, are heritable.²² Confirmed by twin studies, there was a significant correlation between IL-6 and depression indicators, confirming that 66% of the covariance could be explained by shared genetic factors.²² Due to the positive relationship between inflammation and depression, it is important to take into account the mental health of patients with arthrofibrosis and manage inflammation at early stages. Depression can also inhibit patients from recovery. Studies have noted that patient motivation and state of mind play a crucial role in patients' participation in physical therapy after ligamentous surgeries. “In a study by Fisher and colleagues, patients who were depressed or had a low pain tolerance were less likely to properly perform rehabilitation activities, resulting in delayed recovery and an increased likelihood of developing arthrofibrosis.”⁶

6. TREATMENT METHODS

One of the leading causes of failure and hospital readmission is arthrofibrosis.²³ In the USA alone, nearly 3 million individuals undergo procedures that address arthrofibrosis yearly.¹¹ Presently there is no cure for arthrofibrosis, although some discoveries have been made to ameliorate the debilitating effects.¹¹ Treatment for arthrofibrosis includes non-invasive and invasive methods, ranging from non-operative strategies such as physical therapy and manipulation under anesthesia (MUA), to surgical treatment such as lysis of adhesions (LOA) or revision arthroplasty.^{3,24} Rates of arthrofibrosis requiring manipulation under anesthesia (MUA) or arthroscopy within 6 months of the initial surgery were up to 8%.⁵ Many investigations utilizing rabbit models of surgically induced knee joint contractures have identified that myofibroblast proliferation is significantly elevated at two weeks post-surgery and not at later points.¹² For this reason, prevention interventions should be implemented at early stages of recovery.

6.1. Non-Invasive Treatment Methods

6.1.1 Manipulation Under Anesthesia:

Manipulation Under Anesthesia (MUA) has been proven to be an effective treatment method to improve ROM in patients who suffer from stiffness from arthrofibrosis post ligament reconstruction.¹² It is performed by applying progressive flexion to the joint to break up the scar tissue, and is often performed concurrently with arthroscopy.¹² In the United States, 7.3% of patients who undergo a total knee arthroplasty require a MUA, and 11.3% of patients who undergo a knee arthroscopy require a MUA.³ MUA can be done on its own or following the lysis of adhesions, which has been seen to maximize the range of motion.⁵ Arthroscopic lysis of adhesions before MUA permits focal areas of scarring to be addressed while limiting the potential for complications of MUA.¹² The decision to have an MUA depends on patient progress and degree of joint stiffness.¹²

The risk of MUA increases if joint stiffness persists beyond 6 weeks in non-arthroplasty patients.³ The literature has shown that the MUA can be less effective the longer one waits to have one. The amount of ROM regained can be inversely proportional to the time between diagnosis and MUA. MUA's positive utility has been seen to diminish when performed more than 3 months after TKA.²⁴ It is not recommended to have a MUA after 12 weeks from the initial surgery, as there is a high risk of rupturing the patellar tendon or fracturing the patella due to the increased tensile strength of the adhesions.¹¹⁹ Before deciding to have an MUA, all risk factors should be evaluated, such as sex, age, BMI, smoking, and autoimmune diseases. The outcomes following MUA are variable, and 4-26% of individuals experience poor outcomes and continued knee stiffness.²⁵ One study reported that MUA was only 74% successful in reaching 90° of knee flexion.³ However, there is a risk that comes with a MUA since it can provoke an inflammatory response that may aggravate rather than alleviate fibrogenesis.⁵ Other risks of MUA are rare and include heterotrophic ossification, tibial plateau fracture, damage to a prosthesis, patellar ligament avulsion and blood clots, pulmonary embolism, and death.^{3,5}

6.1.2 Physical Therapy:

Physical therapy is often the initial approach in terms of non-surgical treatment for arthrofibrosis.¹² If arthrofibrosis is diagnosed early, physical therapy can be useful as a preventative measure along with stretching, bracing, and non-steroidal anti-inflammatory drugs. (NSAIDs)²⁴ The target of physical therapy is often to target muscle imbalances in order to prevent bone and joint deformity.²⁶ Through passive and active assisted stretching and various exercises, physical therapy can delay and alleviate contractures, as well as help with mobility. Other examples of effective physical therapy treatments include gentle physical exercises, dry needling, cupping, and massage.²³ Additionally, the application of continuous passive motion (CPM) has been discussed widely, and many studies have found it to be highly effective in increasing the range of motion of patients with arthrofibrosis.²⁷ According to one study, patients who used a CPM machine had a significant mean range of motion increase in flexion of 34.6°+ -17°.²⁷ CPM can be a highly effective option along with the use of physical therapy to help counter the debilitating effects of arthrofibrosis.

It is important to note the risks associated with physical therapy if approached too aggressively. Overly aggressive ROM exercises employed before sufficient healing can create inflammation and fibrosis in the affected joint.¹² According to many studies, rigorous physical therapy when the knee joint is irritated can be harmful and lead to arthrofibrosis: “Pain, pain during physiotherapy, and muscle building training when the joint is irritated can significantly increase the risk of arthrofibrosis.”⁸

6.2. Invasive Treatment Methods

6.2.1 Arthroscopic Lysis of Adhesions

Arthroscopic lysis is the most commonly performed treatment for arthrofibrosis with rates up to 8% for patients who required to have an arthroscopy within 6 months of initial surgery.⁵ The decision for surgical intervention is determined by patient progress and the degree of stiffness.¹² Arthroscopic Lysis can be highly effective, as it not only entails the removal of ECM and the physical restriction of ROM, but also removes bound pro-fibrotic mediators, including TGF- β .⁵ Nonetheless, the benefits of lysis should be tempered with an understanding of the problems and risks associated with the procedure.⁵ Most notably, the tissue damage from surgery can exacerbate inflammation and ECM proliferation, the very processes it seeks to ameliorate.⁵ In order to successfully address the adhesions, surgery should draw attention to the specific areas where scarring is present.

To address flexion deficits, the surgeon should direct attention to the suprapatellar pouch, the patellar femoral joint, and the anterior interval, as well as the intercondylar notch which can be associated with flexion deficits when scarring is present.¹² To resolve extension deficits, attention should be drawn to the medial and lateral gutters as well as the posterior joint space. The outcomes of arthroscopic lysis vary from patient to patient as severity of symptoms, degree of inflammation, and extent of fibrosis within and around the joint can differ.⁵ Many studies have

reported significant improvement in ROM following LOA with increased flexion ranging from 45 to 68 degrees in post-traumatic injuries.¹² Of 207 knees treated for postoperative stiffness, 202 responded to conservative management or arthroscopic surgical LOA in a series published by Noyes et al.¹² It is important to note that the accuracy of these rates could be higher than they appear due to reporting bias. Following the surgery, patients will immediately follow up with physical therapy sessions for 5 days per week for 6 weeks.

6.2.2 Revision Arthroscopy

It is important to note that invasive procedures such as RA and LOA are not recommended in the first six months of the procedure, and often occur a year after initial surgery.¹⁹

Many patients fail to regain ROM after LOA, which leaves RA as a final procedure.²⁴ revision has been reported to be a reasonable option, although the reported results seem to be modest in regard to pain, function, and ROM.²⁸ Prior to performing RA, it is important to know the reason for failure and loss of motion because the failure to identify the cause of stiffness may result in recurrence of the problem.²⁸ Outcomes following RA vary from study to study due to the lack of definition of stiffness after surgery makes it difficult to compare results.²⁸ One study by Hartman et al. reported a significant reoperation rate as 49% of their patients required a further intervention for stiffness or postoperative complication, while Moya-Angeler reported that 9.1% required a further intervention for stiffness or sustained a complication.²⁸ Approximately 25% of patients with arthrofibrosis treated with a motion restoring surgical procedure, required multiple surgeries after.¹⁹ Additionally, only 37% of patients surgically treated for more diffuse arthrofibrosis reported satisfactory results.¹⁹ In summary, 21,000 patients each year are at risk of requiring an additional surgery and an additional 54,000 patients per year are at risk of an unsatisfactory outcome, demonstrating either “abnormal” or “severe abnormal” results according to the International Knee Documentation Committee (IKDC), including the arc of motion, joint stability, patient-reported subjective symptoms, and the physician’s clinical findings.¹⁹

7. PAIN MANAGEMENT

Inflammation and pain can make patients’ recovery even more debilitating and extend the recovery time. It is important to treat pain and discomfort at early stages in order to achieve maximum recovery in patients with arthrofibrosis

7.1. Non-Steroidal Anti-Inflammatory Drugs (NSAIDs)

NSAIDs help reduce pain, which allows patients to participate in physical therapy.¹² NSAIDs such as aspirin and ketorolac are effective in treating pain by reducing inflammation through suppression of COX-1 and COX-2, and are typically given at early stages of arthrofibrosis.¹¹ Due to the harmful side effects that come with NSAIDs, they should only be

used in small doses in the early stages of arthrofibrosis as a short-term treatment.¹¹ Possible side effects include increased risk of heart failure, stroke, and renal and gastrointestinal toxicity.¹¹

7.2. TNF- α Antibody

Some studies have identified the possibility of TNF- α antibody treatment as an effective treatment method to manage pain associated with Arthrofibrosis. TNF- α antibodies have been shown to reduce lung fibrosis in mice, and mice lacking TNF- α signaling pathways are protected from lung fibrosis.⁵ TNF- α induces peripheral pain sensitization, so it is expected that TNF- α antibodies will assist in pain management.⁵

7.3. Hyaluronic Acid

Hyaluronic acid can be used as an effective anti-inflammatory and regulates cell proliferation.²⁹ It is considered the principal component of the ECM and plays a key role in tissue regeneration processes.³⁰ Depending on its molecular weight, hyaluronic acid can be an effective anti-inflammatory treatment method for arthrofibrosis. High molecular weight hyaluronic acid has anti-inflammatory properties, while low molecular weight hyaluronic acid is proinflammatory. Hyaluronic acid also reduces the levels of proinflammatory cytokines such as prostaglandins, leukotrienes, IL-1, and IL-6.³¹

The anti-inflammatory aspect of hyaluronic acid has been well documented in osteoarthritis. As reported by many studies, hyaluronic acid has been seen to down regulate the production of IL-8 and iNOS gene expression in unstimulated human fibroblast synoviocytes (FLS) and aggrecanase-2, and tumor necrosis factor alpha (TNF α) gene expression in IL-1-stimulated FLS.³⁰ Additionally, it significantly diminished TLR4, TLR2, MyD88 and NF-kB expression and protein synthesis in synoviocytes in murine model of osteoarthritis.³⁰ They have also observed reduced mRNA expression and protein production for TNF α , IL-1 β , IL-17, MMP-13 and inducible nitrous oxide synthase gene in arthritic mice treated with high molecular size hyaluronic acid.³⁰ Although, it is important to note, though, that these results were an outcome of hyaluronic acid given to osteoarthritis patients at early inflammatory phases.³¹ Similar to arthrofibrosis, the synovial fluid of osteoarthritis is orchestrated by macrophages and inflammatory cytokines. The use of an intra-articular injection in the synovial fluid can be effective in reducing the inflammatory levels of arthrofibrosis and regulating the myofibroblast proliferation. The use of hyaluronic acid to reduce pathogenic cytokines and immune cells was tested on patients with osteoarthritis and was seen as highly effective.³¹ For this reason, hyaluronic acid can be an effective treatment method to reduce knee joint inflammation and myofibroblast proliferation.

7.4. Personalized Bayesian Statistical Approach

Currently, no methods exist to determine patients' individualized risk levels for developing arthrofibrosis. To prevent arthrofibrosis, it is crucial to evaluate patients' personal risk

levels at different stages-before and after injury, pre-surgery, and post surgery. Bayes' Statistical Theorem allows for a perfect solution to this problem, as we can essentially update a previously anticipated probability given new independent information to conjure an updated, more accurate posterior probability.

7.5. Continuously Updated Risk Assessment of Scarring (CURAS)

CURAS is a continuously updated risk assessment of scarring that allows patients to understand their individualized risks of scarring. The process of scarring is longitudinal and at multiple stages during that process, additional risk factors emerge and can be used to update the likelihood of scarring in the near or distant future. In the pre-injury period, the mechanism of injury, the severity of injury, the number of ligaments torn/injured, and muscle imbalances can modify this baseline risk of scarring. In the post-injury period, factors such as previous surgeries, pre-existing stiffness or inflammation, early-onset OA, childhood adversity, female gender, autoimmune diseases, and BMI, can all additionally modify the previous risk of scarring. In the early and late postoperative periods markers of inflammation, mobility, pain, early utilization of physical therapy, time to intervention, etc., all subsequently act to further modify one's personalized risk of scarring.

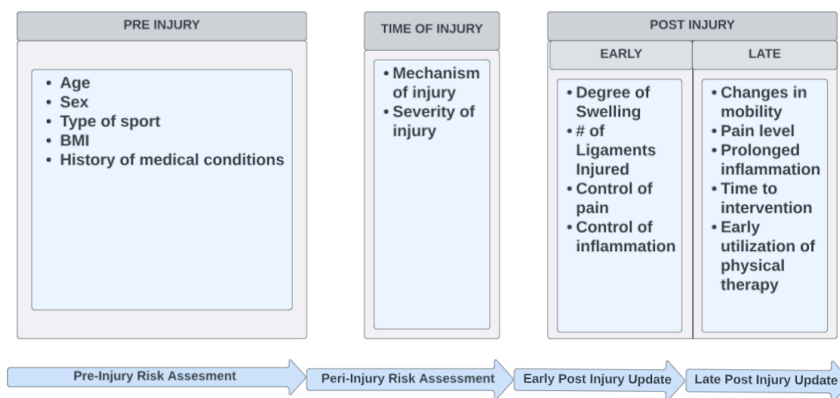


Figure 2 Continuously Updated Risk Assessment of Scarring The figure above depicts risk factors that emerge at each stage (pre injury, time of injury, early post injury, and late post injury) that should be used to assess and continuously update individualized risk of scarring

7.6. Personalized Care

There are no organized, personalized protocols for patient care that synthesize multidimensional risk factors for scarring. There are numerous opportunities, at multiple time

points, to personalize arthrofibrosis prevention and treatment. For example, during the pre injury period, protocols such as informing patients early on of the factors that put them at higher risk of scarring and making them aware that if an injury were to occur they are cognizant that they are at high risk of scarring due to those risk factors. If those risk factors are modifiable, suggesting solutions to the patient and working with them to find a plan in order to reduce the risk of scarring is crucial. Although not much can be done during the inter injury stage, much can be done in the early and late post injury stages. In the early post injury stages, it is important to tackle modifiable risk factors such as pain and inflammation quickly and efficiently. For example, if a patient shows high levels of inflammation and high inflammatory biomarkers, an effective treatment could be to add more immunosuppressants like a steroid injection. For patients who might complain of strong pain levels, treating this pain first and foremost is crucial so that they can be most efficient in physical therapy. The personalization of care in patients is something that has been widely discussed, although it has failed to make its way into clinical care.³² Personalized care can provide unprecedented insight and treatment to patients, allowing each patient to be evaluated based on their characteristics and history, as opposed to a “one size fits all approach.” During each stage of arthrofibrosis pathogenesis, there is potential to implement simple measures that provide more tailored/personalized care, thus improving prognosis and recovery.

8. CONCLUSION

Current approaches to treat arthrofibrosis are non-invasive, conservative, and focus on managing post-operative scarring. In contrast, few approaches target the complex mechanism of the scarring itself, particularly prior to surgery. Above, we discussed the clinical symptomatology as well as the molecular pathophysiology of arthrofibrosis and proposed CURAS, a novel individualized risk assessment model incorporating Bayesian statistics to offer personalized risk grading of scarring longitudinally. The next steps would be to implement a single cohort prospective study incorporating CURAS to understand its effectiveness and how interventions may be offered in a specific or individualized basis. Additional work needs to be done to elucidate additional mechanisms of scar formation.

Works Cited

- Fackler N, Chin G, Karasavvidis T, et al. 2022. Outcomes of Arthroscopic Lysis of Adhesions for the Treatment of Postoperative Knee Arthrofibrosis: A Systematic Review. *Orthop J Sports Med* 10(9):23259671221124911.
- Ly T-D, Sambale M, Klösener L, et al. 2023. Understanding of arthrofibrosis: New explorative insights into extracellular matrix remodeling of synovial fibroblasts. *PLoS One* 18(5):e0286334.
- Stinton SK, Beckley SJ, Branch TP. 2022. Efficacy of non-operative treatment of patients with knee arthrofibrosis using high-intensity home mechanical therapy: a retrospective review of 11,000+ patients. *J. Orthop. Surg. Res.* 17(1):337.
- Rushdi I, Sharifudin S, Shukur A. 2019. Arthrofibrosis Following Anterior Cruciate Ligament Reconstruction. *Malays Orthop J* 13(3):34–38.
- Usher KM, Zhu S, Mavropalias G, et al. 2019. Pathological mechanisms and therapeutic outlooks for arthrofibrosis. *Bone Res* 7:9.
- Thompson R, Novikov D, Cizmic Z, et al. 2019. Arthrofibrosis After Total Knee Arthroplasty: Pathophysiology, Diagnosis, and Management. *Orthop. Clin. North Am.* 50(3):269–279.
- Cheuy VA, Foran JRH, Paxton RJ, et al. 2017. Arthrofibrosis Associated With Total Knee Arthroplasty. *J. Arthroplasty* 32(8):2604–2611.
- Mayr HO, Stöhr A. 2014. [Arthroscopic treatment of arthrofibrosis after ACL reconstruction. Local and generalized arthrofibrosis]. *Oper. Orthop. Traumatol.* 26(1):7–18.
- Kambhampati SBS, Gollamudi S, Shanmugasundaram S, Josyula VVS. 2020. Cyclops Lesions of the Knee: A Narrative Review of the Literature. *Orthop J Sports Med* 8(8):2325967120945671.
- Li CY, Ng Cheong Chung KJ, Ali OME, et al. 2020. Literature review of the causes of pain following total knee replacement surgery: prosthesis, inflammation and arthrofibrosis. *EFORT Open Rev* 5(9):534–543.
- Blessing WA, Williamson AK, Kirsch JR, Grinstaff MW. 2021. The Prognosis of Arthrofibroses: Prevalence, Clinical Shortcomings, and Future Prospects. *Trends Pharmacol. Sci.* 42(5):398–415.
- Lee DR, Therrien E, Song BM, et al. 2022. Arthrofibrosis Nightmares: Prevention and Management Strategies. *Sports Med. Arthrosc.* 30(1):29–41.
- Patel NK, Lian J, Nickoli M, et al. 2021. Risk Factors Associated With Complications After Operative Treatment of Multiligament Knee Injury. *Orthop J Sports Med* 9(3):2325967121994203.
- Zhang L, Fan S, Ye J, et al. 2021. Gender differences in knee kinematics during weight-bearing knee flexion for patients with arthrofibrosis after anterior cruciate ligament reconstruction. *J. Orthop. Surg. Res.* 16(1):573.
- Xia Y, Sokhi UK, Bell RD, et al. 2021. Immune and repair responses in joint tissues and lymph nodes after knee arthroplasty surgery in mice. *J. Bone Miner. Res.* 36(9):1765–1780.

- Kendall RT, Feghali-Bostwick CA. 2014. Fibroblasts in fibrosis: novel roles and mediators. *Front. Pharmacol.* 5:123.
- Stone RC, Pastar I, Ojeh N, et al. 2016. Epithelial-mesenchymal transition in tissue repair and fibrosis. *Cell Tissue Res.* 365(3):495–506.
- Chen X, Wang Z, Huang Y, et al. 2021. Identification of novel biomarkers for arthrofibrosis after total knee arthroplasty in animal models and clinical patients. *EBioMedicine* 70:103486.
- Stephenson JJ, Quimbo RA, Gu T. 2010. Knee-attributable medical costs and risk of re-surgery among patients utilizing non-surgical treatment options for knee arthrofibrosis in a managed care population. *Curr. Med. Res. Opin.* 26(5):1109–1118.
- Schairer WW, Vail TP, Bozic KJ. 2014. What are the rates and causes of hospital readmission after total knee arthroplasty? *Clin. Orthop. Relat. Res.* 472(1):181–187.
- Raison CL, Capuron L, Miller AH. 2006. Cytokines sing the blues: inflammation and the pathogenesis of depression. *Trends Immunol.* 27(1):24–31.
- Bufoalino C, Hepgul N, Aguglia E, Pariante CM. 2013. The role of immune genes in the association between depression and inflammation: a review of recent clinical studies. *Brain Behav. Immun.* 31:31–47.
- Flick TR, Wang CX, Patel AH, et al. 2021. Arthrofibrosis after total knee arthroplasty: patients with keloids at risk. *J. Orthop. Traumatol.* 22(1):1.
- Rockov ZA, Byrne CT, Rezzadeh KT, et al. 2023. Revision total knee arthroplasty for arthrofibrosis improves range of motion. *Knee Surg. Sports Traumatol. Arthrosc.* 31(5):1859–1864.
- Rauzi MR, Foran JRH, Bade MJ. 2022. Multimodal conservative management of arthrofibrosis after total knee arthroplasty compared to manipulation under anesthesia: a feasibility study with retrospective cohort comparison. *Pilot Feasibility Stud* 8(1):71.
- Martinez-Lozano E, Beeram I, Yeritsyan D, et al. 2022. Management of arthrofibrosis in neuromuscular disorders: a review. *BMC Musculoskelet. Disord.* 23(1):725.
- Aspinall SK, Bamber ZA, Hignett SM, et al. 2021. Medical stretching devices are effective in the treatment of knee arthrofibrosis: A systematic review. *J Orthop Translat* 27:119–131.
- Moya-Angeler J, Bas MA, Cooper HJ, et al. 2017. Revision Arthroplasty for the Management of Stiffness After Primary TKA. *J. Arthroplasty* 32(6):1935–1939.
- Pakhomova AS, Veřdner K, Reperan Z, Veselkin NP. 1987. [Participation of anterior thalamic nuclei in the visual afferentation of the hippocampus in the rat]. *Neirofiziologija* 19(2):278–280.
- Litwiniuk M, Krejner A, Speyrer MS, et al. 2016. Hyaluronic Acid in Inflammation and Tissue Regeneration. *Wounds* 28(3):78–88.
- Jin L, Xu K, Liang Y, et al. 2022. Effect of hyaluronic acid on cytokines and immune cells change in patients of knee osteoarthritis. *BMC Musculoskelet. Disord.* 23(1):812.
- Ho D, Quake SR, McCabe ERB, et al. 2020. Enabling Technologies for Personalized and Precision Medicine. *Trends Biotechnol.* 38(5):497–518.

Who is A Genius? What is Beautiful? Who is a Terrorist?: Gender, Religious, and Racial Biases in Image Search Algorithms By Sema Akkaya

Abstract

Our lives have been increasingly surrounded by the transformative waves of Artificial Intelligence (AI). AI systems have the potential to have numerous positive impacts on our daily lives. However, they also have the potential to maintain and further existing stereotypes and biases in society. This study, focusing on the search engines, examines the way they shape and are shaped by user's realities. More specifically, this study analyzes the image search results of popular search engines, Google and Bing, as a lens to understand the biases in these search engines. To this aim, this study analyzed the image search results for "genius," "beautiful," "terrorist," "woman terrorist," "Muslim woman," and "Muslim man," and found the way existing gender, racial, political, and religious biases make their way to these search engines. Documenting the gender, racial, and ethnic biases inherent in web search engines, I argue that it is vital for these companies and their researchers to think critically about the way biases could enter into these systems. To create more inclusive web search systems, it is important to develop more inclusive AI algorithms and automated bias detection methods.

Introduction

Our lives have been increasingly surrounded by the transformative waves of Artificial Intelligence (AI). AI systems have the potential to have numerous positive impacts on our daily lives. However, they also have the potential to maintain and further existing stereotypes and biases in society (Angwin and Kirchner, 2016; Karpa et al., 2022). Since AI systems primarily rely on machine learning algorithms, hence numbers, they might be thought of as neutral systems. Nevertheless, machine learning algorithms' reliance on existing data for training purposes could bring existing historical and societal biases into these systems. More specifically, these AI systems can be biased based on who builds them, the way they are developed, and by those who used them. One AI supported system in which we can find all these factors perpetuating existing biases is web search engines (Trielli, 2022).

Web search engines, through searching the Internet, filter and rank information in response to user queries. In doing so, they provide information and determine what users learn about specific topics. However, a growing number of research studies have shown that search engines can perpetuate existing racial and gender biases and hence can have negative social effects (DeVos et al., 2022; Rozado, 2023; Steiner et al., 2022). These biases can be originated from several factors, including the search engine's algorithm, the developer's implementation, and user's stereotypes and ideologies. In this research, I primarily focus on the search engines and examine the way they shape and are shaped by the user's realities. More specifically, I analyze the image search results of popular search engines to demonstrate the way gender, racial, political, and religious biases make their way to these search engines.

Literature Review

As more AI systems started to become mainstream, an increasing number of studies began to explore the perils of these systems in relation to their societal impacts, including but not limited to bias, government surveillance, and ethics (Ferrer et al., 2021; Karpa et al., 2022; Noble, 2018; Zajko, 2021). For instance, Ferrer et al. (2021) demonstrated that AI-based systems that are used in automated decisions, from credit scores to insurance payouts to health evaluations, can discriminate based on income, education, gender, or ethnicity. Similarly, Howard and Borenstein (2017) found that existing social biases have infused themselves into current AI and robotic systems, such as robot peacekeepers, self-driving cars, and medical robots. They argued that the bias in these systems primarily came from learning algorithms and their reliance on large datasets for training and crowd-sourced masses for labeling.

The AI bias can also be found in translation tools. Bolukbasi et al. (2016), studying Google translation from Turkish to English, demonstrated that even though Turkish does not have gendered pronouns and markings, when Turkish sentences are translated into English, gender neutral Turkish words are assigned a gender that reflects the existing gendered stereotypes. They further showed that this was primarily due to the tools trained on Google News articles. They developed an algorithm to “debias” the embedding.

Another study on AI bias was conducted by Kay et al. (2015) on web search engines. They primarily looked at the image search results for occupations. They found the existing gender stereotypes and systematic underrepresentation of women in their search results. They further found that users rated search engines as higher when the search results matched the existing societal stereotypes for a career. In their study, Introna and Nissenbaum (2000), argued that part of the bias found in search engines came from the search engines' reliance on popular, wealthy, and powerful sites instead of less popular and smaller sites. Since the search engines are not designed to search the whole web, they will primarily collect and summarize the data from the popular sites instead of the less popular ones.

Similarly, in her book “Algorithms of Oppression,” Noble (2018), exploring the way hidden structures shapes how we get information, demonstrated that search engine algorithms are not neutral. For instance, when she entered the term “black girls,” the results came back with a lot of pornographic images. She continued her analysis by typing “why are black women so” in Google search engine and the suggested results completed the phrase with “angry,” “loud,” and “mean.” She concluded that AI’s reliance on data for training purposes brought the existing racial and gender biases into these search engines. After their study, Google has fixed some of these biases in their search engines and the translation tool. However, similar results could still be found for some other minority groups, such as Asians, Muslims, and Latinos.

Methods:

The data has been collected from Google and Bing Search Engines over a week period through three different laptops to control for personalization and randomization that might influence the search output results. To collect data, the researcher entered the key words, such as

“genius,” “beautiful,” “terrorist,” “Muslim woman”, and “women terrorist” into the search box and clicked on search. Afterwards, the image option has been selected to see the results for images only. Each search has been repeated from three different devices and at three different dates. The results only differed from each other by two to six images. Thus, this study primarily relied on the initial search results, noting the difference when it is significant. For certain searches, especially for the words “terrorist,” and “genius,” the results yielded several AI generated pictures. To narrow it down to real images, the option “people” has been selected.

For each search result, only the first page was taken into consideration. I primarily used qualitative data elicitation methods in collecting and analyzing data. I also used descriptive statistics tools to summarize the data for analysis. From the initial page, only the images with real people and clear faces have been taken into consideration (See Table-1 for number of images analyzed). For some search results, there were more AI-generated images than usual, i.e. for “genius” and “genius people.” Therefore, the total number of images were lower for those words.

Key Word Searched	Bing	Google
Terrorist	114	390
Women Terrorist	106	351
Muslim Women	136	393
Muslim Men	131	395
Beautiful	132	395
Genius	134	91
Genius People	134	158

Table-1: Number of Images For Each Search from Bing and Google

Results:

Genius: When the key word “genius” was entered into the Google search engine, the results yielded 421 images in total. However, most of these images are either clipart or wallpaper. Only the images with clear faces were taken into account in this research. Among 91 images with real faces, only 11 of them were women whereas 80 of them were men. When we changed the key word to “genius people,” the results were similar. Only 21 of the pictures with clear faces were women, whereas 137 were men, totaling 158. As it could be seen from Image-1, the first 15 pictures were all men in Google search. When the same search was repeated in Bing, similar results were observed. Among the 134 images, 129 were men and 5 were women. The first 15 images from Bing also primarily showed men, only in one of the pictures a woman was portrayed with other men (Image-2).

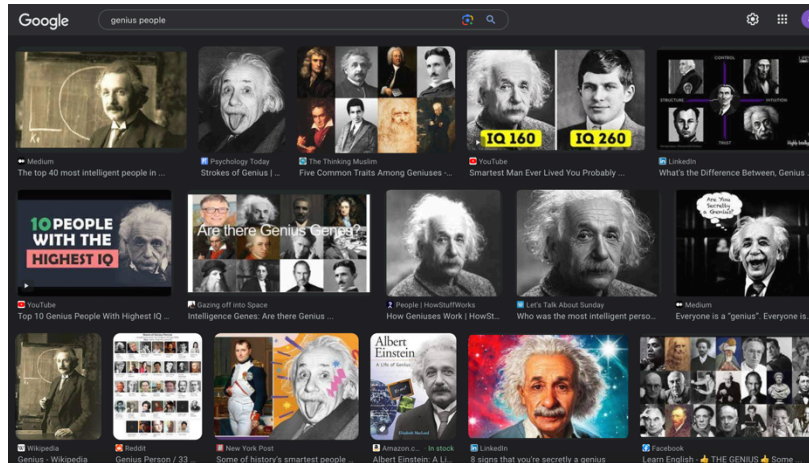


Image-1: Image Search Results for “Genius People” from Google

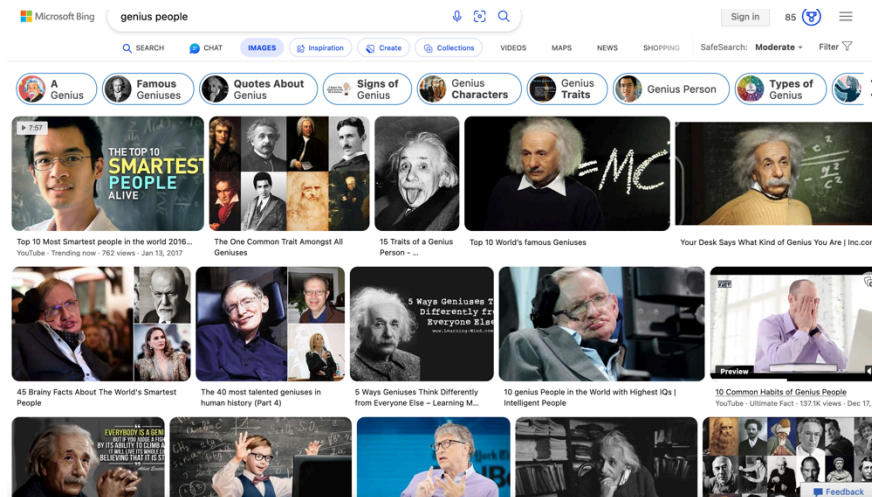


Image-2: Image Search Results for “Genius People” from Bing

Beautiful: When the key word “beautiful” was searched in Google, the image results yielded 339 women out of 395 images. Among the 56 images that do not show a woman, 48 were images of nature, 1 was an architectural building, 2 were sayings, 1 was a movie, 1 was a song, and 3 were men. Of 3 male images, one was portraying a male singer who sings the song “Beautiful girl,” and one man was from the TV series “The Bold and the Beautiful.” Only one man was portrayed as being beautiful among the 395 images. Below is a screen shot of the first two rows of image search results for “Beautiful” from Google search engine (Image-3). Even

though compared to past research results, there is more diversity among women, still, all first 15 pictures portrayed a female as being beautiful. The search results yielded more extreme results in Bing. When the key word “beautiful” was searched in Bing and narrowed down to “people,” all 132 images portrayed a woman figure in the pictures.

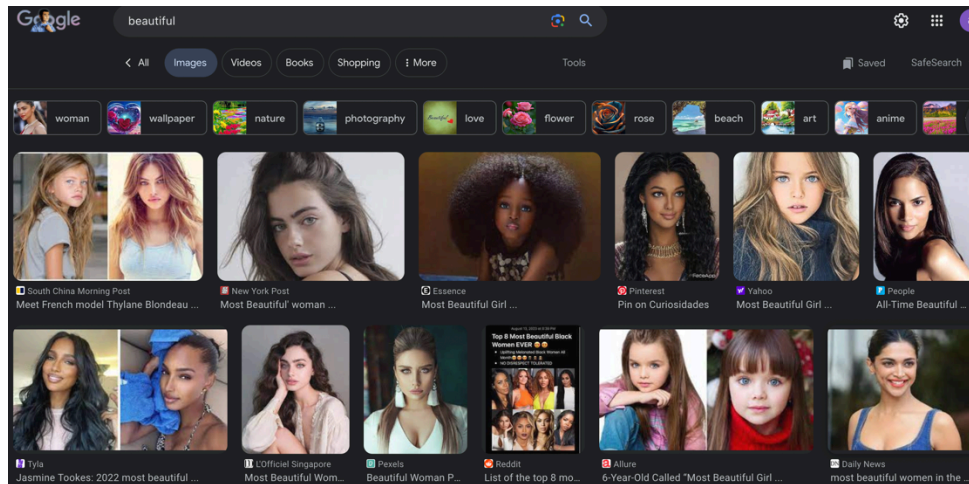


Image-3: Image Results for “Beautiful” from Google Search Engine

Terrorist: The search has been repeated for the chosen keyword “terrorist” in Google search engine. Even though the key word was a noun and did not include an adjective depicting any ethnic, religious, and/or gender information, such as “Muslim,” “men”, or “Middle Eastern,” most of the search results for images portrayed a stereotypical Middle Eastern Muslim man. The initial search results yielded over 350 images. However, only 141 of them were images of real people and the rest were AI generated images. Among the 141 pictures, 119 were Muslim. 3 of the images showed members from “Proud Boys.” However, 1 of them only showed the member from the back with no visible face, and 1 showed a person’s face from an angle. Only 1 of the pictures showed a full face. The same search, when repeated in Bing, yielded similar results for “terrorist.” 108 of 114 pictures were portraying a Muslim person in the pictures as a terrorist.

Woman Terrorist: Since the majority of the images were portraying a Muslim man (90%) in response to the “terrorist” word search, the search has been repeated for “women terrorist” to see if the results would differ from that of the “terrorist.” The results for “women terrorist” in Google brought 349 images in its first page results. About 99 of the images were either blurred, AI generated, or included a book cover with no visible signs and hence eliminated from data analysis. Among the 250 images, 206 portrayed a Muslim woman or women with headscarves. The remaining 44 portrayed 3 Muslim women with no headscarf (2 Kurdish women), 1 Muslim man, 1 Maoist woman, 1 Jewish man, and 3 Hindu women. 5 of the 44 images were about counter-terrorism news, whereas 4 of them were the same book cover portraying the first female terrorist organization in the U.S. (See Image-4, picture 8). 2 of them depicted a sorority girl, one with a gun and one without a gun. The other images were a little ambiguous and some of them

still have some elements that could link them to the Middle Eastern region. When “women terrorist” was searched in Bing, the results yielded similar results to Google. 97 of the women were Muslim women out of 106 images.

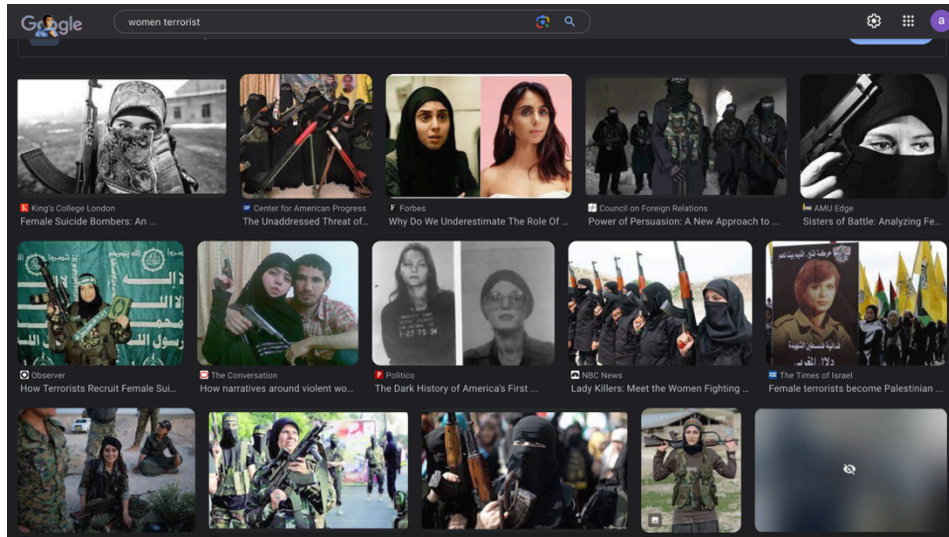


Image-4: Image Results for “Women terrorist” from Google Search Engine

Muslim Woman: The search has been repeated for “Muslim women” starting with the Google search engine. The first page results yielded 393 images. 387 of them showed a Muslim woman with a head covering. Six of the images that did not have a Muslim woman with a head covering included 1 Ahmadi Muslim man, 1 Muslim woman with a scarf on her shoulders instead of over her head, and 4 Muslim women with no head scarf. Among the 387 images, 7 of them showed a Muslim woman with a head covering with other non-Muslim women and 4 Muslim women with no head covering along with other Muslim women with a head covering. Among the first 393 images on Google search results, the first woman with no head scarf came in picture 122. When the search for “Muslim women” was repeated in Bing, 136 of the 136 images showed a Muslim woman with a head covering. Bing did not show any Muslim with no head covering in its first page results. Another interesting finding is that 76 of the 387 images showed a Muslim woman not only with a head covering but also with a face covering (veil). Below is an image from the Google search engine results (Image-5). The first 11 images showed a Muslim woman with a head covering. 4 of the 11 also came with a veil as well.

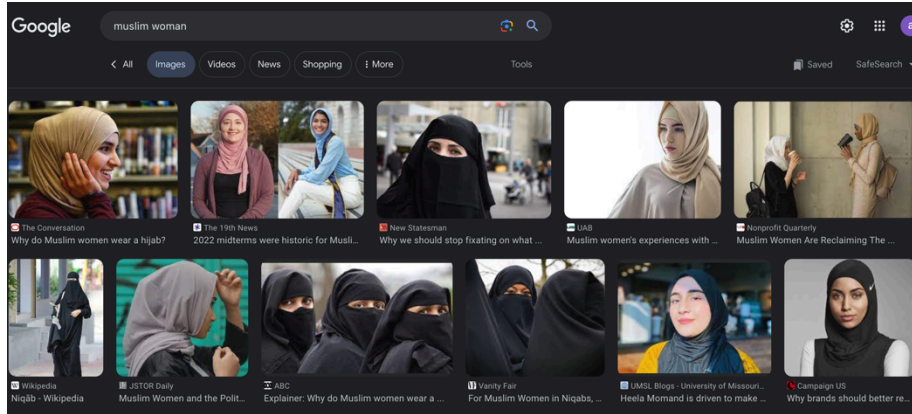


Image-5: Image Results for “Muslim woman” from Google Search Engine

Muslim Men: When the search was repeated for “Muslim men,” the results yielded 336 images, among which 308 was portraying a Middle Eastern Muslim man with a beard and a form of head covering. 6 of the 28 of the images showed an African American Muslim man with no clear visible signs of their religious identity. 4 of the images showed a Muslim man marrying a Muslim woman. Only 16 of the images showed a man with no visible signs of their religious identity. The image below shows the first 12 images that came on Google search engine for “Muslim men” (See Image-6). As it could be seen from the screen shot of the search, only one of the images did not have a clear stereotypical element of being a Muslim and/or a Middle Eastern.

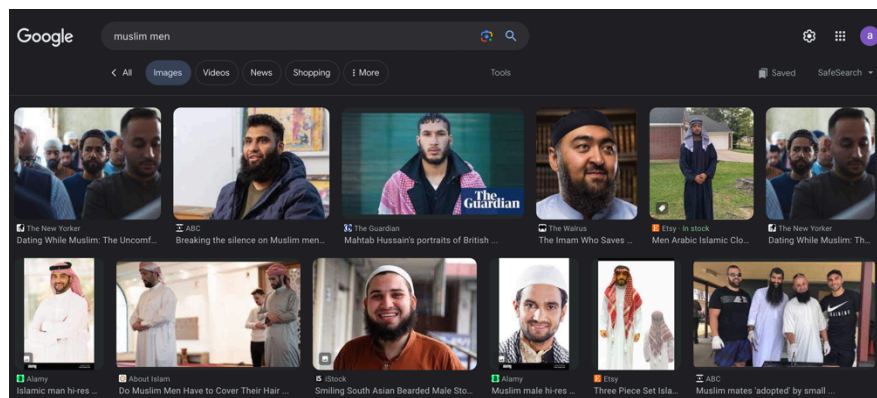


Image-6: Image Results for “Muslim woman” from Google Search Engine

Discussion of Results:

As the analysis demonstrated, web search engines are not neutral tools. Behind these search engines, one can find an evolving algorithm. Since these algorithms have been created by a human, in its developmental stage, they could come with certain parameters to filter and rank information and hence could be skewed in the way they decide which websites to include and exclude when searching the web. Furthermore, as these search tools train on existing data, the

existing biases in these data make their way into these search engines. If they are neutral systems, the search results for “beautiful,” and “genius” should yield more diverse results. Instead, the search engines decide that women are beautiful whereas men are geniuses. Similarly, even though the search engine was only provided with the word “terrorist,” it assigned a gender, religious, and ethnic identity to it.

Some might argue that an educated person could navigate through these search results and find reliable information. However, research shows that images whether we agree with them or not get into our mental library and may shape our thoughts and hence judgements (Eberhardt 2020, Lebeouf 2020). In addition, not all people will have the necessary critical skills to search through the web and eliminate unreliable information. Documenting the gender, racial, and ethnic biases inherent in web search engines, I argue that it is vital for these companies and their researchers to think critically about the way biases could enter into these systems. To create more inclusive web search systems, it is important to develop more inclusive AI algorithms and automated bias detection methods.

When conducting this research, I also realized that Google fixed some of the existing biases in their search engines and translation tools after the published work. For instance, after Noble (2018) published her work, they addressed some of the racial biases in Google search engine. Similarly, after Bolukbasi et al. 's study (2016) on Turkish-English translations, they fixed some of the gendered stereotypes that entered into the translation tool. However, as we have seen in the aforementioned examples, they just came up with case-by-case revisions instead of fixing the root problem of the bias that we see in search engines and translation tools. Thus, it is important to go to the root cause of these AI biases and develop better and more inclusive algorithms in addition to conducting training on diverse data.

Works Cited

- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias. ProPublica. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? debiasing word embeddings. *Advances in neural information processing systems*, 29.
- DeVos, A., Dhabalia, A., Shen, H., Holstein, K., & Eslami, M. (2022, April). Toward User-Driven Algorithm Auditing: Investigating users' strategies for uncovering harmful algorithmic behavior. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1-19).
- Eberhardt, J. L. (2020). *Biased: Uncovering the hidden prejudice that shapes what we see, think, and do*. Penguin.
- Karpa, D., Klarl, T., & Rochlitz, M. (2022). Artificial Intelligence, Surveillance, and Big Data. In *Diginomics Research Perspectives: The Role of Digitalization in Business and Society* (pp. 145-172). Cham: Springer International Publishing.
- Ferrer, X., van Nuenen, T., Such, J. M., Coté, M., & Criado, N. (2021). Bias and discrimination in AI: a cross-disciplinary perspective. *IEEE Technology and Society Magazine*, 40(2), 72-80.
- Howard, A., Borenstein, J. The Ugly Truth About Ourselves and Our Robot Creations: The Problem of Bias and Social Inequity. *Sci Eng Ethics* 24, 1521–1536 (2018). <https://doi.org/10.1007/s11948-017-9975-2>.
- Introna, L. D., & Nissenbaum, H. (2000). Shaping the Web: Why the politics of search engines matters. *The information society*, 16(3), 169-185.
- Kay, M., Matuszek, C., & Munson, S. A. (2015, April). Unequal representation and gender stereotypes in image search results for occupations. In *Proceedings of the 33rd annual acm conference on human factors in computing systems* (pp. 3819-3828).
- Leboeuf, C. (2020). The embodied biased mind. *An introduction to implicit bias: Knowledge, justice, and the social mind*. Routledge.
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.
- Rozado, D. (2023). The political biases of chatgpt. *Social Sciences*, 12(3), 148. <https://doi.org/10.3390/socsci12030148>.
- Steiner, M., Magin, M., Stark, B., & Geiß, S. (2022). Seek and you shall find? A content analysis on the diversity of five search engines' results on political queries. *Information, Communication & Society*, 25(2), 217-241.
- Trielli, D., & Diakopoulos, N. (2022). Partisan search behavior and Google results in the 2018 US midterm elections. *Information, Communication & Society*, 25(1), 145-161.
- Zajko, M. (2021). Conservative AI and social inequality: conceptualizing alternatives to bias through social theory. *AI & SOCIETY*, 36(3), 1047-1056

Bitcoin's Carbon Footprint: Utilizing Artificial Intelligence to Analyze Future Energy Consumption and Environmental Impact By Varsha A, Eesha V

Abstract

Did you know that the daily energy consumption of Bitcoin transactions is equivalent to the electricity usage of over 7 million US households? In this project, we utilized artificial intelligence (AI) models, including Linear Regression, AutoRegressive Integrated Moving Average (ARIMA), and Prophet, to predict the future number of Bitcoin transactions. We used these predictions to analyze the future energy consumption and environmental impact of those transactions. The models were trained on historical Bitcoin transaction data, taking into account whether renewable or non-renewable sources were used to generate electricity to power these transactions. The best-fit model was determined based on statistical performance measures like Root Mean Square Error (RMSE), and we predicted the future Bitcoin transaction volume using the selected model. Based on the predicted transaction volume, we calculated the corresponding energy consumption and environmental impact of Bitcoin transactions. We found that for the 1st quarter of 2023 (from Jan 1st to March 31st), the average daily energy consumption of Bitcoin transactions is 211,652,518 kWh. Additionally, we estimate that 98,178.09 metric tons of CO₂ are released into the environment. This energy consumption is equivalent to the energy consumed by 7,267,045 households in the US per day, and the CO₂ released by these daily Bitcoin transactions is equivalent to the emissions from 21,848 gasoline-powered passenger vehicles driven for one year. Through this project, we have demonstrated the capacity of AI models to predict future trends in the context of Bitcoin transactions and their economic and environmental impacts. Furthermore, we have highlighted the significance of sustainable development and environmental protection within the digital currency ecosystem. The results of this study are valuable for policymakers, researchers, and industry practitioners who are working towards reducing the carbon footprint of Bitcoin transactions and promoting sustainable practices.

Introduction

"Crypto is harming the world," remarked Yasmin Dahnoun, an editor from The Ecologist, on April 12, 2023 (Dahnoun). But what exactly is crypto? Crypto is a virtual currency designed to function as money, known as a cryptocurrency. Bitcoin stands as the most prominent cryptocurrency. Presently, cryptocurrency remains unregulated, obviating the need for third-party involvement in financial transactions, but this also renders the currency somewhat vulnerable as investing in cryptocurrency entails several risks due to the lack of government regulations, exposing the entire network and its investors to fraud and theft (Analytics Insight). The term "crypto" alludes to the usage of cryptographic algorithms to secure transactions. At present, there are 19 million Bitcoins (Blockchain Council) in circulation that have been mined. However, a cap of 21 million Bitcoins (Blockchain Council) exists on the quantity that can be mined. The alleged creator of Bitcoin, Satoshi Nakamoto, conceived the cryptocurrency with a capped supply to restrict the number of Bitcoins available. This limitation engenders tension between

Bitcoin's demand and its price over time. As per the Blockchain Council, the final Bitcoin is anticipated to be mined in the year 2140 (Blockchain Council). Currently, Bitcoin transactions employ computers to solve intricate mathematical problems for transaction verification. This mechanism of transaction verification is energy-intensive, consuming 707 kWh per transaction, equivalent to the power required to charge 37,204 phones (Forbes Advisor; U.S. Environmental Protection Agency). While predictions about Bitcoin's price abound, there is limited research concerning the projection of daily Bitcoin transaction volume and its accompanying environmental impact, which is equally crucial to address. This project contributes to the existing body of research on the environmental repercussions of Bitcoin and aids policymakers, researchers, and industry professionals striving to mitigate the carbon footprint of cryptocurrency transactions while promoting sustainable practices in the realm of virtual currency. In this endeavor, we harnessed artificial intelligence (AI) models to predict the forthcoming volume of Bitcoin transactions. This projected volume was employed to compute future energy consumption, considering the utilization of renewable versus non-renewable energy sources. Through this research, we also scrutinized the impending economic and environmental implications of Bitcoin transactions.

Materials and Methods:

1. Forecast Future Bitcoin Transaction Volume Using AI Models

- a. Collect data: Find data, from a Nasdaq dataset (Nasdaq), on the number of Bitcoin transactions per day from January 2, 2009, until April 15, 2023. Figure 1 shows the daily Bitcoin transactions from 2009 until the end of 2022. The transaction volume was steady for a couple of years until mid-2012 and then it has gradually increased over the years. We can also notice some fluctuations in 2017 and 2018 but overall, there is a steady increase in the transaction volume.

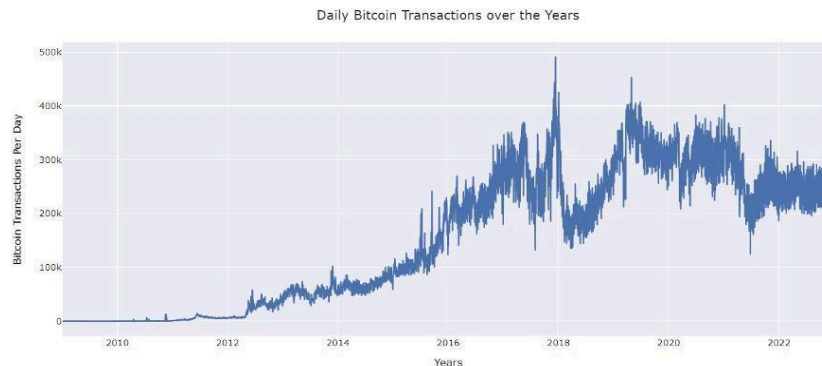


Fig 1: it shows the daily Bitcoin transactions from 2009 until the end of 2022.

- b. Preprocess data: Clean up data by checking for null values and data discrepancies.
- c. Split data: Split the data set into train and test data for the different models.
- d. Train the AI models: Using Linear Regression, AutoRegressive Integrated Moving Averages, and Prophet, train the models using the training data.
- e. Test and evaluate the models: Using test data, see if the model is working as expected. If there are errors, fix them. Figure 2 shows the ARIMA model predictions from December 31st, 2022, until April 1st, 2023. We split our data into 90% for training and 10% for testing. Based on these sets, to see how visually accurate the model is, we gave the same time period for the test predictions (green) as the test set (orange). After seeing that the test predictions were pretty accurate, we decided to forecast daily Bitcoin transactions from December 31st, 2022, to April 1st, 2023, which is shown in the graph as the red color.

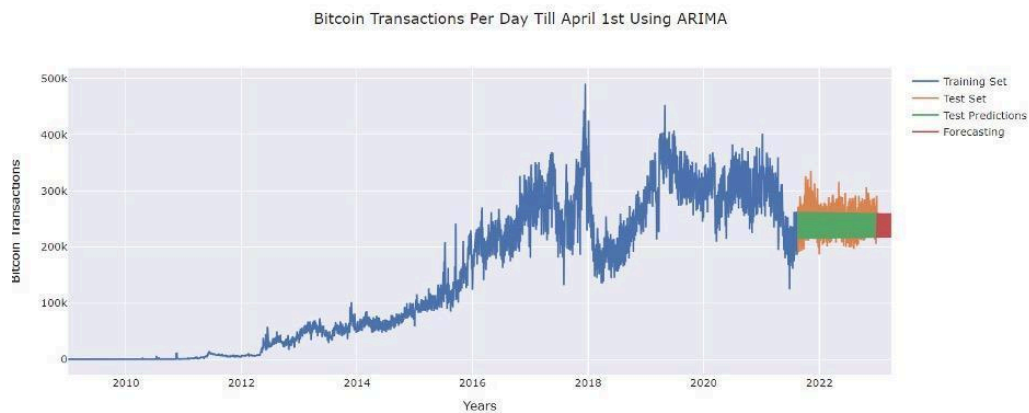


Fig 2: This figure shows the ARIMA model predictions from December 31st, 2022, until April 1st, 2023. The blue color shows the training data set from Jan. 2, 2009, until December 31st, 2022. The orange shows the test set.

- f. Export Data: Export the data from all 3 models into a spreadsheet.
- g. Conduct Analysis: Compare the values for the best-fit model using the Root Mean Square Error (RMSE). In our research models, we found Prophet to be the most accurate in predicting the future Bitcoin transaction values because the RMSE for the Prophet model is the lowest compared to Linear Regression and ARIMA models. We made a graph to show the visual differences between each model's predicted values versus the actual (Figure 3). As you can see on the graph, Linear Regression did not have very accurate predictions as it is a straight line indicating that the transaction volume is constant over the years. Both ARIMA and Prophet started with predictions closer to the actual values

while they were on a short-term prediction date range. However, with long-term date range predictions, there is more variability as the predicted values are farther away from the actuals. This indicates that these models may not be accurate for long-term predictions.

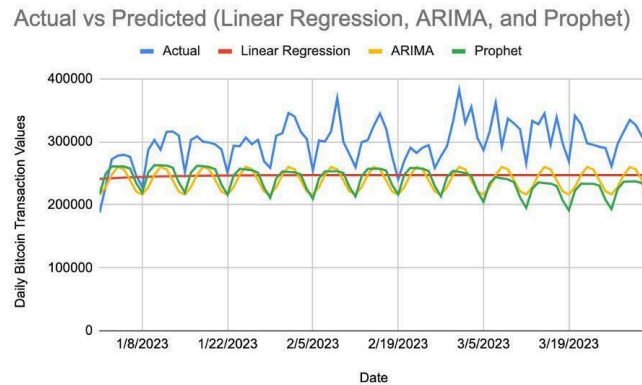


Fig 3: it shows the actual daily Bitcoin transaction values versus the values each of the models predicted.

- h. Analyze best-fit model: The first step in this piece of code (see Figure 4) was to train the model with the daily Bitcoin transaction data. Then we programmed it to predict for 180 days in the future. Next, we used the `MinMaxScaler()` function to transform the actual transaction data and predicted transaction data in the range $[0, 1]$. Our next step was the Root Mean Square Error (RMSE) analysis, which is used to determine the accuracy of the predictions with this model.

```

model = Prophet(
    seasonality_mode="multiplicative"
)
model.fit(train_data)
future = model.make_future_dataframe(periods = 180)
forecast = model.predict(future)

scaler = MinMaxScaler()
scaled_actuals = scaler.fit_transform(df.to_numpy().reshape(-1, 1))
scaled_predictions = scaler.fit_transform(forecast.to_numpy().reshape(-1, 1))
RMSE = np.sqrt(np.mean((np.asarray(scaled_actuals) - np.asarray(scaled_predictions))**2))
print("Root Mean Square Error (RMSE) = ", RMSE)

```

Fig 4: it shows our code for the Prophet ML model.

- i. Forecast: Predict future transaction values using the best-fit model. The first step for arriving at the predictions shown in Figure 5 was splitting our historical data. The data in green shows our training set, or real daily Bitcoin transaction data from Jan. 2, 2009, until Aug. 17, 2021. The data in pink shows the testing set, or the data that we used for testing the Prophet model predictions. The data in blue shows our test predictions, or the number of

daily Bitcoin transactions predicted by the Prophet model that we trained. The data in red is our forecasted daily Bitcoin transactions. As you can see on the graph, Linear Regression did not have very accurate predictions as it is a straight line indicating that the transaction volume is constant over the years. Both ARIMA and Prophet started with predictions closer to the actual values while they were on a short-term prediction date range. However, with long-term date range predictions there is more variability as the predicted values are farther away from the actuals. This indicates that these models may not be accurate for long-term predictions.

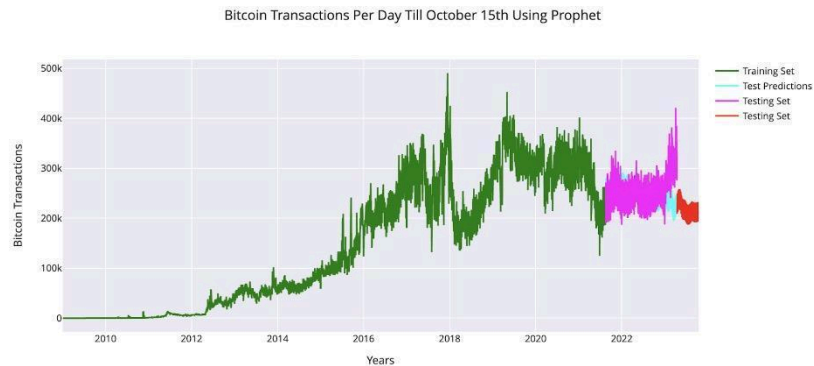


Fig 5: shows the Prophet model predictions for daily Bitcoin Transactions from Apr. 15, 2023, until Oct. 15, 2023.

2. Calculate Economic Impact: Calculate the estimated cost of electricity consumption using renewable and non-renewable energy sources for the future forecasted daily Bitcoin transactions.
 - a. Find the average cost per kWh for six different energy sources. We used gas, oil, and coal for our non-renewable sources. We used solar, hydro, and wind as our renewable sources (International Energy Agency).
 - b. Create a table (Figure 6) and input all the costs. Using the predicted number of Bitcoin transactions per day (X) calculated from step 1h and the amount of electricity needed for a single Bitcoin transaction (Forbes Advisor), we calculated the estimated amount of electricity needed per day for the Bitcoin transactions. The different rows of data depend on the percentage of usage of renewable to non-renewable energy. We start by using the current percentage split (39:61%) and graduate to a balanced 50:50% split. Our findings show that the current split demands less money for daily Bitcoin transactions but more trees to offset the CO_2 emissions. However, a balanced percentage split requires more money but fewer trees.

Economic vs Environmental Impact of Bitcoin Transactions						
Percentage of electricity coming from Renewable vs Non-Renewable	Total cost of generating electricity for daily Bitcoin transactions from renewable source	Total cost of generating for daily Bitcoin transactions from non renewable source	Total cost of electricity consumed by Bitcoin transactions per day	Cost of electricity per Bitcoin transaction	Amount of CO2 released per day (in metric tonnes)	Equivalent # of trees to be planted to offset the emissions
39:61	\$5,579,819.06	\$5,733,068.03	\$11,312,887	\$44.54	83,304.79	3,332,192
40:60	\$5,722,891.34	\$5,639,083.31	\$11,361,975	\$44.73	82,021.72	3,280,869
41:59	\$5,865,963.62	\$5,545,098.59	\$11,411,062	\$44.92	80,738.65	3,229,546
43:57	\$6,152,108.19	\$5,357,129.14	\$11,509,237	\$45.31	78,172.51	3,126,900
45:55	\$6,438,252.76	\$5,169,159.70	\$11,607,412	\$45.70	75,806.37	3,024,255
50:50	\$7,153,614.17	\$4,699,236.09	\$11,852,850	\$46.66	69,191.01	2,767,641

Fig 6: it shows a table that we made depicting environmental and economic data about predicted daily Bitcoin transactions.

- c. Split the estimated amount of electricity into a certain percentage for non-renewables and the remainder for renewables. Originally, we used 39% for renewables and 61% for non-renewables as that is the current split in the United States (U.S. Department of Energy). We also used a 40:60, 41:59, 43:57, 45:55, and 50:50 percentage split to compare the values.
 - d. We took the average cost per kWh for each of the three renewable and each of the three non-renewable sources. We added them to get one average cost per kWh for renewable and one for non-renewable. Assuming that all energy sources were used in equal proportions, we multiplied the average costs by the electricity consumed from that source. We added those two numbers together to get the total cost of electricity consumed by daily Bitcoin transactions (Y). Finally, we divided the sum (Y) by the number of daily Bitcoin transactions (X) to get the cost of electricity for a single Bitcoin transaction.
3. Calculate Environmental Impact: Calculate the estimated CO₂ emissions using renewable and non-renewable energy sources similar to the calculations done in 2a - 2d.
 4. Compare and analyze the cost and environmental impact by changing the percentage split values as shown in step 2c.

Results

In our study, we found that the amount of electricity needed for daily Bitcoin transactions is 211,652,518 kWh. If we multiply that number by 365, it is equivalent to 60.8% of Virginia's annual electricity consumption (U.S. Department of Energy; JustinTools.com).

Economic

We extracted various economic outcomes from our study. For instance, we determined that the electricity cost per Bitcoin transaction rises when employing higher percentages of renewable sources in comparison to non-renewable sources of energy. At the current 39:61 ratio, a transaction costs \$44.54, which increases to \$46.66 with a 50:50 ratio. Thus, the electricity cost per individual Bitcoin transaction equates to the cost of about 14 gallons of milk in Virginia (Zippia).

Environmental

One noteworthy environmental result stemming from our study is that Bitcoin transactions produce greater CO₂ emissions when using higher percentages of non-renewable sources compared to renewable sources of energy. Employing a 39:61% ratio, the split between renewable and non-renewable energy used in current US Bitcoin transactions, we computed that 137,158 trees are necessary to offset CO₂ from renewable energy sources (8 Billion Trees), and 4,199,580 trees are required for non-renewable energy sources (8 Billion Trees). Remarkably, as the balance between renewable and non-renewable energy sources becomes more equal, transitioning to a 50:50% split, the number of trees needed to offset total CO₂ emissions decreases by over half a million.

Discussion

In our project, we underscored the importance of sustainable development and environmental protection within the digital currency ecosystem. This project adds to the existing body of research on the environmental implications of Bitcoin, assisting policymakers, researchers, and industry practitioners in addressing the carbon footprint of cryptocurrency transactions, and promoting sustainable practices in the realm of virtual currency. While predictions on Bitcoin's price abound, there is limited research concerning the projection of daily Bitcoin transaction volume and its corresponding environmental impact, which is equally important to discuss.

Energy Source Impact on CO₂ Emissions

Bitcoin transactions reveal a distinct connection between the energy source used and the resulting CO₂ emissions. Transactions using higher proportions of non-renewable sources exhibit elevated CO₂ emissions compared to those utilizing renewable sources. Interestingly, a trend emerges wherein the number of trees required to offset total CO₂ emissions decreases by over half a million as the balance between renewable and non-renewable sources approaches parity. These findings underscore the intricate interplay between Bitcoin transactions, energy consumption, economic costs, and environmental repercussions. The implications of these results extend to energy policies, economic considerations, and environmental sustainability efforts.

Assumptions and Limitations

While our research offers valuable insights, it's essential to acknowledge its limitations. Future studies could delve deeper into the complexities of optimizing energy mixes, exploring ways to enhance and extend the sustainability of Bitcoin transactions without compromising functionality, economic viability, or environmental impact. In this context, sustainability involves striking a balance between effective Bitcoin transactions and environmental preservation. Using energy sources with lower pollution potential to power Bitcoin could contribute to Earth's health.

As we progressed with this project, we made several assumptions. In our economic research, for instance, we assume that energy sources are used in equal proportions, thus

calculating average costs. Additionally, factors like new regulations, Bitcoin's price, the stock market, supply, and demand, etc., are not considered when forecasting the daily number of Bitcoin transactions. Our summary table of economic impact assumes 254,015 Bitcoin transactions per day, as the ML model with the lowest RMSE, Prophet, projected 254,015 transactions on October 15th, 2023. Furthermore, we chose to forecast daily Bitcoin transactions only six months ahead, not long-term, due to the numerous variables impacting volume, which reduces the accuracy of long-term predictions.

Challenges

One challenge in this project related to an ML model we intended to employ. Long Short-Term Memory (LSTM) is highly effective for modeling sequential data but demands extensive hyperparameter tuning, which is beyond our expertise in ML models. Consequently, we focused on alternative models.

Future Research

Numerous avenues exist for expanding our research. In the future, we aim to analyze different cryptocurrencies besides Bitcoin, such as Ethereum, Tether, etc., and assess their environmental impact. We would also explore the utilization of diverse machine learning algorithms, such as LSTM, to ascertain if different AI algorithms could predict future Bitcoin transaction values more accurately. Subsequent research could investigate strategies for optimizing energy sources in Bitcoin transactions to mitigate environmental impact while managing costs.

Conclusions

The rise of digital currency began with Bitcoin's launch in 2009 and has taken the world by storm. However, in recent years, the focus has shifted to Bitcoin's price speculation and the pronounced volatility characterizing this popular cryptocurrency. The widespread adoption of this digital currency has led to increased demand and amplified transaction volumes. Each of these transactions consumes a substantial amount of electricity because each cryptocurrency utilizes a network of computers through Blockchain technology to solve complex problems. Depending on the sources of electricity used to meet the power demand of these crypto transactions, the rapid increase in volume potentially hampers broader efforts to achieve net-zero carbon pollution commitments by different nations. Bitcoin transactions can have several negative impacts on local communities due to pollution, noise, and even criminal activity, often facilitated by the encrypted nature of Bitcoin. Similar to the patterns observed a few decades ago with the surge in credit card usage and more recently with the adoption of payment apps such as Venmo and Apple Pay, we believe cryptocurrency has the potential to become a mainstream form of currency in the foreseeable future. If so, the number of daily crypto transactions will be orders of magnitude greater than what we have shown in this paper. Given that climate change stands as one of the consequential challenges of our era, we urge scientists, researchers, experts,

and other stakeholders to persistently delve into the environmental repercussions linked to Bitcoin. This research paper underscores the importance of policymakers considering incentives for energy-efficient practices, while researchers could further expand upon this groundwork by investigating novel energy solutions tailored to the digital economy toward a greener future.

Works Cited

- 8 Billion Trees. How Many Trees to Offset 1 Ton of CO₂? Available from:
<https://8billiontrees.com/carbon-offsets-credits/how-many-trees-to-offset-1-ton-of-co2/>
- Analytics Insight. 10 Typical Risks That Cryptocurrency Investors Encounter. Available from:
<https://www.analyticsinsight.net/10-typical-risks-that-cryptocurrency-investors-encounter/#:~:text=The%20major%20problem%20with%20bitcoin,investing%20in%20cryptocurrencies%20in%202023.>
- Blockchain Council. How Many Bitcoins Are Left? Available from:
<https://www.blockchain-council.org/cryptocurrency/how-many-bitcoins-are-left/#:~:text=Summary%E2%80%8B,million%20left%20to%20be%20mined.>
- Dahnoun Y. Crypto is killing the planet. The Ecologist. 2022 Oct 26. Available from:
<https://theecologist.org/2022/oct/26/crypto-killing-planet>
- Forbes Advisor. Bitcoin's Energy Usage, Explained. Available from:
<https://www.forbes.com/advisor/investing/cryptocurrency/bitcoins-energy-usage-explained/#:~:text=In%20fact%2C%20Bitcoin%20uses%20707,make%20it%20an%20easy%20scapegoat.>
- International Energy Agency (IEA). Projected Costs of Generating Electricity 2020. Available from:
<https://www.iea.org/reports/projected-costs-of-generating-electricity-2020>
- JustinTools.com. Energy Unit Conversion. Available from:
<https://www.justintools.com/unit-conversion/energy.php?k1=kilowatt-hour&k2=terawatt-hour>
- Nasdaq. Bitcoin Number of Transactions. Available from:
<https://data.nasdaq.com/data/BCHAIN/NTRAN-bitcoin-number-of-transactions>
- U.S. Department of Energy. Energy Sector Risk Profile. Available from:
https://www.energy.gov/sites/prod/files/2016/09/f33/VA_Energy_Sector_Risk_Profile.pdf
- U.S. Environmental Protection Agency (EPA). Greenhouse Gas Equivalencies Calculator. Available from:
<https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator>
- Zippia. How Much Does a Gallon of Milk Cost in Each State? Available from:
<https://www.zippia.com/advice/gallon-of-milk-costs-each-state/>

Navigating the Landscape and Ethical Considerations of AI-Generated Visual Content

By YunLe (Hanbin) Wang¹ and Brian Huh²

Abstract

In recent months, Artificial Intelligence (AI) has steadily risen through generative AI such as ChatGPT (generative engine). This engine is capable of creating plays and essays in the style of a specific author while talking and writing in Shakespearean dialect. Meanwhile, other generative AI engines have shared the mainstream spotlight with ChatGPT. Artificial networks like Dall-E, Midjourney, Planfinder AI, and even Photoshop's "Generative Fill" have become more advanced and popular in society today. These artificial engines all share a commonality in their relationships to art and design. They can learn from artworks (whether that be style or content) and create hyperrealistic images from human text prompts. However, the widespread integration of AI into society warrants discussions and debates about many aspects of Artificial Intelligence and whether the images it creates can be deemed as "art."

Introduction

It has often been said that language was derived from cave art. Subsequently proving that art was one of man's earliest creations. Art has evolved and adapted to the constantly changing world and has often defined generations. For example, the Renaissance is defined by its art and artists similarly to how the late 20th century has been defined with its challenge towards traditional art. This challenge towards art has caused many questions like *What is art?* and *what makes art art?* These questions were brought into mainstream media with pieces like Marcel Duchamp's *Urinal* or his other readymade sculptures. His art pushed the boundaries as to what can be defined as art. Similarly, Chihuly, an infamous glass sculptor, has often been the center of debate for who deserves credit for art. Chihuly designs his sculptures but allows glass artisans to create said sculptures; should Chihuly receive the credit or should the artisan (Johnson 2017)? These questions were further brought into question with the growing popularity of artificially generated images.

Artificially generated images first began in the 1970s with AARON. Harold Cohen, an American computer scientist, would use this machine to generate abstract pieces of "artwork.": For example, he would tell it to paint blue lines, and blue lines would appear on a screen. Generative images took a huge leap in the 2010s with Generative Adversarial Networks (GANs). It allowed for the prompts to be more complicated, and it also began learning from previous art styles. Although it is hard to find the first AGI, the first AGI that sold was a *Portrait of Edmond de Belamy*, which was sold in 2018 at around 430 thousand dollars. Some see it as a step towards a more expansive definition of art, while others see it as a setback for artists around the world.

The sale of an AGI brings forth many questions regarding art. It references the question brought up by Duchamp (What is art?) and the question brought up by Chihuly (Who deserves the credit). An observer can argue both ways to either of these questions in their respective categories, but can humans do the same when it deals with a machine instead of a human artist?

Should the credit for the art be allocated to the person who came up with the prompt or the machine? How would credit even be allocated to the machine? Should the artist's artwork that the Machine learned off of be mentioned? How can human artists compare to the speed of GAN Machines?

Through interviews and conversations with different artists and art collectors, they stated their different opinions regarding AGI. Most of these artists stated that their opinions on AI Art were complicated. Some believed that AGI was truly art, while others believed that it was a new category of art that cannot yet be defined. These sentiments were largely similar to those in reaction to Duchamp's Readymade sculptures. Some artists expressed their concerns about the potential of Artificial Intelligence potentially being able to overtake jobs in design like architecture and 3D rendering. However, other artists have embraced this change and development with Artificial Intelligence and have incorporated some aspects in their own art through Photoshop or 3D modeling software. In conclusion, it is fair to say that all of these artists have adapted to AGI in different and personal ways.

The emergence of AI-generated art prompts important philosophical questions about the nature of creativity and authorship. While machines can now produce novel images, their capability for true creativity and imagination remains limited. The AI system is ultimately dependent on its human designers and the data it is trained on. Though striking, AI art lacks a human spirit or intent behind it. At the same time, the role of the prompt engineer in directing the output challenges romantic notions of artistic genius emerging *ex nihilo*. Moving forward, we must strike a balance that respects both human creativity and technological innovation. AI art is neither a radical break from the past nor the death knell of human artistic expression. Rather, it represents a new frontier prompting us to reexamine our assumptions about art in light of advancing technology. There are challenges but also opportunities for collaboration between humans and machines. The future will likely see both human and AI artists thriving as long as we positively guide progress in a way that benefits society.

Analysis

A. Areas of Art Impacted by Artificially Generated Visual Content

In terms of artificially generated art, there are many subtopics. For example, music and videography have become a field of art that has become more and more touched by AI. Drake and Taylor Swift's songs have been generated and released online. Some hardcore fans have had difficulties distinguishing these artificially generated songs compared to real music. Similar to music, subjects like poetry can be generated in the style of authors like Shakespeare or Edgar Allen Poe about any subject. However, this development also reinforces the notion that AI needs these original human examples to learn from. For example, an Edgar Allen Poe-styled essay would cease to exist if Edgar Allen Poe, as a human writer, did not have that style himself (iScience 2020). The notion of replicating art forms or literature styles also brings up the debate that these styles cannot truly be accurate due to the wide range of art styles a single artist could have. For example, even though Van Gogh is often known for his expressionist paintings and

swirling pieces, it is difficult to pinpoint him to one specific style due to his vast mastery of the arts (Manovich 2019).

This advancement in the generation and replication of writing and music styles has also led to discussion over other generated videos or audio clips. These “deep fakes” can be created for famous people like the President of the United States, which could potentially lead to very dangerous situations or misunderstandings. These fake videos could cause discord or panic in many people resulting in problems on a large scale.

However, AI-generated art can also lead to positive outcomes. It fosters a level of creativity and expression that hasn’t happened before at speeds unachievable by man (D’Inverno, McCormack 2015). Artificial Intelligence and its generated images should, therefore, be seen as a complement to human creativity and thought (Kodithuwaku, 2023). Artificial intelligence cannot create thoughts by itself and requires human input in order to create. Similarly to Chihuly, the artisans who create these sculptures would not have a design to create if it were not for Chihuly. However, this once again brings back the debate of who should receive credit for an Artificially Generated Image. Should it be the human who provided the prompt? Or the machine or engine that created the final result? Or should it even be allocated to the artist that the machine learned from?

Aside from simply generative functions, the machine learning portion of Artificial Intelligence also leads to other results. For example, due to its fast learning speeds, it can analyze and organize past arts into specific categories and forms. For example, art history can be simplified as a piece of art can be fed into an AI system and it can instantly analyze the time period, author, or style of the painting and categorize it accordingly. There are obvious errors, but an AI performing this task can be more accurate than a human since AI takes a more machine process than a human being able to analyze a piece of art in many different ways.

B. Current Knowledge and Research on Artificially Generated Visual Content

The area specifically regarding Artificially Generated Images and its authenticity hasn’t been officially researched as it is still a new topic, however, there has been research into how humans perceive AI-generated images. For example, research has shown that when people know a piece of art was generated by AI, they decrease its value due to its perceived inauthenticity (Cojean, Martin, Ragot 2020). Similarly, lots of people and artists see art as a statement or depiction of a current time in history. For example, Keith Haring stated that “an artist is a spokesman for a society at any given point in history [and] his language is determined by his perception of the world we all live in” (Haring). An artificially generated image cannot accomplish this level of reflection in the world; however, someone who gave the machine prompts to generate specific images could. Subsequently, this nags at the question of who should receive credit for artificially generated images.

Artificially Generated Images are also not self-sufficient. Take the first sold GAN Image, for example, *Edmond De Balemey*. This artificially generated artwork could not have been generated if it did not have previous human art to learn from. Similarly, that piece of art could

not have been auctioned or sold without human intervention and help. In the case of this piece of art, the money went to the auction house, but what if an individual decides to sell an artificially generated image? Does the credit get allocated to the individual who sold the image? The computer who created the image? The artist who created the artwork that the machine learned of?

This debate could similarly be seen in generated music. It is a form of human expression, but at the end of the day, the generated art is just an algorithm. More specifically, it is an algorithm that is often looked down upon by people as people tend to favor realistic and human expression compared to machine-generated images (Hong, Curran, 2019). In terms of the most effective and uncontroversial integration of AI and art, art analysis is most likely to be the least debated. The parameters and categories are set by humans. Art analysis and categorization is the softest integration of AI in the field of art. However, AI remains one of the more debated utilities in our modern era.

C. Impact of Artificially Generated Visual Content

In the future, if AI Art is truly to become more widespread and continue growing as it is now, there needs to be sufficient regulations and boundaries set in place. Although questions like “What is art?” may never be defined, the question of what AI Art is must be listed in order to protect a creative future. The thin line between creativity and plagiarism for AI art must be clearly defined to protect human artists from a competitor that can produce art at speeds impossible for humans. The allocation of funds for AI Art sold also needs to be regulated and created to ensure there are no debates and/or conflicts over credit allocation.

If a writer uses AI to generate sentences or to help revise, should the AI Component be recognized? If an artist uses AI to render some of his/her ideas, should AI be given full credit? If an architect uses AI to create floor plans, who should receive credit? Even if these questions were not applied to AI, artists like Chihuly could also warrant these discussions about whether his art is truly his art, considering that he has different workers and artisans who create his models.

AI Art also brings forth another concern about jobs. For example, those working in the film industry often have to hire artists to help render thumbnails for scenes in a movie. However, if AI can accomplish this task at a faster and cheaper rate, it might put many people out of the workforce. Artists in general, might lose at jobs due to the fierce competition of AI being able to quickly create original artworks that can be influenced by any artistic time period in history.

Conclusion

Artificially generated visual content is having a profound impact across many areas of art and design. From music and video to visual art and graphic design, AI systems are demonstrating increasing capabilities to produce novel and convincing artistic works. However, significant philosophical questions remain regarding the true creativity of these systems and the appropriate attribution of credit.

While AI art presents opportunities for collaboration and expanded creative possibilities, ethical guidelines and regulations will need to be established regarding issues of plagiarism, labor displacement, and financial compensation. Moving forward, the role of AI in art should be oriented toward complementing and augmenting human creativity rather than fully automating or replacing it. By thoughtfully shaping the development and application of these technologies, humans can retain artistic agency and meaning while also exploring new creative frontiers with artificial intelligence. This will require an open and ongoing dialogue between technology developers, artists, philosophers, and policymakers. If guided responsibly, AI promises to be an inspiring new partner in humanity's eternal quest for beauty and self-expression.

Limitations

While providing a broad overview of key questions surrounding artificially generated visual content, the current analysis has focused predominately on the visual arts. Examining implications across other artistic domains, including music, videography, writing, and performance, represents an area for expanded scholarly inquiry. Additionally, limited empirical data exists on public attitudes and emotional reactions to different forms of AI art as compared to human-generated creative works. As advancements in neural-net creativity systems accelerate, further research must address these gaps.

Future Directions

Moving forward, cross-cultural examinations of perceptions towards non-human creative expression would provide greater insight into philosophical themes of authorship and imagination. Studying how revealing the AI origins of an artwork shifts its reception and value also represents a rich area for future psychology-focused experiments. Moreover, the proliferation of AI art compels efforts to model and quantify potential economic impacts across artistic professions and industries. These interdisciplinary projects call for blended methodologies synthesizing technology criticism, analytical philosophy, cultural anthropology, and econometrics. Furthermore, constructing ethical frameworks to guide the trajectory of AI art will require expanded dialogues between computer scientists, artists, philosophers, policy analysts, and legal experts—conversations that can yield recommendations for balancing creative traditions and technological progress. Identifying these future directions will drive vital scholarship and policy discussions regarding artificial intelligence's growing influence across an expanding range of artistic exploration.

Author Affiliations

YunLe (Hanbin) Wang¹, Yorba Linda High School

Brian Huh², Marshall School of Business, University of Southern California

Works Cited

- Browne, K. (2022). Who (or What) Is an AI Artist? *Leonardo*, Vol 55 (Issue 2), pp 119-124.
https://doi.org/10.1162/leon_a_02092
- Cetinic, E., She, J. (2022). Understanding and creating Art with AI: Review and Outlook. *ACM Transactions on Multimedia Computing, Communications, and Applications*, Vol 18 (Issue 2), pp1-22. <https://doi.org/10.1145/3475799>
- d’Inverno, M., McCormack, J. (2015). Heroic versus Collaborative AI for the Arts. *24th International Joint Conference on Artificial Intelligence (IJCAI)*.
<http://ijcai.org/papers15/contents.php>
- Daniele, A., Song Y. (2019). AI + Art = Human, *AIES ‘19: Proceedings of the 2019 AAI/ACM Conference on AI, Ethics, and Society*, pp 155-161.
- Elgammal, A. (2019). AI Is Blurring the Definition of Artist: Advanced algorithms are using machine learning to create art autonomously. *American Scientist*, Vol 107 (Issue 1), 18+.
<https://link.gale.com/apps/doc/A579092374/AONE?u=anon~92b997b7&sid=googleScholar&xid=e6485d17>
- Epstein, Z., Levine, S., Rand, D., Rahwan, I. (2020). Who Gets Credit for AI-Generated Art? *iScience*, Vol 23 (Issue 9), <https://doi.org/10.1016/j.isci.2020.101515>
- Hong, J., Curran, N. (2019). Artificial Intelligence, Artists, and Art: Attitudes Toward Artwork Produced by Humans vs. Artificial Intelligence. *ACM Transactions on Multimedia Computing, Communications, and Applications*, Vol 15 (Issue 2), pp 1-16.
<https://doi.org/10.1145/3326337>
- Huang, Z. (2022). Analysis of Text-to-Image AI Generators. *IPHS300 AI for the Humanities*. Retrieved on October 22, 2023, from
- Johnson, K. (2017, August 21). Who Is Really Making ‘Chihuly Art’? *The New York Times*. Retrieved October 22, 2023, from
<https://www.nytimes.com/2017/08/21/arts/design/chihuly-glass-bipolar-court-moi.html>
- Kodithuwakku, G. (2023, February 28). *AI in Art and Creativity*. Medium. Retrieved October 22, 2023, from
<https://medium.com/nerd-for-tech/ai-in-art-and-creativity-25a71d4341db#:~:text=AI%2Dgenerated%20art%20is%20that%20it%20allows%20for%20the%20creation,images%20in%20a%20unique%20way>.
- Manovich, L. (2019, June). Defining AI Arts: Three Proposals. Retrieved October 22, 2023, from
<http://manovich.net/content/04-projects/107-defining-ai-arts-three-proposals/manovich.defining-ai-arts.2019.pdf>
<https://doi.org/10.1145/3306618.3314233>
- Nightcafe. (2022, July 5). What Is the First AI Art And When Was it Created? Retrieved October 22, 2023, from
<https://nightcafe.studio/blogs/info/what-is-the-first-ai-art-and-when-was-it-created>
- Ragot, M., Martin, N., Cojean, S. (2020). AI-generated vs. Human Artworks, A Perception Bias Towards Artificial Intelligence? *CHI EA ‘20: Extended Abstracts of the 2020 CHI Co*

Comparing Intrinsic and Extrinsic Motivation: An Overview of Experiments and Future Areas of Research By Juwon Hwang

Abstract

Motivation affects our daily lives, and it plays a pivotal role in one's behaviors towards a goal. Therefore, several scholars have studied two important types of motivation: intrinsic and extrinsic motivation. This paper will go over the definitions and the effects of intrinsic versus extrinsic motivation and provide an overview of the key experiments related to intrinsic and extrinsic motivation. Finally, areas that require more research will also be discussed.

Keywords: Intrinsic motivation, Extrinsic motivation, Self-Determination Theory

Introduction

Why is it that waking up in the morning, sticking to a workout routine, and doing homework is hard for some people while it is easy for others? This is because of the different levels of motivation they have. Motivation refers to a condition that triggers an action toward an objective (National research council, 1999). Scholars, therefore, have studied and categorized motivation in different ways. One of the most widely studied types of motivation are intrinsic motivation and extrinsic motivation. For instance, while intrinsic motivation is led by one's genuine interest or internal goals, such as the promotion of health, extrinsic motivation is caused by external goals, such as physical attractiveness, social recognition, and grades. Based on studies related to motivation, we can advise coaches to use effective measures to motivate athletes, encourage people to exercise, and provide people with an understanding of how they can motivate themselves.

Therefore, in this paper, I will discuss the scholarly definition and the effect of intrinsic vs. extrinsic motivation. Furthermore, I will give an overview of the experiments that were conducted on the effects of intrinsic versus extrinsic motivation. Specifically, I will be focusing on the effect of intrinsic vs. extrinsic motivation in an academic learning setting, in the context of promoting one's health, and in a job environment. Finally, I will discuss the potential areas of research including specific research topics that could be studied in the future.

Definition of Extrinsic and Intrinsic Motivation

Human behavior can be characterized by various motivations. Various scholars have researched the different motivations behind human actions and some of the most widely studied motivations are internal and external motivation. Intrinsic motivation refers to motivation based on one's internal needs (Lin et al., 2003). Therefore, it is based on a desire to pursue an activity that does not entail external rewards (Lepper et al., 1973). For instance, intrinsic motivation could include a genuine interest in the subject or a sincere desire to promote health. The other type of motivation extensively studied by scholars, extrinsic motivation, is based on an outward orientation (William et al., 2000). Extrinsic motivation is triggered by rewards or punishments

(Lin et al., 2003). For example, actions done by external factors, like financial benefits or social recognition could be considered as actions caused by extrinsic motivation.

Effect of Extrinsic vs. Intrinsic Motivation

Many scholars have studied intrinsic vs. extrinsic motivation in relation to Self-Determination Theory. According to Ryan and Deci, Self-Determination Theory explains that all humans have three basic psychological needs—autonomy, competence, and relatedness—that control growth and evolution (2017). Scholars have shown that intrinsic motivation satisfies the basic needs of autonomy, competence, and relatedness. For example, intrinsic motivation could help to make people feel competent by pushing them to pursue challenges. Lin et al., also point out that intrinsic motivation increases competence and autonomy because intrinsic motivation leads to “interest in the subject matter, enjoyment of challenge, or a sense of making progress, and increasing mastery” (Lin et al., 2003, p. 252). Also, if one has an internal motivation to contribute to the community, they will satisfy their needs for relatedness, because they would genuinely be concerned about other people and empathize with other individuals. In essence, intrinsic motivation reflects people’s natural growth tendencies, develops satisfying relationships, strengthens talents and potential, helps people connect with the larger world, and maintains a sense of physical fitness and health (Vansteenkiste et al., 2007).

However, extrinsic motivation does not fully satisfy the basic needs defined by Self-Determination Theory. In fact, extrinsic motivation may even deter an individual from satisfying their basic needs. For instance, it might deter individuals from satisfying their competence by promoting a “narrow-focused” approach to their goals (Vansteenkiste et al., 2005, p.774). Furthermore, extrinsic motivation is limited to the person’s engagement and ambitions towards that goal. Therefore, there is a limitation to how much individuals will push themselves for a certain goal compared to people with intrinsic goals. Unfortunately, extrinsic motivation may promote shallow learning because extrinsic goals distract people from the task and drive people’s attention to the external factors that they are worried about (Vansteenkiste et al., 2005). Moreover, extrinsic motivation doesn't satisfy autonomy because extrinsic motivation is based on extrinsic goals imposed by other people, weakening one’s willingness for self-determination and choices (Deci & Ryan, 1991). Lastly, extrinsic motivation does not satisfy relatedness because extrinsic motivation leads people to objectify other people they interact with (Vansteenkiste et al., 2005). In essence, people who are influenced by extrinsic goals have lower life satisfaction, lower self-actualization, and higher anxiety (Kasser & Ryan, 2001).

Experiments Related to Extrinsic vs. Intrinsic Motivation

Intellectual Tasks

One of the first studies that were conducted related to intrinsic and extrinsic motivation is a study conducted by Dr. Harlow. Dr. Harlow placed puzzles in cages and observed monkeys’ behaviors. Even without any extrinsic reward, the monkeys voluntarily played with the puzzles

and eventually learned how to crack the puzzle (Harlow, 1950). However, when the experimenters introduced raisins and started to reward monkeys using raisins on a different group of monkeys, the monkeys demonstrated poor performance in the puzzle-solving task (Harlow, 1950). Other scholars have investigated the effects of extrinsic vs. intrinsic motivation on academic learning tasks. For instance, a study conducted by Vansteenkiste et al., indicates that intrinsic goal framing enabled students to do better in conceptual learning compared to those with an extrinsic goal. The study showed that the harmful effect of extrinsic goal framing on conceptual learning is a result of students focusing more on social image and recognition rather than learning about the conceptual material (2005). This attitude led them to get distracted from the activity itself. Also, the results indicate that extrinsic goal framing is not effective in helping students retain a conceptual learning topic for a long term (Vansteenkiste et al., 2005). Furthermore, a study conducted by Lin et al., showed similar results (2003). In their experiment, the researchers used questionnaires to measure students' intrinsic and extrinsic motivation using separate items for both intrinsic and extrinsic motivation. Specifically, students were asked to rate the following statements to measure their intrinsic motivation: "I prefer work that is challenging" and "I prefer coursework that arouses my curiosity, even if it is difficult" (Lin et al., 2003). To measure students' extrinsic goal motivation, they were asked to rate the degree to which they agree with the following statements: "Getting good grades is my main goal for this course" and "A good grade in this course is more important to me than what I learn from this course" (Lin et al., 2003). The results confirmed that students who had high intrinsic motivation and medium level of extrinsic motivation had better grades than those who had low or high extrinsic motivation (Lin et al., 2003).

Fitness & Wellness

Moreover, the effect of intrinsic and extrinsic motivation was studied in the context of fitness and wellness related behaviors. Vansteenkiste et al. conducted an experiment on fourth to sixth grade secondary school students (2004). The participants were given either an extrinsic or intrinsic motivation before learning an Asian sport, tae bo. Specifically, students in the intrinsic motivation group were told that tae bo was good for their health. On the other hand, the extrinsic motivation group was told that tae bo helped them remain attractive to others. The researchers predicted that intrinsic goals would produce better performance due to its closer relationship with basic psychological needs such as autonomy, competence, and relatedness. The results aligned with their prediction: people in the intrinsic motivation group performed better than those in the extrinsic motivation group (Vansteenkiste et al., 2004). Also, the students demonstrated more persistence and voluntariness towards the activity, tae bo (Vansteenkiste et al., 2004). Another study conducted by Gagne and Deci showed a similar result. The researchers assigned grade 5s and 6s with either intrinsic (promoting health) or extrinsic motivation (physical attractiveness) after giving them a book that promoted physical exercise. They checked with the students after 10-weeks, 30-weeks, 1-year and 2 years. The result showed that the students' eating habits were positively affected by the task regardless of the type of motivation they were assigned to.

However, in the long run, the results were different. Students with intrinsic motivation had kept the same behavior for 2 years while students who were assigned with extrinsic motivation had not continued to improve their eating habits.

Job Environment

Lastly, intrinsic and extrinsic motivation also affects people's actions in the job environment. For instance, employees who were assigned intrinsic motivation showed better job performance compared to those who had extrinsic motivation (Broeck et al., 2008). Furthermore, a study by Malik et al. on 220 employees in a company also showed similar results (2019). The results indicated that intrinsic motivation increased radical creativity, a set of ideas that drastically differ from existing ideas, but not incremental creativity, a set of ideas that are based on minor changes and improvements of current practices. However, this effect was moderated by the personal characteristics of the employees. Employees with high learning goal orientation were significantly affected by intrinsic motivation while employees with low learning goal orientation were not as greatly affected by intrinsic motivation. On the other hand, extrinsic motivation increased incremental creativity while it did not increase radical creativity. Specifically, extrinsic motivation increased incremental creativity when the person had a high-performance goal orientation.

Future Areas for Research

There are several topics related to intrinsic versus extrinsic motivation that could be researched in the future. One potential area of research could be the effect of intrinsic versus extrinsic motivation on different types of sports. For instance, we can examine the effectiveness of the two motivations in a team sports environment versus an individual sports environment. I would hypothesize that extrinsic motivation would be more effective in a team sports environment compared to an individual sports environment because an extrinsic goal allows players to work towards the same goal, improving communication and teamwork among players. On the other hand, for individual sports, intrinsic motivation would lead to more successful outcomes because players are working towards goals that will satisfy their own needs and do not have to share the goal with other people. Also, we can study the different implications of intrinsic and extrinsic motivations in performance-based sports such as dancing versus result-based sports such as soccer. I would predict that for dancing, extrinsic motivation would be more effectual because the goal of the sport is to impress the judges, which is an external goal.

Furthermore, more research related to specific circumstances that may alter the effect of intrinsic and extrinsic motivation should be conducted. For example, if there is a ballet dancer that needs to earn money for a living, would extrinsic motivation be a better way to promote performance?

Another area of future research could be about whether the entities that give the intrinsic and extrinsic motivation moderate the efficacy of the motivation. Some questions we might ask

include: “Can the personality of a coach or a teacher influence the impact of extrinsic goals?” For instance, for a harsh coach, extrinsic goal promotion may be more beneficial than intrinsic goal promotion because extrinsic goals promoted by harsh coaches can induce fear that would improve performance. However, extrinsic goals promoted by kind coaches will not induce much fear which would decrease the effectiveness of the extrinsic goal.

Lastly, we can also learn if the influence of intrinsic vs. extrinsic motivation could be different based on people’s personalities. Individuals with greater focus on social image might be more influenced by extrinsic motivation and vice versa. These research questions would be ideal to examine for a deeper understanding of both intrinsic and extrinsic motivation.

Work Cited

- Gagné, M., & Deci, E. L. (2005). Self-determination theory and work motivation. *Journal of Organizational Behavior, 26*(4), 331-362.
- Harlow, H. F. (1950). Learning and satiation of response in intrinsically motivated complex puzzle performance by monkeys. *Journal of Comparative and Physiological Psychology, 43*(4), 289.
- Ilardi, B. C., Leone, D., Kasser, T., & Ryan, R. M. (1993). Employee and supervisor ratings of motivation: Main effects and discrepancies associated with job satisfaction and adjustment in a factory setting 1. *Journal of applied social psychology, 23*(21), 1789-1805.
- Kasser, T., & Ryan, R. M. (2001). Be careful what you wish for: Optimal functioning and the relative attainment of intrinsic and extrinsic goals, 116-131
- Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis. *Journal of Personality and Social Psychology, 28*(1), 129.
- Lin, Y. G., McKeachie, W. J., & Kim, Y. C. (2003). College student intrinsic and/or extrinsic motivation and learning. *Learning and individual differences, 13*(3), 251-258.
- National Research Council. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. National Academies Press.
- Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: basic psychological needs in motivation, development, and wellness*. The Guilford Press
- Deci, E. L., & Ryan, R. M. (1991). A motivational approach to self: Integration in personality. *Nebraska symposium on motivation, 38*(1), 237-288.
- Malik, M. A. R., Choi, J. N., & Butt, A. N. (2019). Distinct effects of intrinsic motivation and extrinsic rewards on radical and incremental creativity: The moderating role of goal orientations. *Journal of Organizational Behavior, 40*(9-10), 1013-1026.
- Van den Broeck, A., Vansteenkiste, M., De Witte, H., & Lens, W. (2008). Explaining the relationships between job characteristics, burnout, and engagement: The role of basic psychological need satisfaction. *Work & Stress, 22*(3), 277-294.
- Vansteenkiste, M., Lens, W., & Deci, E. L. (2006). Intrinsic versus extrinsic goal contents in self-determination theory: Another look at the quality of academic motivation. *Educational psychologist, 41*(1), 19-31.
- Vansteenkiste, M., Matos, L., Lens, W., & Soenens, B. (2007). Understanding the impact of intrinsic versus extrinsic goal framing on exercise performance: The conflicting role of task and ego involvement. *Psychology of sport and exercise, 8*(5), 771-794.
- Vansteenkiste, M., Simons, J., Lens, W., Soenens, B., & Matos, L. (2005). Examining the motivational impact of intrinsic versus extrinsic goal framing and autonomy-supportive versus internally controlling communication style on early adolescents' academic achievement. *Child development, 76*(2), 483-501.

- Vansteenkiste, M., Simons, J., Soenens, B., & Lens, W. (2004). How to become a persevering exerciser: The importance of providing a clear, future goal in an panicies associated with job satisfaction and adjustment in a factory setting 1. *Journal of applied social psychology, 23*(21), 1789-1805.
- Vansteenkiste, M., Soenens, B., & Vandereycken, W. (2005). Motivation to change in eating disorder patients: A conceptual clarification on the basis of self-determination theory. *International Journal of Eating Disorders, 37*(3), 207-219.
- Vansteenkiste, M., Soenens, B., & Vandereycken, W. (2005). Motivation to change in eating disorder patients: A conceptual clarification on the basis of self-determination theory. *International Journal of Eating Disorders, 37*(3), 207-219.
- Williams, G. C., Hedberg, V. A., Cox, E. M., & Deci, E. L. (2000). Extrinsic Life Goals and Health-Risk Behaviors in Adolescents 1. *Journal of applied social psychology, 30*(8), 1756-1771.

How to Become a Sports Agent By Daniel Contreras

Abstract

Agents in the NFL play a crucial role in the careers of their clients, ensuring fair assessment of their value, suitable contracts, and easing the transition from college to the NFL. The purpose of this research is to gain a deeper understanding of what it takes to become a successful sports agent. To gain this understanding, I will research the best practices employed by NFL agents in supporting their clients, while highlighting both strong and weak examples from recent NFL history. The key findings of this research indicate that agents must have a mixture of qualitative and quantitative skills to support their clients. By doing this research, I will better understand the tools I need to pursue a career as a sports agent.

Introduction

Businessmen, lawyers, contractors: these are all words that can be used to describe a sports agent. Sports agents have a rich arsenal of talents that provide them with myriad opportunities for success. Additionally, experienced agents can be paid handsomely while still enjoying numerous new experiences, resulting in “sports agent” becoming one of the most highly sought after professions today (SportsManagementDegrees.com). Despite a plethora of people hoping to earn the job, many fail to achieve the title. This is because there are many key steps to be taken before one can even be considered for the job.

For as long as I can remember, sports have been a substantial part of my life. My dad played both high school and college football and always hoped that my brother and I would do the same one day. My brother and I were both lucky enough to do so at Mater Dei, the top high school football program in the nation. The past two years at Mater Dei have only further fueled my passion for football. The majority of players on our varsity team play college football, and many of them go Division 1. One of the reasons I feel so lucky to go to this school is because of the friendships and connections I have built. These are connections I believe I will be able to utilize in the future for not only my own benefit, but also for the benefit of my teammates as well. It would be a dream come true to be able to represent my own friends either in college football or even in the NFL, which is why I am writing this research paper. I hope to analyze real-life situations and learn lessons from successful agents of the past and present. In doing so, I will identify what traits athletes look for when hiring an agent and, eventually, be able to mimic them.

Both of my parents have backgrounds in business and seem to genuinely enjoy their jobs. This is yet another reason why I am drawn to the idea of becoming a sports agent. I am also interested in becoming a lawyer and feel sports agents represent the best of both worlds. Although I have played football for the majority of my life, I have not always enjoyed watching games. In the past couple years, however, I have become captivated with the Los Angeles Chargers and have a blast watching Justin Herbert make plays. During the 2022-2023 season, I began to wonder when Herbert was going to get paid. I started to do some research and noticed

the significant upward trend in QB contracts. I really enjoyed doing this research and began to wonder what it would be like to do such research for a living.

Last year, I was lucky enough to attend a Chargers mini-camp with my family and was fascinated by how seamless the college to NFL transition seemed for the rookies. I am interested in learning more about that transition, which is why I will analyze how agents handle such situations in this paper. Lastly, this mini-camp sparked one more question in my mind: How do coaches and scouts assess player value? This research paper will be broken down into three sections. These three sections will include the transition from college to NFL, how to assess player value, and how to negotiate contracts, ultimately aiming to tackle the question: What does it take to be a successful NFL agent, and how do you get there? In order to become a successful sports agent in the NFL, one must master the transition from the amateurs to the pros, the ability to assess player value, and the ability to negotiate player contracts.

Transition from College to NFL

The transition from college football to the NFL can prove challenging for many athletes. It is during this time, more than ever, that athletes need someone to guide them through the draft process. Athletes need somebody who can give them advice on bolstering their draft stock and securing potential endorsement deals. They need an agent. Agents come in all different shapes and sizes. Some agents come from large, well-known agencies, such as CAA, Wasserman, and Athlete's First, while others come from less-known, boutique agencies, such as World Class Sports and Landmark Sports.

One example of a popular agency is Athlete's First. Athlete's First has managed to produce eight superbowl MVPs and 115 first round picks, while still maintaining a 12:1 player to agent ratio. Furthermore, Athlete's First is in control of \$2 billion worth of athlete contracts, making it the most substantial agency in the entire league. On the other hand, one of the largest agencies in college football goes by the name of Powerhaus and is run by Trae Smith, a former UCLA quarterback and wide receiver. This college football experience helps Smith create a connection with athletes as well as build an agency that both understands players' struggles and tailors their services to players' individual needs.

The key difference between college and NFL agents is that college agents now have to adjust to the new Name Image and Likeness (NIL) rules. As Nelson Granado of Forbes magazine explains, modern-day agents must master skills that the agents of the past did not, "such as how to build fans across social media platforms and how to monetize influencer status" (Granados 2). Athletes no longer want to remain limited to success on the field. Through growing their presence on social media, they can influence millions of fans and set themselves up for a prosperous life after football.

In just one year, Trae Smith has managed to recruit 40 athletes, including quarterback Jake Haener, who was just recently drafted in the 4th round by the Saints. Athletes like Caleb Williams are now recognizing the significance of NIL and enlisting the help of marketing and communication firms, such as Smith & Company, to develop plans for their future. Just recently,

Williams landed a multi-million dollar deal with Neutrogena, officially completing the second largest NIL deal in college football history. Deals like these are not easy to accomplish, however, and social media presence is not the only skill agents must master. Smith also discusses the importance of personalization, financial literacy, and evidence-based representation decisions. Any deal an athlete makes will in some way impact their public image, whether it is for the better or worse. Personalization focuses on finding endorsement deals for athletes that both interest them and reflect their values. Additionally, athletes may not be well-educated on the terms and clauses contained within legal documents, and this is why it is crucial agents educate their athletes on these matters. Using this knowledge, athletes can better gauge the potential benefits of any deal and determine whether or not they are interested. Lastly, evidence-based representation decisions are essentially when agents use data to calculate the current NIL value of an athlete and project any potential growth. Through this process, agents can determine their strongest potential clients.

Some agents even work on their own as freelancers. This is why it is so important to understand the pros and cons of each kind of agent. While a popular, widely-praised agent may seem attractive, they may not have as much time for their clients. This is why many athletes flock to smaller agencies that boast higher agent-to-client ratios, making the athletes feel more important. Despite agent-to-client ratios, however, these smaller agencies may not provide the same opportunities that a larger agency might. For example, a larger agency may have more connections outside of football and give players opportunities to land big-time endorsement deals, as well as grow their public reputation. All of these factors are important for the athletes to consider, as they not only want to succeed in the present but also hope to set up their families for a future after football.

Once an agent has agreed to partner with a client, they must begin preparing him for the NFL. Any good agent must be able to bolster their client's draft stock, secure a reasonable contract, and ensure their client is in contact with teams that would best put his talents to use. Additionally, any agent who worked with an athlete in the NCAA during their college career may not force athletes to retain their services in the NFL (Frieser 10). NIL is essentially creating more opportunities for sponsors to pay athletes while they are still in school. As a sports agent explains to Front Office Sports, "If they give you \$100,000 in August, and you don't become a client when they turn pro, you have to pay that \$100,000 back" (Perez 13). Because of these new NIL regulations, agents are trying to ensure that they end up maintaining their clients throughout their transition to the NFL. This is why it is during the draft process, more than ever, that an agent must prove they can help their client succeed.

As soon as players see their names on a draft board, they become anxious. They worry about what team will draft them, what round they might be drafted in, and how much money they will make. This is why it is key for an agent to comfort their client and ensure he does not become caught up in the pre-draft buzz. In an interview with Front Office Sports, CJ LaBoy, a senior vice president at Wasserman Media Group, emphasizes that "the best way to manage those expectations is just being very honest and transparent about the information that you have and

the realities of the draft and how things could unfold” (Perez 17). At the end of the day, athletes need to hear the truth, and a skilled agent should be able to provide them with that. Once an agent has done this, athletes will have realistic expectations and won’t be as upset if they don’t get drafted as high as they would’ve liked. The draft should be an enjoyable experience for athletes, and it is an agent’s job to make sure it is.

Another key skill an agent must have is the ability to build connections with NFL teams. Most NFL teams do not know or even trust most agents. This is why it is so imperative that an agent can build a trusting relationship with teams. According to Erik Burkhardt, co-head of Roc Nation Sports’ football division, “You can’t help every time, but if you can influence a pick by one spot, that covers the agent’s fees for a player’s entire career” (Perez 28). Clearly, these connections can pay off significantly and set up both players and agents to be successful for a long time. Furthermore, this strengthens the relationship between a player and his agent because a player will recognize his agent’s ability to help him succeed. Thus, through keeping a player’s draft expectations grounded and building strong relationships with teams, agents can accomplish great things for players and for themselves.

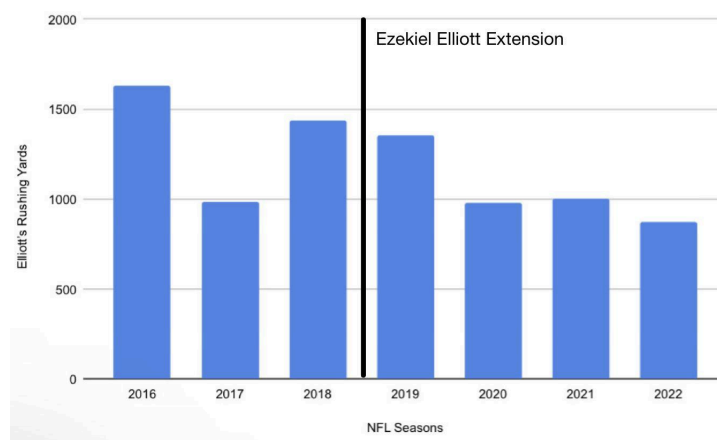
Assessing Player Value

The ability to assess player value is an essential skill that all agents must master. Through assessing player value, an agent can determine a player’s talents and find a team that requires those talents. For example, a strong, bull-dozing running back like Derrick Henry may be extremely useful to a team like the Tennessee Titans; however, his skills may not be as useful on a team like the Los Angeles Chargers, who run a more pass-heavy scheme. Through matching players with the right teams, agents can help their clients earn the money they deserve and have their skills be put to use at the same time. Despite this, it is important to remember that looks are not everything when assessing player value. At the end of the day, statistics are what truly matter. While players such as DeSean Jackson can make a big play every now and then, if they fail to put up consistent numbers, then a team cannot truly rely on them.

To aid in the evaluation of such players, analysis systems such as the Approximate Value (AV) have been developed. AV, which was developed by Pro-Football-Reference (PFR) founder Doug Drinen, takes both the number of offensive points scored and the number of defensive points allowed and compares them to the league averages. 45% of the offensive points are allocated to the offensive line, while the rest are allocated to skill positions. On the other hand, 67% of defensive points are allocated to the defensive line, with the remaining 33% left for the secondary. The main difference for the defensive point system is that All-Pro awards, tackles, interceptions, fumble recoveries, and defensive touchdowns are also taken into account. Lastly, AV compares an individual player’s contribution to the team’s success and estimates the overall influence of their position. Although AV is a very efficient system, it is certainly not perfect. The defensive point system raises some important questions such as: “Do specific position values become skewed based on greater opportunities for tackles at positions such as linebacker?” and “Does the NFL value linebackers as much as AV suggests they should?”

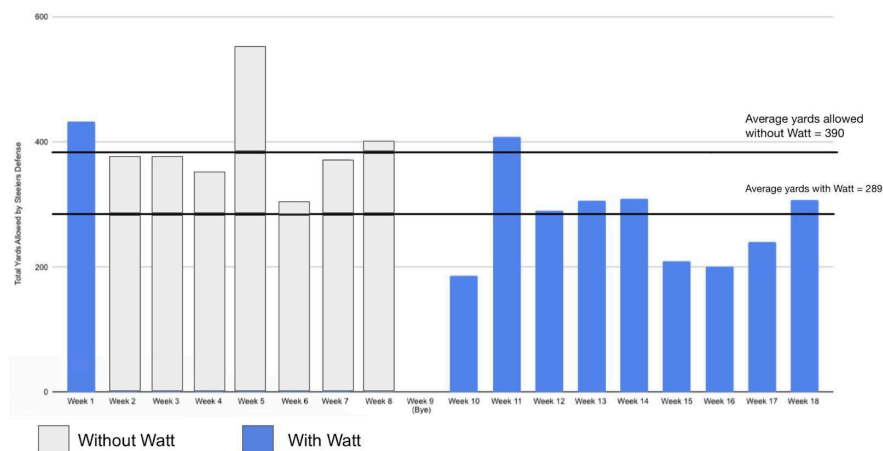
Running backs and linebackers have been undervalued for a long time now, and they are starting to fight back. Modern running backs and linebackers have evolved greatly from those of the past. Running backs, for example, are no longer tackling dummies whose sole purpose is to push the pile. Two prime examples of these evolved running backs are Dalvin Cook and Austin Ekeler. Both possess the ability to make plays on the ground and in the air, as well as provide solid pass protection for their quarterbacks. Despite these skills, neither player has received a large extension yet. On June 1, 2023, Cook was released by the Minnesota Vikings and remained a free agent for much longer than expected by anyone in the league. Moreover, even after clearly expressing his desire for an extension or at least a restructure in his contract, Ekeler has received neither. The halt in negotiations has led Ekeler to set up a Zoom meeting with the NFL's top running backs to discuss the deflated market. The reason teams are so hesitant to offer running backs large contracts is because most running backs have short-lived primes. In a study conducted by Brown University, head acceleration and hit direction were used to calculate an overall score of exposure called HITsp that can help predict concussions. On average, running backs had the highest HITsp at 36.1, followed by quarterbacks at 34.5 and linebackers at 32.6. Running backs are simply not built to last forever, and it is difficult for them to perform at a high level for more than a few seasons. Moreover, running backs are less likely to want to stay in the league for long because of the potential mental health issues that concussions can pose.

After Ezekiel Elliott's stellar performance in the 2018 season, he earned himself a six-year, \$90 million contract. Following this massive extension, Elliott had a considerable decline in his overall performance and struggled to stay on the field due to a multitude of injuries. He went from 12 touchdowns in 2019 to only six in 2020. Furthermore, he rushed for 1,357 yards in 2019, but failed to pass 1,000 in 2020. When a team gives a running back a large extension, they are placing a lot of faith in him, and the reality is that most running backs just cannot stay reliable for a long time. Running backs such as Cook and Ekeler recognize this, which is why they are so eager to secure large contracts as soon as possible.



**Note that Elliott was limited to just 10 games in the 2017 season due to suspension*

Linebackers have always played a key role in any defense, and their impact has only grown more substantial with the increase in linebacker/defensive end hybrids such as T.J. Watt. Although quarterbacks are widely considered the most valuable position to any team, there is certainly an argument to be made that linebackers like Watt are just as valuable, if not moreso. According to NFL research, “Pittsburgh allowed the fewest total yards per game (272.9) in the NFL from Weeks 10-18 after Watt's return from a torn pectoral and knee injury -- more than 100 yards less than the Steelers allowed without him” (Gordon 3). Watt clearly plays a paramount role on the Steelers defense and can be the deciding factor between a win and a loss. His ability to stuff the run game forces teams to pass against a strong Steelers secondary, ultimately increasing the likelihood of an interception. Watt doesn't just impact the run game, however. NFL research also demonstrates that, with Watt on the field, the Steelers allowed just 16.9 points per game and 288.8 total yards, posting 32 sacks and 18 takeaways in 10 games. On the other hand, without Watt on the field, the Steelers allowed 25.3 points and 389.9 yards a game, and tallied just eight sacks and five takeaways in seven games (Gordon 7). Evidently, Watt impacts multiple aspects of the Steelers defense and is without a doubt their most valuable player. This is why Watt believes he deserves a substantial contract and to earn just as much as the league's premier quarterbacks.



Every player has strengths and weaknesses, and it is an agent's job to identify those strengths and take advantage of them. Malik Willis was one of the top quarterback prospects in the 2022 draft, and many expected him to be selected with the 8th overall pick by the Atlanta Falcons. After all, the Falcons were in need of a quarterback, and Willis is a former Atlanta resident. Despite the two seeming like a perfect fit, Willis ended up falling to the third round, where he was ultimately selected by the Tennessee Titans. This was because, although Willis put up eye-popping statistics in college, he never competed on the high level that other quarterbacks, such as Kenny Pickett, played on. Following Willis's selection in the third round, NFL sports writer Ian Rappaport explained, “Everybody loves the talent, but the offense he plays in is so simple. There are questions [about] how quickly could he get acclimated to the NFL?” (Camenker 6). When assessing player value, it is important to understand the scheme one's client plays under. This ultimately allows an agent to find an environment in which their client will

thrive. Fundamentally, it is an agent's job to ensure their client feels comfortable on his team, is in a position where he can maximize his potential, and is earning the money he deserves.

Negotiating Player Contracts

Negotiating player contracts is a key aspect in the success of any sports agent. An agent must not only be thoroughly versed in the clauses and terms contained within each contract, but they must also demonstrate remarkable negotiation skills. Once an agent can juggle signing bonuses, along with incentives and guarantees, they can begin working towards a compromise. Ideally, an agent can create a situation that leaves both athletes and organizations satisfied, along with room for restructuring in the future.

For an agent to master their negotiation skills, they must have a first-rate education. This is why many agents choose to not only study business, but law as well. Both fields of study play a significant role in the negotiation process. Sports law encompasses a variety of coursework in trademarks, endorsements, labor negotiations, and, of course, contracts. Using a system that allots points to each school based on average student-to-teacher ratios, as well as average graduate tuition rates, SportsManagementDegrees.com has created a top 20 ranking of sports law programs. The Thomas Jefferson School of Law earned first place with 8 points, boasting a 9:1 student to faculty ratio and an average graduate tuition of \$26,400/year, while the University of Mississippi earned second place with six points and a 17:1 student to faculty ratio, as well as an average tuition cost of \$25,000/year. These exceptional student-to-teacher ratios make it significantly easier for students to contact their teachers and receive help in areas where they may be struggling. Additionally, these reasonable tuition costs make sports law education much more accessible to students and reduce the likelihood of student debt. To add on, within the Mississippi J.D. (Juris Doctor) school, there is a program that focuses specifically on sports and entertainment law. This equips students with a considerable edge on other schools that follow more broad law programs. Fundamentally, because of their sensible tuition costs, focused programs, and outstanding student to faculty ratios, the Thomas Jefferson School of Law and the University of Mississippi are certainly schools worth considering for any student hoping to become an agent.

When it comes to negotiating player contracts, timing is one of the most important skills an agent must possess. In an interview with CNBC, Kim Miale, an NFL agent who, back in 2020, helped offensive tackle Ronnie Stanley secure a substantial contract with the Ravens, explains, "So you want to go in and ask for a raise after you've just had a huge win because it creates leverage (...) or, if you just finished a bunch of big projects for the company then you can go in and ask for a raise then" (Connley 8). Stanley's contract came at the perfect time, as he had just finished a solid season and played a large role in elevating the Ravens' run game. Furthermore, Miale understood that, because of the increased number of injuries in the 2021 season, which were largely due to the pandemic delaying training camp, the NFL salary cap was likely going to take a considerable hit in the coming season. This is why it was so crucial that Stanley secure his contract before this cap hit came into effect. Secondly, Miale goes on to

highlight the importance of research and past examples of an athlete's contributions, "It's critical that you enter any salary negotiation meeting with ample knowledge about the pay range for your role, as well as concrete examples of the big wins you've had at the company" (Connley 11). If an agent is able to provide specific statistics, such as the fact that Stanley allowed zero sacks in 2019, and remind teams of any awards an athlete may have earned, such as Stanley's Pro-Bowl and First-Team All-Pro selections, teams will likely be far more inclined to offer a reasonable contract. In addition, an agent must ensure they are familiar with recently signed contracts at their client's position, to avoid making unreasonable demands and discrediting not only themselves, but their client as well. Lastly, Miale stresses the value of flexibility. She points out that, although average yearly salary is important to consider, "the most important thing in terms of the contract was going to be the upfront money and the guaranteed money" (Connley 16). Contracts contain a multitude of key clauses, and, while annual salary is usually the number that makes the headlines, if these other clauses are ignored, a contract may not be as exceptionally negotiated as it seems. An injury guarantee, for example, can help ensure Stanley's financial security in the event of an injury. Moreover, the fear of Stanley being cut due to a potential decline in his play is alleviated by the cap guarantee that the contract stipulates. Essentially, through taking timing, player accomplishments, and guaranteed money into consideration, an agent can make certain that their client will earn the money they deserve and remain financially secure.

Despite the vast pool of experience and education many agents boast, there are still a number of athletes that believe they are better off representing themselves. Lamar Jackson, for example, has represented himself since he entered the league. Throughout the earlier part of this offseason, Jackson and the Ravens had reached a deadlock in contract negotiations. While the Ravens simply wanted to franchise tag him, Jackson believes the fact he won MVP, paired with the eye-popping numbers he has put up the past couple seasons, merit a substantial extension. However, Jackson was not helping his odds of reaching such an agreement. Rather than focusing his energy on negotiations, he spent his time clapping back to hateful comments on Twitter. As Michael Rosenberg of Sports Illustrated writes, even if Jackson is right on every account, "it doesn't matter. He is fighting the wrong fights with the wrong people, and all he is really proving is that this is why players hire agents" (Rosenberg 1). Twitter outbursts like these portray Jackson as a volatile player and a potential locker room cancer, which can make the Ravens hesitant to offer him a large extension. On the other hand, if Jackson had an agent, they would have advised against such tweets and to stay patient throughout the process, increasing his odds of receiving the contract he wants. In the end, Lamar did end up receiving a massive five-year \$260 million with \$138 guaranteed; however, he likely could have earned even more money with the assistance of an experienced agent.

Conclusion

Over the course of this research paper, a multitude of topics have been discussed, with each topic playing a vital role in the success of hopeful sports agents. The transition from college

to NFL section, for example, started off by going over the qualities athletes look for in an agency. Some of the essential qualities included a high agent-to-athlete ratio as well as financial opportunities outside of football. NIL, for instance, is a great outlet through which athletes can monetize their talents and grow in popularity. It is an agent's job to aid in this process through educating athletes on financial literacy and help ensure they are accepting partnerships with organizations that reflect their values. This section then went on to emphasize the need for an agent to set realistic pre-draft expectations. In doing this, agents can help make the draft an enjoyable experience for their players, rather than something to stress about.

The next section on assessing player value discussed the integration of data analysis systems such as Approximate Value (AV) into the NFL and how they can help teams determine which players are making the most substantial contributions on both sides of the ball. These systems are flawed however, and do not give enough credit to particular positions, leading one to wonder which positions are currently undervalued. For instance, running backs and linebackers have evolved from the former hit-men and tackling dummies they used to be. The issue, however, lies in the fact of injury risk. Linebackers and running backs often face a plethora of injuries that inhibit their success and career longevity, which is why teams are fearful to offer them sizable contracts. Additionally, this section examined Malik Willis's fall in the 2022 draft and arrived at the conclusion that it was likely a result of the conference in which he played in college. Liberty University, where Willis played, competes outside of the Power Five conferences and does not face many fierce opponents or run a particularly complex scheme. Understandably, teams doubted whether or not his astonishing statistics would translate, causing him to fall significantly in the draft.

The final section of this paper illustrated fundamental negotiation skills along with examples of their use in the league. These three key skills - timing, background knowledge, and flexibility - allowed agent Kim Miale to secure an excellent contract for her client, Ronnie Stanley. Furthermore, this section demonstrated the potential damage an athlete, such as Lamar Jackson, can do to his career without the counsel of an experienced agent. Although Jackson ended up coming to terms with the Ravens on an extension, many athletes are not as fortunate and require an agent's assistance. In addition, this section reviewed some of the top sports law schools in the nation, as well as the highly regarded programs they offer. This information will help establish a path for hopeful agents to follow.

Throughout my time writing this research paper, I have certainly learned a lot. I have found a new appreciation for the sports agency world and now know for sure that this is a career I would like to pursue. Moreover, I feel like I have a clear path set before me and have a better idea of which schools I would be interested in attending. To add on, I am now well-educated on some of the largest and smallest agencies across the country and can begin working towards potentially earning an internship at one of them. I will take Miale's advice and master my research skills so I will easily be able to learn all there is to know about any potential client I will have in the future.

My hope is that this research will inspire readers to pursue a career in sports business and help break down the necessary steps to becoming successful. I do not believe, however, that this paper can only benefit those interested in sports business. I believe that this paper verses readers in a number of life skills, such as networking, researching, and negotiation, which all go beyond sports business. Ultimately, writing this research paper has only reaffirmed my dream of becoming a sports agent, and I hope it can do the same for anyone reading it.

Works Cited

- “The Art of NFL Contracts Part 1: The Basics - Sports Illustrated Kansas ...” Bleacher Report, 21 May 2020, www.si.com/nfl/chiefs/gm-report/the-art-of-nfl-contracts-part-1-the-basics.
- “Caleb Williams Signs Nil Deal with Neutrogena.” NBC Sports, 24 Apr. 2023, www.nbcsports.com/nfl/profootballtalk/rumor-mill/news/caleb-williams-signs-nil-deal-with-neutrogena#:~:text=His%20NIL%20valuation%20reportedly%20has,the%20NFL%20draft%20in%202024.
- Camenker, Jacob. “Why Did Malik Willis Fall in 2022 NFL Draft? Concerns about Liberty Offense Led to QB’s Free Fall to Titans.” Sporting News, 30 Apr. 2022, www.sportingnews.com/us/nfl/news/malik-willis-nfl-draft-titans-liberty/gz7qgezenplwwzbumegn8ix3.
- Courtney Connley. “This NFL Agent Secures Multimillion-Dollar Contracts for Athletes-Here’s Her Best Negotiating Advice.” CNBC, 21 Dec. 2020, www.cnbc.com/2020/12/18/nfl-agent-kim-miales-best-negotiating-advice.html.
- Cody Benjamin. Jul 1. “NFLPA Permits NFL Agents to Represent College Football Players in Nil Marketing Agreements, per Reports.” CBSSports.Com, 1 July 2021, www.cbssports.com/nfl/news/nflpa-permits-nfl-agents-to-represent-college-football-players-in-nil-marketing-agreements-per-reports/.
- English, AS. “How Do NFL Players Get Paid? Per Game or per Week?” Diario AS, 26 Oct. 2022, en.as.com/nfl/how-do-nfl-players-get-paid-per-game-or-per-week-n/#.
- “Ezekiel Elliott Stats, Height, Weight, Position, Draft, College.” Pro, www.pro-football-reference.com/players/E/ElliEz00.htm. Accessed 9 Aug. 2023.
- Frieser, Joshua. “Representing Athletes: The Legal Implications for Nil Agents.” Frieser Legal | Sports Law | NIL, 2 Dec. 2022, frieserlegal.com/representing-athletes-the-legal-implications-for-nil-agents/#:~:text=If%20the%20state%20you%20are,to%20fines%20and%20personal%20liability.
- Gordon, Grant. “Will Steelers Be NFL’s Top Defense with Healthy T.J. Watt?” NFL.Com, 30 June 2023, www.nfl.com/news/will-steelers-be-nfl-s-top-defense-with-healthy-t-j-watt?campaign=Twitter_atn.
- Granados, Nelson. “The Sports Agent of the Nil Era: A Social Media Savvy Life Coach.” Forbes, 30 May 2023, www.forbes.com/sites/nelsongranados/2023/05/23/the-sports-agent-of-the-nil-era-a-social-media-savvy-life-coach/.
- “Creating an Analytic Data Set of NFL Player Value - Philly Cover Corner.” Philly Cover Corner - Eagles News, Analytics, and Draft Content ProFootball Reference’s Approximate Value (AV) Is a Great NFL Analytics Measure of Player Value. AV, Similar to Baseball’s WAR or Basketball’s PER Metrics, Puts a Single Value on a Player’s Season. Here I Use Python to Create an NFL Data Set., 20 Mar. 2021, phillycovercorner.com/2021/03/creating-an-analytic-data-set-of-nfl-player-value/.

- “How to Pick Your Agent.” NFL Players Association, nflpa.com/active-players/how-to-pick-your-agent. Accessed 9 Aug. 2023.
- “Lamar Jackson’s Contract Negotiations Prove Why Players Need Agents ...” Sports Illustrated, www.si.com/nfl/2023/03/30/lamar-jackson-contract-negotiations-need-agents. Accessed 9 Aug. 2023.
- Lillibridge, Marc. “How NFL Teams Determine a Player’s Value.” Bleacher Report, 3 Oct. 2017, bleacherreport.com/articles/1583998-how-nfl-teams-determine-a-players-value.
- Perez, A.J. “How Agents Navigate the NFL Draft Process.” Front Office Sports, 27 Apr. 2023, frontofficesports.com/how-agents-navigate-the-nfl-draft-process/.
- “Record Breaking Contracts.” Athletes First, athletesfirst.net/. Accessed 9 Aug. 2023. September 7, 2011.
- “Running Backs Take Hardest Hits, Linemen Take Most.” Running Backs Take Hardest Hits, Linemen Take Most | News from Brown, 7 Sept. 2011, news.brown.edu/articles/2011/09/headblows.
- “Top 20 Sports Law Schools Programs.” Sports Management Degree Guide, 27 Oct. 2022, www.sports-management-degrees.com/law-programs/.

Exploring the Role of the Vagus Nerve in the Microbiota-Gut-Brain Axis By Kangyun Kim and Gudisa Tufa

Abstract

The vagus nerve is a large cranial parasympathetic nerve involved in motor and sensory functions in visceral organs from the pharynx down to the colon. It plays an instrumental role as a liaison in the microbiota-gut-brain axis, where it is involved bi-directionally through afferent and efferent fibres that interact with both peripheral and central nervous systems. This paper is an exposition of current research on the functions performed by the vagus nerve in the microbiota-gut-brain axis, drawing information from literature reviews and original experimental studies in mammals. We review the neuroanatomical pathways of the nerve, its mechanisms of signalling information from the microbiota in the gut through afferent fibres, and finally explore the cholinergic anti-inflammatory pathway through efferent fibres and serotonin and GABA act as mood modulators. Recent research indicates that efferent fibres significantly contribute to anti-inflammatory responses through acetylcholine and the modulation of mood-regulating neurotransmitters like serotonin and GABA from the gut to the central nervous system; the four nuclei of the medulla oblongata and the nodose ganglion serve as key neuroanatomical pathways, facilitating the multifaceted sensory and motor functions of the vagus.

Introduction

The human body is an intricately interconnected system where various physiological processes orchestrate a harmonious symphony. Among these many interwoven harmonies, the bidirectional flow of signals across the gastrointestinal (GI) tract, the brain, and the nuanced roleplay of microbiota constitutes a remarkable avenue of study. The microbiota-gut-brain (MGB) axis, a complex network of interactions of three major pathways endocrine, immune, and, within the focus of this paper, neural, has garnered substantial attention from the scientific world due to its pivotal role in maintaining physiological homeostasis and influencing various aspects of human health and behaviour.

The MGB axis is a communication between the trillions of microbes in the GI tract with the gut and the brain via multiple pathways, particularly that of nerves in the autonomic nervous system (ANS) (Bonaz, Bazin, et al.). Including the vagus nerve and the enteric nervous system (ENS), significant afferent signals are sent from the GI tract to the brain's ganglia, transmitting the signals to the central nervous system (CNS). As communication in the MGB axis is bidirectional, efferent signals are sent to the CNS, which indirectly or directly changes the levels of digestive inflammation and intestinal permeability in the gut.

Situated at the epicentre of the sophisticated web that makes up the MGB axis, the critical vagus nerve is located, which is a form of neural connectivity that allows the bidirectional communication between the GI tract and the CNS (Carabotti et al.). The vagus nerve also known by its scientific name the tenth cranial nerve, the vagus nerve emulates from the medulla oblongata which possesses afferent and efferent fibres extending through the visceral organs

(Cryan et al.). The vagus nerve is closely associated with autonomic control, although its impact and role goes beyond cardiac and respiratory control to include gastrointestinal neural networks. Here, it embeds itself and forms a link with the ENS of the gut, acting as a medium for the conversion of microbial stimuli into neural impulses that are then transmitted to higher centres (Carabotti et al.).

The vagus nerve's sensory fibres engage with various signals from the gut's microbial inhabitants. Here, microbial metabolites which can range from short-chain fatty acids to bioactive molecules such as substance P, travel through afferent fibres, ascending to the brainstem for the process of integration into autonomic and cognitive networks (Bruning et al.). This transduction of microbial cues into neural signals precipitates multifaceted outcomes from immune modulation to mood regulation. Therefore, the vagus nerve emerges as an important component of the monitoring of the gut's microbial milieu and the relay of information to the CNS (Breit et al.).

In the scope of our paper, we explore the literature available on the vagus nerve in the context of the MGB axis. This includes a general explanation of the structure and function of the vagus nerve as well as how certain neurotransmitters impact the outcomes of both afferent and efferent pathways. Foremost, it is necessary to outline the neuroanatomical pathways of the vagus nerve. This is to elaborate on its origins and motor nucleus, which is instrumental in projections to organs like the heart, lungs, and hose of the digestive system. Secondly, we explore vagal signalling in MGB communication because of the interaction with the ENS and interplay with neuropeptides. Lastly, we investigate the vagus-microbiota-brain (VMB) axis and neurotransmitter modulation as the release of specific neurotransmitter has certain effects on inflammation, and the modulation of some of these neurotransmitters potentially affect mood and behaviour.

The first section discusses neuroanatomical pathways of the vagus nerve. Here, it is apparent that the vagus nerve, the largest cranial nerve, is meant for sensory and motor functions as it is 80% afferent and 20% efferent (Bonaz, Bazin, et al.; Perelló et al.). It also becomes evident that the nerve has four nuclei in the medulla oblongata that extend their efferent fibres to different organs for motor function purposes. This segues into how the cyclical relationship in the MGB axis depends on afferent and efferent fibres of the vagus nerve to transmit signals between the three via the various nervous systems which allow for changes in one neural direction to affect the other. Notwithstanding, changes made in the afferent direction are given more focus because the vagus nerve plays a more significant role in transmitting signals from the microbiota to the brain due to being mostly composed of afferent fibres.

In the next section of this review, the neurological components supporting the discussion around the MGB axis will be investigated. The ENS is a network within the GI tract that exhibits bidirectional interactions with the vagus nerve—a neurological nexus of fundamental significance—and is the foremost of these neurological components. The interactions of the ENS occur over numerous vagal afferent channels which enable the transfer of a collection of gut microbial stimuli to the brainstem which acts as a neurological convergence point. Here, the

vagus nerve utilises significant neuron activators as mediators within this afferent transfer of information, converting the gut-originating signals into a neurological vocabulary that can be understood by the CNS to determine bodily responses such as inflammation in the case of acetylcholine (Forsythe et al.).

Expanding on acetylcholine's particular significance, we explore the VMB axis, a neurological network characterised by neurotransmitter regulation and the ensuing cholinergic anti-inflammatory pathway. The efferent fibres of the vagus nerve act as skillful conductors of the immune response. Within this system, the vagus nerve's impact on immunomodulation is mediated through the release of acetylcholine which can only occur through the cholinergic anti-inflammatory pathway (Pavlov et al.). The molecule of acetylcholine has a wide variety of functions including attenuating inflammatory cascades in the gut. However, the vagus nerve itself extends beyond immunological areas encompassing emotional and behavioural areas. The vagus nerve's control over these aspects is reflected in the modulation of neurotransmitters like serotonin and gamma-aminobutyric acid (GABA) which are both crucial players in mood regulation (Carpenter et al.; Pavlov et al.). These particular interactions highlight the ability of the VMB axis to have direct areas of impact that span not only the immunological domains but also the physiological and physiological domains.

Neuroanatomical Pathways of the Vagus Nerve

The vagus nerve emerges from the medulla oblongata's nucleus ambiguus and dorsal motor nucleus, creating a vast neurological circuit with multi-purposed sensory and motor functions. Its significance within the parasympathetic nervous system denotes a crucial route for transferring sophisticated signals between the peripheral organs and the CNS, especially within the context of the GI tract (Kenny and Bordoni). From an anatomical perspective, the vagus nerve presents an unusual duality with 80% of its fibres being afferent indicating that it is primarily a sensory nerve (Bonaz, Bazin, et al.). These fibres, which are widely distributed throughout visceral organs, form a highly specialised pathway essential for transmitting various signals coming from the gut. This sensory pathway becomes essential to the functionality of the MGB axis, serving as a channel for transporting microbial byproducts that travel through the afferent fibres of the vagus nerve and in turn arise to the brainstem. In turn, the culmination of this transduction process can be derived from the integration of the signals within autonomic and cognitive networks (Carabotti et al.).

In the remaining 20% of the vagal nerve composition, efferent fibres originate from specific medullary nuclei and also extend widely to various organs, including but not limited to the heart, lungs, and GI system (Bonaz, Bazin, et al.). The vagus nerve's function as a master regulator of autonomic activities, as well as its significant influence on the ENS, which is intricately intertwined into the GI landscape, are both highlighted by this complex efferent network. Due to this connectedness, the vagus nerve is positioned as the MGB axis' critical epicentre, facilitating the bidirectional communication essential for maintaining physiological homeostasis.

Conclusively, the unique afferent and efferent fibres of the vagus nerve, which have their origin in the brain, provide a neurological scaffold that is essential for coordinating intricate interactions between the stomach, brain, and microbiota. Therefore, understanding the neurological underpinnings of the MGB axis and its numerous effects on human behaviour and health requires an understanding of the anatomical structure and functional significance of the vagus nerve.

Next, the vagal motor nucleus of the medulla allows for the few efferent fibres of the vagus nerve to be involved in cardiovascular-respiratory regulation, among other automatic processes that include afferent fibres of both the vagus nerve and others (Baker and Lui). The vagal nucleus contains four nuclei: the dorsal motor nucleus, the nucleus ambiguus, the nucleus tractus solitarius (solitary nucleus), and the less vagally influenced spinal trigeminal nucleus. These nuclei are required for chemoreception in motor and sensory functions of the vagus nerve that dictate processes like gustation.

The dorsal motor nucleus, which acquires afferent messages from the vagus and the CNS, has visceral efferent fibres that send parasympathetic signals to the heart and lungs and innervate GI smooth muscles and glands (Baker and Lui). Roughly 80% of the dorsal motor nucleus neurons allow the parasympathetic preganglionic fibres to innervate GI organs (Mussa and Verberne). Regarding the cardiovascular-respiratory system, the ample parasympathetic signals from the dorsal motor nucleus allows for involuntary regulation of heart rate, ventilation, and blood pressure (Cleveland Clinic, “Medulla Oblongata”). Conversely, GI organs including the stomach, intestines, and pancreas require the motor innervation of the dorsal motor nucleus to operate.

Located in the reticular formation of the medulla, the nucleus ambiguus is similar to the dorsal motor nucleus in that its projecting nerves are more or less efferent. Nucleus ambiguus’ tangible distinction is that its efferent special visceral branchiomotor fibres supply specifically the pharynx and larynx with motor innervation for swallowing and phonation. This supply of motor innervation to the two body parts is partially done by the dorsal motor nucleus extending into small efferent portions of glossopharyngeal and spinal accessory nerves, respectively (Baker and Lui).

There is also the solitary nucleus, responsible for receiving sensory signals from the heart and sending motor signals for physiological reflexes like vomiting reflexes in unison with the other nuclei. It receives general afferent signals from chemoreceptors and mechanoreceptors of the carotid body and sinus, innervated by the glossopharyngeal nerve; and the aortic and sinoatrial bodies, innervated by the vagus nerve. Along with the taste information received separately from parts of the tongue, the sensory and motor innervations of the solitary nucleus operate to create gag, carotid sinus, and cough reflexes (Baker and Lui). The vomiting reflex, however, includes the other medullary nuclei because sensory information from different nerve systems and the GI tract is sent to the solitary nucleus which acts as a distributor of the needed motor response of emesis to these nuclei which relay the appropriate motor response to their respective areas of innervation (Zhong et al.).

The vagal motor nucleus and its vast projection through its various nuclei allow it to play sensory and motor functions in the involuntary and voluntary workings of the heart, lungs, throat, and digestive organs. It is this nucleus that makes the pathways of the vagus a valuable nerve to regulate bodily processes, including those that relate the inner workings of the gut to the brain.

With that said, this nature of intrinsic value to the body and the vastness of its projections enables the vagus nerve to be vital in the MGB axis. Through its afferent chemoreceptors and sensory ganglia, the vagus nerve sends information from the microbiome of the GI tract to the CNS through both direct and indirect ways. Vagus afferent fibres become stimulated at their chemoreceptors by gut endocrine cells at the epithelial layer of the GI tract—where the vagus does not pass—and by the microbiota in the lumen which diffuse their metabolite compounds across that layer. Afferent fibres stimulated at their receptors exert stimuli onto the CNS via the intermediary the central autonomic network (CAN) and simultaneously stimulate efferent vagal fibres via inflammatory reflex, and in turn, these fibres reduce digestive inflammation and intestinal permeability via tight junction reinforcement, indirectly modulating microbiota composition alongside vagus-independent forms of bi-directional communication (Bonaz, Bazin, et al.).

At this juncture, focusing on the gut-brain direction or, in other words, the afferent interactions of the vagus nerve from the GI tract to the CNS through the sensory ganglia of the CAN is paramount. Beginning at the GI tract, it is noteworthy that the microbiota number in trillions and majorly reside in the colon, to which the vagus nerve extends. As aforementioned, the connection between the vagus nerve and these microbiotas is, in most cases, indirect because the vagus nerve does not pass through the epithelial layer, and thus, its chemoreceptors receive a portion of the various neuroactive compounds released by the bacteria through diffusion, making it ambiguous which microbiota release which compounds. These compounds include gut hormones like ghrelin and cholecystokinin (CCK); neurotransmitters and neuromodulators like GABA, acetylcholine, dopamine, and serotonin; and neuropeptides such as neuropeptide Y (NPY) and substance P (Bonaz, Bazin, et al.).

Now understanding that various neuroactive compounds at its chemoreceptors stimulate the afferent fibres of the vagus nerve, the responsibility of sensory ganglia in getting the message to the CNS must be addressed. In this pathway, the vagal afferent fibres connect the gut and the solitary nucleus, wherein the nodose ganglion acts as an intermediary station to excite the neuron that transmits the sensory signals to the solitary nucleus in the CNS. A ganglion is a collection of neuronal bodies in the PNS that transfer information from one neuron to the next (Guillaume de Lartigue). Once the CNS receives the signal, the gut-brain communication is complete, and the CNS may respond in various ways depending on the neuroactive compound sent by the microbiota. A cyclical response may be created, where the gut secretes a hormone such as ghrelin so that the CNS regulates satiety-related energy homeostasis, or it can end at the CNS and cause changes in mood and behaviour, as in the case of dopamine being released (Fry and Ferguson; Jaber et al.).

Vagal Signaling in Microbiota-Gut-Brain Communication

The ENS system, made up of linked neurons and glial cells, has an important responsibility in vagal signaling with respect to the MGB axis, controlling processes such as motility, secretion, and sensory perception (Chanpong et al.). Using the vagus nerve as a communication method with the CNS, the ENS system effectively creates a vital connection which becomes essential in the nodose ganglion, a crucial relay site where vagal afferent fibres interact with sensory signals from the ENS (Breit et al.; Powley). Furthermore, the nodose ganglion delivers visceral sensory input to the CNS, initiating a process where vagal afferent fibres originating in the gut congregate to transmit various signals from the gut's milieu. The various signals originate from the neuroactive substances that the microbiota as a byproduct of this extensive process, a process of sensory transduction occurs which allots the ENS to transfer an array of available data on the gut's physiological and microbiological condition to the core component of the CNS - the medulla oblongata (Zhuo et al.).

The interaction between the ENS and the vagus nerve extends beyond pure local GI regulation; it serves as a mediator for the connection between microbial cues and the CNS. This is achieved through the means of regulating gut processes and relaying information connected to neurological factors such as mood and behaviour (Margolis et al.).

Returning to the vagus nerve, some examples of the MGB axis in action in afferent pathways are yet to be discussed. Of interest are the gut hormones, ghrelin and CCK; neurotransmitters, serotonin and dopamine; and neuropeptides, NPY and substance P for their slightly different ways of signalling and the effects they have on the MGB axis, as well as their effects on the body. Although both take relatively the same pathway to the CNS, they differ in how they factor into specific processes and how they are impacted by microbiota and the gut.

Ghrelin, a gut hormone secreted by epsilon cells that line the stomach and pancreas, is commonly known as the "hunger hormone." Additionally, it has other properties, such as promoting fat storage and the release of growth hormones in the pituitary gland (Cleveland Clinic, "Ghrelin Hormone"). Furthermore, its regulation of GI functions, such as gastric acid secretion and motility, is suggested to be performed through mechanisms dependent on the vagus nerve. Though it is unclear whether the hormone crosses the blood-brain barrier, the metabolites that affect it seem to be able to do so (Date; Torres-Fuentes et al.). It also strengthens this mechanism by stimulating neurogenesis in the medulla oblongata's dorsal motor nucleus, enabling efferent parasympathetic outflow to the GI tract (Zhang et al.). Notwithstanding, ghrelin acts through multiple chemoreceptors, particularly the vagus's G-protein-coupled growth hormone secretagogue receptor (GHSR) generated by neurons in nodose ganglia that are found at the ends of the vagus in the gut (Okada et al.). Current studies suggest that the gut microbiome produces metabolites that directly or indirectly influence ghrelin secretion at this junction (Leeuwendaal et al.; Trivedi et al.; Wellman and Abizaid.). The ghrelin stimulates the GHSR of the vagus nerve, and the signal is taken to the nodose ganglia and then to the CNS as sensory input to regulate processes such as glucose homeostasis in mice (Okada et al.). Essentially, this all means that microbiota in the gut can induce appetite growth hormone production and

maintain GI functions through their influence on the secretion of ghrelin: it is how the microbiota enact homeostasis on bodily functions through the gut and brain neural pathways.

Another gut hormone, CCK, is secreted by the small intestine during digestion and found in the CNS. In terms of digestion, the enteroendocrine cells of the duodenum's mucosal lining detect proteins and fats and, in response, trigger the release of CCK which in its namesake directly stimulates the gallbladder to contract and release bile into the duodenum via vagal afferent fibres (Chandra and Little). Simultaneously, it reduces appetite by activating vagal nerves in the stomach wall and stimulating the pancreas to release enzymes (Cleveland Clinic, "Cholecystokinin"). However, its role in the CNS is less understood. It is apparent in anxiety and panic disorders as research suggests that CCK in the brain is positively correlated with anxiety (Rotzinger and Vaccarino); nevertheless, more research is required on its function in the CNS.

Regarding the vagus nerve, the CCK A receptor, a primary chemoreceptor of CCK, is localised in the nodose ganglion and dorsal nucleus that mediate the hormone's satiety effects (Wank). Through CCK, enteroendocrine cells, which also receive input from microbiota toll-like receptors that discern certain bacterial products, communicate with the CNS to produce different outputs (Bonaz, Bazin, et al., Bogunovic et al.). This includes participating in the CNS and other neurotransmitters in the brain to modulate behaviour with regard to anxiety and panic on top of its digestive influence on satiety (Moran and Schwartz). CCK is involved in both digestive and emotional functions because of its versatility as a peptide hormone in the gut and a neuroactive compound in the brain.

Similarly, the neurotransmitters serotonin and dopamine also play a notable role in the MGB axis through afferent vagal fibres. Though both are categorised as 'pleasure hormones,' the former is associated with many physiological outcomes while the latter correlates with happiness and reward-based motivation. Apart from their general characteristics, how they operate slightly differs due to the nature of their production and their bodily functions.

Like CCK, types of enteroendocrine cells called enterochromaffin cells also use serotonin to convert luminal stimuli into the chemoreceptors of such fibres to reach the CNS (Hansen and Witte). Serotonin is produced 95% in the gut because it is made from an essential amino acid called tryptophan that the body must ingest due to being incapable of producing it (Terry and Margolis). Though known for its wide array of functions, those of significance to the MGB axis are related to mood and digestion (Cleveland). Within its multiple pathways to the brain, serotonin is secreted by enteroendocrine cells and microbiota, whereby the 5-hydroxytryptamine receptors (serotonin receptors) and also G-protein-coupled receptors present in vagal afferent fibres transmit the signal up to the CNS. Moreover, intestinal serotonin receptors modulate the ENS and GI development, and CNS-derived serotonin receptors aid in modulating mood and sleep, among other processes (Terry and Margolis). The variety of gut microbiota and their metabolites can modulate serotonergic pathways in the vagus. For instance, introducing *Lactobacillus rhamnosus* has been found to increase the firing frequency of the afferent vagus and increase serotonin transporter levels in the GI tract of rats (Cao et al.). This demonstrates that

serotonin, like the next neuroactive compound of discussion, is active in the afferent pathway of the MGB axis via the vagus nerve.

Dopamine, a neurotransmitter and neuromodulator, is the primary driver of reward-based motivation and influences cognition and satiety, among many other behaviours performed in the CNS and PNS (Cleveland Clinic, “Dopamine”). Like serotonin, it can also act as a hormone, but it is secreted by the adrenal gland and as a neurotransmitter by the hypothalamus. Despite this, specific microbiota in the respective genera of *Prevotella*, *Bacteroides*, *Lactobacillus*, *Bifidobacterium*, *Clostridium*, *Enterococcus*, and *Ruminococcus* are known to have a modulatory impact on its metabolism, synthesis, and breakdown throughout the dopaminergic pathway because of their enzymatic activity vis-a-vis dopamine receptors and transporters (Hamamah et al.). For example, *Lactobacillus johnsonii* induces stimulation in the vagal sensory neurons that innervate the GI tract, meaning that microbiota changes vagal neurochemistry to allow for dopaminergic pathways to be activated (Tanida et al.). These microbiota play this role in the vagus nerve and other physiological systems that affect the MGB axis, including the immune system.

Lastly, NPY and substance P come into the picture as neuropeptides modulate all bodily functions, including the neuroactivators mentioned thus far. Their importance in the MGB axis cannot be overstated as they perform similar roles to these other neuroactivators and are found everywhere and modulate those very neuroactivators.

NPY is part of a group of substances in the bi-directional MGB axis that mimics neurotransmitters as well as gut hormones because they can attach to G-protein-coupled and other receptors in addition to being secreted into the bloodstream (Du et al.). It is found at all levels of the MGB axis with over five different receptor sites. Located mainly in the innervating nerves of the GI tract, including afferent vagal fibres, NPY is synthesised and released following their stimulation. In the GI tract, NPY is known to have a complex myriad of binding sites that it binds to for different functions. For instance, NPY can inhibit CCK-induced contractions, stop peristalsis in guinea pigs, and stimulate contractions of the duodenum and colon in rats (Cadieux et al.; Dumont et al.; Holzer et al.; Krantis et al.). Not only does it impact intestinal motility, but as a neuromodulator, NPY reduces gastric acid secretion by inhibiting acetylcholine found at postganglionic cholinergic neurons via Y receptor pathways (Herring et al.; Tsai and Cheng).

Conversely, NPY is a significant component in the brain as one of the most abundant peptides of the CNS. Here, it modulates noradrenaline in cortical areas of the brain under certain conditions and plays a role in the “interoceptive regulation of anxiety and mood” (Dumont et al.; Gareau). Within the broader context of the MGB axis, it has implications on the vitality of some gut bacteria, behaviour, stress, food intake, and energy homeostasis (Gareau).

Distinctively, substance P is not present universally in the MGB axis as it functions only as a neurotransmitter in the afferents of CNS and PNS, including vagal afferents. Though it has a direct antimicrobial effect in the gut like NPY, research suggests it also has a pro-inflammatory role in both neurons and the gut where it is highly concentrated in the intestines; it may increase the intensity of diseases such as Inflammatory Bowel Disease and pancreatitis (Harrison and

Geppetti). However, it is not that abundant, so its effects are not as significant as they would be if it were as abundant as NPY. Additionally, where NPY can contract some muscles in the GI tract in particular mammals, substance P can contract all parts of the GI tract in humans and other mammals using the enteric motor neurons innervating the tract to release acetylcholine (Harrison and Geppetti; Graefe et al.). Substance P interacts with vagal nerves innervating the GI tract and produces effects in the gut and brain. However, more research is needed to understand its relationship with microbiota in the gut.

All in all, many neuroactivators play important and particular roles through the afferent fibres of the vagus nerve, with the entire tract known as the MGB axis. From gut hormones ghrelin and CCK to neurotransmitters serotonin and dopamine to neuropeptides NPY and substance P, all are influential actors in the gut and each acts with microbiota at some level, as well as interacts with different chemoreceptors of the sensory vagal nerves and also secreted into the blood as hormones.

Vagus-Microbiota-Brain Axis and Neurotransmitter Modulation

Particularly looking at the VMB, we discuss the vagus and how it influences the brain using the various mechanisms present in the MGB axis and vice versa. Here, we discuss the cholinergic anti-inflammatory pathway and how it involves the efferent fibres of the vagus, then build up to investigating the broader effects of releasing acetylcholine in the gut in terms of inflammation, and lastly, explore and elaborate on the potential of the vagus nerve in modulating serotonin and GABA to affect mood and behaviour.

The cholinergic anti-inflammatory pathway is a neuro-immunological pathway involved in weighing the contradictory mechanisms of immunity and anti-inflammation. Efferent fibres of the vagus influence the pathway because of their prominence in the body's visceral organs and their connection with the brain. This newly emerging area of research seems to show the possibility of neuro-immunomodulation, where the brain recognizes the body in a state of inflammation as a normal immune response and makes a decision relayed through efferent fibres and certain neurotransmitters to neutralise and counteract that inflammation to fight off pathogens and other bodily threats (Pavlov and Tracey). In the case of efferent vagal fibres, one study involving rats suggests that a protective cholinergic pathway is operative under myocardial ischemia/reperfusion (Mioni et al.). Sufficient research is indispensable to assert whether or not this efferent cholinergic anti-inflammatory pathway is operative in circumstantial immune inflammation or more universalized cases of inflammation. Still, the evidence suggests that such a mechanism exists that sheds light on the VMB and MGB axes and might have more significant implications in healthcare and pharmaceuticals.

Expanding, a future may be sought in the modulation of neurotransmitters, which is a critical method of applicable communication between the brain, stomach, and subsequent microbiota. One such neurotransmitter is acetylcholine. Released primarily by the efferent fibres that make up the vagus nerve, it serves as a potent anti-inflammatory substance within the gut, responding with a gut-specific balancing method (Bonaz, Sinniger, et al.). This primary response

is heavily reliant on the interaction with local immune systems and immune cells (Cox et al.). Prominently within these immune cells, the specialised immune cell receptor alpha-7 nicotinic acetylcholine receptor (7nAChR) acts as the catalyst for the responses of acetylcholine and its anti-inflammatory nature. 7nAChR, highly expressed within various immune cells throughout the GI tract is activated by acetylcholine, setting up a chain reaction of anti-inflammatory movements. This often starts with the boost in the synthesis of anti-inflammatory cytokines like interleukin-10 (IL-10); it inhibits the release of pro-inflammatory cytokines, including interleukin-6 (IL-6), interleukin-1 beta (IL-1 β), and tumour necrosis factor-alpha (TNF- α) (Chen, Deng, et al.; Mizrahi et al.). Effectively, acetylcholine counteracts inflammation in the gut, acting as a balancing neurotransmitter in the gut. Additionally, acetylcholine impacts the regulation of GI motility and secretory processes in addition to immunological modulation. Peristaltic waves are triggered by their interactions with muscarinic receptors on the smooth muscle cells of the GI tract, allowing chyme to flow through the digestive system effectively (Browning and Travagli). Acetylcholine will also interact with enteroendocrine cells which in turn trigger the release of hormones and digestive enzymes, attributing to the health of the gut.

As previously mentioned, serotonin is a neurotransmitter well known for its function in mood modulation that is mainly synthesised within the enterochromaffin cells of the gut's epithelial lining (Berger et al.). Here the vagus nerve takes a more proactive role, with modern research suggesting that the vagus nerve may act as a direct pathway for serotonin impulses to travel to the brain (Bergland). These signals would in theory then be received by serotonin-receptor-equipped vagal afferent fibres, which then send them to the brainstem for incorporation into more extensive emotional and cognitive networks.

GABA emerges as a crucial role in the VMB axis, particularly in anxiolysis, in contrast to serotonin's influence over mood. The primary inhibitory neurotransmitter in the CNS, GABA, has a calming effect by reducing neuronal excitability. Research also indicates that the vagus nerve's afferent fibres can modify GABAergic pathways, hence affecting anxiety and stress responses (Hou et al). These signals affect emotional states and behavioural responses once they interact with GABAergic circuits in the brainstem (Jie et al.).

Conclusively, the use of the vagus nerve as the key to unlocking the modulation of neurotransmitters such as acetylcholine, serotonin, and GABA can become a source of medical and technological advancements in the practical medical field. Within these applications, research will continue to develop on the cholinergic anti-inflammatory pathway along with the potential that the vagus nerve holds, providing an optimistic outlook on future research in the practical application of neuromodulation as a medical science.

Conclusion

In this exposition of the vagus nerve and its function in the MGB axis, we looked into the neuroanatomical pathways of this cranial nerve. In examining the distinctive efferent and afferent fibres that comprise its intricate structure and cerebral origins, we also shed light on the broad extensions of the vagal motor nucleus, explaining how it coordinates physiological responses in

the heart, lungs, and digestive organs. Mainly focused on the sensory ganglia that receive signals from the GI tract and send them to the CNS, their critical function in modulating visceral feelings was highlighted. The next phase of our investigation delved into vagal signalling in the communication between the microbiota, gut, and brain. After discussing the grand network of the ENS and its reciprocal interactions with the vagus nerve, it was imperative to describe its vital function in transmitting information about gut microbes to the brainstem. The role played by neuroactivators, including ghrelin, CCK, serotonin, dopamine, NPY, and substance P, were also investigated, highlighting their importance in sending signals generated from the microbiota in the gut and further demonstrating the close relationship between the gut and brain. Finally, we shed light on the brain-to-gut anti-inflammatory response by examining the role of efferent vagal fibres in the cholinergic anti-inflammatory pathway and the roles of serotonin and GABA on mood modulation where the signalling is vice versa.

Moreover, we explored various functions of the vagus nerve alongside all parts of the MGB axis. Upon reviewing current research, it is establishable that afferent fibres contribute a large role in the function of the vagus MGB axis than efferent fibres. Neuroanatomical pathways provided by the four nuclei of the medulla oblongata allow for the multiple motor and sensory functions of the vagus and the nodose ganglion, which acts as a relay site between the vagus and the CNS. Furthermore, vagal signalling in the MGB axis follows a pathway that originates from the microbiota and gut, which release neuroactivators that bind to the innervating ENS and afferent vagal fibres to signal the CNS through the nodose ganglion. Lastly, newly emerging research suggests that the efferent fibres of the vagus play a noteworthy role in the anti-inflammatory responses stimulated by acetylcholine and the vagus mediating mood modulators serotonin and GABA from the gut to reach the CNS.

Notwithstanding, more research and investigation are still required for the above claims to be helpful in modern medicine and pharmacology for humans. We recommend more research on particular microbiota and their metabolites vis-a-vis their secretion of neuroactivators to stimulate functions throughout the gut and brain, with a particular emphasis on the mood-modulating capabilities of the microbiota and gut and how mood affects digestive processes.

Works Cited

- Baker, Elliott, and Forshing Lui. "Neuroanatomy, Vagal Nerve Nuclei (Nucleus Vagus)." *PubMed*, StatPearls Publishing, 2020, www.ncbi.nlm.nih.gov/books/NBK545209/.
- Berger, Miles, et al. "The Expanded Biology of Serotonin." *Annual Review of Medicine*, vol. 60, no. 1, 2009, pp. 355–66, <https://doi.org/10.1146/annurev.med.60.042307.110802>.
- Bergland, Christopher. "The Vagus Nerve May Carry Serotonin along the Gut-Brain Axis | Psychology Today." *www.psychologytoday.com*, 6 Oct. 2019, www.psychologytoday.com/us/blog/the-athletes-way/201910/the-vagus-nerve-may-carry-serotonin-along-the-gut-brain-axis#. Accessed 28 Nov. 2023.
- Bogunovic, Milena, et al. "Enteroendocrine Cells Express Functional Toll-like Receptors." *American Journal of Physiology. Gastrointestinal and Liver Physiology*, vol. 292, no. 6, June 2007, pp. G1770–83, <https://doi.org/10.1152/ajpgi.00249.2006>. Accessed 29 Aug 2023.
- Bonaz, Bruno, et al. "The Vagus Nerve at the Interface of the Microbiota-Gut-Brain Axis." *Frontiers in Neuroscience*, vol. 12, no. 49, Feb. 2018, <https://doi.org/10.3389/fnins.2018.00049>.
- . "Therapeutic Potential of Vagus Nerve Stimulation for Inflammatory Bowel Diseases." *Frontiers in Neuroscience*, vol. 15, Mar. 2021, <https://doi.org/10.3389/fnins.2021.650971>.
- Breit, Sigrid, et al. "Vagus Nerve as Modulator of the Brain–Gut Axis in Psychiatric and Inflammatory Disorders." *Frontiers in Psychiatry*, vol. 9, no. 44, Mar. 2018, <https://doi.org/10.3389/fpsy.2018.00044>.
- Browning, Kirsteen N., and R. Alberto Travagli. "Central Nervous System Control of Gastrointestinal Motility and Secretion and Modulation of Gastrointestinal Functions." *Comprehensive Physiology*, vol. 4, no. 4, Sept. 2014, pp. 1339–68, <https://doi.org/10.1002/cphy.c130055>.
- Bruning, Jessica, et al. "Gut Microbiota and Short Chain Fatty Acids: Influence on the Autonomic Nervous System." *Neuroscience Bulletin*, July 2019, <https://doi.org/10.1007/s12264-019-00410-8>. Accessed 5 Sep. 2023.
- Cadieux, Alain, et al. "Pharmacological Actions of Neuropeptide Y and Peptide YY in Rat Colon." *Annals of the New York Academy of Sciences*, vol. 611, no. 1 Central and P, Wiley-Blackwell, Nov. 1990, pp. 372–75, <https://doi.org/10.1111/j.1749-6632.1990.tb48954.x>. Accessed 8 Oct. 2023.
- Cao, Ya-Nan, et al. "Effect of Lactobacillus Rhamnosus GG Supernatant on Serotonin Transporter Expression in Rats with Post-Infectious Irritable Bowel Syndrome." *World Journal of Gastroenterology*, vol. 24, no. 3, Jan. 2018, pp. 338–50, <https://doi.org/10.3748/wjg.v24.i3.338>. Accessed 19 Sep. 2023.
- Carabotti, Marilia, et al. "The Gut-Brain Axis: Interactions between Enteric Microbiota, Central and Enteric Nervous Systems." *Annals of Gastroenterology*, vol. 28, no. 2, Hellenic Society of Gastroenterology, 2015, pp. 203–9, www.ncbi.nlm.nih.gov/pmc/articles/PMC4367209/.

- Carpenter, Linda L., et al. "Effect of Vagus Nerve Stimulation on Cerebrospinal Fluid Monoamine Metabolites, Norepinephrine, and Gamma-Aminobutyric Acid Concentrations in Depressed Patients." *Biological Psychiatry*, vol. 56, no. 6, Sept. 2004, pp. 418–26, <https://doi.org/10.1016/j.biopsych.2004.06.025>. Accessed 21 Sep. 2023.
- Chandra, Rashmi, and A. Rodger Little. "Cholecystokinin." *Pancreapedia: Exocrine Pancreas Knowledge Base*, 2018, <https://doi.org/10.3998/panc.2018.18>.
- Chanpong, Atchariya, et al. "Recent Advances in Understanding the Roles of the Enteric Nervous System." *Faculty Reviews*, vol. 11, Mar. 2022, <https://doi.org/10.12703/r/11-7>.
- Chen, Linlin, et al. "Inflammatory Responses and Inflammation-Associated Diseases in Organs." *Oncotarget*, vol. 9, no. 6, 2018, pp. 7204–18, <https://doi.org/10.18632/oncotarget.23208>.
- Cleveland Clinic. "Cholecystokinin: Hormone Function & Definition." *Cleveland Clinic*, 25 May 2022, my.clevelandclinic.org/health/body/23110-cholecystokinin.
- . "Dopamine: What It Is, Function & Symptoms." *Cleveland Clinic*, 23 Mar. 2022, my.clevelandclinic.org/health/articles/22581-dopamine.
- . "Ghrelin Hormone: Function and Definition." *Cleveland Clinic*, 21 Apr. 2022, my.clevelandclinic.org/health/body/22804-ghrelin.
- . "Medulla Oblongata: What It Is, Function & Anatomy." *Cleveland Clinic*, 2022, my.clevelandclinic.org/health/body/23001-medulla-oblongata.
- . "Serotonin: What Is It, Function & Levels." *Cleveland Clinic*, 18 Mar. 2022, my.clevelandclinic.org/health/articles/22572-serotonin.
- Cox, M. A., et al. "Beyond Neurotransmission: Acetylcholine in Immunity and Inflammation." *Journal of Internal Medicine*, vol. 287, no. 2, Dec. 2019, pp. 120–33, <https://doi.org/10.1111/joim.13006>.
- Cryan, John F., et al. "The Microbiota-Gut-Brain Axis." *Physiological Reviews*, vol. 99, no. 4, Oct. 2019, pp. 1877–2013, <https://doi.org/10.1152/physrev.00018.2018>.
- Date, Yukari. "Ghrelin and the Vagus Nerve." *Methods in Enzymology*, vol. 514, 2012, pp. 261–69, <https://doi.org/10.1016/B978-0-12-381272-8.00016-7>. Accessed 31 Aug 2023.
- Du, Ping, et al. "Modeling the G-Protein-Coupled Neuropeptide Y Y1 Receptor Agonist and Antagonist Binding Sites." *Protein Engineering Design & Selection*, vol. 10, no. 2, Oxford University Press, Feb. 1997, pp. 109–17, <https://doi.org/10.1093/protein/10.2.109>. Accessed 27 Nov. 2023.
- Dumont, Yvan, et al. "Neuropeptide Y and Neuropeptide Y Receptor Subtypes in Brain and Peripheral Tissues." *Progress in Neurobiology*, vol. 38, no. 2, 1992, pp. 125–67, [https://doi.org/10.1016/0301-0082\(92\)90038-g](https://doi.org/10.1016/0301-0082(92)90038-g). Accessed 2 Oct. 2023.
- Forsythe, Paul, et al. "Vagal Pathways for Microbiome-Brain-Gut Axis Communication." *Advances in Experimental Medicine and Biology*, 2014, pp. 115–33, https://doi.org/10.1007/978-1-4939-0897-4_5.
- Fry, Mark, and Alastair V. Ferguson. "Ghrelin: Central Nervous System Sites of Action in Regulation of Energy Balance." *International Journal of Peptides*, vol. 2010, 2010, pp. 1–8, <https://doi.org/10.1155/2010/616757>. Accessed 25 Sep. 2023.

- Gareau, Mélanie G. "Microbiota-Gut-Brain Axis and Cognitive Function." *Advances in Experimental Medicine and Biology*, vol. 8, no. 17, 2014, pp. 357–71, https://doi.org/10.1007/978-1-4939-0897-4_16.
- Graefe, Steven B., et al. "Biochemistry, Substance P." *PubMed*, StatPearls Publishing, 2023, [www.ncbi.nlm.nih.gov/books/NBK554583/#:~:text=Substance%20P%20\(SP\)%20is%20a](http://www.ncbi.nlm.nih.gov/books/NBK554583/#:~:text=Substance%20P%20(SP)%20is%20a)n. Accessed 19 Sep. 2023.
- Guillaume de Lartigue. "Putative Roles of Neuropeptides in Vagal Afferent Signaling." *National Institute of Health*, vol. 136, Sept. 2014, pp. 155–69, <https://doi.org/10.1016/j.physbeh.2014.03.011>. Accessed 6 Sep. 2023.
- Hamamah, Sevag, et al. "Role of Microbiota-Gut-Brain Axis in Regulating Dopaminergic Signaling." *Biomedicines*, vol. 10, no. 2, Feb. 2022, p. 436, <https://doi.org/10.3390/biomedicines10020436>.
- Hansen, M. Berner, and A. -B . Witte. "The Role of Serotonin in Intestinal Luminal Sensing and Secretion." *Acta Physiologica*, vol. 193, no. 4, Aug. 2008, pp. 311–23, <https://doi.org/10.1111/j.1748-1716.2008.01870.x>.
- Harrison, Selena, and Pierangelo Geppetti. "Substance P." *The International Journal of Biochemistry & Cell Biology*, vol. 33, no. 6, June 2001, pp. 555–76, [https://doi.org/10.1016/s1357-2725\(01\)00031-0](https://doi.org/10.1016/s1357-2725(01)00031-0). Accessed 27 Nov. 2023.
- Herring, Neil, et al. "Neuropeptide Y Reduces Acetylcholine Release and Vagal Bradycardia via a Y2 Receptor-Mediated, Protein Kinase C-Dependent Pathway." *Journal of Molecular and Cellular Cardiology*, vol. 44, no. 3, Elsevier BV, Mar. 2008, pp. 477–85, <https://doi.org/10.1016/j.yjmcc.2007.10.001>. Accessed 27 Nov. 2023.
- Holzer, Peter, et al. "Neuropeptide Y Inhibits Excitatory Enteric Neurons Supplying the Circular Muscle of the Guinea Pig Small Intestine." *Gastroenterology*, vol. 92, no. 6, Elsevier BV, June 1987, pp. 1944–50, [https://doi.org/10.1016/0016-5085\(87\)90628-7](https://doi.org/10.1016/0016-5085(87)90628-7). Accessed 8 Oct. 2023.
- Hou, Xueqin, et al. "GABAergic System in Stress: Implications of GABAergic Neuron Subpopulations and the Gut-Vagus-Brain Pathway." *Neural Plasticity*, vol. 2020, Aug. 2020, pp. 1–11, <https://doi.org/10.1155/2020/8858415>.
- Jaber, Mohamed, et al. "Dopamine Receptors and Brain Function." *Neuropharmacology*, vol. 35, no. 11, 1996, pp. 1503–19, [https://doi.org/10.1016/s0028-3908\(96\)00100-1](https://doi.org/10.1016/s0028-3908(96)00100-1).
- Jie, Fan, et al. "Stress in Regulation of GABA Amygdala System and Relevance to Neuropsychiatric Diseases." *Frontiers in Neuroscience*, vol. 12, Aug. 2018, <https://doi.org/10.3389/fnins.2018.00562>.
- Kenny, Brian J., and Bruno Bordoni. "Neuroanatomy, Cranial Nerve 10 (Vagus Nerve)." *PubMed*, StatPearls Publishing, 2020, www.ncbi.nlm.nih.gov/books/NBK537171/.
- Krantis, Anthony, et al. "Peptide YY (PYY) Stimulates Intrinsic Enteric Motor Neurons in the Rat Small Intestine." *Naunyn-Schmiedeberg's Archives of Pharmacology*, vol. 338, no. 3, Sept. 1988, <https://doi.org/10.1007/bf00173402>. Accessed 18 Oct. 2023.

- Leeuwendaal, Natasha K., et al. "Gut Peptides and the Microbiome: Focus on Ghrelin." *Current Opinion in Endocrinology, Diabetes & Obesity*, vol. 28, no. 2, Feb. 2021, pp. 243–52, <https://doi.org/10.1097/med.0000000000000616>. Accessed 24 Oct. 2023.
- Margolis, Kara G., et al. "The Microbiota-Gut-Brain Axis: From Motility to Mood." *Gastroenterology*, vol. 160, no. 5, Jan. 2021, <https://doi.org/10.1053/j.gastro.2020.10.066>.
- Mioni, Chiara, et al. "Activation of an Efferent Cholinergic Pathway Produces Strong Protection against Myocardial Ischemia/Reperfusion Injury in Rats*." *Critical Care Medicine*, vol. 33, no. 11, Nov. 2005, pp. 2621–28, <https://doi.org/10.1097/01.ccm.0000186762.05301.13>. Accessed 4 Nov. 2023.
- Mizrachi, Tehila, et al. "Neuroinflammation Modulation via $\alpha 7$ Nicotinic Acetylcholine Receptor and Its Chaperone, RIC-3." *Molecules*, vol. 26, no. 20, Oct. 2021, p. 6139, <https://doi.org/10.3390/molecules26206139>.
- Moran, Timothy H., and Gary J. Schwartz. "Neurobiology of Cholecystokinin." *PubMed*, vol. 9, no. 1, National Institutes of Health, Jan. 1994, pp. 1–28. Accessed 28 Nov. 2023.
- Mussa, Bashair M., and Anthony J. M. Verberne. "The Dorsal Motor Nucleus of the Vagus and Regulation of Pancreatic Secretory Function." *Experimental Physiology*, vol. 98, no. 1, Oct. 2012, pp. 25–37, <https://doi.org/10.1113/expphysiol.2012.066472>. Accessed 1 Sep. 2023.
- Okada, Tadashi, et al. "Analysis of Peripheral Ghrelin Signaling via the Vagus Nerve in Ghrelin Receptor–Restored GHSR-Null Mice." *Neuroscience Letters*, vol. 681, Aug. 2018, pp. 50–55, <https://doi.org/10.1016/j.neulet.2018.05.035>. Accessed 25 Oct. 2023.
- Pavlov, Valentin A., et al. "The Cholinergic Anti-Inflammatory Pathway: A Missing Link in Neuroimmunomodulation." *Molecular Medicine*, vol. 9, no. 5-8, 2003, pp. 125–34, www.ncbi.nlm.nih.gov/pmc/articles/PMC1430829/. Accessed 25 Oct. 2023.
- Pavlov, Valentin A., and Kevin J. Tracey. "The Cholinergic Anti-Inflammatory Pathway." *Brain, Behavior, and Immunity*, vol. 19, no. 6, Nov. 2005, pp. 493–99, <https://doi.org/10.1016/j.bbi.2005.03.015>. Accessed 22 Nov. 2023.
- Perelló, Mario, et al. "The Controversial Role of the Vagus Nerve in Mediating Ghrelin's Actions: Gut Feelings and Beyond." *IBRO Neuroscience Reports*, vol. 12, Mar. 2022, pp. 228–39, <https://doi.org/10.1016/j.ibneur.2022.03.003>. Accessed 19 Sep. 2023.
- Powley, T. L. "Vagal Input to the Enteric Nervous System." *Gut*, vol. 47, no. 90004, Dec. 2000, pp. 30iv32, https://doi.org/10.1136/gut.47.suppl_4.iv30. Accessed 30 Sep. 2023.
- Rotzinger, Susan, and Franco J. Vaccarino. "Cholecystokinin Receptor Subtypes: Role in the Modulation of Anxiety-Related and Reward-Related Behaviours in Animal Models." *PubMed*, vol. 28, no. 3, National Institutes of Health, May 2003, pp. 171–81. Accessed 8 Oct. 2023.
- Tanida, Mamoru, et al. "Effects of Intraduodenal Injection of *Lactobacillus Johnsonii* La1 on Renal Sympathetic Nerve Activity and Blood Pressure in Urethane-Anesthetized Rats." *Neuroscience Letters*, vol. 389, no. 2, Dec. 2005, pp. 109–14, <https://doi.org/10.1016/j.neulet.2005.07.036>. Accessed 14 Sep. 2023.

- Terry, Natalie, and Kara Gross Margolis. "Serotonergic Mechanisms Regulating the GI Tract: Experimental Evidence and Therapeutic Relevance." *Handbook of Experimental Pharmacology*, vol. 239, 2017, pp. 319–42, https://doi.org/10.1007/164_2016_103.
- Torres-Fuentes, Cristina, et al. "Short-Chain Fatty Acids and Microbiota Metabolites Attenuate Ghrelin Receptor Signaling." *The FASEB Journal*, vol. 33, no. 12, Dec. 2019, pp. 13546–59, <https://doi.org/10.1096/fj.201901433r>. Accessed 2 Oct. 2023.
- Trivedi, Arjun, et al. "Ghrelin, Ghrelin O-Acyltransferase, and Carbohydrate Metabolism during Pregnancy in Calorie-Restricted Mice." *Hormone and Metabolic Research*, vol. 49, no. 01, Oct. 2016, pp. 64–72, <https://doi.org/10.1055/s-0042-116117>. Accessed 29 Aug. 2023.
- Tsai, Li Hsueh, and Juei-Tang Cheng. "Neuropeptide Y (NPY) Inhibits Dimethylphenylpiperazinium (DMPP)-Induced Gastric Acid Secretion in Isolated Rat Stomach." *Neuroscience Research*, vol. 8, no. 1, Elsevier BV, Apr. 1990, pp. 21–28, [https://doi.org/10.1016/0168-0102\(90\)90053-h](https://doi.org/10.1016/0168-0102(90)90053-h). Accessed 27 Nov. 2023.
- Wank, S. A. "Cholecystokinin Receptors." *The American Journal of Physiology*, vol. 269, no. 5 Pt 1, Nov. 1995, pp. G628-646, <https://doi.org/10.1152/ajpgi.1995.269.5.G628>. Accessed 11 Sep. 2023.
- Wellman, Martin, and Alfonso Abizaid. "Knockdown of Central Ghrelin O-Acyltransferase by Vivo-Morpholino Reduces Body Mass of Rats Fed a High-Fat Diet." *Peptides*, vol. 70, Aug. 2015, pp. 17–22, <https://doi.org/10.1016/j.peptides.2015.05.007>. Accessed 10 Sep. 2023.
- Zhang, Weizhen, et al. "Ghrelin Stimulates Neurogenesis in the Dorsal Motor Nucleus of the Vagus." *The Journal of Physiology*, vol. 559, no. 3, Sept. 2004, pp. 729–37, <https://doi.org/10.1113/jphysiol.2004.064121>. Accessed 16 Oct. 2022.
- Zhong, Weixia, et al. "Mechanisms of Nausea and Vomiting: Current Knowledge and Recent Advances in Intracellular Emetic Signaling Systems." *International Journal of Molecular Sciences*, vol. 22, no. 11, May 2021, p. 5797, <https://doi.org/10.3390/ijms22115797>.
- Zhuo, H., et al. "Neurochemistry of the Nodose Ganglion." *Progress in Neurobiology*, vol. 52, no. 2, June 1997, pp. 79–107, [https://doi.org/10.1016/s0301-0082\(97\)00003-8](https://doi.org/10.1016/s0301-0082(97)00003-8). Accessed 3 Nov 2023.

The Application of Enzymes in Industries By Preeti Vadlamani

Chapter 1: Introduction to enzymes

Proteins are a very important and diverse group of biological molecules found in living organisms. They are made by linking tens, hundreds, or thousands of building blocks called amino acids. There are 20 different types of amino acids in the structure of proteins and one protein can differ from another one in the number, type, or the order of amino acids put together. Each and every protein has a unique three-dimensional shape, which allows it to perform its specific task in the cell. In order to simplify the study of proteins, their structure is broken down to four levels of organization called the primary structure, secondary structure, tertiary structure, and quaternary structure. Primary structure is amino acids linked together to create a polypeptide chain by formation of peptide bonds. Secondary structure is formed when parts of the polypeptide strand are folded into structures called Beta pleated sheets, or alpha helices, where non-adjacent amino acids are bonded by hydrogen bonds. Tertiary structure is formed by folding of the polypeptide, with some parts containing secondary structures into a three-dimensional shape, by the help of hydrogen and ionic bonds and also hydrophobic interactions. Quaternary structure is seen only in proteins that have more than one polypeptide chain, with each chain being called a subunit. The same type of bonds and interactions responsible for tertiary structure leads to the formation of quaternary structure.

Enzymes are a very important group of proteins which act as biological catalysts that help speed up chemical reactions. Enzymes are made up of amino acids linked together in different ways to form proteins. An enzyme has a three-dimensional structure which includes an active site. An active site is the part of the enzyme where the substrate can attach to. A substrate is any chemical or group of chemicals on which an enzyme performs catalysis. The substrate(s) will perfectly fit in the active site, similar to how a key fits into a lock. It is important to mention that enzymes are highly specific for their substrates, meaning they will act only on one type of chemical and not a different one. After the active site tightly grasps the substrate, this forms an enzyme-substrate complex, which lowers the activation energy and therefore speeds up the reaction billions of times. This model is known as the induced fit model. In the human body there are around 75,000 different types of enzymes! The three most important enzymes in the human body are trypsin, lactase, and maltase. Trypsin is a digestive enzyme which helps with digesting protein. It is produced in the pancreas where it is known as trypsinogen, it breaks down the proteins in the small intestine. Lactase is active in the small intestine where it breaks down lactose which is a sugar in milk. It is produced by some cells in the small intestine called intestinal epithelial cells. Some people do not make enough lactase which makes them lactose intolerant. Maltase is also a digestive enzyme which breaks down a sugar called maltose, which is formed by joining two glucose molecules. It also plays a huge role in the final steps of digestion of starches. [1]

Chapter 2: Regulation of enzyme activity

For almost each and every enzyme, there are some molecules that may increase or decrease the enzyme's activity rate (measured in number or concentration of substrate molecule(s) turned into product(s)). These molecules are called activators or inhibitors. In allosteric regulation the inhibitor or activator binds somewhere else other than the active site of the enzyme. This process can either increase or reduce the enzyme activity by changing the conformation of the active site, and the enzymes that are regulated due to this mechanism are called allosteric enzymes. What makes these enzymes different from other enzymes is that they have an active site and allosteric sites while other enzymes only have the active site. Under normal conditions the substrate is the one that binds to the active site, however inhibitors known as competitive inhibitors which have similar shapes as the substrate bind to the active site so the substrate won't be able to and decrease enzyme activity rate in this fashion, while the non competitive inhibitors bind to a different site on the enzyme. This does not block the substrate from binding to the active site but it slows down the reaction due to the active site becoming distorted. An example of a competitive inhibitor would be methotrexate which is an antineoplastic drug that has a similar form as the vitamin folic acid that prevents regenerating dihydrofolate from tetrahydrofolate. An example of a non competitive inhibitor would be cyanide which can affect the metabolic pathway reaction rates. Activators increase the enzyme's activity by binding to it at their specific site different from the active site. Enzyme activity can also be affected by pH and temperature. The enzyme activity decreases when the pH is decreased below or increases above the optimum pH. When temperature is increased the enzyme activity is usually sped up (as long as the enzyme is not denatured), and when the temperature decreases the reaction is slowed down. [2]

Chapter 3: Enzymes Used in Food Industry

Did you know that enzymes have an important role in the food industry? A few important ones would be glucoamylase, protease, lipase, esterase, lactase, and isomerase.

Glucoamylase:

In our bodies glucoamylase belongs to the glycoside hydrolase family, it is released in the small intestine, and takes part in digestion. It breaks the bonds between glucose molecules in polymers like starch. Glucoamylase is used throughout the food industry to produce glucose syrup; it also plays a role in the fermentation process for ethanol and beer. Glucoamylase was discovered around the 19th century when there was a shortage in cane sugar, since Napoleon stopped countries he occupied to get goods from England. Around the 1960s disaccharidase research had increased, which led to the discovery of two complex enzymes, and one of those enzymes was glucoamylase. [3]

Protease:

In our body protease is made in the pancreas and also by fungi and bacteria on our skin. It breaks down the proteins on our skin or in our small intestine, this can help with digestion or is

associated with swelling and pain. In the food industry, it is extracted from corn, wheat, rice, etc. Protease is used to tender meat in the food industry, to improve the textures of baked goods (by regulating the gluten strength), and to remove bitterness from coffee. Protease is also extracted from fungal species. Protease was discovered in 1735, it was named after the Greek God Proteus by Carl Linnaeus.[4]

Lipase:

In our body, lipase is produced in the mouth, stomach, and pancreas. It breaks down the fats in our food so that they can be absorbed by the intestine. Lipase is used to flavor butter, margarine, and many other baking products in the food industry, but it is also used to improve the texture of cheese. Lipase is extracted from animals, and some plant sources. Lipase was discovered by Claude Bernard in 1856, and it was discovered in pancreatic juice.[5]

Esterase:

In our bodies esterase is found in various tissues including the liver, skin, and the plasma, and it hydrolyzes compounds containing ester, thioester bonds, and amides. Esterase also causes prodrug detoxification or activation. It is also used to catalyze the conversion of an ester to an alcohol and Esterase was first discovered in 1906, after 100 years of molecular biology, and biochemistry.[6]

Lactase:

In our bodies lactase is used to help with the digestion of lactose, which is a sugar that is found in many dairy products including milk. Lactase is used in the food industry to add flavor to things including frozen yogurt, sweetened condensed milk, etc. Lactase is produced by the cells lining the wall of the small intestine, it is also extracted from animals, and plants. Lactose was discovered in milk by Bartoletti in 1619. In 1780 it was identified as a sugar by Scheele.[7]

Isomerase:

In our body the function of isomerase enzymes is to catalyze the process in which functional groups can be transferred through a molecule, this leads to the production of isomer forms. These enzymes are produced in various tissues and organs. Isomerase has a huge role in the food industry to produce low calorie sweeteners and high-fructose–corn-syrup. It was discovered in 1830 by Jacob Berzelius, and its name comes from the Greek words “isos” and “meros” that mean “equal parts”. [8]

Chapter 4: Enzymes Used in Clothing Industry

Not only are enzymes used in the food industry they are also used throughout the clothing industry. A few important enzymes would be amylases, catalase, laccase, cellulase, pectinases, and protease.

Amylases:

Amylases in our bodies are used to digest starch, and it also takes part in our stomach, it is produced in salivary glands and in the pancreas. In the clothing industry the main purpose of amylases is to desize and to remove starch from fabrics, and it is extracted from plants. Amylase was discovered in 1833 by Anselme Payen.[9]

Catalase:

In our bodies catalase has a role in producing water by regulating the breakdown of cellular hydrogen peroxide. In the clothing industry it can be used to degrade lignin, and to bleach clothing. Catalase is extracted from aerobic microorganisms. Catalase was discovered in 1818 by Louis Jacques when he discovered hydrogen peroxide.[10]

Laccase:

Laccase is not produced in our bodies. Laccase in the clothing industry is used to bleach clothing and to finish or dye other clothing. Laccase is extracted from fungi. Laccase was discovered by Hikorokuro Yoshida in 1833, and its bacterial form was discovered in 1933.[11]

Cellulase:

Cellulase in our bodies is used to control sugar levels, and to keep a balanced cholesterol level and support lowering cholesterol levels. In the clothing industry it is used to eliminate the growth of bacteria in fabrics. Cellulase is produced by fungi and bacteria. Cellulase was discovered by several scientists working independently.[12]

Pectinase:

In our bodies pectinase is used to promote digestion, especially with plant based foods, it helps with a better digestion of veggies and fruits. In the clothing industry Pectinase is used for the formation, and the pretreatment of cotton, it helps with the absorption of the product. Pectinase is produced from bacteria and fungi. Pectinase was first seen to be reported by Z. L. Kertesz in 1930.[13]

Protease:

As already mentioned in the food industry section, protease enzymes break down proteins. In the clothing industry protease is used to help with removing stains from clothing. It is extracted from corn, wheat, rice, etc. It was discovered in 1735 by Carl Linnaeus and was named after the Greek God Proteus. [4]

Chapter 5: Enzymes Used in Medical Industry

Enzymes are not only used in food and clothing industries, but they are also used in the medical industry. A few important enzymes are protease, lipase, and trypsin, bromelain, Streptokinase, Peroxidase, and Pepsin .

Protease:

Protease in our bodies is used to break down the proteins on our skin and in our body. In the medical industry protease can be used to treat cardiovascular diseases. Protease is extracted from fungal species, corn, wheat, rice, etc.. Protease was discovered in 1735, it was named after the Greek God Proteus by Carl Linnaeus.[4]

Lipase:

The role of lipase in our bodies is to break down the fats in the food so they are able to be absorbed by the small intestines. In the medical industry lipase is used to produce enantiopure drugs. It is extracted from animal enzymes. Lipase was discovered by Claude Bernard in 1856, it was discovered in pancreatic juice.[5]

Trypsin:

The role of trypsin in our bodies is to help us with the digestion of proteins. In the medical industry trypsin is used to help with treating swelling, and pain. Trypsin is extracted from the pancreas. Trypsin was discovered in 1876 by Wilhelm Kuhne.[14]

Bromelain:

Even though our body is not able to produce bromelain, it still helps with creating substances for fighting against pain, it can help with relieving pain in our bodies, it can also help with slowing down blood clotting. In the medical industry Bromelain can help with improving drug absorption, it can help with the treatment of many cardiovascular diseases/conditions, and wound debridement. Bromelain is extracted from the stem of pineapple. Bromelain was discovered in 1891 by Vicente Marcano.[15]

Streptokinase:

Streptokinase does not naturally have a role in the human body. But in the medical industry it is helpful with restoring blood flow and lyse fibrin clots. It helps with dissolving blood clots that are formed in the blood vessels, especially arteries in the heart during a heart attack. Streptokinase is extracted from beta hemolytic streptococcus which is a group of aerotolerant bacteria. It was discovered in 1933 by Dr. William Smith Tillett who discovered it through sheer serendipity.[16]

Peroxidase:

In the human body peroxidase is found in saliva, and it breaks down hydrogen peroxide. In the medical field peroxidase is used in the field of biotechnology, where it can catalyze many oxidative reactions. Peroxidase can be extracted from many organisms including plants, humans, and bacteria. Peroxidase was discovered in 1855 by Christian Friedrich Schonbien during the treatment of guaiacol and hydrogen peroxide.[17]

Pepsin:

In the human body pepsin is produced in the stomach and it helps with the partial digestion of proteins found in food. In the medical industry pepsin is used to help with the treatment of digestive disorders. Pepsin is extracted from swine stomachs. Pepsin was discovered in 1836 by Theodor Schwann who was a German doctor and physiologist.[18]

Chapter 6: Enzymes Used in Environmental Industry

Enzymes are used in many industries including Environmental, some popular environmental enzymes are cytochrome p450 monooxygenase, dehalogenase, oxidoreductase, nitrilase, arsenic reductase, and arsenite oxidase.

Cytochrome p450 monooxygenase:

In the human body cytochrome p450 monooxygenase plays a key role in cellular metabolism, homeostasis, and the detoxification of xenobiotics. In the environmental industry it is responsible for many functions including the drug metabolism in our environment. It is extracted from the liver. Cytochrome p450 monooxygenase was first discovered in 1955 by Estabrook, Rosenthal, and Cooper.[19]

Dehalogenase:

Dehalogenase is not used in the human body. In the environmental industry dehalogenase helps with degrading halogenated compounds. Dehalogenase is extracted from bacteria, archaea, and eukaryotes. Dehalogenases are a fairly large family of enzymes and each was discovered by a different scientist, the most notable one is Christopher T. Walsh.[19]

Oxidoreductase:

In the human body oxidoreductases are found within liver cells, and take part in the pathways of oxygen metabolism. In the environmental industry, oxidoreductase takes part in the catalyzing of reactions involving transfer of electrons. It is extracted from the liver. c. Aromatase was discovered by Malcolm J. C. Paine and Harry W. Habenicht.[20]

Nitrilase:

In the human body nitrilase is used to catalyze hydrolysis of a group of chemicals called nitriles to ammonia and corresponding carboxylic acids. In the environmental industry nitrilase is used to detoxify cyanides, and nitriles. Nitrilases are extracted from bacteria, fungi, archaea, and plants. Nitrilase was discovered in 1960 by Mahadevan and Thimann.[21]

Arsenate reductase:

Humans are not able to produce arsenate reductase so it does not have a role in the human body. Arsenate reductase catalyzes the conversion of arsenate (a highly toxic substance) to arsenite, a

less toxic substance. It is used in the environmental industry for bioremediation, and environmental cleanup regarding arsenic. It is extracted from bacteria, mostly E. coli, and also yeast. It was discovered by Barry P. Rosen.[22]

Chapter 7: Enzymes Used in Agricultural Industry

There are also many enzymes that are used throughout the agricultural industry. Some of the popular enzymes include phosphatases, dehydrogenases, urease, carbohydrase, and phytase.

Phosphatases:

In the human body phosphatases help with the regulation of cellular metabolism involving phosphorylation processes and cell signaling. In the agricultural industry phosphatases play a role in the phosphorus cycle and help with plant growth.[23] They can be extracted from the bacteria and fungi, plants, or be produced by recombinant DNA technology in the lab. The first phosphatases were discovered by Edwin G. Krebs and Emond H. Fischer during the 1950s.[24]

Dehydrogenases:

In the human body dehydrogenases are used to catalyze a wide variety of reactions, from conversion of pyruvate to lactate, to detoxification of alcohol. In the agricultural industry dehydrogenases are involved in the biological oxidation of soil.[23] In agriculture, these enzymes are not directly extracted from any organism but are rather measured in soil samples. Dehydrogenases were discovered in 1937 by Hugo Theorell.[25]

Urease:

In the human body urease takes part in the maintenance of the bacterial cells in tissues. In the agricultural industry urease takes part in breaking down the urea-based fertilizers within soil.[23] Urease is extracted from the Jack bean meal. Urease was discovered in 1897 by Wilhelm Kuhne.[26]

Carbohydrase:

Carbohydrase refers to a large family of enzymes that break down carbohydrates into simple sugars. In the human body they are produced by salivary glands, pancreas, and small intestine. They help with digestion of food by breaking down starch and other large sugars into small sugars. In the agricultural industry, these enzymes are used in swine and poultry diets, examples include Xylanase and glucanase, which break down non-starch polysaccharides found in cereal grains (wheat, barley or corn).. The first carbohydrase to be discovered, which is also the first enzyme ever to be discovered, was called diastase, specifically amylase, by Anselme Payen and Jean-Francois Persoz.[27]

Phytase:

In humans phytase takes part in our digestion and breaks down the phytic acid.

In the agricultural industry, specifically aquaculture, phytase is used to improve the bioavailability and utilization of plant phosphorus by fish. Phytase is extracted from plants, animals, and microorganisms. Phytase was discovered in 1907 by Suzuki.[\[28\]](#)

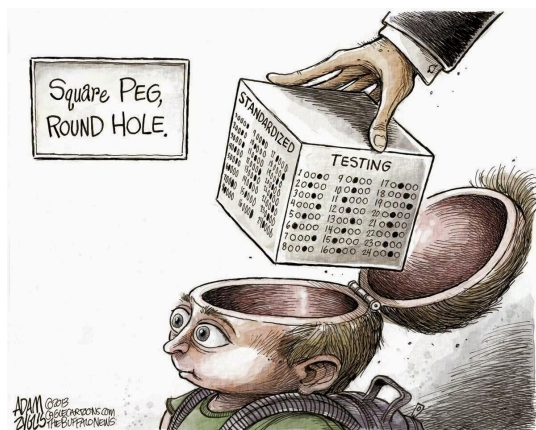
Works Cited

- [1]: <https://byjus.com/biology/enzymes/>
- [2]: <https://mgcub.ac.in/pdf/material/202004260026250a08fd94fd.pdf>
- [3]: <https://infinitabiotech.com/blog/glucoamylase-enzyme/#:~:text=Glucoamylase%20has%20significant%20importance%20in,in%20the%20production%20of%20glucose>
- [4]: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2576539/#:~:text=Thus%2C%20proteases%20regulate%20the%20fate,transduce%2C%20and%20amplify%20molecular%20signals.>
- [5]: <https://microbialcellfactories.biomedcentral.com/articles/10.1186/s12934-020-01428-8>
- [6]: <https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/esterase>
- [7]: [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8410156/#:~:text=Lactase%20is%20usually%20applied%20to,dairy%20products%20\(Table%201\)](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8410156/#:~:text=Lactase%20is%20usually%20applied%20to,dairy%20products%20(Table%201))
- [8]: <https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/isomerase>
- [9]: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3769773/#:~:text=Amylases%20are%20one%20of%20the,food%2C%20fermentation%20and%20pharmaceutical%20industries>
- [10]: <https://www.britannica.com/science/catalase>
- [11]: <https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/laccase#:~:text=Laccases%20are%20lignolytic%20enzymes%20that,They%20occur%20widely%20in%20nature>
- [12]: <https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/cellulase>
- [13]: <https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/pectinase>
- [14]: <https://www.healthgrades.com/right-care/digestive-health/trypsin>
- [15]: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3529416/>
- [16]: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4362735/#:~:text=Streptokinase%20is%20an%20enzyme%20produced,of%20the%20heart%20and%20lungs>
- [17]: <https://www.sciencedirect.com/topics/neuroscience/peroxidase>
- [18]: <https://www.britannica.com/science/pepsin>
- [19]: <https://www.hindawi.com/journals/jchem/2021/8849512/>
- [20]: <https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/oxidoreductase>
- [21]: [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3815905/#:~:text=Nitrilase%20enzymes%20\(nitrilases\)%20catalyse%20the,of%20cyanide%20and%20toxic%20nitriles](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3815905/#:~:text=Nitrilase%20enzymes%20(nitrilases)%20catalyse%20the,of%20cyanide%20and%20toxic%20nitriles)
- [22]: <https://www.sciencedirect.com/topics/pharmacology-toxicology-and-pharmaceutical-science/arsenate-reductase#:~:text=In%20one%2C%20the%20arsenate%20reductase>
- [23]: <https://infinitabiotech.com/blog/enzymes-used-in-agriculture/#:~:text=The%20most%20important%20enzymes%20used,phytase%2C%20sulfatases%2C%20and%20amylases>
- [24]: <https://www.tocris.com/pharmacology/phosphatases#:~:text=Phosphatases%20catalyze%20dephosphorylation,those%20of%20kinases%20and%20phosphorylases>
- [25]: <https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/glucose-6-phosphate-dehydrogenase>
- [26]: <https://www.britannica.com/science/urease>

[27]: <https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/carbohydrase>

[28]: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4368368/#:~:text=Microbial%20phytase%20is%20the%20most,P%20excretion%20to%20the%20environment>

Standardizing Testing: Systematically unfair? By Justin Kang



"Square Peg, Round Hole," By Adam Zyglis

We have all found ourselves in a testing room, sometimes surrounded by hundreds of others filling in bubbles and scratching their heads. However, what we often fail to pay attention to is how this test-taking meritocracy has made us blind and complicit in the economic disparities surrounding us. The College Board, for one, has blindsided us, making society complicit in paying hundreds and sometimes thousands of dollars for tests. Instead of

highlighting knowledge, a high standardized test score can often showcase privilege.

One issue of standardized testing is the cost of these exams. In the United States, Advanced Placement Exams cost [\\$98](#), the SAT costs [\\$60](#), and ACTs, at most, cost [\\$93](#). These prices account for just one exam. Additional costs can arise from repeating to take certain exams to achieve a better score, getting specialized tutoring, or, in some cases, even cheating on the exam. Regardless, all of these options come with a significant financial burden, one that not everyone can afford, aside from the wealthy few.

Secondly, a significant problem after finishing an Advanced Placement Exam is that we can't even check our answers. After taking the exam in May and then waiting until July for the score, one not only has to endure the wait but even after receiving the score, they'll never know what they fell short on and what concepts they should improve. However, thankfully, the SAT offers to showcase your incorrect answers ([with a small fee of \\$16](#)). Thus, not only do you have to pay to take these exams, but as an additional cost, you need to pay more money to see what answers you got wrong for certain exams. This feels bizarre but society has become complicit in accepting certain natures of standardized testing.

Finally, it is well-known (or should be well-known) that there's an unfortunate correlation between [higher scores and family income, race, and ethnicity](#). Those with low income and who are considered POC are statistically more likely to get a lower standardized testing score compared to their wealthier counterparts. Additionally, let us not forget the hundreds of admission scandals that occur nationally, like the [Varsity Blue Scandal](#) at USC. Scandals like those at USC offer certain individuals an unfair advantage to bolster their chances of getting into a competitive school and limit higher-level educational opportunities for many.

So, with all the systematic inequalities that exist through standardized testing, what can you do? I strongly believe high schoolers and post-high school students should consider taking community college courses. For certain students, it can be completely free, and if not, it's very inexpensive regardless. Instead of paying for a test to perpetuate inequality, why not enroll in a college-level

course where you can gain universal credit, learn from an intelligent professor with credentials like a master's degree, and have flexibility?

When you buy a standardized test, you buy a piece of paper. When you enroll in a community college class, you're gaining knowledge.

Harnessing AI for Cancer Therapeutics: The Predictive Power of Machine Learning in Anticancer Peptide Discovery

By Cheuk Ling Choi

Abstract

This review explores the dynamic intersection of oncology and artificial intelligence, specifically focusing on the application of machine learning in the development of novel anticancer peptides (ACPs) for breast cancer therapeutics. ACPs are gaining traction due to their selective cancer cell targeting and potential to act as adjuvants, enhancing the therapeutic outcomes of other agents. However, challenges in ACP stability and synthesis pose significant obstacles. The advent of machine learning offers a promising solution, providing predictive insights into peptide properties such as stability, specificity, and toxicity which can be leveraged to optimize the generation of new peptide sequences. Furthermore, this paper presents a comprehensive analysis of ACPs presently under clinical trials to evaluate their tangible impact, while also reflecting on the limitations and prospective trajectories of machine learning within peptide design. Ultimately, the fusion of ACPs and machine learning's predictive prowess opens up a new frontier in oncology. It signals a shift towards a more data-driven, predictive model of cancer research and treatment strategies. This shift could potentially disrupt the existing paradigms in cancer therapeutics, offering novel solutions that address the escalating challenges with a level of sophistication and precision hitherto unattainable.

Introduction

As the global burden of cancer continues to surge, the urgency for more effective and targeted therapeutics escalates in parallel.¹ One of the captivating areas of exploration within the oncology research sphere is the potential utility of tumor lytic agents in combinatorial therapies. The allure of these agents arises from their unique capability to induce swift and targeted cancer cell apoptosis via physical mechanisms that conventional chemotherapeutics fail to engage. Consequently, anticancer peptides (ACPs), a prominent subclass of these agents, have surfaced as potential contenders in the arsenal of cancer therapeutics, particularly in the escalating battle against drug resistance.

ACPs distinguish themselves through their selective intercalation into and disruption of cancer cell membranes. This selective targeted disruption allows for a rapid and effective response to tumor growth, setting ACPs apart from conventional therapies. Moreover, their potential to act as adjuvants, enhancing the therapeutic outcomes of other agents, further bolsters their appeal. Presently, the scientific community is investing substantial resources in the creation of novel peptides with robust anticancer properties. However, this endeavor is not without challenges. The instability of these molecules in physiological environments and the limitations in their synthesis have become significant hurdles, creating a bottleneck in the progress of ACP research.

In the face of these challenges, the advent and evolution of machine learning offer a beacon of hope. Machine learning models, by harnessing large pools of data and sophisticated

algorithms, have the potential to offer predictive insights about the structural and functional properties of peptides. These insights could range from their stability and specificity, to their toxicity. In turn, this predictive power can optimize the generation of new peptide sequences tailored for specific purposes while also enhancing their characteristics and activities.

Importantly, the application of machine learning in peptide design has the potential to drastically reduce both time and cost associated with peptide development, thereby accelerating the pace of discovery and translation into clinical practice. However, we also acknowledge and will discuss the limitations that come with the development of peptides made with machine learning, completing our review with an analysis of future prospects in this exciting field.

As part of our analysis, we will examine current ACPs in clinical trials to provide a comprehensive overview of advances made in breast cancer treatments with novel ACPs over the past five years. This will include therapies that have been developed using machine learning techniques, allowing us to assess the real-world impact of this technology in the field of oncology. Ultimately, our study suggests an exciting and promising pathway towards the next generation of anticancer peptides, aided by the power and precision of machine learning. This innovative approach at the intersection of oncology and artificial intelligence holds the promise of advancing the development of targeted therapeutics and addressing the growing challenges in cancer treatment.

Breast Cancer Peptides in Clinical Trials

Table 1. Breast Cancer Peptides in Clinical Trials

Study Title	Study URL	Study Status	Phase	Related Publications
Pilot Study of a Breast Cancer Vaccine Plus Poly-ICLC for Breast Cancer	https://beta.clinicaltrials.gov/study/NC T01532960	TERMINATED	PHASE 1	(Dillon et al., 2012)
LTX-315 in Patients With Transdermally Accessible Tumors as Monotherapy or Combination With Ipilimumab or Pembrolizumab	https://beta.clinicaltrials.gov/study/NC T01986426	COMPLETE	PHASE 1	(Sveinbjörnsson et al., 2017)
MUC1 Vaccine for Triple-negative Breast Cancer	https://beta.clinicaltrials.gov/study/NC T00986609	COMPLETE	PHASE 1	(Gao, Cen, & Lei, 2020)

Type I-Polarized Autologous Dendritic Cell Vaccine With Tumor Blood Vessel Antigen-Derived Peptides in Metastatic Breast Cancer Patients	https://beta.clinicaltrials.gov/study/NC T02479230	COMPLETE D	PHASE1	(Baar et al., 2015)
CT-011 and p53 Genetic Vaccine for Advanced Solid Tumors	https://beta.clinicaltrials.gov/study/NC T01386502	COMPLETE D	PHASE2	N/A
ImmunoTEP au 68-Ga-IMP-288 for Patients With a Recurrence of HER2 Negative Breast Carcinoma Expressing CEA	https://beta.clinicaltrials.gov/study/NC T01730612	COMPLETE D	PHASE1	(Yu et al., 2017)
START: Safety and Anti-Tumor Activity of PeptiCRAd-1 in Treatment of Cancer	https://beta.clinicaltrials.gov/study/NC T05492682	RECRUITING	PHASE1	(Ylosmaki et al., 2019)
Vaccine Therapy Plus Interleukin-2 in Treating Women With Stage IV, Recurrent, or Progressive Breast or Ovarian Cancer	https://beta.clinicaltrials.gov/study/NC T00019916	COMPLETE D	PHASE1	(Ylosmaki et al., 2019)
Phase I Study to Evaluate Safety, Tolerability, Anti-Tumour Activity and PK Profiles of Foxy-5 in Metastatic Breast, Colon	https://clinicaltrials.gov/study/NCT02020291	COMPLETE D	PHASE1	(Säfhholm et al., 2008)
Galinpepimut-S in	https://beta.clinicaltrials.gov/study/NC T00019916	ACTIVE_NO	PHASE	(SELLAS Life

Combination With Pembrolizumab in Patients With Selected Advanced Cancers	rials.gov/study/NC T02020291	T_RECRUITING	2	Sciences Group, 2023)
TH1902 in Patients With Advanced Solid Tumors	https://beta.clinicaltrials.gov/study/NC T04706962	ACTIVE_NOT_RECRUITING	PHASE1	(Demeule et al., 2021)
Phase II Study of Allo LMI Vaccine With IL-2 for Stable Metastatic Breast Ca	https://beta.clinicaltrials.gov/study/NC T00784524	TERMINATED	PHASE2	(Demeule et al., 2021)
Vaccine Therapy in Combination With Rintatolimod and/or Sargramostim in Treating Patients With Stage II-IV HER2-Positive Breast Cancer	https://beta.clinicaltrials.gov/study/NC T01355393	COMPLETED	PHASE1	N/A
First-in-Human Study of RLY-5836 in Advanced Breast Cancer and Other Solid Tumors	https://beta.clinicaltrials.gov/study/NC T05759949	RECRUITING	PHASE1	N/A
Vaccine Therapy With or Without Polysaccharide-K in Patients With Stage IV HER2 Positive Breast Cancer Receiving HER2-Targeted Monoclonal Antibody Therapy	https://beta.clinicaltrials.gov/study/NC T01922921	COMPLETED	PHASE2	(Bionity.com, 2023)

The world of breast cancer research sees a continual quest for more effective and less

toxic treatments. Hence, ACPs, which can selectively target cancer cells and show less toxicity to normal cells, have been making their way into clinical trials. A search was conducted on <https://clinicaltrials.gov/> by selecting "breast cancer" under "conditions and diseases" and "peptides" under "interventions/treatments". Given that there was no specific dropdown option for anticancer peptides. Analysis of the data obtained revealed a total of 100 studies that involved the use of peptides in the treatment of breast cancer. Out of these, 15 were selected for further analysis based on their relevance and the stage of their trial. It was observed that a significant majority—nine out of the 15 studies—only advanced as far as Phase 1. The remaining six managed to reach Phase 2. This trend is reflective of the rigorous demands and structured nature of clinical trials. Phase 1 trials, which primarily assess the safety, side effects, optimal dosage, and timing of a new treatment, involve a smaller cohort of participants (typically 20-80). The main goal during this phase is to ensure safety and monitor side effects. Meanwhile, Phase 2 trials involve more participants (usually 100-300) and aim to further evaluate safety and efficacy, helping to determine the best dosage and investigate the potential benefits of the treatment. However, the progression from Phase 2 to Phase 3 is often impeded by multiple factors. In the context of anticancer peptides, these can include:

1. **Efficacy:** Peptide therapies are highly targeted, which can be a double-edged sword. While they often have fewer side effects because they are designed to interact only with specific cancer cells, this specificity can also limit their effectiveness. If the cancer mutates or if not all the cancer cells express the target of the peptide, the therapy may not be effective enough to justify moving to Phase 3.
2. **Safety and Side Effects:** While peptides are often safer than traditional chemotherapy drugs, they can still cause side effects, especially at higher doses. If these side effects are serious or if they significantly impact a patient's quality of life, a trial may not progress to Phase 3.
3. **Delivery Challenges:** Peptides are susceptible to rapid degradation in the body and may not reach the tumor site in sufficient quantities. Researchers are still exploring ways to effectively deliver anticancer peptides to tumor sites. If a promising peptide therapy does not have an effective delivery method, it may fail to show efficacy in Phase 2 trials.
4. **Financial and Logistical Considerations:** Phase 3 trials are much larger and more expensive to run than Phase 1 or 2 trials. If a peptide therapy shows only modest benefit in Phase 2, it may not attract the funding necessary to move to Phase 3.
5. **Competition from other therapies:** There are many types of cancer therapies being investigated, including other targeted therapies, immunotherapies, and traditional chemotherapy drugs. If other therapies for the same type of cancer show more promise, it may be difficult for a peptide therapy to progress to Phase 3.

Out of the 15 clinical trials under review, three were terminated due to varying reasons. A common thread was the treatment's inability to yield anticipated improvements or benefits for the

patients. This lack of efficacy resulted in the termination of these trials as the treatments under investigation did not live up to expectations. Take, for instance, the Phase 2 Study of Allogeneic Large Multivalent Immunogen (LMI) Vaccine and IL-2, being tested for treating Stable Metastatic Breast Cancer. Half of the 14 participants involved in the trial experienced hair loss, a condition known as Alopecia. Similarly, nearly 43% of these subjects encountered Peripheral neuropathy, a nerve-related condition that often leads to weakness, numbness, and pain, usually in your hands and feet. Joint pain and flu-like symptoms were other common experiences, affecting half of the trial participants. Aside from these, there were several instances of other health issues including nausea, fatigue, anxiety, and insomnia.¹⁴ These side effects, in combination with the insufficient efficacy of the treatments, led to the termination of these clinical trials. These cases highlight the inherent challenges in drug development and underscore the crucial role of clinical trials in ensuring the effectiveness and safety of new treatments.

Despite these challenges, some trials remain active, and several have been completed, providing valuable data on the safety, tolerability, and potential efficacy of peptide-based therapies. This continued interest in peptides in cancer therapy underscores their potential as a promising avenue in the fight against breast cancer. More research, however, is needed to overcome the identified hurdles and fully realize their potential in clinical settings.

Machine Learning Algorithms In The Prediction And Identification Anticancer Peptides

Machine learning techniques have become invaluable assets for predicting and identifying ACPs, which hold great potential in the quest for new cancer treatments. These computational methods, including support vector machines (SVM), ensemble strategies, and iACP-GAEnsC, use diverse peptide characteristics like amino acid composition, dipeptide composition, and binary profiles to build accurate models. As the number of known ACPs continues to grow, machine learning-driven approaches play a critical role in effectively screening potential candidates, supporting experimental researchers in virtual screening, and pinpointing mutations that could enhance anticancer efficacy. iACP-GAEnsC in particular, has demonstrated notable success in ACP prediction and recognition—boasting accuracies of 96.45%.¹⁵

One such powerful tool harnessed in this endeavor is SVM, a powerful supervised learning algorithm that can be used for classification, regression, and other tasks. In the context of ACP prediction, SVM is utilized to find the optimal hyperplane that separates ACPs from non-ACPs based on their peptide characteristics.¹⁶ This method has been employed in various studies and has contributed to the development of successful ACP prediction, such as in AntiCP3.

Ensemble strategies, on the other hand, combine multiple learning algorithms to improve the performance of individual methods. By leveraging the strengths of different algorithms, ensemble methods can achieve better prediction accuracy and generalization. The main idea behind ensemble methods is that a group of weak learners can be combined to form a stronger learner. Different ensemble strategies include bagging, boosting, stacking, and voting.¹⁷ For

example, bagging involves training multiple base models, typically decision trees, on different subsets of the training data, which are obtained by random sampling with replacement. The final prediction is obtained by averaging (in regression) or voting (in classification) the predictions of all the base models. Boosting, on the other hand, is an iterative process where each new model attempts to correct the errors made by the previous model. The most common boosting algorithms include AdaBoost and Gradient Boosting. Stacking is where multiple base models are trained independently, and their predictions are used as input for a higher-level model, often referred to as the meta-learner or second-level learner, the idea behind stacking is to leverage the strengths and offset the weaknesses of multiple diverse models to improve overall prediction performance. In the context of ACPs, each base model can be trained to recognize different features or patterns in the peptide sequences that are indicative of anticancer properties. For example, one model might be particularly good at identifying certain amino acid compositions, while another might excel at recognizing specific peptide compositions. The meta-learner can then take these diverse predictions into account and learn how to best combine them to achieve a more accurate and robust prediction of ACPs. Finally, voting aggregates the predictions of multiple models, enhancing the reliability and accuracy of final predictions. It comes in two forms: hard voting, where the class with the most votes from all models is selected, and soft voting, which considers the probability estimates for each class from each model.

Transitioning from the SVM-based and ensemble approaches, another innovative method that has exhibited significant promise in ACP prediction is the iACP-GAEnsC. This method is special in its adaptability to diverse and complex datasets. Leveraging the power of genetic algorithms allows for a dynamic and flexible approach to classifier selection and weighting. Genetic algorithms are heuristic search algorithms inspired by the process of natural selection, where the fittest individuals are selected for reproduction to produce offspring for the next generation. In the context of iACP-GAEnsC, this translates to the selection and combination of the most effective classifiers and features for ACP prediction.¹⁸ Moreover, the ensemble learning approach of iACP-GAEnsC enhances the robustness and generalization capability of the model.¹⁹ By combining multiple classifiers, the model can leverage the strengths of each to achieve better predictive performance. This is particularly valuable in situations where the data may be noisy or contain outliers, as the ensemble model can compensate for the weaknesses of individual classifiers.²⁰ The hybrid feature space used in the iACP-GAEnsC method is another significant factor contributing to its effectiveness. By integrating amino acid composition, dipeptide composition, and binary profiles, the method captures a comprehensive representation of the peptides. This diverse feature space allows for more nuanced and accurate modeling of peptide characteristics, leading to more reliable predictions. For validation, the iACP-GAEnsC method uses either cross-validation or an independent test dataset. Cross-validation is a robust method that provides an unbiased estimate of model prediction performance.²¹ Alternatively, testing on an independent dataset ensures that the model's predictions are generalizable to new, unseen data. Hence, iACP-GAEnsC, by combining genetic algorithms, ensemble learning, and a hybrid feature space, provides a powerful and adaptable method for ACP prediction. Future research

can explore further enhancements to this model, such as incorporating additional types of features or using more advanced ensemble learning techniques.²²

Examples of Anticancer Peptides designed from Machine Learning Algorithms

(1) Machine Learning Guided Discovery of Non-Hemolytic Membrane Disruptive Anticancer Peptides²³

One study successfully identified new non-hemolytic ACPs through the combination of a generative Recurrent Neural Network (RNN) model and a Peptide Design Genetic Algorithm (PDGA), both paired with activity and hemolysis classifiers. The new ACPs displayed properties similar to Lasioglossum-III (LL-III), a known natural ACP, forming amphiphilic α -helices, disrupting membranes, showcasing significant antibacterial activities, and targeting cancer cells by disruption of the outer membrane. Particular attention was given to the peptides performance against breast cancer cell lines, with three peptides showing promising results. Both generative methods and the peptides generated are described below.

Approach 1

The first approach employed a generative machine learning model that was pre-trained on a dataset of AMPs and ACPs. This model was further fine-tuned with 53 ACPs that demonstrated activity against HeLa cancer cells. From this fine-tuned model, 50,000 sequences were sampled. These sequences were then evaluated using a RNN classifier that was trained to distinguish between active and inactive peptides. From this evaluation, 11,458 sequences were predicted to be active. These sequences were then filtered based on several criteria: they had to be short, novel, within the applicability domain of the classifier, composed only of natural amino acids, and predicted to be α -helical and amphiphilic. Following this filtering, 202 sequences remained. These were further clustered based on sequence similarity, with the requirement that sequences within each cluster could be a maximum of 10 mutations away from each other. Finally, 13 sequences were selected for synthesis. See Fig. 1.

Approach 2

The second approach used a PDGA to find analogs of LL-III, a known non-hemolytic ACP. The PDGA generated 715,658 unique sequences that were similar to LL-III (with a MAP4 Jaccard Distance of 0.6 or less). These sequences were evaluated using the same activity classifier from the first approach and a hemolysis classifier trained to distinguish between hemolytic and non-hemolytic peptides. From this evaluation, 6,300 sequences were predicted to be both active and non-hemolytic. These sequences were then filtered and clustered in a similar manner as in the first approach, and 20 sequences were selected for synthesis. See Fig. 1.

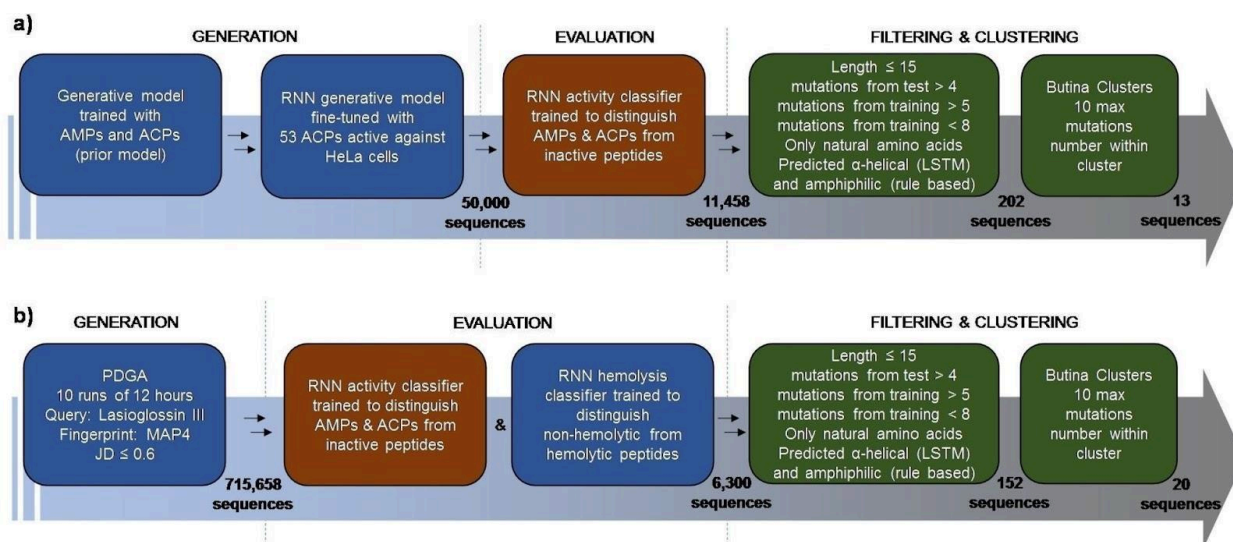


Fig. 1. This flowchart provides a summary of the two distinct machine learning approaches employed in the study. a) The first part of the flowchart describes the process of using a generative machine learning model. This model was pre-trained on a comprehensive dataset, which comprises AMPs and ACPs. This allows the model to learn the underlying patterns and structures in the data, which can then be used to generate new, yet similar data points. b) The second part of the flowchart deals with the use of the PDGA to discover analogs of LL-III. LL-III is a known non-hemolytic ACP. The use of PDGA allows for a systematic and efficient exploration of the vast potential sequence space for such analogs. (Reproduced under Open Access from Zakharova et al., 2022).

Table 2. IC₅₀ profiling of peptides. (Reproduced under Open Access from Zakharova et al., 2022).

nr.	Sequence	HeLa ^a	MCF-7 ^a	MB-MDA-231 ^a	MCF-10a ^b	HEK293 ^a
A1	FAKKFFKKFAKFAFK	8.2±0.5	6.0±0.5	6.4±0.5	17.3±3.5	15.2±0.7
B1	ANWKKWIGKVIKLVK	5.5±0.8	6.1±1.0	5.0±0.6	19.5±2.8	12.1±1.6
B2	NWKKILGKILDHLAC	7.0±1.4	5.4±0.3	5.4±1.2	11.7±0.4	6.7±0.5
LL-III ^a	VNWKKILGKIIVVK	6.0±0.5	7.1±0.3	5.9±0.4	14.8±2.1	15.0±3.0

^a IC₅₀ was determined after 72 h incubation at 37°C in DMEM high glucose medium supplemented with 10 % FBS. ^b IC₅₀ was determined after 72 h incubation at 37°C in serum-free HUMEK ready medium.

Once the ACPs were generated and selected, the research team proceeded with their synthesis and subsequent testing on a variety of cell lines. The three most active ACPs identified - A1, B1, and B2 - displayed promising activity against two distinct types of breast cancer cell lines. The first, MCF-7, is a well-established model for estrogen receptor-positive (ER+) breast adenocarcinoma, while the latter, MB-MDA-231, is widely used as a model for triple-negative breast cancer (TNBC), which is notable for its aggressive nature and lack of hormone receptor expression. These findings suggest a broad spectrum of activity for these peptides across

different breast cancer subtypes.

The synthesis of the 33 selected peptides was executed with semi-automated high-temperature Fmoc solid-phase peptide synthesis. The peptides were then purified by preparative HPLC, resulting in samples that displayed good aqueous solubility — a feature attributed to the presence of two to seven cationic residues (either lysine or arginine) in each sequence. The inclusion of these cationic residues is a common feature in many bioactive peptides due to their role in facilitating interactions with negatively charged cellular membranes. This is particularly significant for peptides designed to interact with or penetrate cells, as is the case with our selected peptides. The number of cationic residues can influence the balance between hydrophilicity and hydrophobicity in a peptide, affecting its solubility, stability, and activity. Therefore, the presence of two to seven cationic residues in each peptide is neither inherently good nor bad but is an important factor that can influence the peptide's overall behavior and efficacy.

The anticancer activity of these peptides was initially evaluated on HeLa cells, a type of cervical cancer cell. After 72 hours of exposure to 50 μM of each peptide, the viability of the remaining live cells was quantified using the Alamar Blue assay. This assay is based on the reduction of resazurin, a non-toxic, cell-permeable compound that is blue in color and virtually non-fluorescent. Living cells metabolically reduce resazurin to resorufin, which is pink and highly fluorescent. The extent of reduction, and hence the color change, directly correlates with the number of viable cells present. Seven of the thirteen peptides generated by the RNN exhibited substantial activity, with IC_{50} values ranging from 8–19 μM . However, these peptides lacked selectivity against non-cancerous cells and were predominantly hemolytic.

Contrarily, in the PDGA-generated series, only three out of 20 LL-III analogs exhibited activity against HeLa cells, again with IC_{50} values in the low micromolar range. These peptides demonstrated approximately 2-fold selectivity against HEK293 cells, akin to the properties of the parent ACP LL-III. Notably, the peptides were all non-hemolytic or weakly hemolytic, presumably reflecting the influence of the hemolysis classifier on the selection.

In the future, the researchers envision employing iterative rounds of machine learning-guided design and experimental validation as a strategy to augment the currently limited repertoire of non-hemolytic ACPs. Such an approach not only promises to yield new and more potent ACPs, but could also deepen our understanding of the underlying principles that govern activity and selectivity in this class of therapeutics. By illuminating these principles, we could more effectively guide the design of future peptides, potentially leading to the development of anticancer therapeutics that are both more effective and selective.

(2) Designing Anticancer Peptides by Constructive Machine Learning²⁴

Another study enlisted a two-step procedure using the RNN model with long short-term memory (LSTM) cells to design ACPs. Overall, a set of twelve distinct and novel peptides were generated and synthesized and their efficacy was tested on MCF7 human breast adenocarcinoma cells, along with their selectivity against human erythrocytes. Out of these, ten peptides

displayed activity against the cancer cells. Notably, six out of the ten active peptides selectively targeted and killed MCF7 cancer cells without any detrimental effect on human erythrocytes, showing a selectivity ratio of at least 3:1. Therefore, this study validates the use of constructive deep learning to design ACPs de novo, eliminating the necessity for explicit molecular design guidelines. The approach and peptides synthesized are described and shown below.

The first step involved training the model on a dataset of 10,000 α -helical cationic amphipathic peptides. These peptides formed the basis for the model to learn the 'grammar' of peptide sequences. The goal was to ensure that the model would learn the generic features of ACPs. In the second step, the model was fine-tuned using a small library of 26 known ACPs with low micromolar activities against human breast adenocarcinoma (MCF7, Michigan Cancer Foundation-7) cells and varying membranolytic effects on human erythrocytes. This fine-tuning approach was based on the observation that amphipathicity (the property of having both hydrophobic and hydrophilic regions) and an overall positive charge are typical but insufficient characteristics of membranolytic α -helical ACPs.

From the fine-tuned model, 1,000 new amino acid sequences were generated, each with six to 28 residue positions. Impressively, 98% of these sequences were unique, with only one identical to a sequence from the training set. None of the new peptides matched the fine-tuning set, the CancerPPD database, or any FDA-approved/investigational therapeutic peptide. This demonstrated the model's ability to generate novel peptide sequences. Then, the generated sequences were then ranked in silico, considering their pharmacophore similarity to known active/selective patented ACPs and their predicted anticancer activity. The top 12 peptides were synthesized and tested on MCF7 cells and their ability to destroy human erythrocytes. Ten out of the 12 peptides showed membranolytic activity in vitro, with varying degrees of anticancer activity and selectivity. Six peptides showed a clear preference for cancer cells, with EC50 values similar to the activity of NK-2, a known selective ACP.

Table 3. Hemolytic activities of designed peptides. (Reproduced under Open Access from Zakharova et al., 2022).

ID	Sequence	EC ₅₀ [μ M] ^[a]	HC ₅₀ [μ M] ^[b]	TI ^[c]
1	KLWKKIEKLIKLLTSIR	47 ± 3	236 ± 13	5.1 ± 0.6
2	YIWARAERWILWWGKFLSL	56 ± 3	inactive	> 7
3	ELAKKLTCLKQLHRIW	inactive	inactive	n.d.
4	DLFQLQRLFLGILYCLYKIW	47 ± 4	132 ± 16	2.8 ± 0.6
5	KLIDQWKKVLYHVE	inactive	inactive	n.d.
6	AIKFFGLAKIVAKV	95 ± 4	inactive	> 4
7	RWNGRIIKGFYNLVKIWKDLKG	42 ± 4	89 ± 6	2.1 ± 0.3
8	KVWIKIKIRLLHGIKRGWKG	34 ± 4	inactive	> 11
9	GFWARIGKVFVAVKNL	101 ± 4	inactive	> 4
10	AFLYRLTRQIRPWWRWLYKW	45.5 ± 0.8	34 ± 5	0.7 ± 0.1
11	RIWGHRSRYIKVRLIQ	50 ± 10	inactive	> 8
12	QIWHKIRKLWQIIKIDGF	16.1 ± 0.3	23 ± 5	1.4 ± 0.3
NegE ^[d]	PAVPPGVLIIS	inactive	n.d.	n.d.
NK-2 ^[e]	KILRGVCKIMRTFLRRISKDILTGGK	30 ± 3	281 ± 48	9 ± 2

[a] Cytotoxicity against the MCF7 cancer cell line; values are the mean ± SEM, N = 3. [b] Hemolysis on human erythrocytes; values are the mean ± SEM, N = 3. [c] Therapeutic index (HC₅₀/EC₅₀). EC₅₀ > 200 μ M (MCF7 cancer cells) and HC₅₀ > 400 μ M (human erythrocytes) were outside the range of detection, and the respective peptides were labeled as inactive. [d] NegE was used as a negative control peptide. [e] The known ACP NK-2^[16,17] was used as a selective positive control. All peptides were synthesized with amidated C-termini; n.d.: not determined.

Therefore, this constructive machine learning technique proved to be applicable for automated peptide design without the need to synthesize and test large sets of peptides. The researchers suggest that this model could be trained on other peptide classes of interest, further

expanding its potential applications.

Limitations of Current Machine Learning Models

Machine learning techniques show great promise in identifying ACPs, but there are limitations in the current models that require further research and development. One primary limitation is the scarcity of large and diverse datasets, which can lead to overfitting.²⁵ Expanding and diversifying the dataset can significantly improve the performance of ACP prediction models. Therefore, a novel ACP prediction model called ACP-DA (Data Augmentation) was recently proposed to address the challenge of insufficient datasets in some cases. ACP-DA employs data augmentation techniques to enhance the prediction performance by generating additional training samples in the feature space. To effectively represent peptide sequences, ACP-DA combines binary profile features and AAindex features, which provide a more comprehensive representation of the peptide properties. The augmented samples are then utilized to train the machine learning model, which in turn is applied for ACP prediction.²⁶

Another limitation is the interpretability and transparency of the models.²⁷ Developing interpretable models or employing explainable AI techniques can improve model trustworthiness and researchers' understanding of the underlying biological mechanisms. Moreover, current models often rely on a limited set of features, which may not capture the intricate relationships between different features, thus constraining their accuracy and predictive power.²⁸ Incorporating more comprehensive feature representations, such as physicochemical properties and structural information, can help improve model performance and provide a deeper understanding of ACPs. Finally, balancing sensitivity and specificity can prove difficult and may necessitate more advanced modeling techniques.²⁹ Utilizing cost-sensitive learning, multi-objective optimization, or ensemble methods can help balance sensitivity and specificity, leading to more reliable and robust ACP prediction models.

Therefore, addressing these limitations by expanding and diversifying datasets, enhancing interpretability, incorporating comprehensive feature representations, and balancing sensitivity and specificity through advanced modeling techniques can improve the performance of machine learning-based ACP prediction models, ultimately accelerating the discovery and development of novel cancer treatments.

Conclusions and Outlook

Machine Learning: Where has it, and where will it take us?

While machine learning has shown immense promise in predicting ACPs and screening potential candidates, it's paramount to remember that these are just predictions. The leap from prediction to clinical application is a vast one, filled with numerous challenges and complexities. The absence of a definitive cure for cancer, despite the growing sophistication of machine learning models, indeed points to some inherent limitations in this approach.

One often seen argument is that machine learning, for all its prowess, is only as good as the data

it is trained on. It's a tool that learns patterns from existing data, and it is these patterns it uses to make future predictions. One of the critical challenges we face in leveraging machine learning for ACP prediction is the quality and quantity of data available. The available datasets are limited and often skewed, with more non-anticancer peptides than anticancer ones. This imbalance can lead machine learning models to develop an inherent bias, reducing their efficacy in identifying potential ACPs. Also, machine learning models, particularly complex ones, can act as “black boxes”. While they can accurately predict whether a particular peptide has anticancer properties, they often don't provide a clear understanding of why a certain peptide is identified as an ACP. This lack of interpretability can make it challenging to translate machine learning predictions into actionable biological insights or therapeutic interventions.

Furthermore, the complexity and variability of cancer itself poses a significant challenge. Even within the same type of cancer, there can be a great deal of heterogeneity, not only from patient to patient but also within different sites of the same tumor. This variability can make it difficult for machine learning models to draw generalizable conclusions applicable to a broad range of patients or cancer types. It's essential to realize that a successful prediction of a peptide as an ACP does not guarantee its clinical efficacy as a cancer treatment. Many promising peptides fail to make it into clinical trials due to issues with bioavailability, stability, toxicity, or delivery methods.

So, where do we go from here? While machine learning approaches have made significant strides in predicting ACPs, there is a clear need to augment these methods with other research avenues. For instance, beyond needing a more comprehensive and balanced datasets for training our models, machine learning should not be viewed as a standalone solution but rather a part of a broader, more integrative approach. Combining machine learning predictions with experimental validation and other computational methods could lead to more robust and reliable identification of potential ACPs. Complementing machine learning with other computational methods, such as molecular dynamics simulations, can also enhance the process of ACP identification. Molecular dynamics simulations can illustrate the intricate details of how a peptide interacts with cancer cells at a molecular level. They can shed light on the structural characteristics of the peptide, its binding affinity, and how it interacts with the target cells. Such insights could provide an additional layer of validation for machine learning predictions, and even guide the design of more effective peptides. Moreover, adding data from genomics, proteomics, and metabolomics into the mix could provide a more comprehensive picture of the peptide's potential anticancer activity. By taking into account the broader biological context, we can better understand and predict how a peptide would behave in the complex and dynamic environment of a living organism. Incorporating methods like wet lab experiments and in vitro and in vivo testing into the process can provide empirical evidence of a peptide's anticancer activity. These experimental validations can help confirm machine learning predictions with real-world data, reinforcing the reliability of the identified ACPs. Furthermore, it can also provide valuable feedback to refine and improve the machine learning models, making them more adept at predicting successful ACPs over time.

In conclusion, while machine learning has proven invaluable in identifying potential ACPs, we must be mindful of its limitations and the challenges inherent in translating these predictions into clinical applications. The journey from a promising peptide to a viable anticancer treatment is a long and complex one, and machine learning is but one tool in our arsenal. By addressing the current challenges and continuing to innovate, we can enhance machine learning's role in this vital area of research and take another step closer to the ultimate goal: effective, targeted cancer treatments.

Acknowledgements

I attest that the ideas, graphics, and writing in this paper are entirely my own unless explicitly stated otherwise. I thank Matthew Aronson from the University of Pennsylvania and the Children's Hospital of Philadelphia and Indigo Research for their mentoring and guidance.

Works Cited

- Cancer.net. (2023). Breast Cancer Statistics. Retrieved from <https://www.cancer.net/cancer-types/breast-cancer/statistics>
- P. M. Dillon, G. R. Petroni, M. E. Smolkin, D. R. Brenin, K. A. Chianese-Bullock, K. T. Smith, W. C. Olson, I. S. Fanous, C. J. Nail, C. M. Brenin, E. H. Hall, Cr. L. Slingluff Jr. A pilot study of the immunogenicity of a 9-peptide breast cancer vaccine plus poly-ICLC in early stage breast cancer. Retrieved from <https://jitc.bmj.com/content/5/1/92>. (2012)
- B. Sveinbjörnsson, K. A. Camilio, B. Erik Haug, Ø. Rekda. LTX-315: a first-in-class oncolytic peptide that reprograms the tumor microenvironment. Retrieved from <https://pubmed.ncbi.nlm.nih.gov/28490192/>. (2017)
- T. Gao, Q. Cen, H. Lei. A review on development of MUC1-based cancer vaccine. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0753332220310805>. (2020)
- J. Baar, W. J. Storkus, J. Finke, L. Butterfield, H. Lazarus, J. Reese, K. Downes, T. Budd, A. Brufsky, P. Fu. Pilot trial of a type I - polarized autologous dendritic cell vaccine incorporating tumor blood vessel antigen-derived peptides in patients with metastatic breast cancer. https://jitc.bmj.com/content/3/Suppl_1/P3 (2015)
- L. Yu , J. Tang, C. Zhang, W. Zeng, H. Yan, M. Li, X. Chen. New Immunotherapy Strategies in Breast Cancer. https://www.scienceopen.com/document_file/e8784502-3d41-4e0d-a427-ff6b75bad9f8/PubMedCentral/e8784502-3d41-4e0d-a427-ff6b75bad9f8.pdf (2017)
- E. Ylosmaki, T. Ranki, P. Priha, C. Backman, M. Vaughn, V. Cerullo, S. Pesonen. Abstract B123: Local treatment with PeptiCRAd-1, a novel cancer immunotherapy approach, mediates a systemic antitumour CD8+ T-cell response and infiltration of CD8+ and CD4+ T-cells into distant untreated tumors in a clinically relevant humanized mouse melanoma model. https://aacrjournals.org/cancerimmunolres/article/7/2_Supplement/B123/469382/Abstract-B123-Local-treatment-with-PeptiCRAd-1-a. (2019)
- E. Ylosmaki, T. Ranki, P. Priha, C. Backman, M. Vaughn, V. Cerullo, S. Pesonen. Abstract B123: Local treatment with PeptiCRAd-1, a novel cancer immunotherapy approach, mediates a systemic antitumour CD8+ T-cell response and infiltration of CD8+ and CD4+ T-cells into distant untreated tumors in a clinically relevant humanized mouse melanoma model. https://aacrjournals.org/cancerimmunolres/article/7/2_Supplement/B123/469382/Abstract-B123-Local-treatment-with-PeptiCRAd-1-a. (2019)
- A. Säfholm, J. Tuomela, J. Rosenkvist, J. Dejmek, P. Härkönen, T. Andersson. The Wnt-5a-derived hexapeptide Foxy-5 inhibits breast cancer metastasis in vivo by targeting cell motility. <https://pubmed.ncbi.nlm.nih.gov/18927296/>. (2008)
- SELLAS Life Sciences Group. Galinpepimut-S by SELLAS Life Sciences Group for Triple-Negative Breast Cancer (TNBC): Likelihood of Approval. <https://www.pharmaceutical-technology.com/data-insights/galinpepimut-s-sellas-life-sciences-group-triple-negative-breast-cancer-tnbc-likelihood-of-approval/> (2023).

- M. Demeule, C. Charfi, J. Currie, A. Larocque, A, Zgheib, S. Kozelko, R. Béliveau, C. Marsolais, B. Annabi. 2TH1902, a new docetaxel-peptide conjugate for the treatment of sortilin-positive triple-negative breast cancer. <https://pubmed.ncbi.nlm.nih.gov/34314556/>. (2021)
- M. Demeule, C. Charfi, J. Currie, A. Larocque, A, Zgheib, S. Kozelko, R. Béliveau, C. Marsolais, B. Annabi. 2TH1902, a new docetaxel-peptide conjugate for the treatment of sortilin-positive triple-negative breast cancer. <https://pubmed.ncbi.nlm.nih.gov/34314556/>. (2021)
- Bionity.com. Polysaccharide K. <https://www.bionity.com/en/encyclopedia/Polysaccharide-K.html>. (2023)
- Phase II Study of Allo LMI Vaccine With IL-2 for Stable Metastatic Breast Cancer. <https://beta.clinicaltrials.gov/study/NCT00784524?tab=results> (2019)
- W. Shoombuatong, N. Schaduangrat, C. Nantasenamat. Unraveling the bioactivity of anticancer peptides as deduced from machine learning. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6123611/>. (2018)
- F. Li, X. Wang. Identifying anticancer peptides by using improved hybrid compositions. Retrieved from <https://link.springer.com/content/pdf/10.1038/srep33910.pdf> (2016).
- J. P. Brownlee. A Gentle Introduction to Ensemble Learning Algorithms. Retrieved from <https://machinelearningmastery.com/tour-of-ensemble-learning-algorithms/> (2021).
- S. Katoch, S. S. Chauhan, V. Kumara. A review on genetic algorithms: past, present, and future. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7599983/>. (2020)
- R. Polikar. Ensemble based systems in decision making. Retrieved from <https://www.semanticscholar.org/paper/Ensemble-based-systems-in-decision-making-Polikar/1a585a498551e9ba55f89207f8a735cfd79cf807> (2006).
- M Alhamid. Ensemble Models. Retrieved from <https://towardsdatascience.com/ensemble-models-5a62d4f4cb0c> (2021).
- A. Joby. What Is Cross-Validation? Comparing Machine Learning Models. Retrieved from <https://learn.g2.com/cross-validation> (2021).
- S. Akbar, M. Hayat, M. Iqbal. iACP-GAEnsC: Evolutionary genetic Algorithm based Ensemble Classification of Anticancer Peptides by utilizing Hybrid Feature space. Retrieved from https://www.researchgate.net/publication/317694554_iACP-GAEnsC_Evolutionary_genetic_Algorithm_based_Ensemble_Classification_of_Anticancer_Peptides_by_utilizing_Hybrid_Feature_space (2017)
- E. Zakharova, M. Orsi, A. Capecchi, J. Reymond. Machine Learning Guided Discovery of Non-Hemolytic Membrane Disruptive Anticancer Peptides. Retrieved from <https://chemistry-europe.onlinelibrary.wiley.com/doi/full/10.1002/cmdc.202200291> (2022).
- F. Grisoni, C. S. Neuhaus, G. Gabernet, A. T. Müller, J. A. Hiss, G. Schneider. Designing Anticancer Peptides by Constructive Machine Learning. ChemMedChem, 13(13),

- 1300-1302. doi: 10.1002/cmdc.201800204. Retrieved from <https://pubmed.ncbi.nlm.nih.gov/29679519/> (2018).
- C. Angermueller, T. Pärnamaa, L. Parts, O. Stegle. Deep learning for computational biology. *Molecular Systems Biology*, 12(7), 878. doi: 10.15252/msb.20156651. Retrieved from <https://www.embopress.org/doi/full/10.15252/msb.20156651> (2016).
- X. Chen, X. X. Zhang, Y. Liu, Z. X. Tan, Z. Shang. ACP-DA: Improving the Prediction of Anticancer Peptides Using Data Augmentation. *Frontiers in Bioengineering and Biotechnology*, 9, 710041. doi: 10.3389/fbioe.2021.710041. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8279753/>. (2021)
- W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, B. Yu. Interpretable machine learning: definitions, methods, and applications. arXiv preprint arXiv:1901.04592. Retrieved from <https://arxiv.org/abs/1901.04592>. (2019)
- J. Chen, X. L. Liu. xDeep-AcPEP: Deep Learning Method for Anticancer Peptide Activity Prediction Based on Convolutional Neural Network and Multitask Learning. *Journal of Chemical Information and Modeling*, 61(7), 3278–3288. doi: 10.1021/acs.jcim.1c00181. Retrieved from <https://pubs.acs.org/doi/10.1021/acs.jcim.1c00181>. (2021)
- Q. Wu, Q. Wang, J. Jiang, T. Zhang, W. Chen, H. Lin, Q. Dai. Recent Progress in Machine Learning-based Prediction of Peptide Activity for Drug Discovery. *Current Medicinal Chemistry*, 26(5), 883-900. doi: 10.2174/0929867325666180123112834. Retrieved from <https://pubmed.ncbi.nlm.nih.gov/30674262/>. (2019)

The Primary Market of Concert Ticketing: Evaluating Live Nation Entertainment's Involvement with Recent Inefficiencies By Parker Upton

Abstract

The record-high prices of concert tickets of 2023 on the primary market have frustrated millions of music fans with their fingers often pointing towards Live Nation Entertainment (LNE), the live entertainment conglomerate that resulted from Live Nation and Ticketmaster's 2010 merger. Since this merger, LNE has maintained a substantial market share that critics argue is the source of recent inefficiencies. However, there are many components contributing to the current conditions of the concert ticketing industry's primary market. After examining the 2010 merger's ramifications and the factors contributing to the aftereffects of the merger, I determined that pent-up demand, a pandemic-weakened supply chain, and rising concert production costs most accurately explain recent events, not LNE's market power. Despite claims, the average ticket price has not severely increased in 2023; the increase in ticket prices is gradual but is heavily concentrated among the top ten music tours of 2023 which is due to dynamic pricing that follows the latest trends in consumer demands. Demand for concerts is higher than ever and outweighs the effect of LNE's potential stronghold over the market. Furthermore, economies of scale in the ticketing market explain why LNE's presence is not detrimental but often beneficial to the live music industry.

Introduction

Concert tickets were not always astronomically expensive. In 2001, Paul McCartney's \$250 concert tickets were considered highly priced (Princeton University News, 2002). But, in 2023, fans are paying well into the thousands to secure seats to their favorite artists' tours, including Taylor Swift and Harry Styles. These record-high ticket prices on today's primary market have frustrated millions of live music fans, many of whom have their fingers pointed towards Live Nation Entertainment (LNE). A heavily documented example of this pin-the-blame-game is Taylor Swift's 'Eras' Tour tickets, primarily moderated through one of LNE's companies, Ticketmaster.³⁸ When music fans attempted to access pre-sale tickets for the Taylor Swift concert, many were locked out of the pre-sale due to the influx in site accesses that overwhelmed Ticketmaster's site; this led to several site glitches, and eventually, Ticketmaster was forced to cancel its general sale due to insufficient ticket inventory. As a result, the Department of Justice (DOJ) has opened an antitrust investigation against LNE. The DOJ previously investigated antitrust behavior of LNE when its companies Ticketmaster and Live Nation announced a merger in 2009, and one of the DOJ's main requisites for the merger was that LNE would improve the primary ticket market's efficiency. For the past decade, LNE kept the concert industry expanding at an exponential pace while increasing live music's economic efficiency through dynamic pricing and artist package deals (aspects that will be further

³⁸ The company LNE is a product of the merger between ticketing company Ticketmaster and artist/venue management company Live Nation. LNE is Ticketmaster's parent company. For the sake of legibility, Ticketmaster and Live Nation will be referred to as LNE unless information specifically pertains to one of the companies.

discussed). However, recent altercations concerning Taylor Swift's concert ticketing have called the merged company's efficiency into question.

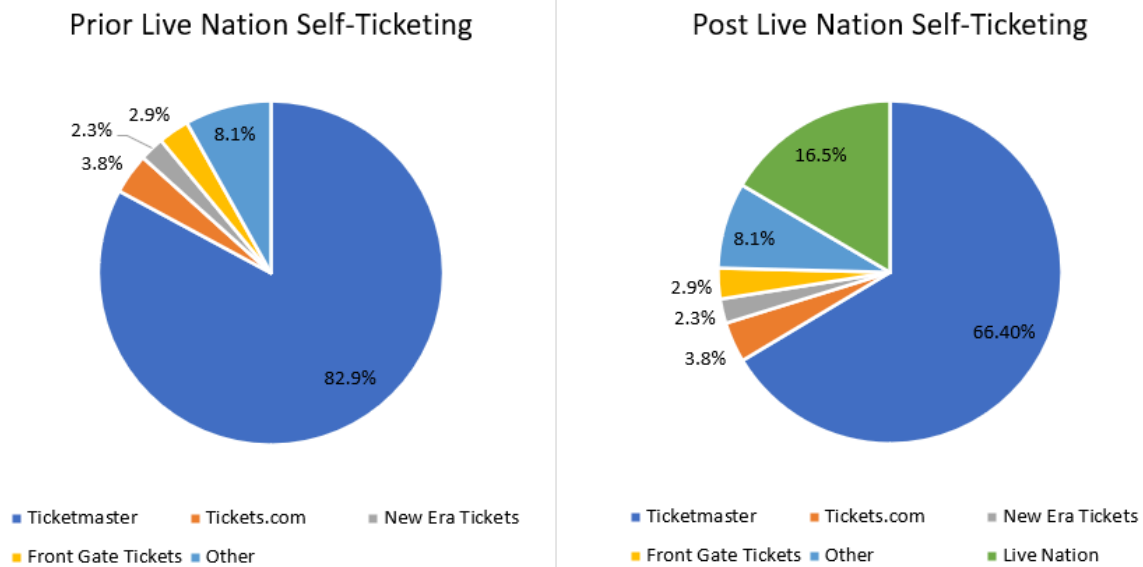
LNE, however, is only one of the many moving gears that was involved recent inefficiencies in live music entertainment. This paper evaluates data on previous concert primary market ticket prices and shows how market trends closely align with two factors: pent-up demand and monopoly. If recent market inefficiencies were catalyzed by LNE's monopoly over concert ticketing, trends in data would more closely align with a decrease in ticket quantity with a simultaneous increase in price. On the other hand, pent-up demand would cause an increase in ticket quantity as well as an increase in price. However, ticket price is only one part of the problem; music fans are also wary of 'hidden' fees. The general public understands these ticket fees as a representation LNE's monopoly, but data on production costs connect fee increases to the COVID-19 pandemic and the trend of more elaborate concerts. This paper examines these ticket fees and ticket prices of recent concerts through LNE and will attempt to identify the principal factor among these variables. By reviewing the key elements of the 2010 merger that created LNE and examined the company's activity ever since, this paper analyzes whether LNE is the main cause for topical issues in the primary market of concert ticketing and if their unbalanced market power must be considered when confronting this problem.

The Merger

To understand LNE's influence on the primary market, we must first examine the company's origin: a merger between Ticketmaster and Live Nation. In the early 2000s, prior to the merger, Ticketmaster was concerned with ticketing whereas Live Nation concentrated on acquiring venues and music artists. Live Nation would use Ticketmaster as a ticketing service for its venues as their largest customer; however, this dynamic significantly changed when Live Nation decided to let their ticketing service contract with Ticketmaster expire. Live Nation saw greater profit from self-ticketing and leased ticketing technology from CTS, a European ticketing service, in order to commence its entrance into the large venue ticketing market in December 2007 (Brandle, 2007).

Ticketmaster took a great hit from its largest client becoming its biggest competitor. Prior to Live Nation practicing self-ticketing, Ticketmaster controlled over 80% of the ticketing services for large concert venues. However, when Live Nation began self-ticketing, Ticketmaster's share dropped to 66.4%, and Live Nation became the second top ticketing service with a 16.5% share achieved nearly overnight (Figure 1).

Figure 1: North American Ticketing Companies



This represents Ticketmaster's market power when ticketing most popular North American venues. On the left is the percentage of most popular venues Ticketmaster ticketed prior to Live Nation entering the industry. On the right is what happened after Live Nation entered.

Source: ©DOJ 2010 Antitrust Case Filings—Complaint 224

After several failed attempts to regain its lost market share through offering alluring contracts to venues, Ticketmaster merged with Live Nation to create Live Nation Entertainment (LNE) in February 2009. This merger sent shockwaves throughout the music industry as Ticketmaster finally encountered a genuine competitor only for that competition to become its parent company. Due to the anti-competitive advantage of LNE, there were skeptics on the ethics behind allowing such a company to originate. Thus, an antitrust investigation was launched in 2009 by the DOJ in order to determine whether Ticketmaster and Live Nation should be permitted to merge.

While the merger was eventually allowed, it is important to note the full extent of the DOJ's assessment surrounding LNE's anti-competitiveness. The fundamental allegations that prompted the antitrust investigation included that 1) Ticketmaster and Live Nation merging would cease healthy competition between two ticketing services, and 2) the merger would allow these companies to both vertically and horizontally integrate. As previously mentioned, Ticketmaster was a ticketing company while Live Nation had a stake in artist management, promotion, venues, and later ticketing. If these two companies merged, LNE would dominate the ticketing industry and have substantial market power in the remaining parts of the live music industry. Critics alleged that a monopoly would originate from this merger. During the investigation, Ticketmaster and Live Nation denied these allegations with four key points (Kwoka, 2014):

- 1) Ticketmaster had significant competition and did not fully control the market; its venue contracts expire after 2-3 years, and it must constantly fight for renewal against other ticketing services during the expiration periods.
- 2) Live Nation was not a true competitor due to the limitations of using ticketing software developed by CTS.
- 3) Vertical integration of Ticketmaster-Live Nation would not be weaponized.
- 4) The merger would solve inefficiencies in the business.

The DOJ's response varied for each of LNE's points. They disregarded the first reason and determined Ticketmaster did, in fact, have market power for ticketing large concerts. The second reason faced a similar future since Live Nation had previously mentioned adapting CTS software to better suit large venue ticketing and that the software already allowed the company to gain a near-20% share of the concert ticketing market. The third reason could not be measured at the time of the investigation and would later be incorporated into the consent decree. On the other hand, the DOJ interpreted the fourth contention in LNE's favor; they acknowledged that Ticketmaster and Live Nation merging would increase efficiency. To defend this point, LNE argued that it would remedy any unnecessary segments within the concert industry that would save businesses time and fans money. LNE would also be able to provide "packages" for artists that would be more efficient than artists having to go to multiple companies for concert production and, most importantly, the introduction of dynamic pricing which will be further elaborated in this paper when discussing recent market trends.

The DOJ was uncertain how the proposed benefits required a merger but logically assumed that a merger would aid in these efficiencies manifesting faster. With the belief that increased efficiency would outweigh anti-competitive aspects, the DOJ concluded that the merger of Ticketmaster and Live Nation was allowed with certain conditions listed under the consent decree. The consent decree is summarized in the following points (Varney, 2015):

1. Ticketmaster would license Host, its ticketing platform, to AEG, the second largest concert promoter in the country as of the merger period, for up to five years with a below-market royalty rate. The DOJ aimed to use the transfer of Host to AEG to add more competition among the American concert ticketing industry.
2. Ticketmaster must relay its Paciolan division, the division that oversaw the specialized software used for ticketing large venues, to Comcast-Spectator, a venue operator and provider of ticketing services, to New Era Tickets (company under Comcast-Spectator). Like the first condition, DOJ hoped to lower the barrier of entry for potential competitors of LNE.
3. Prohibition of A) retaliation against venues that choose a ticketing service other than Ticketmaster B) a venue must use Ticketmaster if it were to use Live Nation artists/concerts C) a venue must use Live Nation artists/concerts if using Ticketmaster D) transferring data from ticketing to non-ticketing related parts of LNE.

As seen, each aspect of the consent decree aimed to create a merger between Ticketmaster and Live Nation that would still allow other companies to compete for a chunk of the market. While the first two parts of the decree were unopposed, music fans questioned how the prohibitions section would weather through the years. It was difficult to determine if the DOJ would effectively enforce these prohibitions. Furthermore, the DOJ explicitly stated in their investigation that Ticketmaster had market power, and thus LNE would dominate concert ticketing if the consent decree did not accomplish what the DOJ had hoped.

Concert Ticketing: Post-Merger

On July 30th, 2010, Ticketmaster and Live Nation officially merged, becoming LNE. The DOJ ended their investigation but urged companies in the ticketing industry to report any anti-competitive behavior of LNE that violated the merger agreement. Such behavior, however, would be unnecessary when one of LNE's to-be adversaries opted to leave the market and the other failed to gain substantial market power.

To elaborate, after Comcast-Spectator acquired Paciolan, there was no sign of competitive movements involving Paciolan,³⁹ and as of today, their ticketing company—New Era Tickets—neither controls a substantial portion of the market nor is among LNE's top competitors. On the other hand, AEG was ready to compete with LNE, but they decided to not use Host; instead, the company teamed up with Outbox Technology, an upstarting ticket company, to create the ticket outlet AXS in 2011. AXS would go on to facilitate a Beyonce concert at the Staples Center in 2013 and become one of LNE's top competitors (Martens and Lowery, 2013). Even as an adversary, however, AEG was never able to replicate Live Nation's pre-merger market share of 16.8%. It is logical to assume that the outcome of the consent decree's first two conditions did not reach the DOJ's expectations; Comcast-Spectator did not even attempt to engage in competition with Ticketmaster, and AEG was not overly successful.

The third condition of the consent decree had different results. The DOJ rarely found LNE dabbling in any prohibited anti-competitive actions. Most claims of LNE violating merger agreements were reported by AEG, but these complaints lacked substance. With that being said, it must also be determined whether the merged company kept their promise of increased efficiency. LNE's efficiencies and inefficiencies are best understood through an examination of market activity in recent years.

The year preceding the pandemic was another growth-filled year for the concert industry. The average gross per show of worldwide top one hundred tours jumped from \$837,000 to \$915,000 while the average number of tickets sold per show went from 9,300 to 9,900 (Pollstar, 2019). 2019, like most years post-merger, saw growth in both concert revenue and attendance. This growth was projected to continue until the COVID-19 pandemic hit the United States. With the Centers for Disease Control and Prevention and the World Health Organization advising people to only venture out for necessities and implementing social distancing policies, live

³⁹ Comcast-Spectator is no longer a large company in the live music industry. The company is presently more involved in live sports and manages a National Hockey League team.

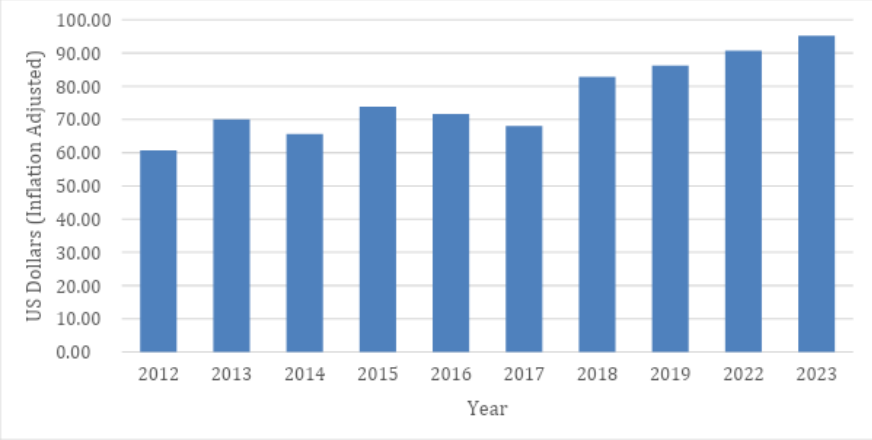
concerts became near-impossible, as well as unsafe, to produce. This is when LNE looked towards alternatives.

Although concert venues seldom opened their doors in 2020, LNE remained busy putting on shows digitally. Live-streamed shows, however, can only imitate the genuine experience so much. Past concert goers yearned for the pre-pandemic world of live music. These people bid their time until the government would give the live entertainment industry the thumbs-up to resume normal activity.

Towards the end of 2021, large venues began to re-navigate show production. During the beginning of 2022, people were still reluctant to attend concerts due to health concerns, but the latter half of the year marked a dramatic surge in ticket sales. Combining low COVID-19 infection rates with several popular artists going on tour, concert goers were leaping at a chance to obtain tickets to concerts of Billie Eilish, Bad Bunny, Coldplay, Bon Jovi, The Weeknd, and Harry Styles. In 2022, fans paid anywhere from \$190 to \$3,000 for a Harry Styles concert ticket (Erdenekhuyag, 2022) and floor-level tickets for the Weekend’s November 26th show were \$489 before fees (Levy, 2022). Price, however, was not the only aspect of the concert industry that exhibited upward trends; attendance also skyrocketed. In quarter three of 2022, LNE reported that its venues reached their highest quarterly attendance, hosting 11,000 concerts for over forty-four million fans while the number of stadium events tripled in amount compared to quarter three of 2019 (Stassen, 2022).

By the end of 2022, it was safe to assume that LNE had bounced back from the pandemic. Excluding the pandemic-affected years of 2020 and 2021, the live music industry is continuing to see an increase in both the price of tickets and concert attendance from the beginning of the merger, all the way to 2023 (Figure 2).

Figure 2: Average Concert Ticket Price in North America



After adjusting for inflation using the FRED’s PCIPE data, this represents the average ticket price of top 100 North American concerts from the years 2012-2023 with 2020-2021 data being excluded due to insufficient concert activity caused by the pandemic.

Source: Pollstar 2018/2022 End of Year Business Analysis and 2023 Mid-Year Business Analysis

Interestingly, this data does not seem to fully match the unreasonable prices reported for recent concerts such as Taylor Swift or Bill Springsteen; the average ticket price is only gradually rising each year. Albeit gradually, prices are still rising, and music fans are left wondering why these prices keep rising if it cannot fully be explained by inflation. There are also 2023’s highest-grossing tour tickets, which have absurdly high prices for the primary market. These aspects have led the public to question LNE’s efficiency; if LNE oversees ticketing for most concerts in the United States, the company should have the capability to release prices that are consumer friendly. There are also concerns regarding Ticketmaster’s site repeatedly malfunctioning when users attempted to buy Taylor Swift tickets; skeptics question whether a company with a “faulty algorithm” maintains market power because they are the best at what they do or if they are utilizing their monopoly to maintain control.

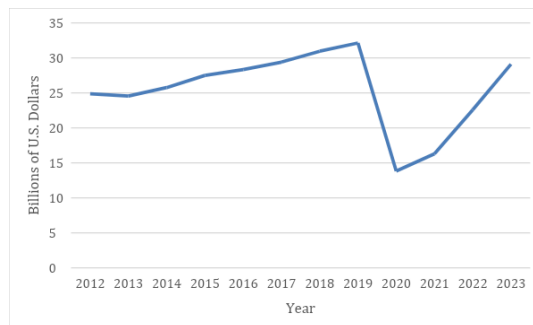
To discern the cause behind these hefty prices for the top concerts of 2023, pent-up demand and monopolization need to be examined.

Factor One: Pent-Up Demand

Pent-up demand is when buyers hold off from spending due to an economic or environmental change and then revert to exaggerated spending habits when there are signs of an economic recovery. This is known as “revenge spending,” or when consumers are previously denied optimal conditions to purchase a commodity and thus spend more than average to make up for lost time.

The recent global pandemic caused a pause in concert ticket purchasing. When the pandemic began in 2020, global ticket sales substantially dropped to 6.7 million (Gotting, 2022). As the situation to attend concerts became ideal come 2022, the live music industry did not just bounce back, it reached an all-time high. In 2023, Live Nation’s (the venue/artist management company under LNE) \$3.1 billion first-quarter revenue was up 71% from 2019 levels, and CTS’s ticketing revenue increased by 163% to \$393 million (Peoples, 2023). The real Personal Consumer Expenditure (PCE) for live entertainment is also presenting signs of pent-up demand (Figure 3).

Figure 3: Real PCE for Live Entertainment, Excluding Sports



Displays Real PCE for Live Events that exclude sports. It is adjusted for inflation using 2012 US dollars. From years 2012-2019, the US live entertainment market exhibited a steady growth until

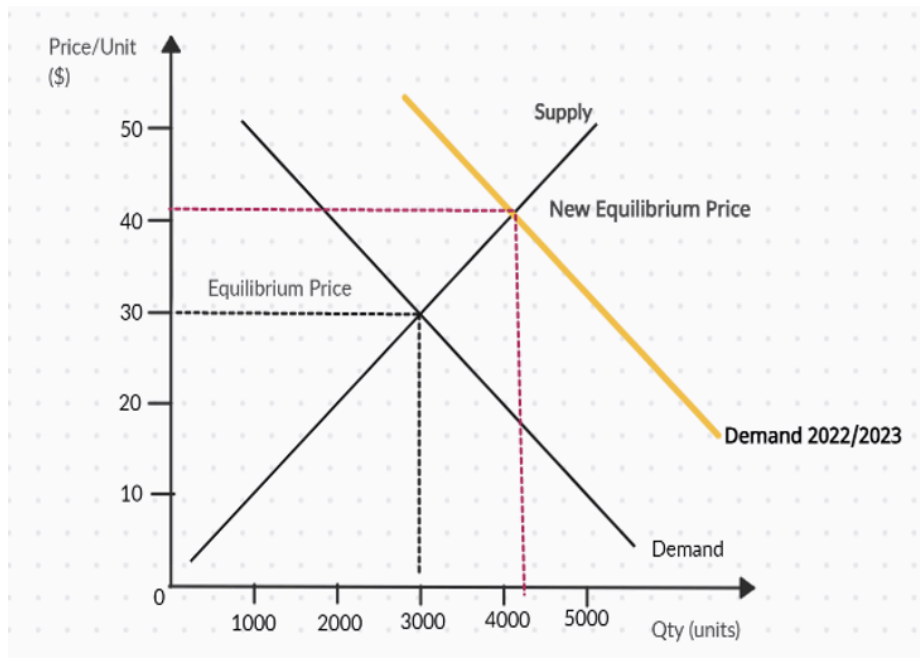
a dip in 2019-2020. From 2020-present, the graph depicts a rapid recovery in the market with a steep trend line.

Source: ©BEA Table 2.4.6U Real Personal Consumption Expenditure

The live entertainment industry steadily developed through the years 2012-2019 but experienced a dip during the pandemic that quickly recovered once the environment returned to a relative normal. Due to the increase at which the rate of the real personal consumption expenditure progressed, latest trends depict possible pent-up demand. This pent-up demand is best understood by comparing the slope of the line segment representing the real PCE of years 2012-2019 and the slope of the 2020-2023 segment. 2012-2019 has an approximate slope of 1.1245 and 2020-2023 with a slope of 5.216; the rate of growth of live entertainment nearly quintupled after COVID-19 no longer posed a threat to consumers attending concerts.

As an effect of this pent-up demand in the live entertainment industry, there is also an increase in the prices of tickets. Consumers, however, are still concerned that these price jumps are being heavily influenced by LNE's market power. To identify the effect of pent-up demand on the live music industry versus a monopoly, we will examine a simplified version of the ticketing market using the relationship between supply and demand. Figure 4 below represents the concert ticketing market prior to the price jumps of 2022-23 and if these price jumps are to be explained by an increase in demand.

Figure 4: Pent-Up Demand



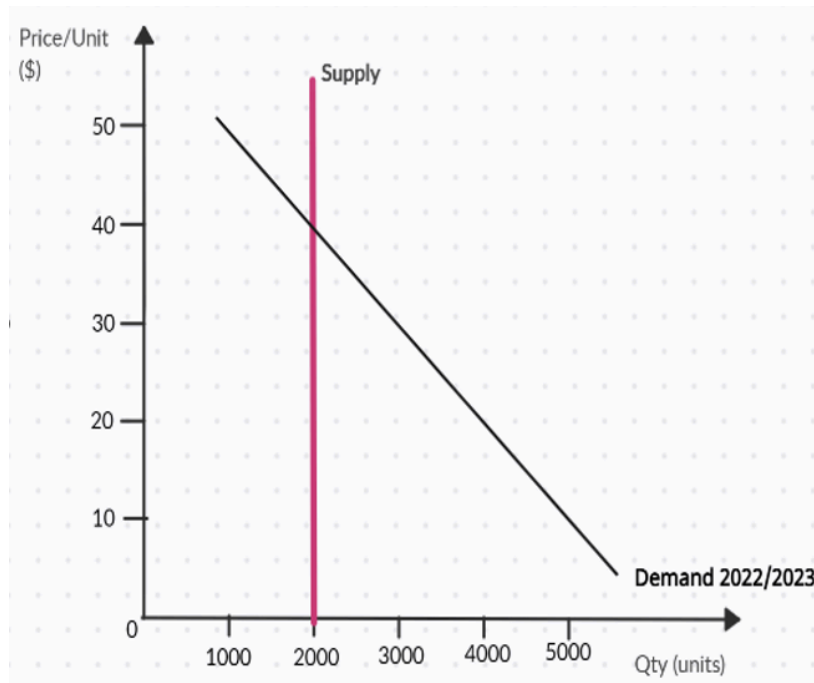
This graph is an illustration of concert ticketing pent-up demand, or a great increase in demand for attending concerts; the numbers in this graph solely function as visual aids and are purely hypothetical. The yellow line represents the new demand of 2023.

Source: Author

When there is a positive shift in the demand curve, both the quantity of tickets and the price of tickets would increase. Correspondingly, recent cases of demand involve music artists adding dates to their tours. In Harry Styles’s “Love on Tour,” Live Nation announced nineteen new Harry Styles tour dates in Europe, more shows in Austin and Chicago, and a venue upgrade in Lima, Peru (Aswad, 2022). This increase in ticket quantity accompanies Harry Styles’s astronomical ticket price range of 2022, as previously mentioned.

On the other hand, if increase of price were explained by a monopoly, the supply and demand graph would appear as follows:

Figure 5: Monopoly



A monopoly is usually illustrated with a supply and demand graph that only has a demand line, which is due to a monopoly being able to control the supply chain, and thus, put out as little or as many units as possible. But for the sake of this visual, we will add a hypothetical supply line that indicates LNE’s ticket supply.

Source: Author

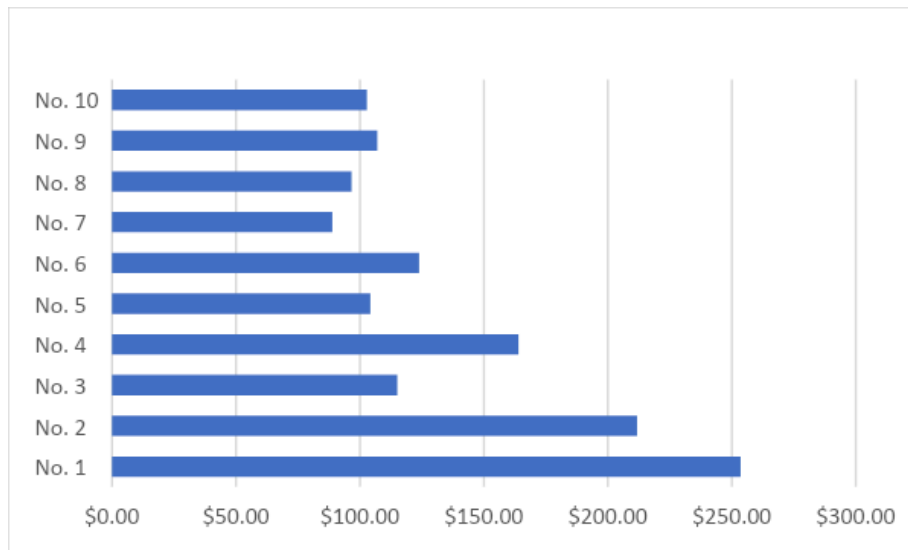
The problem with this visual is the fixed ticket amount. It is already established that artists are adding tour dates, and thus adding additional tickets into the market. If we are to use a monopoly as the primary factor behind recent price increases, it can be logically assumed that if more tickets were added to the market, the price would decrease since there is a greater quantity. Present cases, however, demonstrate that more tickets are selling than ever before with prices indicating zero signs of dropping. Even if we were to account for an increase in demand present during the monopoly, prices would nevertheless drop after artists continued to add tour dates; however, artists like Taylor Swift are only seeing an upward trend in the prices of their tickets even post the increase in ticket supply.

So, how do both prices and quantity rise with pent-up demand? While the rise in quantity is explained by music artists adding tour dates and selling out larger venues, the relationship between pent-up demand and price is still not fully addressed. When it comes to concerts managed by LNE, a ticket price is almost entirely based on the consumer. This is because LNE utilizes a system known as dynamic pricing. Employed by airlines and ticketing companies alike, dynamic pricing allows for a maximization of venue space; it is a system that ensures that most seats are filled. Instead of each seating area having a set price, the price of every seat follows a demand curve. The price of a seat may vary daily or hourly, depending on the concentration of demand for purchasing said seat.

In cases like Taylor Swift, so many people began purchasing seats that dynamic pricing followed suit and raised the initial primary ticket price of \$45 to \$449 (Cohen and Williams, 2023). Since then, the price for tickets on the primary market for the “Eras” tour have skyrocketed, but only because people are willing to pay those prices. This willingness to pay for high priced tickets is, again, a symptom of pent-up demand as well as an indication of the overall demand for these top artists.

The overall demand for these top artists is also a significant part of the equation. While pent-up demand is driving outrageous prices for the highest-grossing artists, previous data indicates that there has only been a steady increase in ticket prices for the top one hundred concerts in North America (revisit Figure 1). That is because pent-up demand is most concentrated among the top ten artists, and even among those artists, ticket prices vary.

Figure 6: Top Ten Most Popular Concert Tours of 2023



The chart represents the most popular tours of 2023 ranked from 1-10 and the average ticket price of each tour. (i.e., No. 1—Taylor Swift—has the most popular tour and the average ticket price of \$253.56).

Source: Pollstar Mid-Year Top Tours 2023

The top-grossing artist of 2023 has the highest average ticket price worldwide, but the dollar amount quickly decreases among these top ten highest-grossing artists (Figure 6). Artist

number 10 in 2023 tour rankings has a ticket price that is 2.5 times less than number 1 and half that of number 2, all the while only three of the top ten artists exceed the average concert ticket price of \$126 (not adjusted for inflation).

What happened here was the pandemic created desperation among live music fans. This desperation turned into pent-up demand. This pent-up demand mostly affected the top five artists because these are the artists a majority of music fans looked to in order to satisfy their desire to attend a concert once again. Since a majority of these music fans either wanted to attend concerts by Taylor Swift, Harry Styles, The Weeknd, or Bruce Springsteen, the pent-up demand most closely followed top artists such as these. With such a high concentration of demand, dynamic pricing gauged the willingness of how much fans were willing to pay for primary market tickets, and those numbers exceeded far into the hundreds.

Factor Two: Monopoly

LNE is a prominent player in the live music industry that, after enduring the pandemic, has continued to increase their revenue. The product of a merger between Ticketmaster and Live Nation, LNE has ultimately cornered the market as a multibillion-dollar company. There is a turning point, however, when cornering the market can be considered a monopoly. To decide if LNE acts as a monopoly, we will first establish their position in the live music industry and then assess whether this position has impacted recent events.

As of 2023, it is reported that LNE controls more than 70% of the ticketing and live events market (Hayes, 2023). This leaves only 30% of room left for competition. With one dominant competitor and little room left, lesser competition has filled the void; there are approximately 3,300 online event ticket sales businesses in the US as of 2023, and these companies together only make up 30% of the industry (IBISWorld, 2023). There are many ticketing companies as of today, yet none of them have managed to replicate the market share Live Nation in 2009 (%16.9) had as a competitor to Ticketmaster. This is when spectators begin to wonder why LNE has kept its presence; is it because they are the most efficient, or because they have a monopoly?

Arguing the monopoly side of the controversy requires analysis of the average concert ticket sold by LNE. A concert ticket has two main parts: the actual ticket and the ticket fees. The basic price of concert tickets for prominent artists is usually determined by the venue and artist. In cases like Taylor Swift, the venues that have contracts with LNE would collaborate with Taylor Swift to produce concert tickets that are fairly priced. These base prices, however, are not what rouses suspicion of monopoly among music fans; it is the fees. While artists and promoters share in the base ticket price, the fees are how venues and ticketing companies make money and, as critics claim, take advantage of fans. The four main fees charged—excluding tax—by LNE include service fees, order processing fees, delivery fees, and facility charge; all of which are only disclosed to the consumer at the final stages of ticket purchasing. Among these fees, Ticketmaster pockets service, order processing, and delivery fees, while venues (which are characteristically LNE venues) take the facility charge.

For years, music fans were irritated how these fees can easily turn a \$150 ticket into \$200. In 2016, there was an average of 21% ticketing fees for concert venues in New York for the primary market; this number went up to 27% in 2018 (Government Accountability Office, 2018), and in 2023, primary market fees have reached an estimate of 32% (Sisario, 2023). There has been a steady increase in fees over the years, and LNE has played an active role. Combining high fees with the company's site crashing when fans tried to purchase Taylor Swift tickets as well as the ever-rising prices of tickets, the public believes LNE's market power is beginning to outweigh the efficiency they bring to the market.

The increasing prices of tickets to high-grossing artists, however, has already been explained better by pent-up demand rather than monopoly. That leaves the other aspects of LNE's so-called inefficiency also open to debate. Ticket fees have risen over the years, but from 2018-2023, the increase is very gradual, similarly to the average ticket price. These trends are clarified by a factor entirely separate from monopoly and pent-up demand: the cost of concert production.

Concert production has become increasingly complex over the years when considering dancers, lighting, set layouts, heightened use of larger venues, special effects, elaborate sound systems, and the overall elevation of concert quality. In 2019, the stage of U2's 360° Tour cost about \$31 million to construct due to its 200-ton metal configuration (BBC News, 2009); the tickets to the 360° Tour were also U2's most expensive tickets of all time with an average price of \$144.93 if inflation-adjusted for 2023 dollars (Waddell, 2013). On the other hand, the average ticket price of that year for the top one hundred music tours was \$94.37 inflation adjusted (Pollstar, 2017).

What happens is the top-performing artists skew the average ticket price higher every year—thus, raising fees as well—because, every year, high-grossing concerts become more elaborate, and thus, more expensive. COVID-19 has only exacerbated this rise in fees; the pandemic caused several key parts of concert production to cost more than they did in past years. Singer/songwriter Santigold canceled her 2022 United States tour because “gas, tour buses, hotels, and flight costs skyrocketed” (Varga, 2023). Santigold is not the only one in the industry that shares this sentiment. The head of the Adams Avenue Business Association, an association that facilitates the live music events in the area, remarked, “Before... we paid \$12,000 dollars for renting five stages. This year, it costs \$24,500 for four stages... Band fees have increased, our port-a-potty rental fees have doubled, and the price we're paying for everything has gone up” (Varga, 2023).

While pent-up demand has been running its course and artists are eager to get back to performing, the technical side of the live music industry is struggling to match this eagerness with fair prices. LNE is not only a ticketing corporation; the company also manages venues and artists, and fees like the facility charges are geared towards helping the venue profit. In times where cost of production is experiencing a 30-300% increase, seeing fees increase should be considered normal when reflecting on the current situation (Eventech, 2023).

With cost and fees explained by cost of production, the current inefficiencies of the concert industry cannot be fully pinned on LNE. As for the other major inefficiency—Ticketmaster’s site malfunction for Taylor Swift tickets—it is more of a case of pent-up demand than anything else. For the first presale day, Ticketmaster sent 1.5 million people Verified Fan codes to join the sale, but there were 3.5 billion total system requests—four times the company’s previous peak—that included bot attacks and fans that did not have codes (Ticketmaster, 2022).

The Taylor Swift case was an unprecedented event in concert history; there is no current ticketing software that could have fully controlled 3.5 billion system requests happening in a day. The only way to determine whether this event is a demonstration of LNE’s monopoly is if the company does not continue to innovate to avoid similar events in the future. Lack of innovation would be a clear indicator that LNE is inefficient and does not control the market purely because they have a superior product; however, this can only be determined when LNE once again faces this issue.

Critics argue that if LNE has more substantial competition, the company would constantly be motivated to improve on the quality and cost of their service. While this assumption is logical for various industries, it falls short when applied to the live music industry due to it being an economy of scale. The prices of concert tickets and rate of fees ultimately hinges on the resources available to the ticketing company and/or venue. Spreading resources between several venues and ticketing companies is more likely to injure the market more than LNE maintaining its stronghold over the market. For example, in 2019, LNE paid artists more than \$6 billion, marking them as the largest financial supporter of musicians (Live Nation Entertainment, 2020). LNE would not be able to provide substantial support if they did not have their current resources. It would be highly unlikely that LNE could maintain constant prices after spending \$6 billion if they were thoroughly competing with other companies; because LNE reaps a vast revenue, the company is able to devote more time to artists which, in turn, benefits consumers. In cases such as LNE, market power only becomes an issue when it is repeatedly abused.

Conclusion

By evaluating recent primary market inefficiencies through the dissection of Ticketmaster and Live Nation’s merger agreement and the factors contributing to the aftereffects of the merger, I determined that LNE’s market power is not a viable explanation. LNE undoubtedly plays an influential role in the live music industry, but recent events cannot be fully explained by their majority market share. Primary market ticket prices of 2022 and 2023 rose significantly among the highest-grossing artists and was caused by pandemic-related pent-up demand and heightened cost of production. Concerts are becoming more expensive, and it is not only LNE that is contributing to this rise in prices.

In the 21st century, concerts have been on the rise in popularity among both fans and artists; however, the amount of people that want to participate in 2023’s live music scene

outmatch the economic capabilities of the concert industry. The economy is still recovering from the pandemic, and thus, prices for stage production are unprecedentedly high which leads to increased ticket prices and/or higher fees. Aside from the pandemic, concerts are also becoming more expensive because more artists are using larger venues, and the venue set up is becoming more costly. With so many moving parts that are causing an increase in ticket prices, it is difficult to pin LNE as the source of the problem. The data in the paper demonstrates consumer trends that best align with pent-up demand over monopoly, and the arguments for LNE having a monopoly are refuted by the increase in production costs. Therefore, the primary market for concert ticketing as of 2023 is not a product of LNE's market power but rather a product of the pandemic and change in demand.

Instead of continuing to blame corporations such as Ticketmaster for all the mishaps that come with ticketing in the live music industry, it is essential that we examine our own role in economic changes. Consumers hold insurmountable influence when it comes to determining the price of goods in a capitalistic market; when there are cases of pent-up demand like in live entertainment, consumers are the driving force behind an increase in prices. People become willing to pay a higher price, and sellers respond by raising prices accordingly. This relationship between buyers and sellers is often ignored in cases like LNE where the explanation of there being an abusive monopoly reigns over all other ideas. We must understand the impact we have on prices before jumping to the conclusion that private corporations are the source of all market inefficiencies. By doing so, we will better comprehend our power as buyers and be able to assess similar events in the future with a more well-rounded perspective.

Works Cited

- Aswad, Jem. "Harry Styles Announces More Love On Tour Dates for 2023." *Variety*, 26 Aug. 2022, [variety.com/2022/music/news/harry-styles-more-love-on-tour-dates-2023-1235351823/](https://www.variety.com/2022/music/news/harry-styles-more-love-on-tour-dates-2023-1235351823/).
- Brandle, Lars. "Live Nation, CTS Team for Ticketing Platform." *Billboard*, 20 Dec. 2007, www.billboard.com/music/music-news/live-nation-cts-team-for-ticketing-platform-1315634/. Accessed 14 Sept. 2023.
- Cohen, Steven and Williams, Jordan. "How to Buy Taylor Swift 'Eras Tour' Concert Tickets." *Business Insider*, 4 Aug. 2023, www.businessinsider.com/guides/streaming/how-to-buy-taylor-swift-eras-tour-concert-tickets#:~:text=In%20the%20US%2C%20general%20ticket%20prices%20for%20Taylor,sit es%20are%20now%20charging%20much%20more%20for%20tickets.
- Erdenekhyag, Enji. "Here's How Much Tickets Are to Harry Styles' 6 Shows in Chicago." *NBC Chicago*, NBC Chicago, 3 Sept. 2022, www.nbcchicago.com/news/local/heres-how-much-tickets-are-to-harry-styles-6-shows-in-chicago/2931037/.
- EvenTech Systems. "Concert Attendance Rising, So Is the Cost of Production." *LinkedIn*, 6 April 2023, www.linkedin.com/pulse/concert-attendance-rising-so-cost-production-eventech-systems#:~:text=Concerts%20across%20the%20United%20States%20saw%20sold%20out,more%20costly%20than%20ever%2C%20ranging%20from%20a%2030-300%25increase.
- Götting, Marie. "Number of Music Tour Tickets Sold Worldwide 2011-2020." *Statista*, 8 June 2022, www.statista.com/statistics/380115/number-of-music-tour-tickets-sold-worldwide/.
- Government Accountability Office. "Management Report: Areas for Improvement in the Bureau of Consumer Financial Protection's Internal Controls and Accounting Procedures." *GAO*, September 2018, www.gao.gov/assets/gao-18-347.pdf.
- Hayes, Adam. "Is Ticketmaster a Monopoly?" *Investopedia*, 18 Feb. 2023, [www.investopedia.com/is-ticketmaster-a-monopoly-6834539#:~:text=How%20does%20Ticketmaster%20make%20money,%2C%20artists%2C%20and%20Ticketmaster%20its" Number of Businesses in Online Event Ticket Sales, United States.](https://www.investopedia.com/is-ticketmaster-a-monopoly-6834539#:~:text=How%20does%20Ticketmaster%20make%20money,%2C%20artists%2C%20and%20Ticketmaster%20its)
- IBISWorld*, 22 May 2023, www.ibisworld.com/industry-statistics/number-of-businesses/online-event-ticket-sales-united-states/#:~:text=There%20are%203,326%20Online%20Event,increase%20of%203.8%25%20from%202022
- Kwoka, John E. *The Antitrust Revolution: Economics, Competition, and Policy*. New York, Oxford University Press, 2014.
- Levy, Matt. "The Weeknd Has 2 Shows at La's SOFI Stadium: Last-Minute Tickets under \$75." *New York Post*, New York Post, 22 Nov. 2022, nypost.com/2022/11/22/how-to-get-tickets-to-the-weeknds-sofi-stadium-concerts-2022/.

- “Life at LN - Live Nation Entertainment.” *Live Nation Entertainment*, 2020, www.livenationentertainment.com/life-at-ln/.
- Martens and Lowery. "AEG's Randy Phillips: Michael Jackson 'just didn't seem to understand' money." *Los Angeles Times*, 4 Feb. 2013, <https://www.latimes.com/entertainment/music/la-xpm-2013-feb-04-la-et-ms-aeg-20130204-story.html>.
- Peoples, Glenn. "Concert Ticket Sales Remain Strong Through 2023." *Billboard*, 26 May 2023, www.billboard.com/pro/concert-ticket-sales-strong-through-2023/.
- “2017 Mid Year Business Analysis - Pollstar.” *Pollstar*, 2017, data.pollstar.com/chart/2017/07/2017MidYearBusinessAnalysis_584.pdf.
- "Mid-Year Analysis." *Pollstar*, July 2019, https://data.pollstar.com/chart/2019/07/MidYearAnalysis_759.pdf.
- "Economist Alan Krueger Examines Pricing Concert Tickets." *Princeton University News*, 24 Sept. 2002, <https://www.princeton.edu/news/2002/09/24/economist-alan-krueger-examines-pricing-concert-tickets>.
- Sisario, Ben. "Concert Ticket Prices Are on the Rise." *The New York Times*, 7 April 2023, www.nytimes.com/2023/04/07/arts/music/concert-ticket-prices.html.
- “Wembley Record ‘Broken by U2 Gig.’” *BBC News*, BBC, 15 Aug. 2009, news.bbc.co.uk/1/hi/entertainment/8202230.stm.
- Stassen, Murray. "Over 44m people attended Live Nation concerts in Q3 – its highest quarterly attendance ever." *Music Business Worldwide*, 4 Nov. 2022, <https://www.musicbusinessworldwide.com/over-44m-people-attended-live-nation-concerts-q3-its-highest-quarterly-attendance-ever12/>.
- Ticketmaster. "Taylor Swift: The ERAS Tour Onsale Explained." *Ticketmaster Business Solutions*, 19 Nov. 2022, business.ticketmaster.com/business-solutions/taylor-swift-the-eras-tour-onsale-explained
- Varga, George. "Cost of Doing Festivals, Concerts, and Tours Soared in 2022: A Year of Big Profits and Daunting Challenges." *The San Diego Union-Tribune*, 8 January 2023, www.sandiegouniontribune.com/entertainment/music/story/2023-01-08/cost-of-doing-festivals-concerts-and-tours-soared-in-2022-a-year-of-big-profits-and-daunting-challenges.
- Varney, Christine. “The TicketMaster/Live Nation Merger Review and Consent Decree in Perspective.” *Department of Justice*, 25 June 2015, www.justice.gov/atr/speech/ticketmasterlive-nation-merger-review-and-consent-decree-perspective.
- Waddell, Ray. “Billboard Power 100: U2 & Paul McGuinness.” *Billboard*, 7 Feb. 2013, www.billboard.com/music/music-news/billboard-power-100-u2-paul-mcguinness-1099191/.

Shattering the Glass Flask: A Women's Battle Against Sexism in STEM By Ashley Hurjak

Abstract

The growth of women in STEM fields has always been a force to be reckoned with. Men have tried to show through countless opportunities that they preside in all fields and have created barriers, such as the female pay gap, to display their dominance and preserve their confidence. Around 27% of women have the confidence to pursue a career in STEM and by breaking down the barriers of sexism, hopefully that number will increase exponentially.

1.0 Introduction

Less than 30% of women work in the fields of science, technology, engineering, and mathematics (STEM) in the globe, even though women make up almost half of the global population (Parks 2023). They publish in the Proceedings of the National Academy of Sciences of the United States of America at a rate comparable to males, and their research has about the same effect (Parks 2023). The representation of women in STEM varies by field. In biological areas, women predominate over males (Parks 2023). But in the fields of engineering, computer science, and physics, men predominate over women (Parks 2023) (**Figure 1**). STEM education and employment have historically been viewed as being exclusively for men. Moreover, sexual harassment, unconscious prejudice, and systemic discrimination can put an early stop to women's STEM careers (Parks 2023).

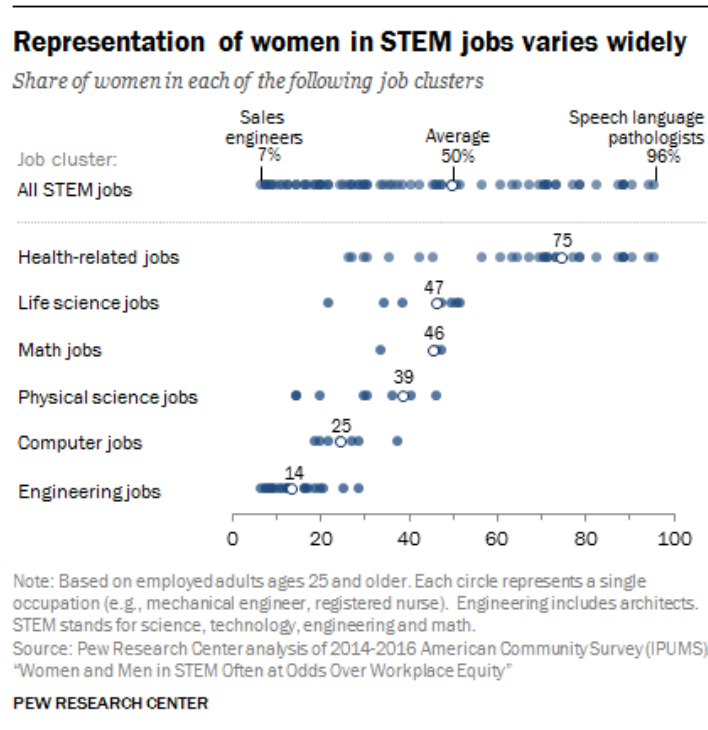


Figure 1 Graph displaying women in different fields of STEM.

1.1 Discrimination in the Field

The underrepresentation of women in STEM fields—science, technology, engineering, and mathematics—has garnered a lot of attention lately, but progress toward gender equality in these areas is lethargic (Casad et al., 2021). Even more concerning, these gender differences get worse when one looks at how many women are represented in academic STEM departments (**Figure 1**). The proportion of women in STEM faculty roles has not changed significantly, despite an increase in the number of women obtaining postgraduate degrees in recent years (Casad et al., 2021). Negative and widespread gender stereotypes, which may encourage hiring discrimination and limit possibilities for women's professional progression, are one reason for the lack of progress toward gender equity (Casad et al., 2021).

Women in STEM fields are also less likely to have access to social capital, such as support networks, which might hinder their chances of obtaining tenure and learning about grant funding sources (Casad et al., 2021). In addition, female STEM faculty members may feel that the academic atmosphere is hostile and intimidating, and they may experience unpleasant conflicts at work, including discrimination and sexual harassment (Casad et al., 2021). Even the mere existence of gender-biased indicators (such as "geeky" décor) in physically male-oriented areas might contribute to a feeling of exclusion from STEM fields (Casad et al., 2021).

2.0 Background on Sexism in STEM

It is possible to pinpoint the root cause of the gender gap in STEM fields as early as middle school (Swaford & Anderson, 2020). In mathematics, boys have historically scored better than girls, but in recent decades, the gender gap has closed and the inequalities are now insignificant (Hyde, Lindberg, Linn, Ellis, & Williams, 2008). According to the U.S. Department of Education's National Center for Education Statistics (NCES), 2007 report, females are now obtaining somewhat higher grades in math and science, nevertheless, boys are currently gaining credits at an equal rate (Swaford & Anderson, 2020).

Once students enter university, the gender disparity widens even further. Among entering freshmen, males outnumbered women by 29% to 15% in the pursuit of STEM degrees, despite the fact that women make up the majority of college students (National Science Foundation, 2009). While women made up 40% of full-time STEM faculty in degree-granting institutions in 2005, they made up fewer than 25% of faculty in the physical sciences (18%), computer and information sciences (22%), mathematics (19%), engineering (12%), and other fields (Swaford & Anderson, 2020).

2.1 Experiencing Imposter Syndrome

From a cultural standpoint, women in higher education encounter several obstacles to becoming leaders (Laux 2018). A lack of gender parity in professor ranks suggests hurdles to opportunity structures since access to higher education is a major factor in class mobility and achieving enhanced professional positions (Cama, Jorge, & Andrades, 2016). Research demonstrates that norms, organizational structures, and values work against women's

development in the academic workforce (Beauregard, 2012; Giesler, Kaminski, & Berkley, 2007).

Gender stereotypes in the workplace usually result in gender expectations for women to participate in more organizational citizenship activities, which are demonstrated by actions like helping coworkers, volunteering, serving on committees, and fostering interpersonal relationships (Beauregard, 2012). According to some (Dlamini & Adams, 2014; O'Meara, 2015), the male-dominated field of academia harbors patriarchal attitudes and ideologies that elevate men to leadership positions and a status of recognition over women while ignoring the abilities and merits of women (Laux 2018). This creates yet another obstacle to women's upward mobility in higher education.

In addition to these external obstacles to professional development, women face internal constraints that also lead to their underrepresentation in senior leadership positions in academia (Cama et al., 2016). Due to feelings and perceptions of being undervalued, overlooked, overworked, and the subjects of unequal treatment in a management climate that can be alienating in the academic setting, academic women's identities are frequently compromised, challenged, and made vulnerable (Vergauwe, Wille, Feys, DeFruyt, & Anseel, 2015). This research investigates the experiences of female academic staff members with impostor syndrome and working in higher education, particularly concerning their quest for tenure and promotion (Laux 2018).

2.2 Anxiety among Women in STEM

A model of the variables influencing career decisions is called social cognitive career theory, or SCCT. Adding occupational interest, career decisions, perseverance, and performance tasks to Bandura's (1986) SCT, Lent et al. (1994) built on it (Lapan, Shaughnesy, & Boggs, 1996). People's choices of majors are enabled by a combination of self-efficacy, interest, and personality (Larson et al., 2010). The claims made by Lent et al. regarding the SCCT of careers were as follows: (a) efficacy and interests are strongly correlated; (b) efficacy expectations mediate the relationship between achievement and interests; (c) efficacy beliefs influence the manifestation of goals and beliefs at the start and end of college; and (d) vocational interests, prior goals, and beliefs predict college major choice. This model was used in the study to investigate the relationship between interest and self-efficacy beliefs among female STEM majors, as well as the relationship between interest and college major choices.

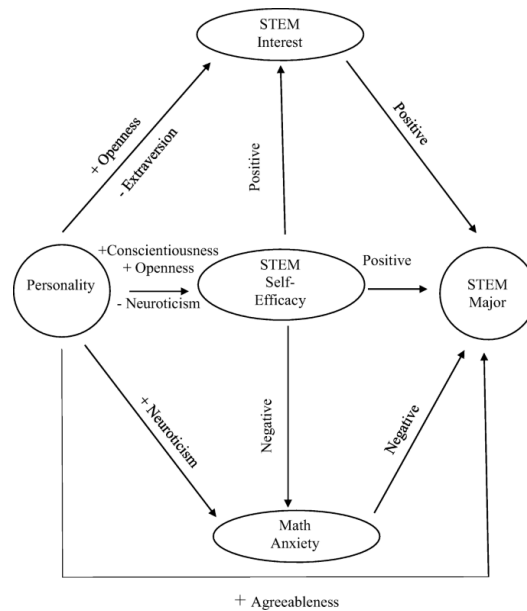


Figure 2: A chart showing the behavioral aspects of a Woman in STEM and how self-efficacy affects a woman’s personality in science.

<https://link.springer.com/article/10.1007/s41979-021-00050-6#Fig3>

2.2.1 Mathematics Anxiety Self-efficacy concerning various individual variables.

Self-efficacy completely enabled the aptitudes for interest relationships, and it was a significant element in STEM ambitions (Lent et al., 1994). The results of Lent et al. (1994) were extended by Lapan et al. (1996). According to Lapan et al. (1996), women's poorer math self-efficacy views led to a decline in their mathematical interests and investigative occupational efficacy beliefs (**Figure 2**). Interests decreased as a result of diminished ability beliefs. Lower efficacies were the cause of women's decreased interest in mathematics (Betz & Hackett, 1997; Lapan et al., 1996) (**Figure 2**).

Other variables were also predicted by self-efficacy. There was a strong correlation between mathematics self-efficacy and past achievement, prior high school mathematics courses, and mathematics anxiety (Hackett, 1985). Moreover, worry during task performance may lower endurance, composure, and contingent self-efficacy (Lent et al., 1994). Hackett and Betz (1982) discovered that mathematics anxiety was mediated by mathematics self-efficacy in a sample of undergraduate college students. According to Simon et al. (2015), there was a significant correlation between the reported reduced negative effect of STEM females and their higher levels of self-efficacy (NA; Simon et al., 2015).

In their 1986 study of secondary school girls, Eccles and Jacobs discovered three main negative effects of anxiety: (a) girls experienced higher levels of anxiety related to mathematics, and anxiety was found to be a significant predictor of mathematics grades and course intention; (b) anxiety hindered STEM intentions in a secondary setting; and (c) anxiety related to

mathematics was more strongly associated with future mathematics intentions (i.e., course taking) than mathematics aptitude and achievement were. Similar to Eccles and Jacobs, but with a broader scope, were the findings of Hackett (1985). According to Hackett, factors such as gender, the quantity of math taught in high school, math anxiety, and self-efficacy all strongly predicted students' choice of STEM-related college major (Mckinney 2016).

2.3 Hegemonic Masculinity in The Sciences

The upholding of behaviors that legitimize men's domination over women is referred to as hegemonic masculinity (Connell and Messerschmit 2007). These kinds of actions may be seen in the narrative titled "Blue Blazer Club." One professional society hosts a one-day specialist pre-conference before the main conference every year (Page et al., 2009). Traditionally, the top academics in the discipline provide their predictions for the field's future at the end of the day. Considering their numerical dominance in academia, it should come as no surprise that all of the distinguished professors selected in the 1970s and 1980s were men (Page et al., 2009). Because all of the distinguished men happened to be dressed in blue blazers and khaki trousers on a specific day, these academics came to be known as the Blue Blazer Club.

This research focuses on the tacit and unquestioning acceptance of institutional procedures and social rituals like the blue blazer club. Despite the language of complete inclusion, there is simply the appearance of full involvement in many sectors, including academics (Page et al., 2009). For instance, 84% of academic posts held by men were tenured or tenure-track, compared to 72% held by women (AAUP 2005). This indicates that males are more likely than women to hold tenured or tenure-track employment. Even while these figures show progress, there is still more to be done. For instance, women are disproportionately represented in non-tenure-track or contingent faculty posts, which are thought to be at the bottom of the academic hierarchy (Bergman and Waltman, 2009).

Gender parity was once believed to be attained after a sufficient number of competent women entered STEM disciplines, or the "pipeline" (Glazer-Raymo 1999). Women are still underrepresented at the full professor rank in science, despite advances in all areas, particularly biology, including doctorate degrees and associate professorships (Page et al., 2009). For instance, women only make up 19% of full professors and only 30% of tenured or tenure-track professors in doctorate awarding institutions (all scientific and engineering areas combined; 2003 figures reported in Burrelli, 2008). Only 5% of full professors in some professions, including engineering, are female. The percentages for math (8.6%) and physical sciences (8.3%) are much lower (Page et al., 2009).

Remarkably, Ash and colleagues (2004) discovered that women's slower rate of progression to full professor was not caused by productivity when they studied medical school faculty (Page et al., 2009). In conclusion, Marschke and colleagues ran several models predicting gender parity in faculty representation to the percentage of women with PhDs (40%) and discovered that there will never be more than 34% of women on faculty if institutions do not take any further action or steps to achieve gender equality; if they only hire women, it will take seven

years to reach 40% and another to reach exact gender parity; if they hire men and women equally (and retain, promote, and retire), it will take 20 years to reach 40% and 57 years to reach 50%.

3.0 Diminishing Sexism in STEM

Sexism in STEM stands as a prevalent issue in today's society. Women have been shunned out of opportunities that they have worked hard towards and have had no choice but to cap their passion and career aspirations.

3.1 Starting Early

Sexism in STEM and other male-dominated disciplines is not usually addressed until the late teen and early adult age. Women step into a science career unaware of the adversities at hand and realize soon their raised hand won't matter in a room full of individuals who believe women do not hold the same amount of value.

One way that this could be remedied is by teaching women at an early age what to expect in a science career and developing tools to face the adversities in the world of science. At elementary schools, a day or class period could be dedicated to learning about the lost voices of women scientists who have made significant contributions to STEM, such as Chien-Siung Wu, a nuclear and particle physicist. Educating young girls of women in STEM could be followed with an arts and crafts activity or a fun experiment such as elephant toothpaste, catalase breaking down hydrogen peroxide into gas and water. If young girls become aware of the discrimination women face, they will become more aware of the challenges included in a prospective STEM career.

3.2 Recognizing the Achievements of Women in the Workplace

At institutions, such as NASA, the amount of women in the workforce has increased exponentially compared to a decade ago. However, women are not being appreciated enough as the pay gap between males and females is significantly different. To exemplify, in 2013 women earned 82% or 87% of the paycheck that men earned in computer and mathematical engineering (Sterling et al., 2020).

To appreciate the contribution of women in STEM, there should be a designated work day in every STEM job for the women in the workforce, as they have also made important research contributions. Notable computer figures were Mary Jackson, Katherine Johnson, and Dorothy Vaughan. Their efforts guaranteed the mission's safety and assisted John Glenn, one of the first American astronauts, successfully orbiting the planet in 1962 (National Geographic). Their contributions have not been recognized as much as deserved to be and women in STEM workday could make the future women of STEM shine.

3.3 Amplifying the Voices of Women from the Past

As previously mentioned, various female scientists have gone unnoticed throughout history. There are barely any platforms that notice these impactful women and the contributions

they made throughout history. Creating a platform or database tailored specifically to these women would allow their contributions to gain the attention they deserve and inspire today's youth in STEM. This platform could be provided to the school's libraries and research projects could be assigned around it to educate young women about the possible contributions that they can make as well. Women all over the world can be integrated into the world of science with one easy click, creating many more meaningful advancements.

4.0 Conclusion

In all, women have experienced numerous struggles in science-related fields, and recognizing the importance of women in STEM fields should be applied to the workforce and society. Experiencing imposter syndrome, skill insecurities, and discrimination in the field; about 30% of women quit their jobs in STEM. A career in STEM is hard work over a myriad of years and it is devastating to see all that hard work going to waste. Incorporating women's appreciation days in the workplace, teaching young children the value of female skills, and recognizing the impactful women of the past could help future women navigating the pitfalls of careers in STEM feel more comfortable and appreciated. Following the indomitable discrimination towards women in science is a call to action for an inclusive future where every scientist, regardless of gender, can contribute to the mosaic of scientific accomplishment.

Works Cited

- American Association of University Professors (AAUP). (2005). Inequities persist for women & non-tenure-track faculty. *Academe*, 91.
- Ash, A.S., Carr, P.L., Goldstein, R., & Friedman, R.H. (2004). Compensation and advancement of women in academic medicine: Is there equity? *Annals of Internal Medicine*, 141, 205-212.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bergman, I., & Waltman, J. (2009, January). Satisfaction and discontent: Voices of non-tenure-track faculty. *On Campus with Women*.
- Casad, B. J., Franks, J. E., Garasky, C. E., Kittleman, M. M., Roesler, A. C., Hall, D. Y., & Petzel, Z. W. (2021). Gender inequality in academia: Problems and solutions for women faculty in STEM.
- Connell, R. W., & Messerschmidt, J. W. (2007). Hegemonic masculinity: Rethinking the concept. *Gender & Society*, 19, 829-859.
- Eccles, J. S., & Jacobs, J. E. (1986). Social forces shape math attitudes and performance. *Signs*, 11(2), 367-380.
- Glazer-Raymo, J. (1999). *Shattering the myths: Women in academia*. Baltimore: The Johns Hopkins University Press.
- Hackett, G., Casas, J. M., Betz, N. E., & Rocha-Singh, I. A. (1992). Gender, ethnicity, and social cognitive factors predicting the academic achievement of students in engineering. *Journal of Counseling Psychology*, 39(4), 527-538.
- Hackett, G., & Betz, N. E. (1982, March). Mathematics self-efficacy expectations, math performance, and the consideration of math-related majors. Paper presented at The Annual Meeting of the American Educational Research Association, New York, New York.
- Hackett, G. (1985). Role of mathematics self-efficacy in the choice of math-related majors of college women and men: A path analysis. *Journal of Counseling Psychology*, 32(1), 47-56.
- Laux, S. E. (2018). *Experiencing the imposter syndrome in academia: Women faculty members' perception of the tenure and promotion process* (Doctoral dissertation, Saint Louis University).
- Larson, L. M., Wu, T. F., Bailey, D. C., Gasser, C. E., Bonitz, V. S., & Borgen, F. H. (2010). The role of personality in the selection of a major: With and without vocational self-efficacy and interests. *Journal of Vocational Behavior*, 76, 211-222.
- Lapan, R. T., Shaughnessy, P., & Boggs, K. (1996). Efficacy expectations and vocational interests as mediators between sex and choice of math/science college majors: A longitudinal study. *Journal of Vocational Behavior*, 49(3), 277-291.

- Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interests, choice, and performance [Monograph]. *Journal of Vocational Behavior*, 45, 79–122.
- Lent, R. W., Brown, S. D., & Larkin, K. C. (1986). Self-efficacy in the prediction of academic performance and perceived career options. *Journal of Counseling Psychology*, 33(3), 265–269. doi:10.1037/0022-0167.33.3.265
- Marschke, R., Laursen, S., McCarl Nielsen, J., & Rankin, P. (2007). Demographic inertia revisited: An immodest proposal to achieve equitable gender representation among faculty in higher education. *Journal of Higher Education*, 78, 1-26.
- McKinney, J. L. (2016). Personality, Interests, Self-Efficacy, and Anxiety of Female STEM Majors: A Description, Comparison, and Prediction of Female STEM Majors.
- Page, M. C., Bailey, L. E., & Van Delinder, J. (2009). The Blue Blazer Club: Masculine Hegemony in Science, Technology, Engineering, and Math Fields. In *Forum on Public Policy Online* (Vol. 2009, No. 2). Oxford Round Table. 406 West Florida Avenue, Urbana, IL 61801.
- Parks, C. (n.d.). Women fighting stereotypes and systemic discrimination in STEM. Education. <https://education.nationalgeographic.org/resource/women-fighting-stereotypes-and-systemic-discrimination-stem/>
- Simon, R. A., Aulls, M. W., Dedic, H., Hubbard, K., & Hall, N. C. (2015). Exploring student persistence in STEM programs: A motivational model. *Canadian Journal of Education*, 38(1), 1–27.
- Sterling AD, Thompson ME, Wang S, Kusimo A, Gilmartin S, Sheppard S. (n.d.). The confidence gap predicts the gender pay gap among STEM graduates. *Proceedings of the National Academy of Sciences of the United States of America*. <https://pubmed.ncbi.nlm.nih.gov/33199594/>
- Women of NASA. Education. (n.d.). <https://education.nationalgeographic.org/resource/women-nasa/>

Investigating the Role of Epigenetics in Cancer By Anushree Choudhury

Abstract

Cancer, a disease as unique as its host, arises from the interplay of genomic and environmental factors, with the epigenome adding a layer of complexity. Epigenetic elements such as DNA methylation, histone modifications, and non-coding RNAs are pivotal in cancer initiation, progression, metastasis, and the development of drug resistance. In the past decade, epigenetics, once considered a contributor to cancer, is now viewed as a potential solution, given its reversible nature. Epidrugs transform the cause into the cure, forging a path to correct epigenetic alterations such as oncogene activation or TSG suppression in order to restore normal cellular activity. Some cancer, notably breast, lung, and prostate cancer, which are particularly susceptible to epigenetically induced carcinogenesis. The unique and individualised nature of each cancer underscores the importance of understanding the epigenetic landscape in order to tailor effective treatment strategies.

Keywords: DNA methylation, carcinogenesis, breast cancer, histone modifications, lung cancer, non-coding RNA and prostate cancer.

Introduction

The term "Epigenetics" was formulated by British developmental biologist & paleontologist C. H. Waddington by combining the two words "epigenesis" and "genetics" (Hurd, 2010). It refers to the relation between an organism's genetic makeup (genotype) and how its genes are expressed, eventually determining observable qualities (phenotype). These epigenetic alterations have no effect on the underlying DNA sequence, but impacts how genes are transcribed. The term 'epigenome' refers to modifications, which include DNA methylation patterns, histone alterations, and microRNA variations. Interestingly, these epigenetic modifications are influenced by extrinsic factors like lifestyle, nutrition, and stress. They have the potential to disrupt the normal regulation of genes, leading to alterations in protein production, which in turn, can contribute to various diseases, including cancer. Epigenetic dysregulation activates the oncogenes and inactivates the tumour suppressor genes (Pathak et al., 2023).

Cancer is characterised by uncontrolled cell growth and the loss of critical features such as contact inhibition and apoptosis (cell death). Its development is driven by the disruption of essential cellular pathways governing processes like the cell proliferation, differentiation, and programmed cell death. Cancer cells outcompete healthy ones by accelerating nutrient consumption, increasing metabolic rates, and promoting the formation of abnormal blood vessels through angiogenesis. To effectively combat cancer, it's essential to address its root causes through epigenetic therapies like epidrugs supplemented by chemotherapy, immunotherapies, etc. While ongoing research strives to unravel the precise mechanisms through which epigenetic alterations initiate and advance carcinogenesis, establishing a clear causal link remains a

challenge. Additionally, many biomarkers influenced by epigenetic changes are not yet definitely recognized as cancer triggers, as they can also result from the presence of cancer. In summary, understanding and targeting the fundamental drivers of cancer. Epigenetic alterations are leads for distinctive markers for cancer detection, therapy and prognosis.

This review addresses a notable knowledge gap in cancer research by shifting the predominant focus from genetics to epigenetics. Historically, cancer research has primarily concentrated on genetic factors, leaving the vital role of epigenetics in the shadows. Epigenetic pathways, which include DNA methylation, histone modifications, and non-coding RNA regulation, are increasingly becoming recognized as key factors to cancer initiation, development, and therapeutic resistance. Understanding the reversible and dynamic nature of epigenetic changes, offers the potential to revolutionise cancer research and therapy, transforming what was once considered a contributing factor into a promising solution. This paper fills an existing gap in the literature by emphasising the essential role of epigenetics in cancer research. By doing so, it broadens our understanding of cancer progression and therapy, highlighting the transformative potential of epigenetic insights, and contributing to the evolving landscape of cancer diagnosis and treatment.

1. Types Of Epigenetics Modifications:

1.1 DNA Methylation

DNA methylation is a crucial epigenetic process that modifies the methyl group at the fifth carbon of cytosine's pyrimidine ring, creating 5-methylcytosine (5mC), which controls the expression of genes (Jacobsen et al., 2010). This covalent alteration recruits proteins or prevents transcription factors from binding, which directly affects gene expression.

One crucial aspect of DNA methylation involves CpG islands, where a methyl group is added to the C5 position of the cytosine ring in the context of cytosine-guanine pairs (CpG). When promoter sequences in CpG islands become hypermethylated, this often leads to the inactivation of tumour suppressor genes. Tumour suppressor genes play a vital role in regulating various cellular pathways involved in cancer prevention, including those related to cell growth, cell proliferation, and apoptosis. DNA methyltransferase (DNMT) is an enzyme that regulates DNA methylation patterns. DNMT1 is primarily involved in maintaining existing DNA methylation patterns, while DNMT3a and DNMT3b establish new methylation patterns. In cancer, hypermethylation of promoter regions is a well-documented epigenetic change found in virtually every type of human neoplasm. This hypermethylation is associated with the inappropriate transcriptional silencing of genes that are important for tumorigenesis. These genes include O6-methylguanine-DNA methyltransferase (MGMT), which encodes a critical DNA repair gene, cyclin-dependent kinase inhibitor 2B (CDKN2B), which encodes p15, a cell-cycle regulator, and RASSF1A (Park et al., 2008).

According to Knudson's two-hit hypothesis, tumorigenesis results from the silencing of both alleles of tumour suppressor genes. One allele is frequently already mutated in those with a hereditary history of cancer, and DNA methylation can mute the other allele. According to

Knudson's model, unless both alleles of a tumor suppressor gene are inactivated, the phenotypic effect of tumour suppressor gene loss will not be seen (Wang et al., 2019). In cancer, the Warburg Effect is a well-known hallmark. It involves cancer cells consuming significantly more glucose and producing lactate even in the presence of oxygen. This metabolic alteration supports the rapid growth and energy demands of cancer cells (Zhu et al., 2020).

Hypomethylated regions in the genome have an unusually open nucleosome configuration and an abnormal acetylation of histone lysines. In contrast, abnormal nucleosomes positioned over the transcription start sites of silenced genes are associated with DNA hypermethylation in promoter CpG islands (Deaton et al., 2011). As per study conducted by Figeroa et al. & Noushmer et al., 2010, whole-exon sequencing of various human cancers, including leukemias, lymphomas, ovarian, renal, and pancreatic cancers, as well as rhabdomyosarcoma, has revealed a high occurrence of mutations in specific genes. Some of the genes include ARID1A, DNMT3A, EZH2, IDH1, MLL, PBRM1, SNF5, and VHL.

1.2 Histone modification

DNA is coiled around basic histone proteins resulting in a DNA-protein complex called chromatin. Both heterochromatin and euchromatin are histone proteins that control the rate of transcription. Euchromatin is a less compact and loosely packed form of chromatin. It is associated with a higher transcriptional rate because genes in euchromatin are more accessible for transcription. On the other hand, Heterochromatin is tightly packed and associated with a reduced transcriptional rate because the chromatin shape hinders access to the underlying DNA (Zhao et al., 2018). Thus, the control of gene expression is greatly aided by histone alterations. Dysregulation can lead to a disruption in homeostasis gain or loss of function of genes, overexpression, suppression by promoter hypermethylation, chromosomal translocation, or mutations of the histone-modifying enzymes/complexes or even the modification site of the histone (Lawrence et al., 2016) & (Audia et al., 2016).

Histone proteins can be modified via acetylation, methylation, and phosphorylation. Acetylation is the process of adding acetyl groups to lysine residues on histone proteins, most notably in the N-terminal tails. Acetylation reduces the interaction between histones and DNA by neutralizing the positive charge of lysine. This results in a more open and accessible chromatin structure, allowing for increased gene transcription. Histone acetyltransferases (HATs) are enzymes that catalyze the addition of acetyl groups to lysine residues.

Methylation of histones can occur on lysine and arginine residues and has varying effects on gene expression depending on the specific histone and the level of methylation. Methylation has the ability to either stimulate or suppress gene transcription. For example, trimethylation of histone H3 at lysine 4 (H3K4me3) is associated with active gene transcription, while trimethylation at lysine 9 (H3K9me3) is associated with gene repression.

Phosphorylation of histones involves the addition of phosphate groups, often at serine or threonine residues. Phosphorylation plays a role in various cellular processes, including DNA

damage repair, cell cycle regulation, and gene transcription. It can modulate the binding of other proteins to histones (Park et al., 2020).

According to a study conducted by Srivastava, et al., 2022 it was discovered that post-translational modifications (PTMs) of histones, such as acetylation, methylation, or phosphorylation, also contributed to the development of tumors.

Studies to understand lymphomagenesis demonstrate that mutations in genes responsible for modifying histones or in related transcription factors can disrupt the normal patterns of histone modifications, contributing to the development of lymphoma. The examination of recurrent mutations in B-cell Non-Hodgkin lymphoma, which comprises more than 90% of all lymphoma cases, has demonstrated that somatic mutations frequently target genes linked to histone modification. The most often mutated gene is the MLL2 gene, often referred to as histone-lysine N-methyltransferase (KMT2D), which has several single nucleotide variants spread out over the entire sequence. MLL2 is a significant histone methyltransferase in mammals, playing a role in mono-methylation H3K4 and functioning as a tumor suppressor gene in non-Hodgkin lymphoma. Another gene affected by mutations in this context is MEF2B, a transcription factor involved in calcium-regulated histone acetylation and methylation (Yang et al., 2022). The most prevalent mutation in MEF2B leads to changes in the amino acid D83V, which doesn't impact DNA interactions but does disrupt interactions with certain repressive complexes, including the HUCA complex and HDAC class IIa members. This disruption results in abnormal expression of histone markers at H3K27 (Yang et al., 2022).

1.3 Non-coding RNA

Approximately 2% of DNA is dedicated to encoding protein-coding genes, while the remaining 98% is considered non-coding, historically viewed by the scientific community as junk or without a recognized function (Pathania et al., 2022). Most non-coding RNA emerges from this junk DNA and does not code for any functional proteins.

Non-coding RNAs (non-coding RNAs) that are not translated into proteins are classified as housekeeping non-coding RNAs and regulatory non-coding RNAs. According to size, RNA with a regulatory function is primarily separated into two groups: long non-coding RNA (lncRNAs) and short-chain non-coding RNAs (siRNAs, miRNAs, and piRNAs) (Zaratiegui et al., 2007) & (Ponting et al., 2007). Recent research has demonstrated that non-coding RNAs are important for epigenetic modification and can control chromosomal and gene expression to regulate cell differentiation (Costa, 2008, Amaral et al., 2008, Ghildiyal et al., 2009 & Yu H, 2009). Non-coding RNAs that are not translated into proteins are classified as housekeeping non-coding RNAs and regulatory non-coding RNAs. Short-chain non-coding RNAs (siRNAs, miRNAs, and piRNAs) and long non-coding RNAs (lncRNAs) are the two main categories of RNA with a regulatory role based on size. (Zaratiegui et al., 2007; Ponting et al., 2007).

According to a 1969 theory of gene regulation by American molecular biologist Roy John Britten and developmental biologist Eric Harris Davidson, repetitive non-coding sequences may code for non-coding RNAs that function as regulatory mediators to transmit signals and

influence gene translation (Wolter et al. 1969). ncRNA have several important functions such as signalling pathways responsible for cancer initiation and progression, gene expression through chromosomal modulation, transcriptional regulation, and post-transcriptional modifications (Zhou et al., 2016). ncRNA are divided into various classes based on their size and functions; miRNA and piRNA (around 22 nucleotides) function in RNA suppression and transposon activity in germ line and somatic cells. lcrRNA modulates gene expression at mRNA stability, transcriptional, and post-transcriptional levels (Yao et al., 2019). Recent research has demonstrated that non-coding RNAs are important for epigenetic modification and can control chromosomal and gene expression to regulate cell differentiation (Costa, 2008, Amaral et al., 2008, Ghildiyal et al., 2009 & Yu H, 2009).

Cancers can develop as a result of either oncogene overexpression or tumor suppressor gene inactivation. The interplay of epigenetics and genetics to give rise to cancer phenotype is called carcinogenesis. This is often linked to uncontrolled cell proliferation, loss of contact inhibition, loss of the ability of apoptosis, and the acquisition of the property of angiogenesis. Tumour suppressor genes can become inactive in three ways: through disabling mutations, gene loss, or somatic heritable epigenetic changes that switch off the gene's function without altering its DNA sequence (Pathak et al., 2023). Interestingly, CpG island hypermethylation has been associated with the silencing of genes involved in cell proliferation and tumor suppressors. Examples of genes where hypermethylation is linked to carcinogenesis are the Glutathione S-transferase P (GSTP1) in prostate cancer and the BRCA1 gene in breast cancer (Esteller et al., 2000).

Epigenetic biomarkers are not only linked to tumor initiation but also have a significant role in cancer progression. For instance, EZH2 (enhancer zeste homolog 2)- a histone methyltransferase and a key part of PRC2, is responsible for adding three methyl groups to histone H3 at Lysine 27 (H3K27), a process known as trimethylation. This modification is crucial for controlling gene expression through epigenetic mechanisms and is linked to increased tumor growth. EZH2 promotes cell survival, proliferation, epithelial to mesenchymal, invasion, and drug resistance of cancer cells and has decisive roles on immune cells affecting the tumor microenvironment (Gan et al., 2018). Long Non-Coding RNA (lncRNA): Long non-coding RNAs, which do not code for proteins, can interact with DNA, RNA, and proteins, influencing various cell processes, including cell differentiation and regulated cell death like ferroptosis. In non-small cell lung cancers, lncRNAs suppress cancer development through ferroptosis pathways.

Epigenetic Alterations in different Cancers

2.1 Breast cancer

Breast cancer encompasses a diverse group of diseases originating from the breast, marked by biological and molecular variations. BRCA1/2, classified as tumour suppressor genes, have associations not only with breast cancer but also prostate and pancreatic cancers. These genes play crucial roles in cell cycle regulation, chromosomal modifications, DNA repair, and

apoptosis (Feng et al., 2018). They are especially important for DNA repair processes, including homologous recombination and nucleotide excision repair, in response to chemotherapy-induced DNA damage.

Furthermore, epigenetic alterations, such as DNA methylation and changes in chromatin structure, are detected early in cancer development.

A trend in global hypomethylation is observed, however, the number of genes which are hypomethylated are significantly lesser such as *FEN1*, *BCSG1*, *PLAU*, *IGF2* and *CDH3*. *On the contrary*, over 100 genes which play involved in various cellular processes, including cell-cycle regulation, apoptosis, tissue invasion, metastasis, angiogenesis, and hormone signalling undergo hypermethylation such as *CCND2* and *p16ink4A/CDKN2A* (cell cycle regulators), *APC*, *TWIST*, and *HOXA5* (apoptosis regulators), and *ER α* and *PR* (hormone regulators).

Secondly, due to histone modification, gene expression is regulated, for example histone demethylases, such as *Pygo2*, *JMJD2B*, and *LSD1*, are implicated in breast tumorigenesis. They influence various cellular processes, including transcriptional activation and cell cycle progression.

Long non-coding RNAs and microRNAs are recognized as pivotal contributors to breast cancer progression, leading to intratumoral heterogeneity. Since miRNA are downregulated, leads to reduced levels of let-7 family and miR-15/16 in breast cancer contribute to heightened tumorigenicity and abnormal *BCL2* expression, while the frequent amplification of miR-21 is associated with increased invasiveness and metastasis to the lungs (Chen et al., 2014). Additionally, exposure to environmental factors like mineral dust, tobacco smoke, arsenic, and silica can stimulate the expression of the mineral dust-induced gene (*MDIG*). *MDIG* promotes cell proliferation, cell cycle progression, and anti-apoptotic behaviours in various cell types, indicating its pro-proliferative properties (Thakur et al., 2022). However, *MDIG* has a dual role in breast cancer, acting as a facilitator in the early stages of cancer development due to its pro-proliferative characteristics but serving as an inhibitor in later stages, where it exhibits metastasis-inhibitory features. Distinct patterns of reprogramming are observable from the early stages of ductal hyperplasia to invasive carcinoma. These epigenetic effects extend beyond cancer cells, also manifesting in normal cells, indicating a broader field effect (Feng et al., 2018).

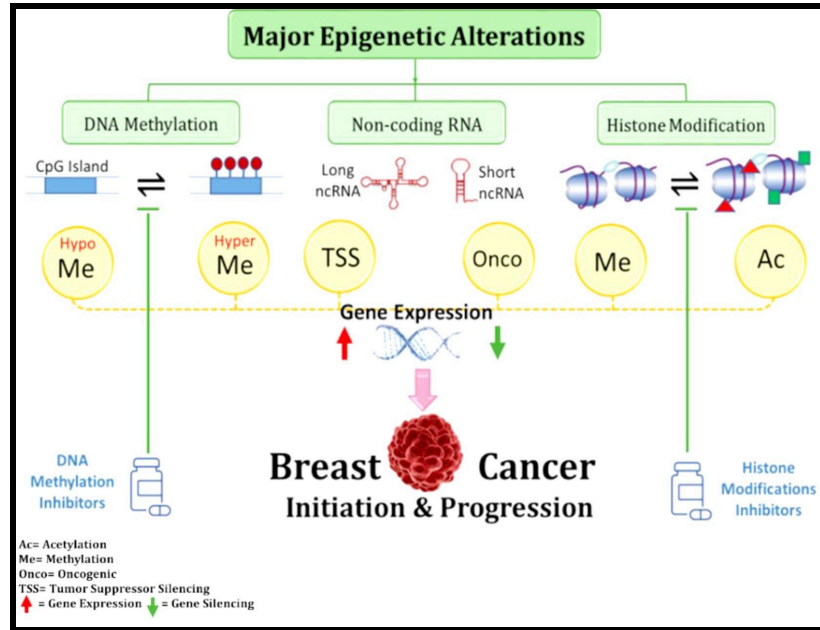


Figure 1: Diagram depicting CpG island methylation and histone changes that result in oncogene and tumor suppressor gene dysregulation. (Source: Gulab Sher et al.,2022).

2.2 Lung cancer

Lung cancer displays significant diversity in its development, attributed to factors like the presThe control of EMT and cancer stem cells in lung cancer is greatly impacted by histone modifications and non-coding RNA alterations. Epigenetic alterations, particularly histone modifications, play a pivotal role in lung cancer, impacting various aspects like cell growth, invasion, metastasis, cell death, and cell cycle regulation. EMT, a process where epithelial cells transform, losing polarity and adopting mesenchymal characteristics, exacerbates cancer and promotes metastasis (Baghel et al., 2023). Furthermore, significant signalling pathways such as the NF- κ B, Hedgehog, and ERK families are impacted by epigenetic dysregulation in lung cancer (Shi et al., 2019).

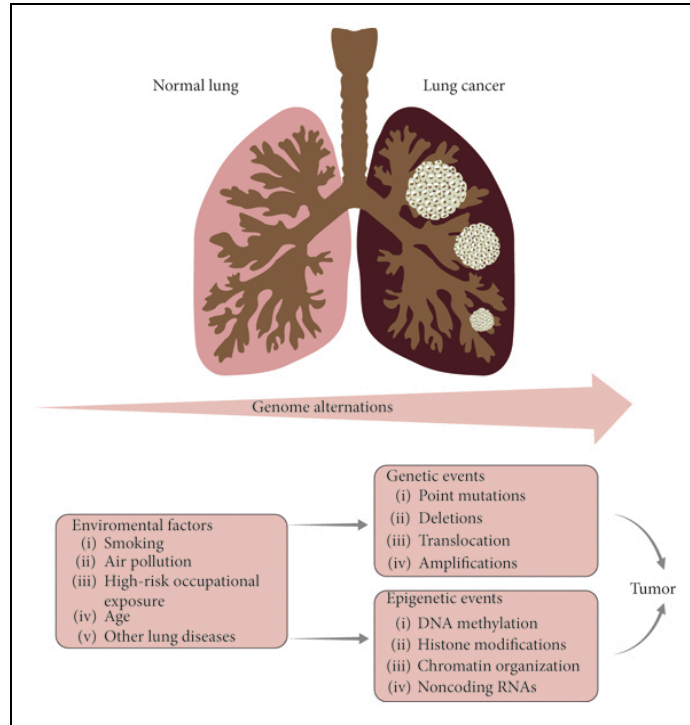


Figure 2: Genetic and epigenetic alterations linked to tumorigenesis
(Source: Shi et al., 2019)

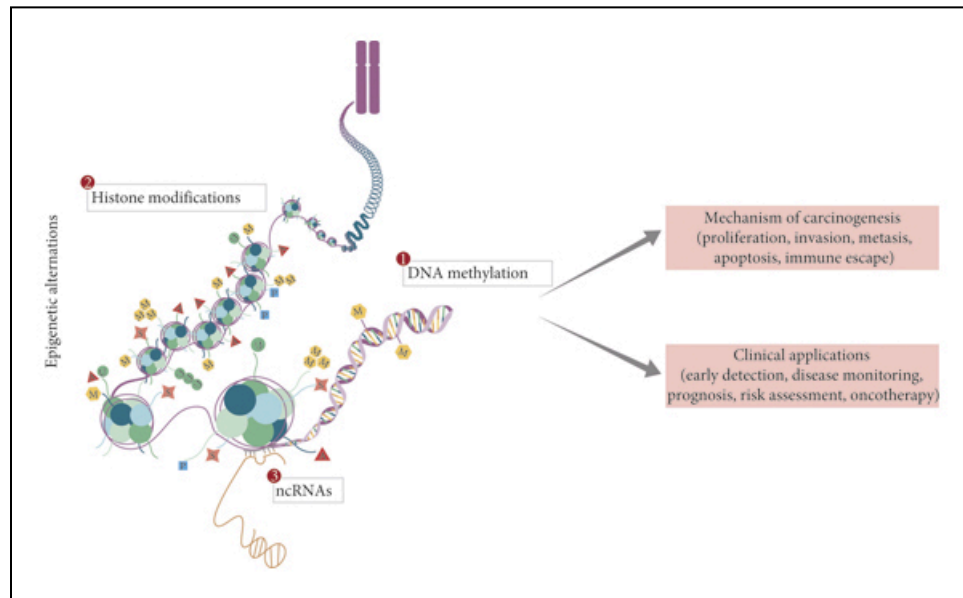


Figure 3: Landscape of epigenetics and its clinical application
(Source : Shi et al., 2019)

2.3 Prostate cancer

Prostate cancer is the uncontrolled proliferation of cells in the prostate, a gland in the male reproductive system located directly below the bladder. Epigenetic modifications, which include changes in DNA methylation, histone modifications, and nucleosome rearrangements, are important at all stages of prostate cancer (PCa) development and progression. These mutations cause tumor-suppressor genes to be silenced, oncogenic drivers to be activated, and therapeutic resistance to emerge (Pathak et al., 2023). In prostate cancer, over 50 genes consistently exhibit abnormal hypermethylation, impacting various cellular functions such as cell cycle regulation, apoptosis, hormonal responses, DNA repair, signal transduction, tumor invasion, and suppression. Noteworthy genes like APC, CCND2, GSTP1, RAR β 2, RASSF1A, and PTGS2 frequently demonstrate promoter methylation in PCa (Lam et al., 2020). A promising avenue in treating prostate cancer involves targeting epigenetic pathways associated with androgen receptors (Sarkissyan et al., 2014). In particular, EZH2, an essential part of the polycomb repressive complex (PRC2) that modifies H3K27me₃ to repress genes, has been found to be a viable target for PCa therapy.

Treatment

Epigenetic modifications do not only initiate carcinogenesis and its progression but also help neoplastic cells develop resistance to drugs hence rendering therapies against cancer futile. In CpG-rich islands, cancer cells like TSG experience hypermethylation and are inactivated. On the other hand, there is a genome-wide hypomethylation activating oncogenes. DNA methyltransferase is the focal target to restore the normal methylation levels by activating TSGs and deactivating oncogenes. There are two classes of DNMT inhibitors- Nucleoside analogs and non nucleosides (Pathak et al., 2023)

Azacitidine and Decitabine (FDA-approved), these two induce methylation to activate TSGs. Considering both the achievements and limitations of azacitidine and decitabine, researchers have developed prodrugs based on these compounds as advanced (DNA methyltransferase) DNMT inhibitors. Nucleoside analogs, featuring altered cytosine rings, can be integrated into newly formed DNA, forming a covalent bond with DNMTs. This interaction hinders the transfer of methyl groups to the hemimethylated DNA in offspring cells, resulting in widespread demethylation and subsequent degradation of DNMTs through the proteasomal pathway.

These next-generation inhibitors offer improved pharmacokinetic characteristics. As drugs are designed to target specific epigenetic abnormalities, they represent a pathway toward the future of personalized medicine. This means that individuals with particular epigenetic errors can receive custom-made epi drug treatments tailored to their specific condition (Furtabo et al., 2019).

According to Rodríguez-Paredes et al. (2011) and Salarinia, R. et al. (2016), epigenetic drugs are chemicals that change the structure of DNA and chromatin, encouraging the disruption of transcriptional and post-transcriptional modifications. They do this primarily by controlling the enzymes required for these modifications' establishment and maintenance, which reactivate

genes that are epigenetically silenced and that repair DNA. The primary method by which epidrugs work on the enzymes required for the creation and preservation of epigenetic changes is by inhibiting DNMTs and HDACs (Paredes, Rodriguez, and others, 2011).

Conclusion

Cancer is a highly individualized disease, making one-size-fits-all treatments ineffective. The causes of cancer are often complex and not well-defined, with epigenetics being one contributing factor. To effectively treat cancer, it is essential to diagnose it early using biomarkers and create a personalized treatment plan, known as precision medicine. Recently, there has been a growing focus on targeting epigenetic changes in cancer treatment. These changes can both trigger cancer growth and contribute to challenges like diverse tumor characteristics and resistance to drugs. However, the reversibility of epigenetic modifications presents both a challenge and an opportunity. They can be targeted and reversed effectively through approaches like combination therapies, using bacteria for therapy, or employing CRISPR technology. Moreover, the topic of epigenetics and cancer research is relatively new and offers exciting research opportunities, such as the creation of more precise biomarkers for early cancer detection and growth prediction.

Works Cited

- Amaral, P. P., Dinger, M. E., Mercer, T. R., & Mattick, J. S. (2008). The eukaryotic genome as an RNA machine. *science*, 319(5871), 1787-1789.
- Audia, J. E., & Campbell, R. M. (2016). Histone modifications and cancer. *Cold Spring Harbor perspectives in biology*, 8(4), a019521.
- Baghel, V. S., Shinde, S., Sinha, V., Dixit, V., Tiwari, A. K., Saxena, S., ... & Bhatt, P. (2023). Inhibitors targeting epigenetic modifications in cancer. In *Transcription and Translation in Health and Disease* (pp. 287-324). Academic Press.
- Costa, F. F. (2008). Non-coding RNAs, epigenetics and complexity. *Gene*, 410(1), 9-17.
- Chen, Q.W., Zhu, X.Y., Li, Y.Y., & Meng, Z.Q. (2014). Epigenetic regulation and cancer (Review). *Oncology Reports*, 31, 523-532. <https://doi.org/10.3892/or.2013.2913>
- Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes & development*, 25(10), 1010-1022.
- Esteller, M., Silva, J. M., Dominguez, G., Bonilla, F., Matias-Guiu, X., Lerma, E., ... & Herman, J. G. (2000). Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. *JNCI: Journal of the National Cancer Institute*, 92(7), 564-569.
- Feng, Y., Spezia, M., Huang, S., Yuan, C., Zeng, Z., Zhang, L., ... & Ren, G. (2018). Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. *Genes & diseases*, 5(2), 77-106.
- Feng, Y., Spezia, M., Huang, S., Yuan, C., Zeng, Z., Zhang, L., ... & Ren, G. (2018). Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. *Genes & diseases*, 5(2), 77-106.
- Figueroa, M. E. et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation (2010).
- Gan, L., Yang, Y., Li, Q., Feng, Y., Liu, T., & Guo, W. (2018). Epigenetic regulation of cancer progression by EZH2: from biological insights to therapeutic potential. *Biomarker research*, 6(1), 1-10.
- Ghildiyal, M., & Zamore, P. D. (2009). Small silencing RNAs: an expanding universe. *Nature reviews genetics*, 10(2), 94-108.
- He, N., Park, K., Zhang, Y., Huang, J., Lu, S., & Wang, L. (2008). Epigenetic inhibition of nuclear receptor small heterodimer partner is associated with and regulates hepatocellular carcinoma growth. *Gastroenterology*, 134(3), 793-802.
- Hurd, P. J. (2010). The era of epigenetics. *Briefings in functional genomics*, 9(5-6), 425-428.

- Karsli-Cebioglu, S., Dagdemir, A., Judes, G., Ngollo, M., Penault-Llorca, F., Pajon, A., & Bernard-Gallon, D. (2014). Epigenetic mechanisms of breast cancer: an update of the current knowledge. *Epigenomics*, *6*(6), 651-664.
- Lam, D., Clark, S., Stirzaker, C., & Pidsley, R. (2020). Advances in prognostic methylation biomarkers for prostate cancer. *Cancers*, *12*(10), 2993.
- Law, J. A., & Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, *11*(3), 204-220.
- Law, J. A., & Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, *11*(3), 204-220.
- Lawrence, M., Daujat, S., & Schneider, R. (2016). Lateral thinking: how histone modifications regulate gene expression. *Trends in Genetics*, *32*(1), 42-56.
- Miranda Furtado, C. L., Dos Santos Luciano, M. C., Silva Santos, R. D., Furtado, G. P., Moraes, M. O., & Pessoa, C. (2019). Epidrugs: targeting epigenetic marks in cancer treatment. *Epigenetics*, *14*(12), 1164-1176.
- Noushmehr, H. et al. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma (2010).
- Park, S. Y., & Kim, J. S. (2020). A short guide to histone deacetylases including recent progress on class II enzymes. *Experimental & molecular medicine*, *52*(2), 204-212.
- Pathak, A., Tomar, S., & Pathak, S. (2023). Epigenetics and Cancer: A Comprehensive Review. *Asian Pacific Journal of Cancer Biology*, *8*(1), 75-89.
- Pathania, A. S., Prathipati, P., Pandey, M. K., Byraredy, S. N., Coulter, D. W., Gupta, S. C., & Challagundla, K. B. (2022, August). The emerging role of non-coding RNAs in the epigenetic regulation of paediatric cancers. In *Seminars in cancer biology* (Vol. 83, pp. 227-241). Academic Press.
- Ponting, C. P., Oliver, P. L., & Reik, W. (2009). Evolution and functions of long noncoding RNAs. *Cell*, *136*(4), 629-641.
- Qi, P., Zhou, X. Y., & Du, X. (2016). Circulating long non-coding RNAs in cancer: current status and future perspectives. *Molecular cancer*, *15*, 1-11.
- Rodríguez-Paredes, M., & Esteller, M. (2011). Cancer epigenetics reaches mainstream oncology. *Nature medicine*, *17*(3), 330-339.
- Salarinia, R., Sahebkar, A., Peyvandi, M., Reza Mirzaei, H., Reza Jaafari, M., Matbou Riahi, M., ... & Mirzaei, H. (2016). Epi-drugs and Epi-miRs: moving beyond current cancer therapies. *Current cancer drug targets*, *16*(9), 773-788.
- Shi, Y. X., Sheng, D. Q., Cheng, L., & Song, X. Y. (2019). Current landscape of epigenetics in lung cancer: focus on the mechanism and application. *Journal of oncology*, *2019*.
- Srivastava, A. K., Guadagnin, G., Cappello, P., & Novelli, F. (2022). Post-Translational Modifications in Tumor-Associated Antigens as a Platform for Novel Immuno-Oncology Therapies. *Cancers*, *15*(1), 138.

- Thakur, C., Qiu, Y., Fu, Y., Bi, Z., Zhang, W., Ji, H., & Chen, F. (2022). Epigenetics and environment in breast cancer: New paradigms for anti-cancer therapies. *Frontiers in Oncology*, *12*, 971288.
- Wang, L. H., Wu, C. F., Rajasekaran, N., & Shin, Y. K. (2019). Loss of tumor suppressor gene function in human cancer: an overview. *Cellular Physiology and Biochemistry*, *51*(6), 2647-2693.
- Wolter, J. (1969). Gene regulation for higher cells: A theory.
- Wu, Y., Sarkissyan, M., & Vadgama, J. V. (2015). Epigenetics in breast and prostate cancer. *Cancer Epigenetics: Risk Assessment, Diagnosis, Treatment, and Prognosis*, 425-466
- Yang, Y., Zhang, M., & Wang, Y. (2022). The roles of histone modifications in tumorigenesis and associated inhibitors in cancer therapy. *Journal of the National Cancer Center*.
- Yao, R. W., Wang, Y., & Chen, L. L. (2019). Cellular functions of long noncoding RNAs. *Nature cell biology*, *21*(5), 542-551.
- Yu, H. (2009). Epigenetics: Advances of non-coding RNAs regulation in mammalian cells. *Yi Chuan= Hereditas*, *31*(11), 1077-1086.
- Zaratiegui, M., Irvine, D. V., & Martienssen, R. A. (2007). Noncoding RNAs and gene silencing. *Cell*, *128*(4), 763-776.
- Zhao, Z., & Shilatifard, A. (2019). Epigenetic modifications of histones in cancer. *Genome biology*, *20*, 1-16.
- Zhu, X., Xuan, Z., Chen, J., Li, Z., Zheng, S., & Song, P. (2020). How DNA methylation affects the Warburg effect. *International Journal of Biological Sciences*, *16*(12), 2029.

Investigating the Role of Epigenetics in Cancer By Anushree Choudhury

Abstract

Cancer, a disease as unique as its host, arises from the interplay of genomic and environmental factors, with the epigenome adding a layer of complexity. Epigenetic elements such as DNA methylation, histone modifications, and non-coding RNAs are pivotal in cancer initiation, progression, metastasis, and the development of drug resistance. In the past decade, epigenetics, once considered a contributor to cancer, is now viewed as a potential solution, given its reversible nature. Epidrugs transform the cause into the cure, forging a path to correct epigenetic alterations such as oncogene activation or TSG suppression in order to restore normal cellular activity. Some cancer, notably breast, lung, and prostate cancer, which are particularly susceptible to epigenetically induced carcinogenesis. The unique and individualised nature of each cancer underscores the importance of understanding the epigenetic landscape in order to tailor effective treatment strategies.

Keywords: DNA methylation, carcinogenesis, breast cancer, histone modifications, lung cancer, non-coding RNA and prostate cancer.

Introduction

The term "Epigenetics" was formulated by British developmental biologist & paleontologist C. H. Waddington by combining the two words "epigenesis" and "genetics" (Hurd, 2010). It refers to the relation between an organism's genetic makeup (genotype) and how its genes are expressed, eventually determining observable qualities (phenotype). These epigenetic alterations have no effect on the underlying DNA sequence, but impacts how genes are transcribed. The term 'epigenome' refers to modifications, which include DNA methylation patterns, histone alterations, and microRNA variations. Interestingly, these epigenetic modifications are influenced by extrinsic factors like lifestyle, nutrition, and stress. They have the potential to disrupt the normal regulation of genes, leading to alterations in protein production, which in turn, can contribute to various diseases, including cancer. Epigenetic dysregulation activates the oncogenes and inactivates the tumour suppressor genes (Pathak et al., 2023).

Cancer is characterised by uncontrolled cell growth and the loss of critical features such as contact inhibition and apoptosis (cell death). Its development is driven by the disruption of essential cellular pathways governing processes like the cell proliferation, differentiation, and programmed cell death. Cancer cells outcompete healthy ones by accelerating nutrient consumption, increasing metabolic rates, and promoting the formation of abnormal blood vessels through angiogenesis. To effectively combat cancer, it's essential to address its root causes through epigenetic therapies like epidrugs supplemented by chemotherapy, immunotherapies, etc. While ongoing research strives to unravel the precise mechanisms through which epigenetic alterations initiate and advance carcinogenesis, establishing a clear causal link remains a

challenge. Additionally, many biomarkers influenced by epigenetic changes are not yet definitely recognized as cancer triggers, as they can also result from the presence of cancer. In summary, understanding and targeting the fundamental drivers of cancer. Epigenetic alterations are leads for distinctive markers for cancer detection, therapy and prognosis.

This review addresses a notable knowledge gap in cancer research by shifting the predominant focus from genetics to epigenetics. Historically, cancer research has primarily concentrated on genetic factors, leaving the vital role of epigenetics in the shadows. Epigenetic pathways, which include DNA methylation, histone modifications, and non-coding RNA regulation, are increasingly becoming recognized as key factors to cancer initiation, development, and therapeutic resistance. Understanding the reversible and dynamic nature of epigenetic changes, offers the potential to revolutionise cancer research and therapy, transforming what was once considered a contributing factor into a promising solution. This paper fills an existing gap in the literature by emphasising the essential role of epigenetics in cancer research. By doing so, it broadens our understanding of cancer progression and therapy, highlighting the transformative potential of epigenetic insights, and contributing to the evolving landscape of cancer diagnosis and treatment.

1. Types Of Epigenetics Modifications:

1.1 DNA Methylation

DNA methylation is a crucial epigenetic process that modifies the methyl group at the fifth carbon of cytosine's pyrimidine ring, creating 5-methylcytosine (5mC), which controls the expression of genes (Jacobsen et al., 2010). This covalent alteration recruits proteins or prevents transcription factors from binding, which directly affects gene expression.

One crucial aspect of DNA methylation involves CpG islands, where a methyl group is added to the C5 position of the cytosine ring in the context of cytosine-guanine pairs (CpG). When promoter sequences in CpG islands become hypermethylated, this often leads to the inactivation of tumour suppressor genes. Tumour suppressor genes play a vital role in regulating various cellular pathways involved in cancer prevention, including those related to cell growth, cell proliferation, and apoptosis. DNA methyltransferase (DNMT) is an enzyme that regulates DNA methylation patterns. DNMT1 is primarily involved in maintaining existing DNA methylation patterns, while DNMT3a and DNMT3b establish new methylation patterns. In cancer, hypermethylation of promoter regions is a well-documented epigenetic change found in virtually every type of human neoplasm. This hypermethylation is associated with the inappropriate transcriptional silencing of genes that are important for tumorigenesis. These genes include O6-methylguanine-DNA methyltransferase (MGMT), which encodes a critical DNA repair gene, cyclin-dependent kinase inhibitor 2B (CDKN2B), which encodes p15, a cell-cycle regulator, and RASSF1A (Park et al., 2008).

According to Knudson's two-hit hypothesis, tumorigenesis results from the silencing of both alleles of tumour suppressor genes. One allele is frequently already mutated in those with a hereditary history of cancer, and DNA methylation can mute the other allele. According to

Knudson's model, unless both alleles of a tumor suppressor gene are inactivated, the phenotypic effect of tumour suppressor gene loss will not be seen (Wang et al., 2019). In cancer, the Warburg Effect is a well-known hallmark. It involves cancer cells consuming significantly more glucose and producing lactate even in the presence of oxygen. This metabolic alteration supports the rapid growth and energy demands of cancer cells (Zhu et al., 2020).

Hypomethylated regions in the genome have an unusually open nucleosome configuration and an abnormal acetylation of histone lysines. In contrast, abnormal nucleosomes positioned over the transcription start sites of silenced genes are associated with DNA hypermethylation in promoter CpG islands (Deaton et al., 2011). As per study conducted by Figeroa et al. & Noushmer et al., 2010, whole-exon sequencing of various human cancers, including leukemias, lymphomas, ovarian, renal, and pancreatic cancers, as well as rhabdomyosarcoma, has revealed a high occurrence of mutations in specific genes. Some of the genes include ARID1A, DNMT3A, EZH2, IDH1, MLL, PBRM1, SNF5, and VHL.

1.2 Histone modification

DNA is coiled around basic histone proteins resulting in a DNA-protein complex called chromatin. Both heterochromatin and euchromatin are histone proteins that control the rate of transcription. Euchromatin is a less compact and loosely packed form of chromatin. It is associated with a higher transcriptional rate because genes in euchromatin are more accessible for transcription. On the other hand, Heterochromatin is tightly packed and associated with a reduced transcriptional rate because the chromatin shape hinders access to the underlying DNA (Zhao et al., 2018). Thus, the control of gene expression is greatly aided by histone alterations. Dysregulation can lead to a disruption in homeostasis gain or loss of function of genes, overexpression, suppression by promoter hypermethylation, chromosomal translocation, or mutations of the histone-modifying enzymes/complexes or even the modification site of the histone (Lawrence et al., 2016) & (Audia et al., 2016).

Histone proteins can be modified via acetylation, methylation, and phosphorylation. Acetylation is the process of adding acetyl groups to lysine residues on histone proteins, most notably in the N-terminal tails. Acetylation reduces the interaction between histones and DNA by neutralizing the positive charge of lysine. This results in a more open and accessible chromatin structure, allowing for increased gene transcription. Histone acetyltransferases (HATs) are enzymes that catalyze the addition of acetyl groups to lysine residues.

Methylation of histones can occur on lysine and arginine residues and has varying effects on gene expression depending on the specific histone and the level of methylation. Methylation has the ability to either stimulate or suppress gene transcription. For example, trimethylation of histone H3 at lysine 4 (H3K4me3) is associated with active gene transcription, while trimethylation at lysine 9 (H3K9me3) is associated with gene repression.

Phosphorylation of histones involves the addition of phosphate groups, often at serine or threonine residues. Phosphorylation plays a role in various cellular processes, including DNA

damage repair, cell cycle regulation, and gene transcription. It can modulate the binding of other proteins to histones (Park et al., 2020).

According to a study conducted by Srivastava, et al., 2022 it was discovered that post-translational modifications (PTMs) of histones, such as acetylation, methylation, or phosphorylation, also contributed to the development of tumors.

Studies to understand lymphomagenesis demonstrate that mutations in genes responsible for modifying histones or in related transcription factors can disrupt the normal patterns of histone modifications, contributing to the development of lymphoma. The examination of recurrent mutations in B-cell Non-Hodgkin lymphoma, which comprises more than 90% of all lymphoma cases, has demonstrated that somatic mutations frequently target genes linked to histone modification. The most often mutated gene is the MLL2 gene, often referred to as histone-lysine N-methyltransferase (KMT2D), which has several single nucleotide variants spread out over the entire sequence. MLL2 is a significant histone methyltransferase in mammals, playing a role in mono-methylation H3K4 and functioning as a tumor suppressor gene in non-Hodgkin lymphoma. Another gene affected by mutations in this context is MEF2B, a transcription factor involved in calcium-regulated histone acetylation and methylation (Yang et al., 2022). The most prevalent mutation in MEF2B leads to changes in the amino acid D83V, which doesn't impact DNA interactions but does disrupt interactions with certain repressive complexes, including the HUCA complex and HDAC class IIa members. This disruption results in abnormal expression of histone markers at H3K27 (Yang et al., 2022).

1.3 Non-coding RNA

Approximately 2% of DNA is dedicated to encoding protein-coding genes, while the remaining 98% is considered non-coding, historically viewed by the scientific community as junk or without a recognized function (Pathania et al., 2022). Most non-coding RNA emerges from this junk DNA and does not code for any functional proteins.

Non-coding RNAs (non-coding RNAs) that are not translated into proteins are classified as housekeeping non-coding RNAs and regulatory non-coding RNAs. According to size, RNA with a regulatory function is primarily separated into two groups: long non-coding RNA (lncRNAs) and short-chain non-coding RNAs (siRNAs, miRNAs, and piRNAs) (Zaratiegui et al., 2007) & (Ponting et al., 2007). Recent research has demonstrated that non-coding RNAs are important for epigenetic modification and can control chromosomal and gene expression to regulate cell differentiation (Costa, 2008, Amaral et al., 2008, Ghildiyal et al., 2009 & Yu H, 2009). Non-coding RNAs that are not translated into proteins are classified as housekeeping non-coding RNAs and regulatory non-coding RNAs. Short-chain non-coding RNAs (siRNAs, miRNAs, and piRNAs) and long non-coding RNAs (lncRNAs) are the two main categories of RNA with a regulatory role based on size. (Zaratiegui et al., 2007; Ponting et al., 2007).

According to a 1969 theory of gene regulation by American molecular biologist Roy John Britten and developmental biologist Eric Harris Davidson, repetitive non-coding sequences may code for non-coding RNAs that function as regulatory mediators to transmit signals and

influence gene translation (Wolter et al. 1969). ncRNA have several important functions such as signalling pathways responsible for cancer initiation and progression, gene expression through chromosomal modulation, transcriptional regulation, and post-transcriptional modifications (Zhou et al., 2016). ncRNA are divided into various classes based on their size and functions; miRNA and piRNA (around 22 nucleotides) function in RNA suppression and transposon activity in germ line and somatic cells. lcrRNA modulates gene expression at mRNA stability, transcriptional, and post-transcriptional levels (Yao et al., 2019). Recent research has demonstrated that non-coding RNAs are important for epigenetic modification and can control chromosomal and gene expression to regulate cell differentiation (Costa, 2008, Amaral et al., 2008, Ghildiyal et al., 2009 & Yu H, 2009).

Cancers can develop as a result of either oncogene overexpression or tumor suppressor gene inactivation. The interplay of epigenetics and genetics to give rise to cancer phenotype is called carcinogenesis. This is often linked to uncontrolled cell proliferation, loss of contact inhibition, loss of the ability of apoptosis, and the acquisition of the property of angiogenesis. Tumour suppressor genes can become inactive in three ways: through disabling mutations, gene loss, or somatic heritable epigenetic changes that switch off the gene's function without altering its DNA sequence (Pathak et al., 2023). Interestingly, CpG island hypermethylation has been associated with the silencing of genes involved in cell proliferation and tumor suppressors. Examples of genes where hypermethylation is linked to carcinogenesis are the Glutathione S-transferase P (GSTP1) in prostate cancer and the BRCA1 gene in breast cancer (Esteller et al., 2000).

Epigenetic biomarkers are not only linked to tumor initiation but also have a significant role in cancer progression. For instance, EZH2 (enhancer zeste homolog 2)- a histone methyltransferase and a key part of PRC2, is responsible for adding three methyl groups to histone H3 at Lysine 27 (H3K27), a process known as trimethylation. This modification is crucial for controlling gene expression through epigenetic mechanisms and is linked to increased tumor growth. EZH2 promotes cell survival, proliferation, epithelial to mesenchymal, invasion, and drug resistance of cancer cells and has decisive roles on immune cells affecting the tumor microenvironment (Gan et al., 2018). Long Non-Coding RNA (lncRNA): Long non-coding RNAs, which do not code for proteins, can interact with DNA, RNA, and proteins, influencing various cell processes, including cell differentiation and regulated cell death like ferroptosis. In non-small cell lung cancers, lncRNAs suppress cancer development through ferroptosis pathways.

Epigenetic Alterations in different Cancers

2.1 Breast cancer

Breast cancer encompasses a diverse group of diseases originating from the breast, marked by biological and molecular variations. BRCA1/2, classified as tumour suppressor genes, have associations not only with breast cancer but also prostate and pancreatic cancers. These genes play crucial roles in cell cycle regulation, chromosomal modifications, DNA repair, and

apoptosis (Feng et al., 2018). They are especially important for DNA repair processes, including homologous recombination and nucleotide excision repair, in response to chemotherapy-induced DNA damage.

Furthermore, epigenetic alterations, such as DNA methylation and changes in chromatin structure, are detected early in cancer development.

A trend in global hypomethylation is observed, however, the number of genes which are hypomethylated are significantly lesser such as *FEN1*, *BCSG1*, *PLAU*, *IGF2* and *CDH3*. *On the contrary*, over 100 genes which play involved in various cellular processes, including cell-cycle regulation, apoptosis, tissue invasion, metastasis, angiogenesis, and hormone signalling undergo hypermethylation such as *CCND2* and *p16ink4A/CDKN2A* (cell cycle regulators), *APC*, *TWIST*, and *HOXA5* (apoptosis regulators), and *ER α* and *PR* (hormone regulators).

Secondly, due to histone modification, gene expression is regulated, for example histone demethylases, such as *Pygo2*, *JMJD2B*, and *LSD1*, are implicated in breast tumorigenesis. They influence various cellular processes, including transcriptional activation and cell cycle progression.

Long non-coding RNAs and microRNAs are recognized as pivotal contributors to breast cancer progression, leading to intratumoral heterogeneity. Since miRNA are downregulated, leads to reduced levels of let-7 family and miR-15/16 in breast cancer contribute to heightened tumorigenicity and abnormal *BCL2* expression, while the frequent amplification of miR-21 is associated with increased invasiveness and metastasis to the lungs (Chen et al., 2014). Additionally, exposure to environmental factors like mineral dust, tobacco smoke, arsenic, and silica can stimulate the expression of the mineral dust-induced gene (*MDIG*). *MDIG* promotes cell proliferation, cell cycle progression, and anti-apoptotic behaviours in various cell types, indicating its pro-proliferative properties (Thakur et al., 2022). However, *MDIG* has a dual role in breast cancer, acting as a facilitator in the early stages of cancer development due to its pro-proliferative characteristics but serving as an inhibitor in later stages, where it exhibits metastasis-inhibitory features. Distinct patterns of reprogramming are observable from the early stages of ductal hyperplasia to invasive carcinoma. These epigenetic effects extend beyond cancer cells, also manifesting in normal cells, indicating a broader field effect (Feng et al., 2018).

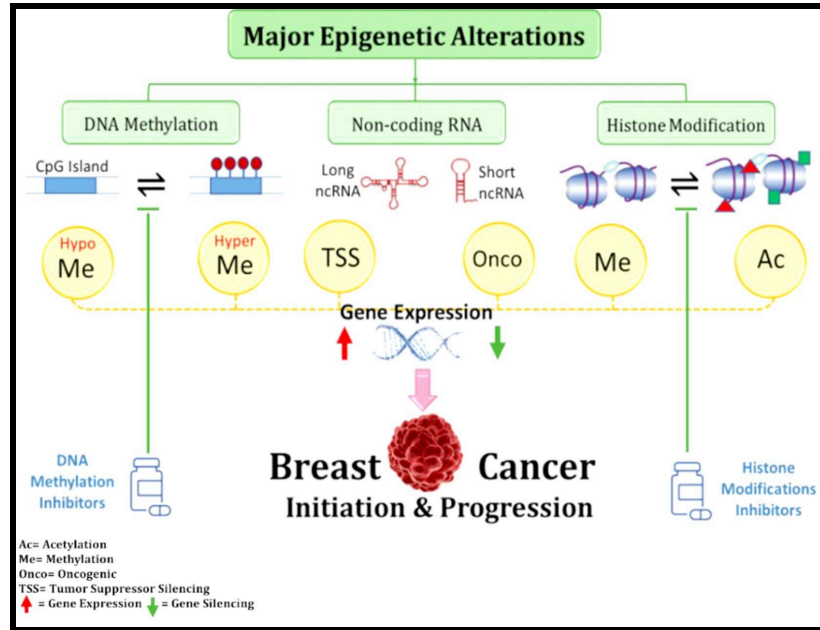


Figure 1: Diagram depicting CpG island methylation and histone changes that result in oncogene and tumor suppressor gene dysregulation. (Source: Gulab Sher et al.,2022).

2.2 Lung cancer

Lung cancer displays significant diversity in its development, attributed to factors like the presThe control of EMT and cancer stem cells in lung cancer is greatly impacted by histone modifications and non-coding RNA alterations. Epigenetic alterations, particularly histone modifications, play a pivotal role in lung cancer, impacting various aspects like cell growth, invasion, metastasis, cell death, and cell cycle regulation. EMT, a process where epithelial cells transform, losing polarity and adopting mesenchymal characteristics, exacerbates cancer and promotes metastasis (Baghel et al., 2023). Furthermore, significant signalling pathways such as the NF- κ B, Hedgehog, and ERK families are impacted by epigenetic dysregulation in lung cancer (Shi et al., 2019).

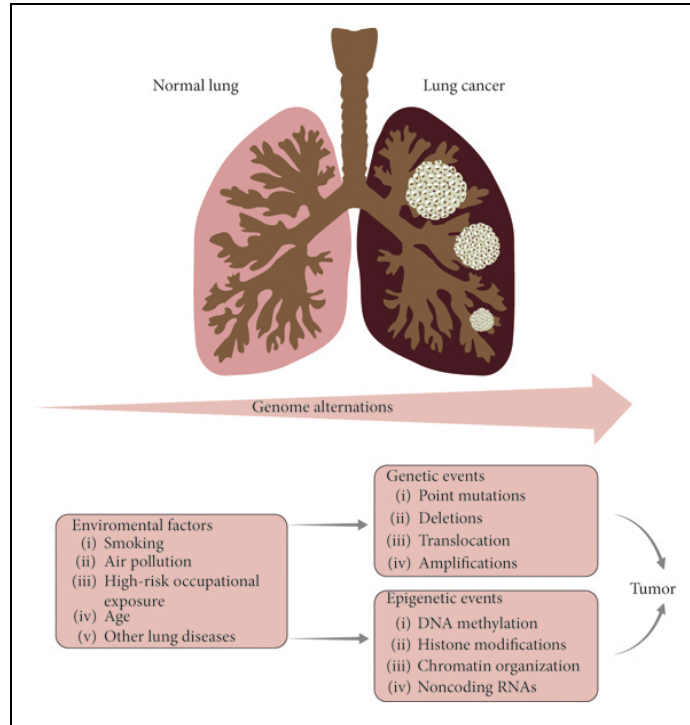


Figure 2: Genetic and epigenetic alterations linked to tumorigenesis
(Source: Shi et al., 2019)

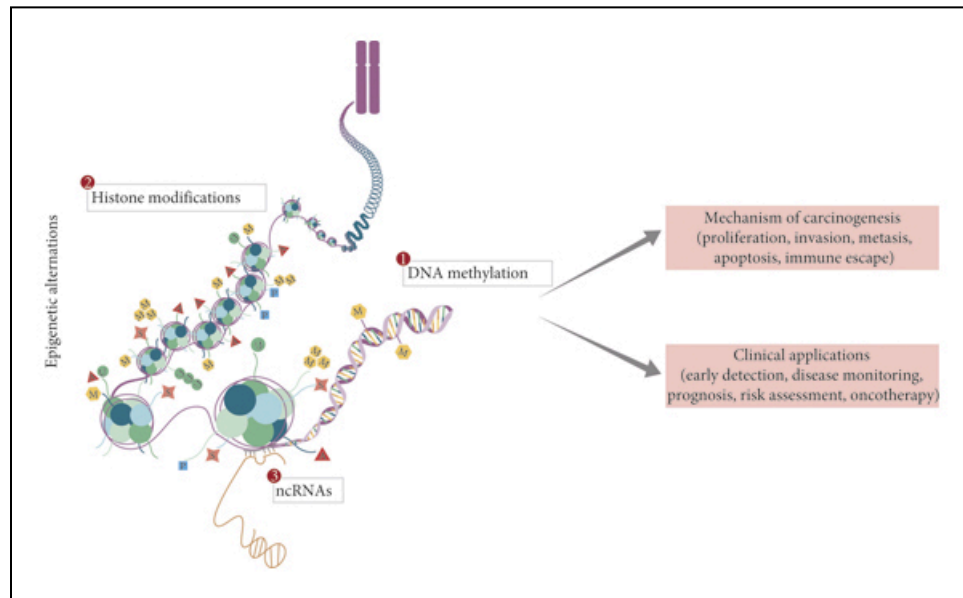


Figure 3: Landscape of epigenetics and its clinical application
(Source : Shi et al., 2019)

2.3 Prostate cancer

Prostate cancer is the uncontrolled proliferation of cells in the prostate, a gland in the male reproductive system located directly below the bladder. Epigenetic modifications, which include changes in DNA methylation, histone modifications, and nucleosome rearrangements, are important at all stages of prostate cancer (PCa) development and progression. These mutations cause tumor-suppressor genes to be silenced, oncogenic drivers to be activated, and therapeutic resistance to emerge (Pathak et al., 2023). In prostate cancer, over 50 genes consistently exhibit abnormal hypermethylation, impacting various cellular functions such as cell cycle regulation, apoptosis, hormonal responses, DNA repair, signal transduction, tumor invasion, and suppression. Noteworthy genes like APC, CCND2, GSTP1, RAR β 2, RASSF1A, and PTGS2 frequently demonstrate promoter methylation in PCa (Lam et al., 2020). A promising avenue in treating prostate cancer involves targeting epigenetic pathways associated with androgen receptors (Sarkissyan et al., 2014). In particular, EZH2, an essential part of the polycomb repressive complex (PRC2) that modifies H3K27me₃ to repress genes, has been found to be a viable target for PCa therapy.

Treatment

Epigenetic modifications do not only initiate carcinogenesis and its progression but also help neoplastic cells develop resistance to drugs hence rendering therapies against cancer futile. In CpG-rich islands, cancer cells like TSG experience hypermethylation and are inactivated. On the other hand, there is a genome-wide hypomethylation activating oncogenes. DNA methyltransferase is the focal target to restore the normal methylation levels by activating TSGs and deactivating oncogenes. There are two classes of DNMT inhibitors- Nucleoside analogs and non nucleosides (Pathak et al., 2023)

Azacitidine and Decitabine (FDA-approved), these two induce methylation to activate TSGs. Considering both the achievements and limitations of azacitidine and decitabine, researchers have developed prodrugs based on these compounds as advanced (DNA methyltransferase) DNMT inhibitors. Nucleoside analogs, featuring altered cytosine rings, can be integrated into newly formed DNA, forming a covalent bond with DNMTs. This interaction hinders the transfer of methyl groups to the hemimethylated DNA in offspring cells, resulting in widespread demethylation and subsequent degradation of DNMTs through the proteasomal pathway.

These next-generation inhibitors offer improved pharmacokinetic characteristics. As drugs are designed to target specific epigenetic abnormalities, they represent a pathway toward the future of personalized medicine. This means that individuals with particular epigenetic errors can receive custom-made epi drug treatments tailored to their specific condition (Furtabo et al., 2019).

According to Rodríguez-Paredes et al. (2011) and Salarinia, R. et al. (2016), epigenetic drugs are chemicals that change the structure of DNA and chromatin, encouraging the disruption of transcriptional and post-transcriptional modifications. They do this primarily by controlling

the enzymes required for these modifications' establishment and maintenance, which reactivate genes that are epigenetically silenced and that repair DNA. The primary method by which epidrugs work on the enzymes required for the creation and preservation of epigenetic changes is by inhibiting DNMTs and HDACs (Paredes, Rodriguez, and others, 2011).

Conclusion

Cancer is a highly individualized disease, making one-size-fits-all treatments ineffective. The causes of cancer are often complex and not well-defined, with epigenetics being one contributing factor. To effectively treat cancer, it is essential to diagnose it early using biomarkers and create a personalized treatment plan, known as precision medicine. Recently, there has been a growing focus on targeting epigenetic changes in cancer treatment. These changes can both trigger cancer growth and contribute to challenges like diverse tumor characteristics and resistance to drugs. However, the reversibility of epigenetic modifications presents both a challenge and an opportunity. They can be targeted and reversed effectively through approaches like combination therapies, using bacteria for therapy, or employing CRISPR technology. Moreover, the topic of epigenetics and cancer research is relatively new and offers exciting research opportunities, such as the creation of more precise biomarkers for early cancer detection and growth prediction.

Works Cited

- Amaral, P. P., Dinger, M. E., Mercer, T. R., & Mattick, J. S. (2008). The eukaryotic genome as an RNA machine. *science*, 319(5871), 1787-1789.
- Audia, J. E., & Campbell, R. M. (2016). Histone modifications and cancer. *Cold Spring Harbor perspectives in biology*, 8(4), a019521.
- Baghel, V. S., Shinde, S., Sinha, V., Dixit, V., Tiwari, A. K., Saxena, S., ... & Bhatt, P. (2023). Inhibitors targeting epigenetic modifications in cancer. In *Transcription and Translation in Health and Disease* (pp. 287-324). Academic Press.
- Costa, F. F. (2008). Non-coding RNAs, epigenetics and complexity. *Gene*, 410(1), 9-17.
- Chen, Q.W., Zhu, X.Y., Li, Y.Y., & Meng, Z.Q. (2014). Epigenetic regulation and cancer (Review). *Oncology Reports*, 31, 523-532. <https://doi.org/10.3892/or.2013.2913>
- Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes & development*, 25(10), 1010-1022.
- Esteller, M., Silva, J. M., Dominguez, G., Bonilla, F., Matias-Guiu, X., Lerma, E., ... & Herman, J. G. (2000). Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. *JNCI: Journal of the National Cancer Institute*, 92(7), 564-569.
- Feng, Y., Spezia, M., Huang, S., Yuan, C., Zeng, Z., Zhang, L., ... & Ren, G. (2018). Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. *Genes & diseases*, 5(2), 77-106.
- Feng, Y., Spezia, M., Huang, S., Yuan, C., Zeng, Z., Zhang, L., ... & Ren, G. (2018). Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. *Genes & diseases*, 5(2), 77-106.
- Figueroa, M. E. et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation (2010).
- Gan, L., Yang, Y., Li, Q., Feng, Y., Liu, T., & Guo, W. (2018). Epigenetic regulation of cancer progression by EZH2: from biological insights to therapeutic potential. *Biomarker research*, 6(1), 1-10.
- Ghildiyal, M., & Zamore, P. D. (2009). Small silencing RNAs: an expanding universe. *Nature reviews genetics*, 10(2), 94-108.
- He, N., Park, K., Zhang, Y., Huang, J., Lu, S., & Wang, L. (2008). Epigenetic inhibition of nuclear receptor small heterodimer partner is associated with and regulates hepatocellular carcinoma growth. *Gastroenterology*, 134(3), 793-802.
- Hurd, P. J. (2010). The era of epigenetics. *Briefings in functional genomics*, 9(5-6), 425-428.
- Karsli-Cebioglu, S., Dagdemir, A., Judes, G., Ngollo, M., Penault-Llorca, F., Pajon, A., & Bernard-Gallon, D. (2014). Epigenetic mechanisms of breast cancer: an update of the current knowledge. *Epigenomics*, 6(6), 651-664.
- Lam, D., Clark, S., Stirzaker, C., & Pidsley, R. (2020). Advances in prognostic methylation biomarkers for prostate cancer. *Cancers*, 12(10), 2993.
- Law, J. A., & Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, 11(3), 204-220.

- Law, J. A., & Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, *11*(3), 204-220.
- Lawrence, M., Daujat, S., & Schneider, R. (2016). Lateral thinking: how histone modifications regulate gene expression. *Trends in Genetics*, *32*(1), 42-56.
- Miranda Furtado, C. L., Dos Santos Luciano, M. C., Silva Santos, R. D., Furtado, G. P., Moraes, M. O., & Pessoa, C. (2019). Epidrugs: targeting epigenetic marks in cancer treatment. *Epigenetics*, *14*(12), 1164-1176.
- Noushmehr, H. et al. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma (2010).
- Park, S. Y., & Kim, J. S. (2020). A short guide to histone deacetylases including recent progress on class II enzymes. *Experimental & molecular medicine*, *52*(2), 204-212.
- Pathak, A., Tomar, S., & Pathak, S. (2023). Epigenetics and Cancer: A Comprehensive Review. *Asian Pacific Journal of Cancer Biology*, *8*(1), 75-89.
- Pathania, A. S., Prathipati, P., Pandey, M. K., Byrareddy, S. N., Coulter, D. W., Gupta, S. C., & Challagundla, K. B. (2022, August). The emerging role of non-coding RNAs in the epigenetic regulation of paediatric cancers. In *Seminars in cancer biology* (Vol. 83, pp. 227-241). Academic Press.
- Ponting, C. P., Oliver, P. L., & Reik, W. (2009). Evolution and functions of long noncoding RNAs. *Cell*, *136*(4), 629-641.
- Qi, P., Zhou, X. Y., & Du, X. (2016). Circulating long non-coding RNAs in cancer: current status and future perspectives. *Molecular cancer*, *15*, 1-11.
- Rodríguez-Paredes, M., & Esteller, M. (2011). Cancer epigenetics reaches mainstream oncology. *Nature medicine*, *17*(3), 330-339.
- Salarinia, R., Sahebkar, A., Peyvandi, M., Reza Mirzaei, H., Reza Jaafari, M., Matbou Riahi, M., ... & Mirzaei, H. (2016). Epi-drugs and Epi-miRs: moving beyond current cancer therapies. *Current cancer drug targets*, *16*(9), 773-788.
- Shi, Y. X., Sheng, D. Q., Cheng, L., & Song, X. Y. (2019). Current landscape of epigenetics in lung cancer: focus on the mechanism and application. *Journal of oncology*, *2019*.
- Srivastava, A. K., Guadagnin, G., Cappello, P., & Novelli, F. (2022). Post-Translational Modifications in Tumor-Associated Antigens as a Platform for Novel Immuno-Oncology Therapies. *Cancers*, *15*(1), 138.
- Thakur, C., Qiu, Y., Fu, Y., Bi, Z., Zhang, W., Ji, H., & Chen, F. (2022). Epigenetics and environment in breast cancer: New paradigms for anti-cancer therapies. *Frontiers in Oncology*, *12*, 971288.
- Wang, L. H., Wu, C. F., Rajasekaran, N., & Shin, Y. K. (2019). Loss of tumor suppressor gene function in human cancer: an overview. *Cellular Physiology and Biochemistry*, *51*(6), 2647-2693.
- Wolter, J. (1969). Gene regulation for higher cells: A theory.
- Wu, Y., Sarkissyan, M., & Vadgama, J. V. (2015). Epigenetics in breast and prostate cancer. *Cancer Epigenetics: Risk Assessment, Diagnosis, Treatment, and Prognosis*, 425-466

- Yang, Y., Zhang, M., & Wang, Y. (2022). The roles of histone modifications in tumorigenesis and associated inhibitors in cancer therapy. *Journal of the National Cancer Center*.
- Yao, R. W., Wang, Y., & Chen, L. L. (2019). Cellular functions of long noncoding RNAs. *Nature cell biology*, *21*(5), 542-551.
- Yu, H. (2009). Epigenetics: Advances of non-coding RNAs regulation in mammalian cells. *Yi Chuan= Hereditas*, *31*(11), 1077-1086.
- Zaratiegui, M., Irvine, D. V., & Martienssen, R. A. (2007). Noncoding RNAs and gene silencing. *Cell*, *128*(4), 763-776.
- Zhao, Z., & Shilatifard, A. (2019). Epigenetic modifications of histones in cancer. *Genome biology*, *20*, 1-16.
- Zhu, X., Xuan, Z., Chen, J., Li, Z., Zheng, S., & Song, P. (2020). How DNA methylation affects the Warburg effect. *International Journal of Biological Sciences*, *16*(12), 2029.

Why Humans Dream: A Review of Sleep and Brain Biology, and Diseases in the Context of Dreams By Tyler Lovejoy

Abstract

While there is much research around sleep and its function, there is a seeming lack of research into the purpose of dreaming that if addressed could open a lot of research and clinical benefits. Specifically, dream research around certain diseases like PTSD and Parkinson's can increase understanding of dreams, and therapies pertaining to dreams. Here we review the biology of sleep in the context of dreams, review some popular dream theories, and dive into some deeper research around brain biology and disease research around dreams. By using the biology of sleep as a frame to understand dream theories we can evaluate the theories validity, and extend those theories into deeper concepts like small parts of the brain, and how disease affects dreams. The goal of the paper is to both assess dream theories for their validity and how certain diseases pertain to that, but also to encourage and prove how further research on dreaming could have widespread impacts in many fields.

Introduction

It is unknown exactly why humans dream, however dreams are a part of a normal healthy night of sleep. Exploring the biological role of dreams could illuminate both an unexplored metric of sleep health and novel insights into certain diseases. While there is not a predominant theory, a multitude of theories exist around the role of dreams, such as dreams as a byproduct of memory consolidation, threat simulation theory, activation synthesis, continual activation theory, dreams for strengthening semantic memory, and reverse learning theory. There is a clear connection between the biological role of sleep and the theories behind dreams, though it is not as often researched. Making that connection is important for finding out why we dream, which in turn could open the door for understanding many facets of general health and sleep quality giving insights into disorders like PTSD, where dreaming has been shown to aid in recovery (Cowdin et al.).

As dreams are a part of sleep, it is important to understand what is known about sleep and current theories about dreams. In this paper I will discuss the architecture and biology of sleep. We will also explain several dream theories, including threat simulation, reverse learning, and continual activation theories. Finally I will cover some more specific biology that pertains to dreams, like certain research around the role of the amygdala and hippocampus in dreams, and some disorders like Parkinson's and PTSD. The goal of the paper is to add to the theories around dreams, and make useful connections to the biological role of dreams.

What is sleep

Dreams occur during sleep. As such, understanding the architecture and components of sleep is essential for understanding theories and biological roles of dreams. Topics to be discussed include the basic biology of sleep, such as the role of synapses, the stages of sleep, and

areas of the brain involved in sleep. There are two sleep types: non-rapid eye movement (NREM) and rapid eye movement (REM). While dreams can happen in both, most occur during REM, so most studies about dreaming are focused on REM.

i. Brain Biology

While the impact of sleep on the brain is still debated, a consensus has been reached on the parts of the brain involved in sleep. There are three main sections of the brain: cerebrum, brainstem, and cerebellum (“Brain Anatomy”). The hypothalamus, which sits above the brainstem, is one of the most important sleep regulating centers and monitors many unconscious bodily functions. It contains neurons that regulate arousal, or wakefulness. Much like the hypothalamus, the brainstem is important for unconscious behaviors, including communicating with the hypothalamus to transition between sleep and wake and relaxing muscles during REM sleep, keeping us from acting out our dreams. The thalamus, which sits above the brainstem in the middle of the brain, has a key role in memory formation during sleep. The thalamus relays information from the senses to the cerebral cortex, in the cerebrum, helping convert information to memory. The thalamus is less active in NREM sleep, but becomes more active in REM sleep, influencing dreams by transmitting images and sensations to the cortex. The amygdala and hippocampus are vital for sleep and dreams (“Brain Basics”). The amygdala is an almond shaped collection of nuclei found in the temporal lobe. The amygdala plays important roles in emotion and behavior. When we are exposed to fearful stimulus, the information is first sent to the amygdala, which then can then send signals to other parts of the brain. In addition to initiating a fear response, the amygdala is important in forming fear and anxiety associated memories. For example, when a tone is played near mice then the mouse is shocked, the mice with an intact amygdala will display a fear reaction when the tone is played; however mice with a impaired amygdala will not. The amygdala is not just concerned with fear; it also is involved in the formation of positive memories so a popular definition is that the amygdala is involved with evaluating environmental stimuli, determining their importance, and generating responses to the stimuli (“Know Your Brain: Amygdala”)(“Know Your Brain: Limbic System”). The hippocampus, which is also found in the temporal lobe plays a critical role in memory consolidation, including spatial navigation memory and orientation (“Know Your Brain: Hippocampus”). In sleep, hippocampal sleep spindles support memory consolidation (Ferrara et al.). This function is important to dreams, as many theories place memory as important for the generation and function of dreams.

Synapses are an essential target for the restorative properties of sleep. A synapse is the region where two neurons connect and communicate through chemical or electrical signals. The space between the two neurons is called the synaptic cleft. The presynaptic neuron is the signaling neuron that releases neurotransmitters into the synaptic cleft. The postsynaptic neuron is the neuron that receives the signal, based on the signals received, a postsynaptic neuron may release an action potential, or a rapid event that causes a neuron to signal to other neurons. The type of neurotransmitter, excitatory or inhibitory, can influence the odds of an action potential. Excitatory neurotransmitters, for instance glutamate, promote the generation of an action

potential, while inhibitory transmitters, like GABA, prevent it (“What Are Neurotransmitters?”). Recent studies demonstrated that sleep directly impacts synaptic strength (Cline). One leading theory summarizing this effect, is the synaptic homeostasis hypothesis (SHY). SHY states that when awake synapses strengthen, but while asleep most synapses weaken with select ones strengthening impacting memory consolidation (Tononi and Cirelli). As many theories propose that dreams strengthen memory formation, synapses may play a role in the creation, presence, or function of dreams. Furthermore specific theories, such as the reverse learning theory, directly discuss the connection between synapses and dreams.

ii. NREM Sleep

Of the two types of sleep, NREM sleep is about 75 to 80 percent of sleep in adults with its three stages, N1, N2, and N3 (Colten). N1 is the lightest stage of sleep and lasts about 7 minutes. During N1, heart rate and breathing slow down, and the brain produces alpha and theta waves. N2 lasts around 25 minutes. In N2 one is much less likely to be awakened, and heart rate and breathing continue to slow, as the body temperature drops. During this stage the brain activity is synchronous and consists of bursts of activity called sleep spindles and k-complexes which are important for memory consolidation, and learning. The last NREM stage is N3, or slow wave sleep. This stage lasts about 40 minutes and is essential for key restorative bodily processes including repairing muscle and tissue, encouraging growth and development, and improving immune function. This stage has no muscle movement, and the brain produces slow delta waves (Pacheco)(Suni)(“What are the Sleep Stages”).

iii. REM Sleep

REM is where the bulk of dreaming happens, although some dreaming does occur in NREM. In REM the brain becomes nearly as active as awake, and the brain activity is asynchronous with theta and alpha waves being frequent and connected to dreaming and successful dream recall. During REM the body has fast and irregular breathing, increased heart rate, blood pressure, and the muscles become paralyzed to prevent acting out dreams. REM is important for concentration and mood regulation, and even helps humans be more emotionally sensitive (Suni)(“What are the Sleep Stages”).

An important concept to note is the connection between REM, NREM, and dreaming. While a majority of dream research surrounds REM, dreaming also happens during NREM which, if researched, could open up a new perspective into dream research as the two types of sleep are very different in their brain and bodily activity. Similarly, understanding different aspects of sleep is important to make useful connections about dreams and the human brain. Some of those facets are the biology of different parts of the brain and what they do, the synapses that make up the brain, and all of the stages of sleep. With these three we can make connections between memory, emotion, and other brain functions that could be important to dreams.

Dream Theories

It is unknown why humans dream; however there are a multitude of popular theories that try to explain why we dream. The three theories I’m touching on are the threat simulation,

reverse learning, and continual activation theories. The threat simulation theory was first introduced in Antti Revonsuo in 2000, the reverse learning was introduced by Francis Crick and Graeme Mitchison in 1983, and the continual activation theory was introduced in Jie Zhang in 2004. All three are fairly prominent theories, but more importantly they each highlight different aspects of sleep, such as memory consolidation and negative emotion. It is important to note that these theories are older and have been expanded on multiple times over the years.

i. Threat Simulation Theory:

The threat simulation theory deals mostly with fear and how humans evolved (Revonsuo). The theory is based on six propositions:

1. Dream consciousness is an organized and selective simulation of the perceptual world.
2. Dream consciousness is specialized in the simulation of threatening events.
3. Nothing but exposure to real threatening events fully activates the threat-simulation system.
4. The threat simulations produced by the fully activated system are perceptually and behaviorally realistic rehearsals of threatening events.
5. The realistic rehearsal of these skills can lead to enhanced performance regardless of whether or not the training episodes are explicitly remembered.
6. The ancestral environment in which the human brain evolved included frequent dangerous events that constituted extreme threats to human reproductive success. They thus presented serious selection pressures to ancestral human populations and fully activated the threat-simulation mechanisms.

The first proposition means that dreams are far too organized and lifelike to be accidental. Adding to this, there is a notable disconnect between daily activities and dream content, with activities like reading and writing not appearing in dreams. Together the first proposition suggests that dreams are a simulation of the real world.

Building on this, the second proposition states dreams are a simulation of specifically threatening events. The first part of this proposition deals with dream content representing threatening elements. Supporting this, most emotions in REM dream reports are negative, with fear, anger, and stress being the most frequent (Hall and Van de Castle)(Strauch and Meier). He states that negative emotional dreams increase fitness in dangerous scenarios, which leads to increased reproductive success. The second part of this proposition suggests that dream content is more aligned with our ancestral environment than our present one. Studies indicate that in dreams the most common aggressors are unknown males and animals, even though such encounters are rarely aggressive in today's world (Hall and Van de Castle). This supports the idea that our dreams are reflecting our evolutionary environment.

The third proposition states that only realistic threatening events can activate the threat simulation system and is mainly supported by studies on REM sleep. Such as, patients with PTSD and anxiety spend more time in REM than control patients and had increased brain activity during REM. Another study shows that children exposed to trauma had better dream recall than non-trauma exposed children, suggesting dreams create stressful or dangerous

simulations of real life events of which the patients of these studies have more of, which activates the threat simulation system.

Proposition four works off of the third one, saying that, to be useful, dreams need to be realistic threat simulations. A supporting study showed that dreams contain vivid, real images and that dreamers lack awareness of their dreaming state which ensures an authentic experience, and enhances the motivation to defend themselves (Boeve et al.). Supporting this, mental imagery of motor actions uses the same neural mechanisms as actual motor action, making dreamed actions realistic, adding to the authenticity of the threat simulations.

Proposition five states the threat avoidance skills performed during dreams enhances real life responses. Because motor imagery and mental training can improve muscular strength, learning motor skills, and sports, then dreams can have similar effects, thus enhancing threat-avoidance skills (Yue and Cole)(Hall et al.)(Lejeune et al.). Supporting this, amnesic patients can learn motor skills and implicit emotional memory, demonstrating that we can learn implicitly, or without consciously remembering the dreams, and still gain threat-avoidance skills.

Proposition six states that our dangerous ancestral environment supported the threat simulation of dreams, through practically natural selection. Our ancestors had an intense pressure to reproduce, as 25% of people died before entering a reproductive age and 70% of those people died before reproducing (Meindl). Being able to recognize and perceive threats would aid in avoiding and coping with danger, all skills gained while dreaming.

The threat simulation theory posits that dreams were evolved as a way to prepare us for threatening situations, ultimately improving our chances of reproductive success. It says that dreams are intentionally realistic and draw from our most negative long-term memories to help implicitly learn skills. More recent reviews continue to support and expand on this theory, such as the review paper published in 2009 Katja Valli and the original author Antti Revonsuo. In the review they found that the original evidence still stood, and the theory was valid. The major strength of the theory is that it accounts for prevalence of threats in dreams. They do note that its largest weakness is that there is no direct evidence for the effects of the threat rehearsal of dreams; however they do note that could be addressed through video game or virtual reality simulations (Valli and Revonsuo).

ii. Reverse Learning Theory

The reverse learning theory says that the purpose of REM sleep is to remove unneeded information and memories, which in excess can lead to the brain not functioning properly, as they call it, ‘parasitic modes of behavior’ (Crick and Mitchison). The theory is based on the idea that our brain is a network of interconnected cells that support many different types of activity, and that its nature leads to it having parasitic modes, or unneeded memories. During REM sleep, these modes are detected and suppressed. Memory in the brain is stored by groups of synapses with different strengths that retain associations, or what we would call memories. The theory posits a few characteristics of memory. First, memory is distributed over many synapses and information will not be lost if a few synapses are removed. Furthermore, one synapse stores several pieces of information. The theory states that humans accumulate vast amounts of

information, including many parasitic modes, which cause a drain on the brain. To maintain functionality, the brain employs a mechanism that sends a signal through the brain that removes the parasitic modes. The proposed mechanism excites parts of the brain and more easily excites parasitic modes than important memories, which require specific signals. The mechanism dampens the modes that are excited, removing the unneeded memories. The theory places dreams as a byproduct of this mechanism. During REM sleep the brain is removed from external inputs, but remains highly active from the random signals that excite memories, which are reflected into dreams. To end, the reverse learning theory states that dreams are a byproduct of the brain removing unneeded memories during REM sleep.

While written in 1983, the theory continues to be supported by some researchers (Langille). A paper written in 2019 by Jesse Langille supports and slightly modifies Crick's theory. Langille notes and supports the idea that sleep has a role in memory consolidation, but also in "adaptive forgetting," supporting Crick. Langille modifies the argument, pointing that there may be a cellular mechanism for sleep forgetting as well. Langille goes on to say that the established sleep stage rhythms consolidate new memories in the cortex and hippocampus, and that the same rhythms simultaneously help the forgetting of older memories in those brain regions. Not only does Langille agree in that regard, Langille also refines the idea of random signals that Crick proposed. Langille proposes that the brain uses its oscillations, specifically sharp waves, to stabilize any adaptive information, and to isolate and remove non-adaptive data.

iii. Continual Activation Theory

The continual activation theory was proposed in 2004 by Jie Zhang. Zhang says that dreams are experienced when the conscious brain hits a certain level of in-activity and, through a certain mechanism, becomes active and retrieves memories. (Zhang)One distinction to understand in the theory is between declarative and procedural memory. Declarative is conscious memory and procedural is non-conscious. The theory also separates brain function into three parts: awake, NREM, and REM.

As shown in figure 1, when awake the sensory memory system is constantly receiving information from the senses. The brain needs to store this information and make sense of the surrounding environment. When sensory information is received, it is temporarily stored in the sensory memory. That information is then passed to the "conscious working memory," or the conscious memory used daily. The conscious working memory retrieves relevant data from either the declarative temporary or long-term memory associated with the stimuli.

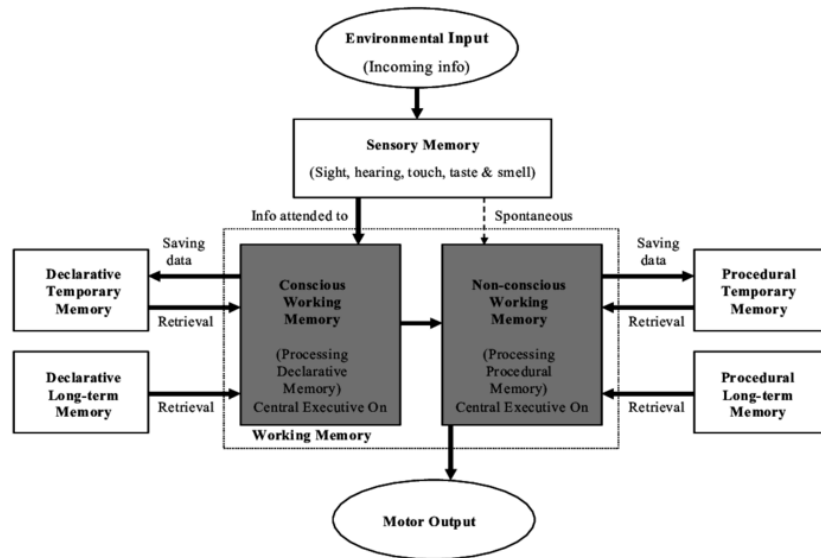


Fig 1: Memory model for the waking brain from Zhang

The processed information is sent to the declarative temporary memory for rapid memory saving. At the same time, the non-conscious working memory retrieves corresponding procedural memory from either the procedural temporary or long-term memories. The ensuing memories are also rapidly stored, similar to how declarative memory works while awake. Notably, procedural temporary memory retrievals influence motor output.

While awake, both the declarative and procedural long-term memories are in retrieval mode only, while the temporary memories are constantly saving and retrieving memories. This means that during waking time, both the conscious and non-conscious systems of working memory are continually activated.

During NREM, the brain is mostly engaged with the conscious working memory. The brain goes through the declarative temporary memory and compares it with the declarative long-term, and removes any unwanted, overlapping, or duplicated data. The new or updated data is transferred to the declarative long-term memory. Differing from wake, both declarative and procedural temporary memories are in retrieval-only mode as there is nearly no sensory memory to be added. If awakened during NREM, a thought-like brain activity happens, which is defined as a type I dream. Because the procedural memory is not getting engaged with by the brain, the continual-activation mechanism is triggered, and random data is retrieved from the procedural memory stores. This data flows through the non-conscious working memory to maintain brain activity. Because our muscles are locked during NREM this data is not acted out, however sometimes the data is too strong or the muscles are not fully locked, so things like sleep walking, sleep talking, and periodic limb movement disorders can happen.

During REM, the brain is mainly engaged with the non-conscious working memory. Similar to the declarative in NREM, the brain files the procedural memory for long-term storage, the process causing the rapid eye movement of REM. While the brain is mainly engaged with the

procedural memory, the continual-activation mechanism randomly retrieves declarative memories. The easiest data, like residue from the day, will go first as it is more salient. This data causes thoughts similar to the type I dream. Soon, as the declarative memories flow through the conscious working memory, the brain engages and attempts to make sense of the data. With the associative thinking system and emotion system of the brain engaged, vivid dreaming (type II dream) starts. Then the level of brain activation from the conscious working memory is increased. Based on the dreamer's thinking, the brain retrieves memories to turn into pictures and realistic events. While the random data retrieved from the continual activation mechanism is still interpreted by the brain, the type II dream is mainly influenced by the dreamer's thinking.

To conclude the continual activation theory, the working memory has two systems: the conscious and non-conscious systems. Both have to be continually activated, and when the flow of data drops below a threshold, the continual activation mechanism randomly retrieves memories. Dreaming as an experience, is the result of the conscious system of the brain getting involved in the continual activation data stream.

Biological Factors Affecting Dreaming and its Theories

While there is much to understand with the basic biology and some of the theories around dreams, even knowing one theory and some of the basics can help to make important connections. This section's purpose is to look into specific factors that may affect dreams and its theories. The first thing to cover merges brain biology with fear and dreams. Then the last connection to cover is how certain disabilities affect dreams. Because dreams cannot be studied in lab animals like mice, research on dreams can be hard to conduct. To counteract this researchers use people with disabilities that highlight certain parts of the body as important for dreams.

i. Brain Biology

One illuminating connection is how the amygdala and hippocampus affect dreams. While it is fairly agreed upon that the hippocampus is important for memory consolidation, a paper about both the hippocampus and amygdala adds support to the idea and connects both to dreams (De Gennaro et al.). This paper shows a solid correlation between hippocampal volume and density and vividness and bizarreness of dreams. Because dreams and REM specifically are established to affect memory consolidation, the fact that the hippocampus affects dreams and is highly active in REM means that it too is important to memory consolidation. This paper also lends support to the idea that the hippocampus is important to spatial navigation and orientation because the bizarreness seen in patients with a smaller hippocampus could be because their hippocampus is less effective in making dreams make spatial sense. The paper connects the amygdala even more concretely than the hippocampus. The research shows that the amygdala is connected to the emotional load of dreams and patients word count in describing them. While this may mean a lot of different things, it places the amygdala as important to dreams and shows that it functions the same awake and asleep. The amygdala's volume also negatively correlated with bizarreness, meaning the higher the volume of amygdala the less bizarre the patients'

dreams were. This could be because the amygdala deals with the processing of fear response and is one of the first parts of the brain to receive stimuli and chooses whether a response is necessary and therefore the larger the amygdala the more the dream emotions would be dealt with sensibly rather than bizarrely.

Another paper connects the amygdala to dreams (Blake et al.). The paper looks at patients with Urbach-Wiethe Disease (UWD) which affects the basolateral amygdala. The dreams of these patients were shorter, less complex, and more pleasant than the control patients. This study is similar to the first however it does add complexity. It agrees with shorter dream recalls, however it says that the amygdala is more important to negative emotions in dreams than positive ones. The dreams of the patients with UWD still had emotions, however they were more positive. This insight could be useful as the amygdala is important for both positive and negative emotions and responses in wake. The studies solidly place the hippocampus and the amygdala as important for dreaming. The papers also show that the amygdala is connected to the emotional content of dreams, and more specifically the basolateral amygdala is important for negative content of dreams.

ii. Dreams and Disease, Disorders

Dreams are a hard thing to study in humans as specific parts of a human cannot be tested like with lab mice. Because of this, many studies around dreams and REM sleep use patients that have disorders like PTSD, Parkinson's, Urbach-Wiethe Disease, and others. Because many of these diseases target specific parts of the brain, scientists are able to look at the effects of losing some function in certain parts of the brain. These studies around diseases can tell us many important things like what the dopaminergic system does to dreams, and how emotions and the amygdala affect dreams.

One disease that can be used for study is Parkinson's Disease. While it affects many parts of the body and brain, like the limbic system, cholinergic system, and the cortex it can be very useful to study in the context of dreams because of its effect on the dopaminergic system (Valli et al.). One study says that a hypodopaminergic state, which is a state of reduced dopamine activity in the brain and is identified by higher doses of dopamine agonists, may be associated with impoverished dream reports, in the sense that they lack bizarreness and emotional load (De Gennaro et al.). This connection is a strong support for the importance of the dopaminergic system in dream experience and generation.

Another disorder that presents a novel way to look at dreams is Post Traumatic Stress Disorder (PTSD) because of its connection to sleep through its connection to memory, and because of the nightmares that are a symptom of PTSD. One study found that the theta frequencies produced in REM are reduced in patients with PTSD (Sopp et al.). The researchers continued by saying that REM theta activity could have a role in emotional memory processing. This was backed up by their findings that REM theta activity is correlated to reduced re-experiencing of trauma. Another study agrees, finding that 'resilient' patients have increased REM theta activity, and that that activity shows the capability of adaptive emotional memory processing (Cowdin et al.). The study goes on to say that because patients with PTSD have

altered activity in the amygdala and thalamus in both wake and sleep, that reduced REM theta activity could mean a weaker communication between the structures. The study also supports the idea that REM sleep functions to integrate and eventually remove traumatic memories.

While some diseases and disorders have fairly large populations and thus are generally more useful to study, there are many other diseases that are less widespread and can be illuminating to the study of dreams. One example is Urbach-Wiethe Disease(UWD). One of UWD's symptoms can be a calcification of the basolateral amygdala. Because of the clear focus on the basolateral amygdala on dreaming, the study found first that the basolateral amygdala is not indispensable for dreaming (Blake et al.). The study also found that the patients with UWD had shorter, more pleasant, and less complex dreams than the control patients. Another infrequent disorder is Charcot-Wilbrand syndrome(CWS) which is dream loss following brain damage. The study focuses on one case in 1997, where a 73 year old woman had a bi-lateral deep occipital stroke (Bischof and Bassetti). The woman had a wildly vivid dream that resembled a hallucination the second night after her stroke. The study postulates that the damaged visual processing areas are responsible for the vivid dream, however that damage is also responsible for the total dream loss found later. This observation supports a link between dreaming and hallucinations, at least in certain clinical situations. The study's final takeaway is that there must be a dissociation between REM sleep and dreaming. The patient continued to sleep with normal REM traits however no recall of dreams ever when awakened during REM. The studies proposed that while REM and dreaming may be linked, they could be generated independently. These two studies show us how researching uncommon diseases and how they affect dreams illuminates new facets of dreams.

Discussion

Proper sleep is known to be essential for health, but what still remains a mystery is the role of dreams. Sleep is known to involve many brain regions, such as the hypothalamus or cortex, and profoundly affect synapse dynamics. We discussed the potential of dreams to play a role in these functions of sleep and theories behind dreams, such as the threat simulation theory, the reverse learning theory, and the continual activation theory. The theories all lend insights into dreams and may suggest what biology is at play.

The hippocampus and amygdala are important brain regions for dreaming. De Gennaro et al. and Blake et al. show that the amygdala is important for, if not drives, the emotional intensity of dreams, and more specifically the basolateral amygdala which is important for negative content of dreams. These findings indirectly support the threat simulation theory, as the amygdala deals with threat response in wake, and is important to dreaming; connecting dreams to threat response. Future studies could examine if experiencing threats improves performance in a given task, possibly using virtual reality technology to test the threat simulation theory.

The papers also lend support to the idea that the hippocampus is essential for cognition such as spatial navigation and orientation memory. A smaller hippocampus increases the bizarreness, or non-realistic people, events, or objects, of patients' dreams. This could mean that

their hippocampus is less effective in having their dreams make spatial sense. The import of the hippocampus could also support the continual activation theory as the hippocampus is crucial for memory and the continual activation theory says that dreams are a byproduct of our memory systems.

The Parkinson's disease studies indicated the importance of the dopaminergic system, suggesting dreams are not a byproduct of certain aspects of sleep like some theories claim. The dopaminergic system is important for our bodies reward system and the motivation for actions (Bromberg-Martin et al.). Connecting the dopaminergic system to dreams could mean dreams are an active choice by the body instead of a byproduct of a mechanism like the continual activation mechanism. This would lend support to threat simulation, and reverse learning, however it is in contention with the continual activation theory. Because Parkinson's is such a widespread disease, there is a lot of research about its effects and causes, but not on its relation to dreams. In the future it would be valuable to see more targeted dream research with Parkinson's as its concrete connection to the dopaminergic system and dreaming.

PTSD allows useful connections to dreams because of its connections to memory and certain parts of the brain. Challenging the belief that sleep deprivation is useful in the aftermath of traumatic events, Sopp et al. shows that sleep reduces overall symptoms and specifically re-experiences of trauma. Cowdin et al. also shows that REM sleep reduces amygdala reactivity to previously encountered emotional stimuli. These findings are an example of why specific research around certain disorders can be so potent. Using dreams and sleep as a basis for research these studies showed that while nightmares and other potent dreams face a challenge to patients with PTSD, ultimately the cure is sleep, and not sleep deprivation which worsens the problem.

The research on less widespread disorders lead to even more potent connections. The research on UWD shows that the basolateral amygdala is not indispensable for dreaming. It also shows that the basolateral amygdala is important for the negative emotions in dreams, which would make sense as it functions with anxiety and defensive behaviors in wake (Blake et al.)(Beyeler and Dabrowska). Based on the paper's findings one could argue that because the loss of the basolateral amygdala, the part of the brain that deals with anxiety and defense in wake, has such a big impact on dreaming, then dreaming has a large function for simulating threats and dealing with anxiety. However, one could also challenge that by the idea that the basolateral amygdala is not necessary for dreams, and emotional content is still present in the UWD patient's dreams. Either way the paper shows the importance of studying disorders as the information on the basolateral amygdala offers insights for a wide range of problems. Similarly Bischof and Bassetti could highlight the importance of dream research. The idea that REM sleep and dreaming are linked is an insight that can be linked to the dream theories. It challenges the continual activation and reverse learning theories as they place dreams as a byproduct of a mechanism of REM sleep, however it would support the threat simulation theory as it proposes a specific mechanism and purpose for dreams.

To end, the connections made between diseases, biology, and sleep can lend helpful insights into dreams. Not only that but dream research, especially in the context of diseases, can illuminate how to help people with those diseases. As targeted dream research helps provide novel ways to look at disease and sleep health, it is an important area that, if researched more, would help different areas of study greatly.

Works Cited

- Beyeler, Anna, and Joanna Dabrowska. "Neuronal Diversity of the Amygdala and the Bed Nucleus of the Stria Terminalis." *Handbook of Behavioral Neuroscience*, 2020, pp. 63–100. <https://doi.org/10.1016/b978-0-12-815134-1.00003-9>.
- Bischof, Matthias, and Claudio Bassetti. "Total Dream Loss: A Distinct Neuropsychological Dysfunction After Bilateral PCA Stroke." *Annals of Neurology*, vol. 56, no. 4, Wiley, Sept. 2004, pp. 583–86. <https://doi.org/10.1002/ana.20246>.
- Blake, Yvonne et al. "The role of the basolateral amygdala in dreaming." *Cortex; a journal devoted to the study of the nervous system and behavior* vol. 113 (2019): 169-183. doi:10.1016/j.cortex.2018.12.016.
- Boeve, B F et al. "REM sleep behavior disorder and degenerative dementia: an association likely reflecting Lewy body disease." *Neurology* vol. 51,2 (1998): 363-70. doi:10.1212/wnl.51.2.363
- "Brain Anatomy and How the Brain Works." *Johns Hopkins Medicine*, 14 July 2021, www.hopkinsmedicine.org/health/conditions-and-diseases/anatomy-of-the-brain.
- "Brain Basics: Understanding Sleep." *National Institute of Neurological Disorders and Stroke*, www.ninds.nih.gov/health-information/public-education/brain-basics/brain-basics-understanding-sleep#:~:text=Anatomy%20of%20Sleep&text=The%20hypothalamus%2C%20a%20peanut%2Dsize,centers%20affecting%20sleep%20and%20arousal.
- Bromberg-Martin, Ethan S et al. "Dopamine in motivational control: rewarding, aversive, and alerting." *Neuron* vol. 68,5 (2010): 815-34. doi:10.1016/j.neuron.2010.11.022
- Cline, John. "Sleep and the Amazing Shrinking Synapse." *Psychology Today*, 30 Mar. 2017, www.psychologytoday.com/us/blog/sleepless-in-america/201703/sleep-and-the-amazing-shrinking-synapse. Accessed 5 Nov. 2023.
- Colten, Harvey R. "Sleep Physiology." *Sleep Disorders and Sleep Deprivation - NCBI Bookshelf*, 2006, www.ncbi.nlm.nih.gov/books/NBK19956/#:~:text=NREM%20sleep%20constitutes%20a%20bout%2075,is%2070%20to%20100%20minutes.
- Cowdin, Nancy et al. "Theta frequency activity during rapid eye movement (REM) sleep is greater in people with resilience versus PTSD." *Experimental brain research* vol. 232,5 (2014): 1479-85. doi:10.1007/s00221-014-3857-5.
- Crick, F., and Mitchison, G. (1983). "The function of dream sleep." *Nature*, 304(5922), 111–114. <https://doi.org/10.1038/304111a0>.
- De Gennaro, Luigi et al. "Amygdala and hippocampus volumetry and diffusivity in relation to dreaming." *Human brain mapping* vol. 32,9 (2011): 1458-70. doi:10.1002/hbm.21120.
- . "Dopaminergic system and dream recall: An MRI study in Parkinson's disease patients." *Human brain mapping* vol. 37,3 (2016): 1136-47. doi:10.1002/hbm.23095.
- Ferrara, Michele et al. "Hippocampal sleep features: relations to human memory function." *Frontiers in neurology* vol. 3 57. 17 Apr. 2012, doi:10.3389/fneur.2012.00057.
- Hall, Calvin, and Robert Van de Castle. "The content analysis of dreams." (1966).

- Hall, Calvin et al. "Imagery and the acquisition of motor skills." *Canadian journal of sport sciences = Journal canadien des sciences du sport* vol. 17,1 (1992): 19-27.
- "Know Your Brain: Amygdala." @Neurochallenged, neuroscientificallychallenged.com/posts/know-your-brain-amygdala.
- "Know Your Brain: Hippocampus." @Neurochallenged, neuroscientificallychallenged.com/posts/know-your-brain-hippocampus.
- "Know Your Brain: Limbic System." @Neurochallenged, neuroscientificallychallenged.com/posts/know-your-brain-limbic-system.
- Langille, Jesse J. "Remembering to Forget: A Dual Role for Sleep Oscillations in Memory Consolidation and Forgetting." *Frontiers in cellular neuroscience* vol. 13 71. 12 Mar. 2019, doi:10.3389/fncel.2019.00071.
- Lejeune, M et al. "Mental rehearsal in table tennis performance." *Perceptual and motor skills* vol. 79,1 Pt 2 (1994): 627-41. doi:10.2466/pms.1994.79.1.627
- Meindl, R. S. "Human Populations Before Agriculture." *The Cambridge Encyclopedia of Human Evolution*, 1992.
- Pacheco, Danielle "What Is NREM Sleep?" *Sleep Foundation*, 3 Nov. 2023, www.sleepfoundation.org/stages-of-sleep/nrem-sleep.
- Revonsuo, A. (2000). "The reinterpretation of dreams: An evolutionary hypothesis of the function of dreaming." *Behavioral and Brain Sciences*, 23(6), 877–901. <https://doi.org/10.1017/s0140525x00004015>.
- Sopp, M. Roxanne, et al. "REM Theta Activity Predicts Re-experiencing Symptoms After Exposure to a Traumatic Film." *Sleep Medicine*, vol. 54, Elsevier BV, Feb. 2019, pp. 142–52. <https://doi.org/10.1016/j.sleep.2018.10.030>.
- Strauch, Inge, and Barbara Meier. "In Search of Dreams." *State University of New York Press*, Jan. 1996, sunypress.edu/Books/I/In-Search-of-Dreams.
- Suni, Eric "Stages of Sleep: What Happens in a Sleep Cycle." *Sleep Foundation*, 3 Nov. 2023, www.sleepfoundation.org/stages-of-sleep.
- "The synapse." *Khan Academy*. <https://www.khanacademy.org/science/biology/human-biology/neuron-nervous-system/a/the-synapse>.
- Tononi, Giulio, and Chiara Cirelli. "Sleep and the price of plasticity: from synaptic and cellular homeostasis to memory consolidation and integration." *Neuron* vol. 81,1 (2014): 12-34. doi:10.1016/j.neuron.2013.12.025.
- Valli, Katja, and Antti Revonsuo. "The threat simulation theory in light of recent empirical evidence: a review." *The American journal of psychology* vol. 122,1 (2009): 17-38.
- Valli, Mikael, et al. "Neuroimaging of Rapid Eye Movement Sleep Behavior Disorder and Its Relation to Parkinson's Disease." *Journal of Neuroscience Research*, vol. 100, no. 10, Wiley, July 2022, pp. 1815–33. <https://doi.org/10.1002/jnr.25099>.
- "What Are Neurotransmitters?." *Queensland Brain Institute - University of Queensland*, 29 July 2022, qbi.uq.edu.au/brain/brain-functions/what-are-neurotransmitters.

- “What Are the Sleep Stages? - National Sleep Foundation.” *National Sleep Foundation*, 3 May 2023, www.thensf.org/what-are-the-sleep-stages.
- Yue, G, and K J Cole. “Strength increases from the motor program: comparison of training with maximal voluntary and imagined muscle contractions.” *Journal of neurophysiology* vol. 67,5 (1992): 1114-23. doi:10.1152/jn.1992.67.5.1114
- Zhang, J. (2016). Continual-Activation theory of dreaming. *Nocuffs*.
https://www.academia.edu/26464861/Continual_Activation_Theory_of_Dreaming#:~:text=The%20theory%20presented%20in%20this,to%20maintain%20proper%20brain%20functioning.

Investigation of High-Performance Sodium-Ion Batteries' Electrode through the Optimization Materials and Manufacturing Processes. By Lam Anh Khoi (Kelvin Lam)

Abstract:

This research paper explores different techniques for approaching an optimal sodium ion battery and identifies a suitable electrode for the cell. Lithium-ion batteries are a well-rounded type of battery used worldwide because of their high efficiency and capacity. However, with climate change, global warming, and all the environmentally deteriorating factors occurring worldwide, the demand for a sustainable energy storage system forces the creation of alternative, better battery technology. Sodium-ion batteries have been shown to be promising candidates to replace lithium-ion batteries. The first part of this research is a basic overview of a battery's internal structure and process, including components like the anode, cathode, electrolyte, and current collector. The mechanism of how electrons are circulated to create an electric current is also presented. The electrode, also known as the electron donor and receiver, is a significant component of this battery structure, which allows electron circulation. The next part of the research is the processes in which the electrode is synthesized, including solid-state reaction, sol-gel, and hydrothermal methods. This research continues with the characterization of the electrode being tested using methods like X-ray diffraction (XRD), transmission electron microscopy (TEM), and energy-dispersive X-ray spectroscopy (EDS). The next part is the fabrication of SIBs, where the optimized electrodes are used. The cells will be tested using cyclic voltammetry (CV), galvanostatic discharge-charge cycling, and electrochemical impedance spectroscopy (EIS). A simulation of an optimized electrode will then be carried out using the Battery Explorer tool of The Material Project with performance data and characteristics. I found two promising materials for the Sodium-Ion Battery electrode: sodium vanadium oxide (NaV_3O_8) and hard carbon (sucrose heated at 1500°C with an inert atmosphere). I looked into their structure, capacity, energy rate, durability, and energy density compared with other materials and found that these materials stand out from the rest. This research successfully creates an insight into manufacturing a Sodium-Ion Battery in replace of a Lithium-Ion Battery with promising electrode makeup and manufacturing processes.

Acknowledgment

I would like to give the warmest attitude to my high school teacher, Dr. Abreu, who suggested this topic and navigated me throughout the process when I had any problems. With his guidance, I devised ideas and organized my research paper. I struggled to find a topic to research, but with his suggestions, I found myself more capable and my path to make this project successful more transparent. I also want to thank my brother, Kevin Lam, for coming up with ideas and helping me organize them in this paper. My experiments and simulations wouldn't have been done without his assistance and research skills.

Table of Contents

ABSTRACT.....	2
ACKNOWLEDGEMENTS.....	3
TABLE OF CONTENTS.....	4
INTRODUCTION.....	5-6
SODIUM-ION BATTERY DESCRIPTION.....	6-9
EXPLANATIONS.....	1-13
SIMULATIONS & DATA.....	14-16
RESULTS.....	16
DISCUSSION.....	17
CONCLUSION.....	18
REFERENCES.....	19-20

Introduction:

Lithium was and is undeniably a practical component in developing an electrode for a rechargeable battery. There has been a lot of competition in the battery industry with strong competitors like Solid State LFP Battery. However, an alternative in this industry is necessary for an era where sustainability is put on top. Sodium-ion batteries (SIBs) are a promising alternative to lithium-ion batteries due to their abundance (less reactive), low cost of sodium (abundant, uncomplicated manufacturing process, lower EV cost), environmental friendliness (less toxic raw materials), and longer lifespan (engine more packed, low operating voltage, which stresses less on electrode). However, the development of high-performance SIBs is hindered by the limited capacity and stability of the currently available electrode materials. With the high power to exceed its electrochemical properties, it becomes stronger in fast charging (80% in 15 minutes). It has high thermal stability and does not heat up like a Li-Ion battery, so it can still work well at different temperatures, even at -20°C. One of the most significant challenges is to find a suitable electrode material with high energy density and the long cycle life required for the practicalness of commercial use. Another is to find an electrode that works well with sodium ions and creates no conflict in the operation process. Therefore, this research project aims to

investigate promising electrodes in developing high-performance SIBs by optimizing the composition and structure of the electrode materials. One of the challenges I found in this research is the low opportunity to have hands-on research using an actual cell because the internal structure of a battery is highly complex, it is difficult to gather precise data, and potentially hazardous. Instead of creating a complete sodium ion battery and evaluating its performance in time, I instead investigated proper techniques in developing a suitable electrode and evaluating its performance using digital tools. The hypothesis of the experiment proposed in this research paper is that the higher and more balanced total gravimetric capacity, energy density, and discharge capacity of the electrodes, the more effective and durable the performance is. The findings and investigations of this research promise a better understanding of the currently developing Sodium-Ion Battery and the potential uses of promising electrodes. This research is a step further in developing a high-performance, commercially workable version of a Sodium-Ion Battery.

Sodium-Ion Battery Description:

- **Components**⁸:

Anode: This is the positive electrode. One side of the SIB accepts anions. It is made of hard-layered carbon material, which can interpolate sodium ions during the charge cycle. This site can accept sodium ions and electrons while charging.

Cathode: This is the negative electrode. The other side of the SIB accepts cations (sodium ions). It is layered with oxide materials like Sodium cobalt oxide (NaCoO_2), Sodium iron phosphate (NaFePO_4), and Sodium vanadium oxide (NaV_3O_8). This site can accept sodium ions and electrons while discharging.

Electrolyte: This can either be a liquid or a solid-state material that sodium ions can travel from the cathode to the anode and back. The electrolyte can be made of a sodium salt like Sodium perchlorate (NaClO_4), an ionic liquid such as 1-ethyl-3-methylimidazolium hexafluorophosphate (BMIM PF_6), or a polymer electrolyte like polyethylene oxide (PEO). There can be dangerous potential problems regarding the electrolyte; for example, when the battery's temperature reaches the electrolyte's boiling point (which ranges from 70°C to 250°C), the electrolyte will evaporate, causing the ions to collapse and cause an explosion. Moreover, the electrons can be attracted to the electrolyte, and a phenomenon called electrolyte degradation occurs, causing capacity loss, impedance increase, short-circuiting, and gas generation. However, this is prevented through an accidental phenomenon when the sodium ions travel through the electrolyte and make contact with the electrons, creating the SEI layer, which can prevent electrolyte degradation.

Current collector: This connects the anode and the cathode, allowing the electrons to flow from one side to another, creating electricity. The electrons will flow from the cathode to the anode during the charge cycle and from the anode to the cathode during the discharge cycle.

Separator: The separator is placed between the anode and the cathode, which serves as a savior when there is a short circuit (when the electrolyte evaporates) and allows the smooth transportation of sodium ions.

- **Mechanism:**⁸

-A sodium ion Battery is a rechargeable battery where the movements of sodium ions and electrons are the primary energy output source. Because of the highly reactive property of Sodium and the desire to be reduced, it immediately looks for electrons when released to nature as a (2+) ion. The battery lets the ions lose and absorb electrons while moving between the anode and the cathode. When the ion finds the electron, it slides back into the stable form with the metal oxide designated in the cathode side of the cell, and this movement releases energy.

-The movements of these Sodium ions across the cell are assisted by the help of the electrolyte, which does not allow electrons to accompany these ions. While charging, the sodium ions move from the cathode side, where the metal oxide is, to the anode, where a hard-carbon compound is placed. On the other hand, the electrons take a longer path through the circuit from the cathode's current collector to the anode's current collector and find themselves in the same place as the ions. When all the ions and electrons are designated to their area, the cell is fully charged, but this is just an unstable temporary state. When discharged, things start happening.

-When discharged, or energy is utilized for electronic activities like phones or torches, the sodium ions move back to the cathode side with the metal oxide through the electrolyte, and the electrons go opposite their charging pathway. When all the ions and electrons are transferred back to their original place, the cell is fully discharged or "out of battery."

-In special, unwanted conditions, when the temperature rises too high, the electrolyte (liquid form) evaporates, and there will be a "short circuit" between the sides, creating an explosion. The separator was created to prevent this phenomenon, allowing only sodium ions to pass due to its microporosity.

-This complex and detailed structure is compacted into copper and aluminum current collector sheets, with an electrolyte sandwiched by these current collectors. These sheets are then coiled around a battery core to be more compact and usable.

Materials and Methods in Developing a suitable electrode

Synthesis of Electrode Materials: A range of sodium-ion electrode materials will be synthesized using different synthesis methods, including solid-state reaction, sol-gel, and hydrothermal methods. The composition and structure of the materials will be optimized to improve their performance in SIBs.

Characterization of Electrode Materials: The synthesized electrode materials will be characterized using various analytical techniques, including X-ray diffraction (XRD), transmission electron microscopy (TEM), and energy-dispersive X-ray spectroscopy (EDS). These techniques will be used to determine the materials' composition, crystal structure, and surface morphology.

Fabrication of Sodium-Ion Batteries: Sodium-ion batteries will be fabricated using the optimized electrode materials and a sodium metal anode. The cells will be assembled in a standard half-cell configuration using a sodium metal anode and the optimized cathode material.

The cells will be tested using cyclic voltammetry (CV), galvanostatic discharge-charge cycling, and electrochemical impedance spectroscopy (EIS).

Explanations:

- **Synthesis of Electrode Materials:**

Solid-state reaction: Solid-state reactions create a new solid with a firmer structure than the reactants. The products could be polycrystalline, crystalline, glasses, or thin-film materials to be used for electronic materials like Sodium-ion Batteries.⁶ For the electrode case, the precursor materials like metal oxides and salts will be weighed and mixed using mortar and pestle. The material is then heated to molten form in a furnace. This heat, combined with specific properties of the reactant, will carry out the desired chemical reaction and phase transformation. Now, this electrode product will be ground into a fine powder and pressed to a desired shape for use.⁷

Sol-gel method⁹: A precursor sol is first prepared by dissolving a metal salt and a metal oxide in water or alcohol. Then, the precursor is transformed into a gel in the gelation process by adding acid or base. This forms a 3-D network structure and can be modified by varying pH. The solvent is then removed from the gel by heating or CO₂ extraction. The following process is calcination, where the dried gel is heated in a furnace to remove organic material and convert it into an electrode. This electrode product is then shaped ideally to be put into practice.

Hydrothermal method¹⁰: The main components of this method are heat and pressure. You start with a metal salt or metal oxide precursor dissolved in water or alcohol. Then, you put it into a vessel and seal it. Then, the vessel is heated to 300°C with high pressure for many hours to days to form electrodes. The reaction vessel is cooled down to room temperature, and the synthesized material is washed to remove any impurities and dried using vacuum drying.

- **Characterization of Electrode Materials:**

X-ray diffraction (XRD)¹¹: XDR determines the electrode's crystal structure. The mechanism uses the properties of light rays, where an X-ray beam is shone into the sample, and the diffraction angle is measured. This way, all the small crystal lattice structures can be determined. A detector will collect the distinctive patterns of spots and lines. These data can further determine the phase composition (binders, conductive additives, or electrolytes), crystal size, and lattice distortions. XRD helps a lot in analyzing capacity, voltage, and cycling capacity.

Transmission electron microscopy (TEM)¹²: This method is similar to XRD, in which a beam is passed through the sample to determine the internal structure. In the TEM case, a beam of electrons is passed through the sample; the electrons are collected on the other side to determine how compact that part of the sample is at a nanoscale. This technique is more accurate because electrons are much smaller than the light beam. TEM can even determine the impurities of the sample. TEM can be used to study the solid-electrolyte interface (SEI), which is extremely important in determining the degradation and seeking future improvements of the cell as this layer prevents the electrons from degrading the electrolyte.

Energy-dispersive electron X-ray spectroscopy (EDS): Instead of seeing how electrons or X-ray beam is passed through the sample. EDS technique uses the X-ray emitted from the sample when

bombarded with electrons to determine the internal structure and composition. It is often accompanied by scanning electron microscopy (SEM). A detector consists of a silicon semiconductor, and a preamplifier is placed close to the sample's surface to generate the most accurate data. When the X-ray emitted from the sample strikes the detector, it creates electron-hole pairs, which generate a small electrical current amplified by the attached preamplifier to be collected as line data.

- **Fabrication of Sodium-Ion Batteries:**

Cyclic voltammetry (CV)¹⁴: This technique fabricates and investigates electrochemical properties, such as redox reactions and charge storage capacity. The cell is analyzed using three connected electrodes (working, reference, and counter electrodes). The voltage applied to the electrodes is collected through a sensor into a computer for accurate data and graphs (potential-current graph). The voltage ranges from negative to positive, going back and forth with the current collected through a voltammogram. Usually, the graph is stable from a distant view; however, when zooming closer, the voltage fluctuates at a specific potential with peaks representing the reductions in the reaction process. Redox properties, kinetics, and capacity are then determined to optimize the electrode's performance for the battery's overall performance.

Galvanostatic discharge-charge cycling¹⁵: This fabricating method is the opposite of Cyclic voltammetry. The battery is charged and discharged at a constant applied current through a supercapacitor. This method can generate a potential vs. time graph to determine capacity, voltage, and electrochemical capacity. However, there can be a deviation in the process because of a series of resistances, causing the cell voltage to drop sharply when switching between charging and discharging states. If the cell is shorted, the parasitic current may be significant. This method reflects the reality of the battery more accurately.

Simulation:

- **Potential use of NaV_3O_8 as an electrode of SIB:**

-Scientists have been researching using NaV_3O_8 as an electrode for SIBs. They put them into the experiment, and the results were undoubtedly promising. When doped with Nb, NaV_3O_8 can be a competent material for SIB electrodes, allowing smoother movement of ions through the cell. VO_x materials contain a layered structure with a large capacity, excellent ion shuttling rate, and stable structure, which make them suitable materials with extensive attraction. However, because of their poor conductivity and low sodium-ion diffusion coefficient, combined with their unstable structure, their performance was not maxed out. Therefore, Nb-coating was used to solve these problems. It expands the lattice space, reduces the energy band gap, and increases the absorption of Sodium-ion. SEM and XRD showed that Nb-coated NVO possesses high structural stability during fast cycling.¹

-When doped with Nitrogen, the crystal structure of NVO is not changed, but the thickness and properties of the surface coating are enhanced. Electrochemical impedance spectroscopy (EIS) proved that coated materials like Nitrogen have a lower resistance than a bulked NVO, and the

Galvanostatic intermittent titration technique (GITT) has a higher sodium ion diffusion coefficient.³

- **Statistics of a simulation of the electrode NaV_3O_8 using “The Materials Project”:**

-Predicted properties:²

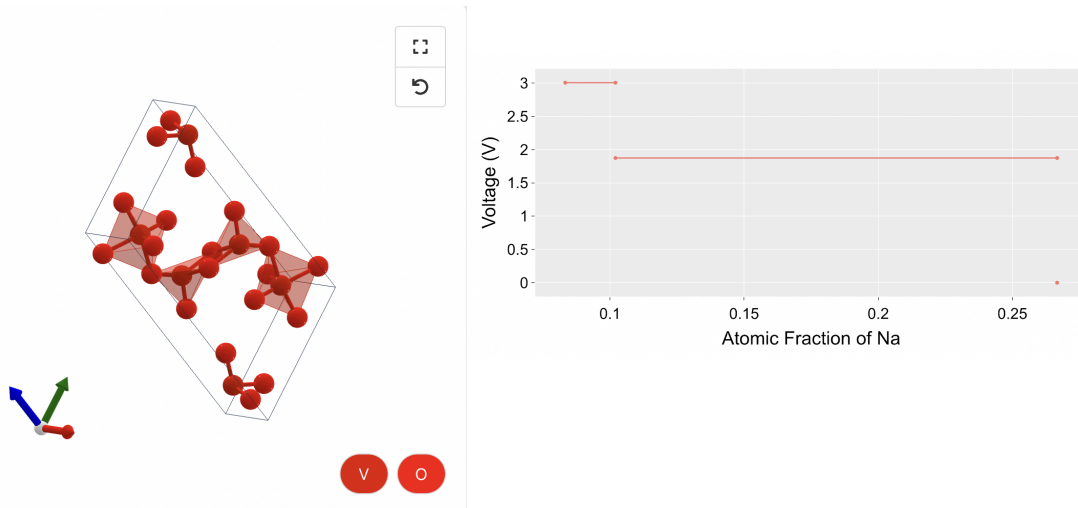
+Total Gravimetric Capacity: 215.69 mAh/g

+Total Volumetric Capacity: 765.13 Ah/l

+Volume change: 0.12%

+Energy Density: 1215.47 Wh/l

+Specific Energy: 240.59 Wh/kg



(Host material diagram of NaV_3O_8 used in a SIB's electrode)²

(Voltage graph of NaV_3O_8 SIB)²

-Discharged capacity and cycles:³

+Discharge capacity: 231.3 mAh/g at 150 mA/g

+Discharge capacity after 50 cycles: 249.5 mAh/g

-Volmmatogram¹³:

+2 peaks corresponding to the two-step reduction process of NaV_3O_8 to V_2O_3 .

+The first peak represents the reduction of NaV_3O_8 to $\text{Na}_3\text{V}_2\text{O}_8$, and the second one shows the reduction to V_2O_3 .

- High capacity, density, and charge/discharge capacity numbers from “The Material Projects” prove that NaV_3O_8 is capable of maintaining a high energy rate even after many cycles. The strongly bonded crystal structure of the specimen also plays a vital role in ensuring the stability of the cell.

- **The current use of “hard carbon” in manufacturing the anode for sustainable SIB:**

Similar to graphite-LIBs, hard carbon works well with SIBs. Its advantage is its low cost, high capacity, low voltage hysteresis (the difference between the charged and discharged voltage),

long cycle life, uncomplicated manufacturing process, and safety. Hard carbon is made of regions of curved graphene sheets and a lot of nanopores in microcrystalline areas. Hard carbon is better than crystalline carbon, whose amorphous structure increases sodium insertion and extraction. Hard carbon promises a high charging rate because of its high surface area and porosity, ensuring fast ion transport. Hard carbon applies to high-power operations. Lastly, hard carbon has a low operating potential, which means it can operate at a lower voltage while still delivering high energy density.⁴

- **Statistics of recent use of Hard Carbon in manufacturing electrodes of SIB:**

-Recorded data:⁴

+Capacity: 120mAh/g

+Cell energy density: 146.1 Wh/g

+Extension rates for over 200 cycles: 90%

-Long cycle test:⁵

+Voltage range: 0.0-0.2 V

+Current density: 25 mA g^{-1}

+Total duration of charge: 14 hours

Results:

We found that NaV3O8 is the optimal electrode for sodium-ion batteries. The optimization of the electrode materials is expected to result in significant improvements in the performance of SIBs. The results of this project will provide valuable insights into the mechanism of sodium-ion insertion/extraction and the factors that govern the performance of SIBs. The optimized electrode materials investigated in this project have the potential to be used in commercial SIBs, providing a cost-effective and environmentally friendly alternative to lithium-ion batteries.

Discussion:

The exploration of NaV3O8 as a potential electrode material for Sodium-Ion Batteries (SIBs) reveals promising insights into addressing challenges associated with conventional materials. Doping NaV3O8 with Nb emerges as a crucial strategy to enhance its competence as an electrode material. The layered structure of VO_x materials, including NaV3O8, offers significant advantages such as large capacity, excellent ion shuttling rate, and structural stability. However, inherent issues like poor conductivity, low sodium-ion diffusion coefficient, and structural instability necessitated innovative solutions for optimal performance.

The simulated statistics obtained from "The Materials Project" showcase NaV3O8's potential, with impressive predicted properties such as a total gravimetric capacity of 215.69 mAh/g, volumetric capacity of 765.13 Ah/l, minimal volume change (0.12%), energy density of 1215.47 Wh/l, and specific energy of 240.59 Wh/kg. The discharged capacity data, along with cyclic stability after 50 cycles, attests to NaV3O8's capability to maintain high energy rates over extended periods.

In parallel, the current utilization of "hard carbon" as an anode material for sustainable SIBs presents a compelling alternative. Hard carbon's distinct advantages, including low cost, high capacity, low voltage hysteresis, long cycle life, uncomplicated manufacturing process, and safety, position it as a viable candidate for SIB applications. The amorphous structure of hard carbon, with curved graphene sheets and abundant nanopores, enhances sodium insertion and extraction, contributing to its high charging rate and applicability to high-power operations. The results collectively indicate a promising direction for optimizing electrode materials in SIBs, with both NaV₃O₈ and hard carbon showing substantial potential. The insights gained from this research contribute to our understanding of sodium-ion insertion/extraction mechanisms and the critical factors influencing SIB performance. These optimized electrode materials hold the promise of providing a cost-effective and environmentally friendly alternative to lithium-ion batteries, offering a significant advancement in sustainable energy storage technologies. The potential commercialization of these materials would mark a pivotal step toward addressing the growing demand for efficient and eco-friendly energy storage solutions.

Conclusion:

A battery cell contains many complex and unstable components, making it difficult to simulate using simple lab instruments. This includes everything from the electrolytes and electrodes to the current collector and separators. The process of creating a Sodium-Ion Battery includes many steps, from synthesizing and characterizing to fabricating. In order to make a sodium ion battery work properly, certain suitable electrode materials are required to maintain the voltage rate and capacity of the battery. In this research, I found two promising materials that can fit into these spots, namely NaV₃O₈ and hard carbon, which have been proven to have high capacity and long charged discharge cycles.

Works Cited

- Author links open overlay panel Limin Zhu a c, et al. “The Improved Cycling Stability and Rate Capability of Nb-Doped NaV3O8 Cathode for Sodium-Ion Batteries.” *Journal of Alloys and Compounds*, Elsevier, 10 Sept. 2021, www.sciencedirect.com/science/article/abs/pii/S0925838821032941.
- “Battery Explorer - Mp-752972_na.” Materials Project, materialsproject.org/batteries/mp-752972_Na?_limit=50&formula=NaV3O8&working_ion=Na
- Author links open overlay panel Chunliang Pan a c, et al. “N-Doped Carbon Coated NaV3O8 Cathodes towards High-Capacity and Ultrafast Na-Ion Storage.” *Ceramics International*, Elsevier, 1 Apr. 2022, www.sciencedirect.com/science/article/pii/S0272884222010574.
- Liu, Liyang, et al. “Hard Carbons as Anodes in Sodium-Ion Batteries: Sodium Storage Mechanism and Optimization Strategies.” MDPI, Multidisciplinary Digital Publishing Institute, 2 Oct. 2022, www.mdpi.com/1420-3049/27/19/6516.
- Solid State Synthesis, www.sigmaaldrich.com/US/en/applications/materials-science-and-engineering/solid-stat
- Palomares, Verónica, and Teófilo Rojo. “Synthesis Processes for Li-Ion Battery Electrodes – from Solid State Reaction to Solvothermal Self-Assembly Methods.” IntechOpen, IntechOpen, 24 Feb. 2012, www.intechopen.com/chapters/29286.
- “Lithium-Ion Battery, How Does It Work?” YouTube, YouTube, 12 Apr. 2019, www.youtube.com/watch?v=VxMM4g2Sk8U.
- Bokov, Dmitry, et al. “Nanomaterial by Sol-Gel Method: Synthesis and Application.” *Advances in Materials Science and Engineering*, Hindawi, 24 Dec. 2021, www.hindawi.com/journals/amse/2021/5102014/.
- Gan, Yong X., et al. “Hydrothermal Synthesis of Nanomaterials.” *Journal of Nanomaterials*, Hindawi, 30 Jan. 2020, www.hindawi.com/journals/jnm/2020/8917013/.
- “X-Ray Diffraction Analysis.” X-Ray Diffraction Analysis - an Overview | ScienceDirect Topics, www.sciencedirect.com/topics/materials-science/x-ray-diffraction-analysis.
- Basak, Shibabrata, et al. “Characterizing Battery Materials and Electrodes via in Situ/Operando Transmission Electron Microscopy.” AIP Publishing, AIP Publishing LLC AIP Publishing, 1 Jan. 1970, aip.scitation.org/doi/10.1063/5.0075430.
- Fang Hu, et al. “Synthesis and Electrochemical Performance of Nav 3 O 8 Nanobelts for Li/Na-Ion Batteries and Aqueous Zinc-Ion Batteries.” *RSC Advances*, Royal Society of Chemistry, 2 July 2019, pubs.rsc.org/en/content/articlehtml/2019/ra/c9ra04339j.
- A Practical Beginner’s Guide to Cyclic Voltammetry. pubs.acs.org/doi/10.1021/acs.jchemed.7b00361.
- “Supercapacitor.” CleanEnergy WIKI, cleanenergywiki.org/index.php?title=Supercapacitor#:~:text=Galvanostatic%20cycling%20is%20the%20experimental,accurately%20reflects%20real%2Dworld%20performance.

Engineering a Deep Learning Model to Correlate Between Single-cell RNA Sequencing Gene Datasets By Ameya Bhide

Abstract

Combining deep learning and single-cell RNA sequencing data has become increasingly popular due to the potential of both scRNA-seq and deep learning. However, creating a neural network in order to correlate single-cell RNA sequencing data is a more relatively unexplored area, and the key to understanding why it is needed and how it is done relies on understanding deep learning and single-cell RNA sequencing separately, and the power of combining them. By being able to predict single-cell RNA gene data sets, time, money and efficiency would be conserved. This paper details the creation of such a model to correlate between scRNA-seq gene data sets, as well as the measures of its performance and future steps.

Introduction

Single-cell RNA sequencing is a technique commonly used to measure the gene expression of thousands of cells at the same time. With easy access to these gene expressions from a population, scientists are able to use this data to answer more questions about the population of cells at large. They can also use that population to learn more about individual genes. The importance of singlecell RNA sequencing to this paper is that it allowed for the collection of 33694 genes from 425 CD14+ monocyte cells. Being able to have access to a large amount of cells and genes created an opportunity to use deep learning to construct gene-to-gene comparisons. Deep learning is a growing and powerful space that continues to have far reaching applications. With the expansion of machine learning, computers can utilize data in order to identify a pattern among said data, allowing researchers to model scenarios and successfully predict future data. At first, this modeling consisted of basic patterns, such as regression, which were able to model many scenarios effectively. However, much data contains patterns that can't be modeled linearly, and require more flexible modeling (Petegrosso et al., 2). Through the use of neural networks, non-linear patterns can be created to model scenarios that couldn't be done feasibly at scale before, hence why deep learning is so capable. Deep learning's predictive aspects are currently aiding researchers in many ways in many fields (Kolodziejczyk et al., 613). Single-cell RNA-sequencing data contains massive potential for the field going forward, through its ability to efficiently give a view into single cells in a new way (Kolodziejczyk et al., 610). Combining deep learning's predictive abilities with single-cell RNA-seq data helps this data have far greater applications in the field of biology, as it can deal with high dimensional data and extract features in a way that was simply not possible before (Brendel et al. 815). Using deep learning to attempt to predict single-cell RNA sequencing data would pose many applications going forward. For instance, taking two different but similar functioning gene groups (for instance cell adhesion and cell junction), if it were possible to predict one side's genes given the other, researchers would be able to do so without extensive testing. This testing is time-consuming and expensive, and so any help that could be done by a model (very quick and

efficient) would be a large boon (Brendel et al., 829). For this paper, deep learning was used to fully utilize the single-cell RNA-sequencing data to try and create a model which could correlate single-cell RNA sequencing gene data sets.

Methods

To validate the data, the model was split into 80% training and 20% testing data. The cell adhesion genes were used in the model to predict the cell junction genes. The overall accuracy was found for each individual cell (Cell Accuracy), each individual gene (Individual Gene Accuracy), and for every instance where a gene existed in the test cell (Special Accuracy). The data collected was a singlecell RNA-sequencing dataset of CD14+ monocytes from a healthy human (InText Citation). This data consisted of 33694 genes (rows) by 425 cells (columns). When creating the model, the 20 cell adhesion genes were used as input, with the end goal of predicting the 20 cell junction genes, which were the output. The architecture for this model was a dense neural network. The layers for the dense neural network were 120, 120, Dropout(0.5), 90, 80, Dropout(0.5), 90, 80, 40, 20. The hardware used to create the model was Dell XPS 15 1920 with a Nvidia GeForce 3080, an 12th Gen Intel Core i9-12000HK and 32 GB of RAM. The software utilized to manage the environments was Anaconda, with Python and Jupyter Notebook used to code the model using Google Tensorflow.

Results

When predicting 20 genes for the test cells, the individual genes were right 93% of the time. However, that is where the difficulties arose. For many of the genes, a constraint placed upon them is the requirement to be doing the same/related cell function. Once these genes were selected together, a key issue was that most of the cells either only had 1 of the genes or none. Even when looking for genes that repeated over many cells, the constraint of related/same cell function prevented the use of cells that had more 1s (to indicate the presence of the gene) than 0s. This led to lower cell accuracy of the whole 12.9%. The special accuracy, which indicates the accuracy in which the test cell's genes that were 1 were correct gave 32.8% accuracy. This demonstrates how the model was only able to predict the special cases around 1/3 of the time, due to the sheer amount of genes that were just 0.

Individual Gene Accuracy: 0.9364705882352942
Cell Accuracy: 0.12941176470588237
Special Accuracy: 0.328125

Fig 1: Results of the accuracies listed in the result section

Limitations

A key limitation for the neural network was the large number of genes that weren't present in the cells. This large number of genes that were not present for any cells caused the

model to implicitly predict the gene not present, even when the gene was there. This might be fixed with better data primed with genes that were far more often to appear inside of the cells, hence helping the high false negative rate. Even though the model was able to get a 93% accuracy for the individual genes, the lower cell accuracy and special accuracy dampened the success of those results. As such, the future potential of this model specifically isn't very good, as it wouldn't be accurate enough to consider actual use in predicting genes with the Special Accuracy being a middling 32.8%. However, by overcoming this limitation, which would be likely given the encouragingly high special accuracy for such a data set, the special accuracy could reach the threshold needed to be used. By improving the selection of gene data in order to ensure that each gene is highly present in the cells, and choosing a data set to match that goal, the model's special accuracy would likely be highly beneficial and improve compared to the results from this model.

Conclusion

The purpose of this paper was to attempt to create a model which could correlate between single-cell RNA sequencing data sets with a high accuracy. The data used was a collection of 33694 genes from 425 CD14+ monocyte cells, using 20 cell adhesion genes for input and 20 cell junction genes for output. A dense neural network was used. In the end, the results were 93.64705882352942% accurate for individual genes, 12.9% accurate for each cell, and 32.8% accurate for each special gene. The model struggled to handle the data consisting mainly of a lack of genes when it was meant to predict the presence of genes, which made this model more unsuccessful. Going forward however, the model could demonstrate promise if it were to be given data geared towards having the genes be more present.

Works Cited

- Brendel M, Su C, Bai Z, Zhang H, Elemento O, Wang F. Application of deep learning on single-cell RNA sequencing data analysis: a review. *Genomics Proteomics Bioinformatics* 2022;20:814–35.
- Baker, Gladys L., Wayne D. Rasmussen, Vivian Wiser, and Jane M. Porter. *Century of Service: The First 100 Years of the United States Department of Agriculture*. [Federal Government], 1996. Print.
- Goudot, Christel et al. “Aryl Hydrocarbon Receptor Controls Monocyte Differentiation into Dendritic Cells versus Macrophages.” *Immunity* vol. 47,3 (2017): 582-596.e6. doi:10.1016/j.immuni.2017.08.016
- Raphael Petegrosso, Zhuliu Li, Rui Kuang, Machine learning and statistical methods for clustering single-cell RNA-sequencing data, *Briefings in Bioinformatics*, Volume 21, Issue 4, July 2020, Pages 1209–1223, <https://doi.org/10.1093/bib/bbz063>

Integrating Artificial Intelligence and Technology for Urban Environment Sustainability

By Aarav Mittal

Introduction

In the contemporary era, urban landscapes are rapidly evolving, driven by a confluence of technological advancements and burgeoning population densities. This transformation, while pivotal for economic and social development, often precipitates significant environmental challenges. Urbanization, a global phenomenon, not only reshapes the physicality of landscapes but also profoundly impacts the ecological balance within these environments. The intricate dance between urban growth and environmental sustainability has become a focal point for researchers and policymakers alike.

Central to this discourse is the role of artificial intelligence (AI) and technology in mitigating the environmental impacts of urbanization. AI, with its unparalleled capacity for data analysis and predictive modeling, offers innovative solutions for sustainable urban management. It presents a paradigm shift in how we approach environmental challenges in densely populated areas. Studies such as "Urban Evolution: How Species Adapt to Survive in Cities" shed light on the adaptive responses of urban ecosystems to human-induced changes. Similarly, research by Sarah Diamond delves into the nuanced interplay between urban development and ecological systems, highlighting the transformative potential of AI in this realm.

However, the integration of AI in urban environmental management is not without its complexities. While AI promises enhanced efficiency and novel solutions, it also brings forth challenges related to implementation, ethics, and equitable access. The potential of AI to contribute positively to urban sustainability is immense, yet its application must be navigated with caution and foresight.

This paper aims to explore the multifaceted role of AI and technology in enhancing urban environmental sustainability. It seeks to examine both the potential benefits and the inherent challenges involved in this integration. By analyzing current applications of AI in urban settings, particularly in conservation and sustainability efforts, and evaluating their effectiveness, this study endeavors to provide a comprehensive understanding of how technological innovation can be harnessed for more sustainable urban futures. The ultimate goal is to contribute to the ongoing dialogue on sustainable urban development, offering insights into how AI and technology can be leveraged to balance urban growth with environmental stewardship.

Literature Review

The integration of artificial intelligence (AI) and technology in urban environmental sustainability is a rapidly evolving field, necessitating a thorough understanding of both urban evolution and the potential of AI to address related challenges. This literature review examines key studies in this domain, focusing on how urban ecosystems adapt to human-induced changes and how AI can be leveraged to foster sustainable urban development.

"Urban Evolution: How Species Adapt to Survive in Cities" provides a foundational perspective on the adaptive responses of species within urban environments. The article emphasizes the significance of understanding these adaptations for improved urban planning and conservation management. It suggests that observing how creatures respond to urban life can enhance conservation efforts and contribute to the planning of cities with robust, functioning ecosystems. This insight is crucial for integrating AI into urban environmental sustainability, as it highlights the need for technology that can monitor and adapt to these evolving urban ecosystems.

Sarah Diamond's research on urban evolution further explores this theme, particularly focusing on how urbanization impacts ecological systems. Her work sheds light on the potential future of urban ecosystems in the Anthropocene, suggesting that urban evolution studies can inform strategies to create a more harmonious coexistence between humans and other forms of life in urban settings. Diamond's research underscores the importance of predictive and adaptive approaches in AI applications, aligning technological innovation with ecological needs.

Expanding on this, Peng et al. (2020) in their study on "Spatial–Temporal Evolution and Regional Difference Decomposition of Urban Environmental Governance Efficiency in China," delve into the efficiency of urban environmental governance. Their findings highlight the potential of AI in enhancing the effectiveness of environmental management in urban areas, particularly through spatial-temporal analysis. This study illustrates how AI can be utilized to assess and improve governance strategies, ensuring more sustainable urban environments.

Xiang et al. (2022) in "Spatiotemporal Evolution and Coupling Pattern Analysis of Urbanization and Ecological Environmental Quality of the Chinese Loess Plateau," analyze the relationship between urbanization and eco-environmental quality. Their research employs AI-driven tools like city nighttime light indexes and remote sensing ecological indexes, demonstrating how AI can aid in understanding the complex dynamics between urban development and environmental quality. This study is particularly relevant in highlighting how AI can be used to monitor and manage the ecological impacts of urbanization.

In "Study on the Spatial and Temporal Evolution Patterns of Green Innovation Efficiency and Driving Factors in Three Major Urban Agglomerations in China" by Hu et al. (2022), the focus shifts to green innovation. This paper explores the efficiency of ecological environmental protection and economic development in urban agglomerations, emphasizing the role of AI in promoting sustainable urban development. The study's findings on spatial and temporal patterns of green innovation provide valuable insights into how AI can support ecological and economic sustainability in urban settings.

Lastly, Nguyen et al. (2019) in their study on "The Environmental Effects of Urban Development in Hanoi, Vietnam," provide empirical evidence of the environmental impacts of urbanization. Utilizing remote sensing technologies, this research highlights the critical need for AI-driven solutions in urban environmental management, particularly in addressing issues like urban sprawl and heat islands.

In summary, these studies collectively illustrate the diverse applications and significant potential of AI in understanding, monitoring, and managing urban ecosystems. They underscore the necessity of integrating AI and technology in urban environmental sustainability efforts, offering innovative solutions to the complex challenges posed by rapid urbanization.

The Role of AI Urban Environmental Management

The advent of artificial intelligence (AI) has ushered in a new era in urban environmental management, offering innovative solutions to the complex challenges posed by rapidly evolving urban ecosystems. AI's role in this domain is multifaceted, encompassing the monitoring and management of urban ecosystems, as well as the conservation of urban wildlife and plant life. This section explores the various applications of AI in these areas, highlighting current implementations and their impact on urban environmental sustainability.

AI in Monitoring and Managing Urban Ecosystems

AI technologies have become instrumental in monitoring urban ecosystems, providing real-time data and predictive insights that are crucial for effective management. For instance, AI-powered sensors and drones are increasingly used for collecting environmental data, such as air and water quality indicators, noise levels, and biodiversity metrics. This data is then analyzed using advanced algorithms to identify patterns and trends, enabling city planners and environmentalists to make informed decisions. For example, AI can predict areas at higher risk of pollution or identify regions where biodiversity is declining, prompting proactive measures.

Machine learning models are particularly effective in processing large datasets from various sources, including satellite imagery and ground sensors, to monitor changes in land use, urban sprawl, and green cover. These models can forecast the environmental impact of urban development projects, helping in the planning of sustainable urban spaces that balance development needs with ecological preservation.

AI-Driven Solutions for Urban Wildlife and Plant Conservation

AI also plays a pivotal role in the conservation of urban wildlife and plants. By analyzing data from various sources, including wildlife tracking systems and environmental monitoring devices, AI can help identify patterns in wildlife movement and habitat use within urban areas. This information is vital for creating urban green spaces that are conducive to the needs of different species, thus promoting biodiversity within cities.

In plant conservation, AI-driven image recognition technologies are used to monitor the health of urban trees and plants. These systems can detect early signs of disease or stress in plants, enabling timely interventions to prevent further deterioration. Additionally, AI models can assist in identifying suitable locations for planting new trees, considering factors such as soil quality, sunlight exposure, and existing urban infrastructure, thereby enhancing urban greenery in a strategic manner.

Examples of Current AI Applications in Urban Environmental Sustainability

Several cities around the world are already harnessing the power of AI for environmental sustainability. For instance, Singapore's use of AI in its 'Smart Nation' initiative includes environmental monitoring through a network of sensors across the city, providing valuable data for urban planning and environmental protection. Similarly, projects like the 'Tree Canopy Project' in Los Angeles utilize AI to analyze aerial imagery and assess tree canopy coverage, guiding efforts to increase green spaces in urban areas.

In wildlife conservation, AI-enabled camera traps and acoustic sensors are being used in cities to monitor the presence and behavior of urban wildlife, aiding in the development of wildlife-friendly urban policies. These technologies not only provide insights into the current state of urban biodiversity but also help in tracking the success of conservation efforts over time.

Challenges & Opportunities in Implementing AI In Urban Environmental Management

The integration of Artificial Intelligence (AI) into urban environmental management presents a complex interplay of challenges and opportunities, each influencing the effectiveness and impact of AI applications in urban settings.

Technical and Infrastructural Challenges: Implementing AI in urban environments necessitates advanced infrastructure and specialized technical expertise. Alastal and Shaqfa (2022) in "GeoAI Technologies and Their Application Areas in Urban Planning and Development" highlight the intricacies involved in adopting GeoAI methods. The challenges are not just limited to the need for advanced data analysis capabilities but also encompass the establishment of robust infrastructure capable of supporting AI technologies. This includes high-speed connectivity, data storage and processing facilities, and the integration of AI with existing urban systems. Furthermore, the lack of standardization in AI technologies can lead to compatibility issues, making it difficult to implement AI solutions on a large scale. The need for continuous maintenance and updates of AI systems also adds to the complexity, requiring ongoing investment and resource allocation.

Security and Privacy Concerns: The integration of AI and robotics in urban settings introduces significant security and privacy challenges. As noted by Neupane et al. (2023) in "Security Considerations in AI-Robotics," these systems are susceptible to various security threats, ranging from data breaches to cyber-attacks. The implications of such vulnerabilities are profound, potentially compromising the safety and privacy of urban residents. The interconnected nature of AI systems means that a breach in one area can have cascading effects across the entire urban network. Additionally, the collection and analysis of large volumes of personal data by AI systems raise privacy concerns, necessitating stringent data protection measures and transparent data handling practices. The challenge lies in developing AI systems that are secure and resilient while respecting the privacy and rights of individuals.

Ethical and Societal Implications: The deployment of AI in urban management must contend with a range of ethical and societal issues. Lanne and Leikas (2021), in "Ethical AI in

the reablement of older people," discuss the multifaceted ethical challenges in AI utilization. These include issues of social trust, where the public's acceptance of AI solutions is contingent on their perceived reliability and fairness. Autonomy is another concern, as AI systems can potentially infringe on individual decision-making processes. Privacy concerns are paramount, especially in the context of surveillance and data collection in public spaces. Furthermore, AI systems can perpetuate existing biases, leading to unequal treatment and discrimination if not carefully designed and monitored. Addressing these ethical challenges requires a concerted effort to develop AI solutions that are transparent, accountable, and aligned with societal values.

Conversely, AI presents significant opportunities for enhancing urban environments:

Enhanced Urban Planning and Management: AI offers transformative solutions for urban challenges. Alastal and Shaqfa (2022) illustrate how GeoAI can revolutionize urban planning and development. AI's ability to analyze complex datasets can lead to more efficient resource allocation, better traffic management, and enhanced public services. AI can also play a pivotal role in environmental monitoring, providing real-time data on pollution levels, water usage, and energy consumption, thereby enabling more sustainable urban practices. Furthermore, AI can assist in disaster management and response by predicting and mitigating the impacts of natural calamities, thus safeguarding urban populations.

Improvement in Quality of Life and Safety: AI's impact on urban living conditions is substantial. As highlighted by Sawhney in "Rivers and Cooperative Urban AI Ecosystems," AI-driven infrastructures can significantly enhance urban mobility, sustainability, and safety. AI can optimize public transportation systems, reduce traffic congestion, and improve air quality, thereby enhancing the overall quality of life for urban residents. In terms of safety, AI can aid in crime prevention and public safety monitoring, ensuring a secure environment for city dwellers. The use of AI in urban healthcare systems can also lead to better health outcomes, with AI-enabled diagnostics and treatment plans tailored to individual needs.

Addressing Urban Crises: AI's role in managing urban crises is increasingly crucial. Sawhney's work underscores the importance of AI in addressing challenges such as climate change, pandemics, and social inequality. AI can provide critical insights into climate change impacts, aiding in the development of resilient urban infrastructures. During pandemics, AI can assist in tracking disease spread, managing healthcare resources, and developing response strategies. Additionally, AI can contribute to addressing social inequalities by identifying and addressing gaps in public services and resources.

In conclusion, the integration of AI in urban environmental management is a multifaceted endeavor, marked by both challenges and opportunities. While AI offers the potential to revolutionize urban living, it also necessitates careful consideration of technical, security, ethical, and societal implications. As urban areas continue to evolve, the strategic implementation of AI will be key in shaping sustainable and equitable urban futures.

Case Studies: Artificial Intelligence in Urban Environmental Sustainability

The effective use of Artificial Intelligence (AI) in urban environmental sustainability can be best understood through real-world case studies. These examples not only demonstrate the practical applications of AI but also offer insights into the success factors and limitations inherent in these technologies.

Case Study 1: Singapore's Smart Nation Initiative: Singapore's Smart Nation initiative stands as a prime example of AI's application in urban sustainability. The city-state has implemented an extensive network of sensors and cameras across the city to monitor various aspects of urban life, including traffic, crowd density, and environmental parameters like air and water quality. AI algorithms analyze this data to optimize traffic flow, reduce energy consumption, and improve public safety. A key success factor in this initiative is the government's commitment to integrating technology into every aspect of urban planning. However, the extensive surveillance network has raised privacy concerns, highlighting the need for careful consideration of ethical implications in AI deployment.

Case Study 2: Los Angeles' Tree Canopy Project: Los Angeles' Tree Canopy Project utilizes AI to address urban greenery challenges. By analyzing aerial imagery with AI algorithms, the project identifies areas with low tree coverage. This information guides urban planners in strategically increasing green spaces, thereby enhancing urban biodiversity and air quality. The project's success is attributed to its use of AI in providing precise, actionable data, enabling targeted interventions. However, the project faces limitations in terms of scalability and the need for continuous data updating to reflect current urban landscapes accurately.

Case Study 3: Amsterdam's AI-Powered Energy Grid: Amsterdam has implemented an AI-powered energy grid to optimize energy distribution and consumption. The system uses AI to predict energy demand patterns and adjust supply accordingly, thus reducing waste and promoting energy efficiency. The success of this project lies in its ability to integrate AI with existing infrastructure, demonstrating AI's potential in enhancing the efficiency of urban systems. However, the project faces challenges in ensuring the security of the energy grid against cyber threats, a critical aspect given the reliance on digital technologies.

Case Study 4: AI for Water Management in Bengaluru, India: In Bengaluru, India, AI has been employed to tackle water management issues. The system uses AI to monitor and predict water usage patterns, identify leaks, and optimize water distribution across the city. This approach has led to significant improvements in water conservation and reduced water wastage. The project's success is largely due to its innovative use of AI in addressing a critical urban issue. However, the project faces challenges in terms of data accuracy and the need for widespread sensor deployment to cover the entire city effectively.

These case studies demonstrate the diverse applications and potential of AI in enhancing urban environmental sustainability. Success factors include government support, precise data analysis, integration with existing infrastructure, and innovative problem-solving approaches. However, these initiatives also face limitations and challenges, such as privacy concerns,

scalability issues, continuous data updating needs, and security vulnerabilities. These examples provide valuable lessons for future AI implementations in urban sustainability, highlighting the importance of balancing technological innovation with ethical and practical considerations.

Conclusion

This research report has comprehensively explored the integration of Artificial Intelligence (AI) in urban environmental sustainability, revealing a landscape rich with potential yet fraught with challenges. AI's role in urban ecosystems, as evidenced through various studies and real-world applications, is transformative. It offers innovative solutions for monitoring and managing urban environments, enhancing wildlife and plant conservation, and improving overall urban planning and management. The case studies from Singapore, Los Angeles, Amsterdam, and Bengaluru exemplify AI's diverse applications, demonstrating its effectiveness in addressing specific urban sustainability challenges.

This research report has comprehensively explored the integration of Artificial Intelligence (AI) in urban environmental sustainability, revealing a landscape rich with potential yet fraught with challenges. AI's role in urban ecosystems, as evidenced through various studies and real-world applications, is transformative. It offers innovative solutions for monitoring and managing urban environments, enhancing wildlife and plant conservation, and improving overall urban planning and management. The case studies from Singapore, Los Angeles, Amsterdam, and Bengaluru exemplify AI's diverse applications, demonstrating its effectiveness in addressing specific urban sustainability challenges.

Looking ahead, the future of AI in urban environmental sustainability appears promising yet requires cautious optimism. As urban areas continue to grow and evolve, AI will undoubtedly play a pivotal role in shaping sustainable and equitable urban futures. The ongoing development of AI technologies, coupled with an increasing awareness of their potential and limitations, suggests a future where AI not only supports but significantly enhances urban environmental sustainability.

In conclusion, the integration of AI into urban environmental management represents a critical juncture in our approach to urban sustainability. While challenges remain, the opportunities AI presents are vast and varied. By continuing to develop and refine AI technologies, and by addressing the associated challenges head-on, we can harness AI's full potential to create more sustainable, efficient, and livable urban environments for future generations.

Works Cited

- Alastal, Abdelkhalek I., and Ashraf Hassan Shaqfa. "GeoAI Technologies and Their Application Areas in Urban Planning and Development: Concepts, Opportunities and Challenges in Smart City (Kuwait, Study Case)." **Journal of Data Analysis and Information Processing**, vol. 10, no. 2, 2022.
- Bender, Eric. "Urban Evolution: How Species Adapt to Survive in Cities." **Knowable Magazine* Annual Reviews*, 2022.
- Diamond, Sarah. "Urban Evolution Research." **Annual Review of Ecology, Evolution, and Systematics**, vol. 50, 2019, pp. 123-146.
- Hu, Biao, et al. "Study on the Spatial and Temporal Evolution Patterns of Green Innovation Efficiency and Driving Factors in Three Major Urban Agglomerations in China—Based on the Perspective of Economic Geography." **Sustainability**, vol. 14, no. 15, 2022, 9239.
- Lanne, M., and J. Leikas. "Ethical AI in the re-ablement of older people: Opportunities and challenges." **Gerontechnology**, vol. 20, no. 2, 2021.
- Neupane, Subash, et al. "Security Considerations in AI-Robotics: A Survey of Current Methods, Challenges, and Opportunities." **arXiv**, 2023.
- Nguyen, Thi Mai Chi, Lin T., and Hai-Po Chan. "The Environmental Effects of Urban Development in Hanoi, Vietnam from Satellite and Meteorological Observations from 1999–2016." **Sustainability**, vol. 11, no. 6, 2019, 1768.
- Peng, Geng, et al. "Spatial–Temporal Evolution and Regional Difference Decomposition of Urban Environmental Governance Efficiency in China." **Environment, Development and Sustainability**, vol. 23, no. 3, 2020, pp. 3884-3902.
- Sawhney, Nitin. "Rivers and Cooperative Urban AI Ecosystems." Date and publisher unknown.
- Xiang, Kaizheng, et al. "Spatiotemporal Evolution and Coupling Pattern Analysis of Urbanization and Ecological Environmental Quality of the Chinese Loess Plateau." **Sustainability**, vol. 14, no. 12, 2022, 7236.

Interdisciplinary Perspectives on Persuasion: A Comparative Analysis With Insights into Algorithmic Game Theory By Aarav Mittal

Abstract

Persuasion, as a construct, bears significance across a multitude of academic disciplines, each approaching it with unique definitions, methodologies, and objectives. This paper seeks to explore how perspectives on persuasion diverge and converge across five distinct fields: economics, game theory, psychology, marketing, and neuroscience. Utilizing a comparative framework, the study delineates the nuances and commonalities in each discipline's approach to persuasion. Further, the paper extends the insights to the burgeoning field of algorithmic game theory to understand how universal principles of persuasion could be effectively operationalized in computational models. Through comprehensive literature reviews, quantitative analysis, and interdisciplinary synthesis, the research identifies a set of core principles that resonate universally across the examined disciplines. These core principles hold potential not only for theoretical convergence but also for practical applications in algorithmic game theory, thereby opening new avenues for interdisciplinary research and strategy formulation in persuasive technologies. The findings of this study are instrumental for academics, practitioners, and technologists aiming to navigate the multifaceted landscape of persuasion.

Introduction

Persuasion has been the subject of intrigue and study across a myriad of disciplines, each offering its different lens and interpretations. From the economic models that explore market behaviors to the intricate neural pathways illuminated by neuroscience, the art and science of persuasion have been dissected, analyzed, and reconstructed in numerous ways. Yet, a fundamental question emerges from this vast sea of knowledge: How do definitions, methods, and objectives related to persuasion differ and overlap between fields such as economics, game theory, psychology, marketing, and neuroscience? Furthermore, amidst these diverse perspectives, can we discern a set of core principles or strategies universally acknowledged and employed across these disciplines? This research endeavors to unravel these questions, aiming to not only understand the multifaceted nature of persuasion across disciplines but also to explore how these insights might inform and enrich perspectives on persuasion within the realm of algorithmic game theory.

Literature Review

1: Economics - Definitions, Methods, and Objectives in Persuasion

Economics, at its core, is the study of how individuals, businesses, and governments make choices to allocate limited resources to meet their needs and desires. Within this vast domain, persuasion plays a pivotal role, influencing decisions and behaviors across various economic contexts. Stefano DellaVigna and Matthew Gentzkow, esteemed scholars in the field

of economics, have delved deep into the intricacies of persuasion within the economic landscape. Their studies provide a nuanced exploration of economic persuasion, emphasizing the interplay between market dynamics, consumer behavior, and the broader economic environment.

One of the foundational concepts highlighted is the idea of "market signals." These signals, ranging from stock market trends to interest rate shifts, serve as persuasive cues for investors. They act as indicators, guiding individuals and institutions in their investment decisions. For instance, a bullish stock market trend can persuade investors to buy, while a bearish trend might signal them to sell. The credibility of these signals is paramount; they are often anchored in empirical data, expert analysis, and real-time market movements. A credible economic forecast, bolstered by robust data, can significantly sway market sentiments, leading to shifts in investment patterns and consumption behaviors.

The methods employed in economic persuasion are multifaceted. While the analytical dimension is catered to through quantitative data, graphs, and statistical models, there's also a strong emphasis on emotive appeals. Narratives surrounding economic prosperity, downturns, or transformative policies tap into the emotional psyche of economic agents. DellaVigna and Gentzkow's studies shed light on the profound impact of fiscal policies on consumer sentiment. A government's decision to introduce tax cuts or increase public spending can be persuasively framed. Depending on the narrative's objective, it can either be portrayed as a catalyst for economic growth or a potential risk to fiscal stability. Such narratives, backed by data and presented persuasively, can influence public opinion, shape voting patterns, and even drive policy changes.

At its heart, the objective of persuasion in economics is to influence. Whether it's nudging consumers towards specific consumption choices, persuading businesses to adopt sustainable practices, or influencing policymakers to consider particular economic reforms, persuasion is omnipresent. DellaVigna and Gentzkow emphasize the importance of understanding the interplay of data, emotion, and narrative in crafting compelling economic messages. By leveraging credible information and compelling storytelling, economic agents can effectively steer decisions, mold behaviors, and guide societies towards desired economic outcomes.

In essence, the realm of economics, enriched by its intricate blend of data-driven analysis and emotive narratives, offers profound insights into the art and science of persuasion. By understanding and harnessing these dynamics, one can navigate the complex economic landscape, influencing decisions and shaping the future.

2 Game Theory - Definitions, Methods, and Objectives in Persuasion

Game theory, a discipline that studies strategic interactions, has profound implications for the art and science of persuasion. At its core, persuasion is about influencing others to adopt a particular viewpoint or take a specific action. Game theory provides a mathematical framework to understand these interactions, especially in scenarios where communication plays a pivotal role.

One of the most intriguing challenges in persuasion is communicating effectively when there's no shared language or common understanding. The research by Blume, DeJong, Kim, and Sprinkel offers invaluable insights into this domain (Crawford). They explored how individuals can convey persuasive messages even in the absence of a shared linguistic framework. Their findings, rooted in the "evolutionary" refinements proposed by Blume, Kim, and Sobel, suggest that persuasive signaling can emerge when players' interests or objectives align closely (Crawford).

This alignment of interests is a cornerstone of effective persuasion. When individuals or parties have shared goals or mutual benefits, the process of persuasion becomes more streamlined. Even without a common language, the underlying intent and message can resonate, leading to successful persuasion. Techniques like abstract message spaces and label-scrambling, as employed in the aforementioned studies, further emphasize that the essence of persuasion goes beyond mere words. It's about strategically aligning messages with the receiver's preferences and objectives.

Moreover, the dynamics of persuasion are deeply influenced by the context. In game theory, this is evident in the concept of "equilibria," where players' strategies stabilize based on mutual expectations. Persuasion, in many ways, mirrors this concept. For persuasion to be effective, both the persuader and the persuadee need to find a common ground, a mutual understanding that benefits both.

In scenarios with closely aligned goals, the power of persuasion is amplified. However, when objectives diverge, the challenge intensifies. It requires a deeper understanding of the other party's motivations, preferences, and constraints. Game theory suggests that in such situations, the persuader must adopt a more nuanced approach, leveraging indirect cues, and subtle signals to sway the other party.

In conclusion, game theory offers a rich tapestry of insights into the dynamics of persuasion. It underscores the importance of strategic alignment, mutual understanding, and the subtle art of signaling. Whether it's in business negotiations, political discourse, or interpersonal communication, the principles of game theory can provide valuable guidance for effective persuasion.

3: Psychology - Definitions, Methods, and Objectives in Persuasion

Social psychology, a specialized branch of psychology, is dedicated to understanding the intricate ways in which individuals' thoughts, feelings, and behaviors are influenced by others. Within this realm, persuasion emerges as a central concept. At its core, persuasion in social psychology can be defined as the process by which attitudes, beliefs, or behaviors are shaped or reinforced through communication (Smith, 2013).

The effectiveness of persuasion is not solely dependent on the content of the message but also on the source delivering it. The communicator, or the source, plays a pivotal role in this dynamic. Research in social psychology has identified credibility and attractiveness as two primary factors that influence the persuasiveness of a communicator. Credibility, encompassing

perceived expertise and trustworthiness, can have a profound immediate impact. However, its effects may wane over time, giving rise to phenomena like the " sleeper effect," where the message's influence persists even after the source's credibility is forgotten (Smith, 2013). On the other hand, attractiveness extends beyond mere physical appeal. It encapsulates likability, and studies have shown that messages, especially those with emotional undertones, resonate more when delivered by individuals perceived as attractive or likable.

Delving deeper into the methods of persuasion, the content of the message itself can be categorized into rational or emotional appeals. The effectiveness of these appeals often hinges on the nature of the audience. For instance, audiences that are well-educated or analytical tend to be more receptive to logical arguments. In contrast, those less involved or invested might be swayed more by their emotional reactions to the communicator or the message itself (Smith, 2013). Furthermore, the manner in which messages are associated with emotions, be it positive feelings or fear, can amplify their persuasive potential.

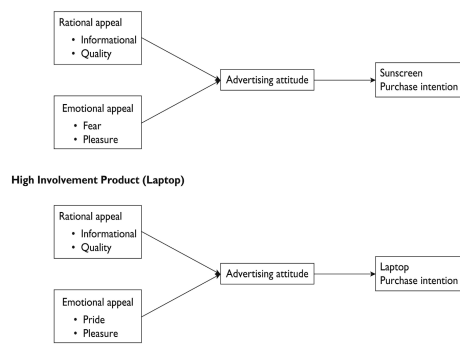


Figure 1. Conceptual Model for Low Involvement Product (Sunscreen)

The ultimate objective of persuasion is to influence, and this influence is deeply intertwined with the characteristics of the audience. Age, for instance, plays a significant role. Younger individuals, especially during their formative teenage and early adult years, are more malleable in their attitudes. These attitudes, once formed, tend to solidify as they transition into middle adulthood (Smith, 2013). Beyond demographics, the cognitive processes of the audience also matter. The way an individual internally responds to a message, be it through acceptance or

counterargument, can dictate the success or failure of a persuasive attempt.

In essence, social psychology provides a comprehensive lens to understand persuasion. It underscores the symbiotic relationship between the communicator, the message, and the audience. By grasping these dynamics, one can craft strategies that effectively leverage the principles of persuasion across diverse contexts.

4: Marketing - Definitions, Methods, and Objectives in Persuasion

Marketing, a multifaceted discipline, has always been about connecting with consumers in meaningful ways. One of the most potent techniques within this domain is persuasion marketing, which seamlessly blends human psychology with strategic communication to



influence consumer behavior. By definition, persuasion marketing is an approach that applies our understanding of human psychology to develop strategies that promote products or services. It focuses on tapping into the subconscious elements of the decision-making process, guiding consumers towards a desired outcome, often a

purchase (Marketing Schools, 2020). This technique is not just about the message but also about understanding and leveraging the underlying psychological triggers that influence consumer decisions.

The methods employed in persuasion marketing are diverse and have evolved with the digital age. In today's interconnected world, this extends to digital platforms, particularly website design. A well-designed website, in the context of persuasion marketing, doesn't just provide information. It strategically combines elements like layout, copy, typography, and promotional messages to guide visitors along pre-defined pathways, subtly influencing their interactions and decisions on the site (Marketing Schools, 2020). This strategic design is rooted in principles that understand and leverage subconscious decision-making processes. Niels Schillewaert (2021) encapsulates these principles in the acronym STEPPS: Social Currency, Triggers, Emotion, Public, Practical Value, and Stories. Each principle serves as a method to enhance the persuasive power of marketing content, from making consumers feel exclusive with Social Currency to crafting memorable narratives with Stories.

The primary objective of persuasion marketing is to influence consumer behavior in a way that aligns with a brand's goals, whether that's making a purchase, signing up for a newsletter, or simply building brand awareness. It's about more than just immediate sales; it's about building lasting relationships with consumers by understanding and catering to their subconscious needs and desires. For instance, the Blendtec campaign, which showcased the CEO blending various objects, wasn't just about demonstrating product durability. It was about crafting a narrative that resonated with viewers, making them more likely to remember the brand and its message (Schillewaert, 2021).

In conclusion, persuasion marketing stands as a testament to the power of understanding human behavior in the realm of marketing. By defining its principles, employing effective methods, and aiming for clear objectives, marketers can craft messages that not only resonate but also drive action. As the landscape of marketing continues to evolve, the principles of persuasion remain a timeless tool, bridging the gap between brands and consumers in meaningful, impactful ways.

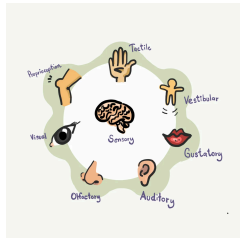
5: Neuroscience - Definitions, Methods, and Objectives in Persuasion

Neuroscience, in the context of persuasion, delves into the intricate neural mechanisms that influence our decisions, attitudes, and behaviors. A foundational principle in this domain is that our brain actively interprets information, rather than passively receiving it. When exposed to persuasive content, our brain processes, filters, and deciphers this information based on past experiences, biases, and our current emotional state (Cacioppo, 2018).

A pivotal principle in the neuroscience of persuasion is the dual-process theory. It suggests that our brain processes information through two distinct pathways: the central (or systematic) route and the peripheral (or heuristic) route. The central route involves deep, analytical thinking, activated when we are highly involved in the persuasive content. Conversely, the peripheral route is more automatic, relying on heuristics or shortcuts. This distinction is

evident in Cacioppo's study's emphasis on the importance of understanding the subconscious and conscious elements of decision-making, where as much as 90% of the decision-making process occurs beyond our conscious reasoning (Cacioppo, 2018).

Emotions play a significant role in persuasion. Neuroscientific research has shown that emotional appeals, such as visual and olfactory cues, can prime particular moods, making messages more impactful. Cacioppo's study underscores this by highlighting the use of visual imagery in website design to improve visitors' responses, emphasizing the power of emotions in shaping decisions (Cacioppo, 2018).



Neurotransmitters, like dopamine associated with pleasure and reward, play a crucial role in persuasion. When we encounter persuasive messages offering a reward or aligning with our values, dopamine release makes us more receptive to the message. Similarly, Cacioppo's study touches upon the concept of "social proof," where testimonials from others can influence our decisions, possibly through the release of oxytocin, the "trust hormone" (Cacioppo, 2018).

Mirror neurons, which fire both when we perform an action and observe someone else doing it, have profound implications for persuasion. This mirroring mechanism can be harnessed by demonstrating behaviors we want others to emulate. Cacioppo's study's mention of the "mirror technique," where mirroring a target's physical posture can create an unspoken bond, is a testament to the power of mirror neurons in persuasion (Cacioppo, 2018).

In conclusion, the neuroscience of persuasion, enriched by insights from Cacioppo's study, underscores the brain's active role in interpreting information, influenced by cognitive and emotional processes. By understanding dual-processing, the influence of emotions, neurotransmitters, and mirror neurons, we can craft persuasive messages that resonate deeply with the audience (Cacioppo, 2018).

6: Algorithmic Game Theory - Definitions, Methods, and Objectives in Persuasion

Algorithmic Game Theory (AGT) is a flourishing field that synergizes the mathematical constructs of game theory with the computational paradigms of computer science. Central to AGT is the exploration of strategic behavior of algorithmically-driven agents in diverse scenarios. Within this ambit, the study of persuasion emerges as a pivotal area, focusing on the strategic structuring and presentation of information to sway agent decisions.

Professors Dughmi and Xu (2017) delve into the intricate algorithmics of information structure design, often referred to as persuasion or signaling. In their work, "Algorithmic Persuasion with No Externalities", persuasion is conceptualized as a scenario wherein a principal or sender, endowed with exclusive knowledge about a particular state of the world, judiciously discloses information to agents or receivers. The crux lies in the sender's ability to meticulously design this information structure, determining the depth of information to unveil, its recipients, and the potential infusion of noise. This orchestrated communication, termed a signaling scheme, is essentially an algorithmic strategy employed by the sender to achieve her objectives. The

research underscores the computational challenges inherent in devising optimal signaling schemes, drawing a clear demarcation between public and private signal models. In scenarios permitting private signals, the findings are predominantly positive, elucidating efficient implementations for supermodular or anonymous objectives and offering approximations for submodular objectives (Dughmi & Xu, 2017).

Concurrently, another research paper titled "Persuasion in Bayesian Games" from arXiv offers an in-depth exploration of the Bayesian persuasion paradigm. Bayesian games, a specialized subset of AGT, encompass players operating with incomplete information, either about the game's parameters or the types of other players. In such intricate settings, a knowledgeable sender can transmit a signal to the receiver, modulating the latter's beliefs and subsequent actions. This study pioneers a novel methodology to the Bayesian persuasion conundrum, harnessing the duality theory of linear programming. This innovative approach not only facilitates the characterization of optimal signaling strategies but also proffers invaluable insights into the strategic maneuvers of the sender in Bayesian contexts.

In conclusion, the realm of Algorithmic Game Theory, fortified by the works of Dughmi and Xu (2017) and the revelations from "Persuasion in Bayesian Games," accentuates the profound interplay of algorithms and strategic behavior in the art of persuasion. By delving into the nuances of signaling schemes, Bayesian paradigms, and the computational intricacies of information design, we gain a deeper understanding of how to craft persuasive strategies that align seamlessly with the sender's objectives. The continuous research endeavors in this domain not only shed light on the multifaceted challenges of persuasion but also emphasize the indispensable role of algorithmic methodologies in shaping strategic communication.

7: Gaps in the Literature

In the academic landscape, individual disciplines have rigorously pursued their specialized inquiries into the multifaceted phenomenon of persuasion. Economics, with its intricate models and theories, primarily emphasizes the dynamics of market behaviors, consumer choices, and the role of incentives in shaping decision-making processes. This economic perspective, while profound in its own right, may not seamlessly align with the neuroscientific paradigm, which delves into the intricate neural mechanisms and cognitive processes activated by persuasive stimuli.

On a parallel note, psychology offers a rich tapestry of insights into the behavioral and cognitive aspects of persuasion, examining the underlying mental processes and external influences that drive individuals towards specific decisions or beliefs. Contrastingly, game theory, rooted in mathematical and computational foundations, approaches persuasion from a strategic vantage point, focusing on the interplay of agents within defined systems and the strategies they employ to achieve desired outcomes.

However, despite the depth and breadth of these individual explorations, the academic literature reveals a visible absence: a comprehensive, interdisciplinary synthesis that bridges these diverse perspectives. The challenge, and indeed the prevailing gap, is the endeavor to

integrate these varied disciplinary insights into a unified framework. Such a framework would not only elucidate the definitions, methods, and objectives related to persuasion across these disciplines but also provide invaluable insights into the potential core principles or strategies that might be universally acknowledged. Moreover, in the evolving domain of algorithmic game theory, there is a pressing need to understand how these interdisciplinary insights can be harmonized and applied, especially given the increasing relevance of strategic communication and persuasion in computational settings.

Comparative Analysis

1: Definitions of Persuasion - Comparison Across Disciplines

Persuasion, as a concept, has been defined and redefined through the lens of various academic disciplines, each bringing its unique perspective to the table. In economics, persuasion is often viewed in terms of influencing consumer behavior, emphasizing the role of incentives and market dynamics. This contrasts with the psychological definition, which delves into the cognitive and emotional processes that underlie the act of persuading or being persuaded. Neuroscience, on the other hand, offers a biological perspective, focusing on the neural pathways and brain regions activated during persuasive interactions. Game theory, with its strategic foundation, defines persuasion in terms of information exchange and strategic communication between players in a system. By juxtaposing these definitions, we can discern both overlaps and divergences, offering a multidimensional understanding of what persuasion entails across different academic terrains.

2: Methods of Persuasion - Comparison Across Disciplines

The methods employed to persuade, much like the definitions, vary considerably across disciplines. Economists might study persuasion through the lens of advertising campaigns, pricing strategies, or market signaling. Psychologists, meanwhile, delve into techniques like social proof, authority, and reciprocity, exploring how these methods tap into human behaviors and biases. Neuroscientists might employ advanced imaging techniques to observe brain activity during exposure to persuasive stimuli, aiming to understand the neural correlates of persuasion. In the realm of game theory, methods of persuasion revolve around strategic communication, signaling, and information asymmetry. By comparing these methods, we gain a holistic view of the diverse tools and techniques harnessed across disciplines to achieve persuasive outcomes.

3: Objectives of Persuasion - Comparison Across Disciplines

The end goals or objectives of persuasion also exhibit variability across academic fields. In economics, the primary objective might be to influence consumer purchasing decisions, maximizing profit or market share. Psychology often seeks to understand the underlying motives behind persuasion, such as the desire for social approval, conformity, or the fulfillment of certain needs. Neuroscience aims to map the brain's response to persuasive stimuli, seeking to identify the neural circuits that drive decision-making. Game theory, with its strategic focus, often centers

on achieving optimal outcomes in multi-agent systems through effective persuasion. By contrasting these objectives, we can identify common threads of intent that run through diverse disciplinary approaches, as well as the unique goals that each field prioritizes in its study of persuasion.

Core Principles and Strategies

1: Identification of Core Principles

Across the disciplines that have delved into the realm of persuasion, certain foundational principles consistently emerge, transcending the boundaries of individual fields. One such principle is the centrality of the audience or the target of persuasion. Whether it's an economic model that emphasizes market signals or a neuroscientific study that delves into the brain's responses to stimuli, the focus invariably remains on the recipient of the persuasive effort. Understanding the motivations, preferences, and biases of this audience becomes paramount.

Another universally acknowledged principle is the credibility of the source or the persuader. From a marketing campaign to a game theoretic strategy, the effectiveness of persuasion often hinges on the perceived authenticity and reliability of the source. This is evident in psychological studies that highlight the role of ethos in persuasive communication, as well as in economic models where credible signaling plays a pivotal role.

Furthermore, the principle of reciprocity, deeply rooted in human psychology, finds its relevance across disciplines. In marketing, for instance, the 'give and take' strategy, where a brand offers something of value to its consumers, expecting loyalty in return, is a direct application of this principle. Similarly, in game theory, the strategies often revolve around mutual cooperation, where reciprocity can lead to optimal outcomes for all players involved.

2: Universality Across Discipline

The universality of persuasion principles across disciplines underscores a fundamental human trait: our susceptibility to influence. Despite the varied methodologies and objectives, each discipline acknowledges the human brain's innate desire for coherence and consistency. This is why narratives, whether crafted by marketers or presented in economic models, resonate deeply. They cater to our inherent need to make sense of the world around us.

Another universal aspect is the emotional component of persuasion. Neuroscientific studies have consistently highlighted the brain's emotional centers' activation in response to persuasive stimuli. This emotional resonance is not limited to just neuroscience. In economics, the concept of 'utility' often incorporates not just tangible benefits but also emotional satisfaction. Similarly, in game theory, players' decisions aren't always driven by cold, calculated logic; emotions and perceived fairness often play a crucial role.

Lastly, the universality of feedback mechanisms in enhancing persuasive efforts is evident across disciplines. Whether it's the iterative strategies in game theory, A/B testing in marketing, or behavioral modifications based on neurofeedback, the importance of feedback in refining and optimizing persuasive strategies is universally acknowledged.

3: Discrepancies & Outliers

While the universality of certain principles of persuasion is undeniable, the approach and emphasis can vary significantly across disciplines. For instance, while neuroscience might focus on the brain's hardwired responses to stimuli, game theory often operates in a more abstract realm, emphasizing strategic interactions devoid of emotional complexities.

Another discrepancy lies in the temporal dimension of persuasive efforts. In marketing, the impact of a persuasive campaign might be immediate, gauged by metrics like sales or customer engagement. In contrast, game theoretic models, especially those in algorithmic game theory, might focus on long-term equilibria, where the effects of persuasive strategies manifest over extended interactions.

Furthermore, the very definition of 'success' in persuasive efforts can vary. In economics, it might be about achieving market equilibrium, while in psychology, it could be about changing a deep-seated belief or behavior. These discrepancies highlight the rich tapestry of persuasion, where each discipline offers a unique perspective, adding depth and nuance to our understanding.

Insights into algorithmic Game Theory

1: Relevance of Core Principles

Algorithmic Game Theory (AGT) offers a structured approach to understanding strategic interactions among rational agents. Central to AGT is the concept of equilibrium, a state where each agent's strategy is optimal given the strategies of others. This equilibrium concept is pivotal in the realm of persuasion. The Bayesian persuasion model, as proposed by Kamenica and Gentzkow, provides a framework for strategic information revelation. For instance, the Vickrey auction, a cornerstone of AGT, employs a second-price payment rule that persuades participants to bid their true valuations. Such mechanisms showcase how AGT principles can be harnessed for effective persuasion. Predicting these equilibrium states in AGT can guide the design of persuasive strategies. Furthermore, the relevance of AGT's core principles in persuasion is underscored by their applicability in real-world scenarios, emphasizing the importance of understanding and leveraging these principles.

2: Interdisciplinary Synthesis in AGT

AGT's interdisciplinary nature, bridging computer science and economics, provides a rich tapestry of techniques for persuasion. Concepts like selfish routing, where the behavior of individual agents affects the overall system, highlight the importance of understanding and influencing agent behavior. The Bayesian persuasion model further exemplifies this synthesis, emphasizing the strategic revelation of information to achieve desired outcomes. By integrating principles from both fields, persuasive strategies can be designed that are not only effective but also grounded in sound economic and computational principles. This synthesis ensures that persuasive techniques are robust, scalable, and adaptable to various scenarios, making them invaluable in diverse applications.

3: Applications and Limitations

Key applications of AGT, such as the Vickrey auction and the VCG Mechanism, inherently involve elements of persuasion. The Vickrey auction persuades bidders to bid their true valuations by ensuring the winner pays the second-highest bid. The VCG Mechanism, a dominant-strategy incentive-compatible mechanism, charges participants based on the externalities they impose, persuading them to make welfare-maximizing decisions. However, real-world applications like spectrum auctions have shown that small design changes can have significant financial implications. While AGT offers robust solutions for persuasion, it's essential to recognize its limitations. The potential consequences of its application, especially in high-stakes scenarios, emphasize the need for careful implementation and continuous refinement.

4: Ethical and Strategic Implications

Strategically, AGT provides tools like the Price of Anarchy to measure the inefficiency of equilibria in games. By understanding this inefficiency, persuasive strategies can be designed to guide agents towards more efficient outcomes. However, the ethical dimensions in AGT-based persuasion are paramount. For instance, the VCG Mechanism, while promoting welfare-maximization, must be implemented with fairness in mind. The strategic and ethical implications of AGT in persuasion underscore the need for a balanced approach that prioritizes both efficiency and equity. As AGT continues to evolve, there's an opportunity to explore the ethical implications of persuasive strategies, ensuring they align with societal values.

5: Suggestions for Future Research

The dynamics of multi-agent systems and their convergence properties present rich avenues for future research at the intersection of AGT and persuasion. Mechanism design, a subfield of AGT, offers a plethora of techniques that can be harnessed for developing novel persuasive strategies. The Bayesian persuasion model suggests that there are rich avenues for future research, especially when considering combinatorial actions and constraints. Research can also delve into the potential of simple and prior-independent auctions, as suggested by the Bulow-Klemperer Theorem, to achieve near-optimal outcomes. As the field of AGT grows, there's a pressing need to explore its broader implications, especially in the realm of persuasion, to ensure that strategies are both effective and ethically sound.

Conclusion

The multifaceted nature of persuasion has been explored and dissected across a myriad of academic disciplines, each offering its unique lens and interpretation. This research embarked on a journey to understand how perspectives on persuasion diverge and converge across five distinct fields: economics, game theory, psychology, marketing, and neuroscience. Through a comparative framework, this study has successfully delineated the nuances and commonalities in each discipline's approach to persuasion.

Central to the findings is the recognition of certain foundational principles that consistently emerge across disciplines. The centrality of the audience or the target of persuasion, for instance, remains a consistent theme. Whether it's an economic model emphasizing market signals or a neuroscientific study focusing on the brain's responses to stimuli, the recipient of the persuasive effort remains paramount. Another universally acknowledged principle is the credibility of the source or the persuader, emphasizing the importance of authenticity and reliability.

The insights derived from these disciplines were further extended to the burgeoning field of Algorithmic Game Theory (AGT). AGT, with its structured approach to understanding strategic interactions among rational agents, offers a rich tapestry of techniques for persuasion. The study showcased how universal principles of persuasion could be effectively operationalized in computational models, particularly in AGT. The relevance of AGT's core principles in persuasion, underscored by their applicability in real-world scenarios, emphasizes the importance of understanding and leveraging these principles.

Furthermore, the interdisciplinary nature of AGT, bridging computer science and economics, provides a unique perspective on persuasion. By integrating principles from both fields, persuasive strategies can be designed that are not only effective but also grounded in sound economic and computational principles.

As we venture further into the future, the importance of combining and contrasting the diverse perspectives on persuasion becomes even more pronounced. By juxtaposing the insights from economics, game theory, psychology, marketing, and neuroscience, we can cultivate a more holistic and enriched understanding of persuasion's multifaceted nature. One potential theory is the development of a unified persuasion model that integrates economic rationality with psychological triggers, optimized through game-theoretical strategies, and enhanced by neuroscientific insights into human decision-making. Such a model could revolutionize marketing campaigns, making them more effective and tailored to individual responses. Another theory could explore the interplay between market signals and cognitive biases, leading to more informed and ethical advertising strategies. By harnessing the strengths of each discipline, we can also anticipate the emergence of hybrid persuasion techniques that leverage both the emotional and rational aspects of decision-making.

Looking ahead, as the field of Algorithmic Game Theory (AGT) expands, the dynamics of multi-agent systems and their convergence properties unveil rich avenues for exploration, especially at the crossroads of AGT and persuasion. The multifaceted landscape of persuasion, illuminated through the diverse lenses of various disciplines and accentuated by AGT, provides profound insights and knowledge. There's an imperative to delve deeper into AGT's broader implications in the realm of persuasion, ensuring that strategies are not only effective but also ethically grounded. This paper stands as a testament to the potential of interdisciplinary collaboration, setting the stage for future studies that aim to enhance our understanding of persuasion and its pivotal role in the contemporary world.

Works Cited

- Cacioppo, J. T., Cacioppo, S., & Petty, R. E. (2018). *The Neuroscience of Persuasion: A Review with Emphasis on Issues and Opportunities*. Routledge: Taylor & Friends Group.
<https://www.tandfonline.com/doi/full/10.1080/17470919.2016.1273851>
- Crawford, V. P. (1998, February). *A Survey of Experiments on Communication via Cheap Talk*, - *ResearchGate*. researchgate.net.
https://www.researchgate.net/publication/4977393_A_Survey_of_Experiments_on_Communication_via_Cheap_Talk
- DellaVigna, S., & Gentzkow, M. (2010, May 7). *Persuasion: Empirical Evidence*. eml.berkeley.edu.
<https://eml.berkeley.edu/~sdellavi/wp/PersuasionAERDellaVignaGentzkowProofsJun10.pdf>
- Dughmi, S., & Feng, H. (2017, June 26). *Algorithmic Persuasion with No Externalities*. dl.acm.org. <https://dl.acm.org/doi/pdf/10.1145/3033274.3085152>
- Fujii, K., & Sakaue, S. (2021, December 12). *Algorithmic Bayesian Persuasion with Combinatorial Actions*. arXiv.org. <https://arxiv.org/abs/2112.06282>
- Hoefler, M., Manurangsi, P., & Psomas, A. (2020, August 31). *Algorithmic Persuasion with Evidence*. arXiv.org. <https://arxiv.org/abs/2008.12626>
- Li, Y. X., & Schipper, B. C. (2020, February 8). *Strategic Reasoning in Persuasion Games: An Experiment*. Burkhard C. Schipper - University of California, Davis.
<https://faculty.econ.ucdavis.edu/faculty/schipper/>
- Nisan, N., Roughgarden, T., Tardos, E., & Vazirani, V. V. (2007). *Algorithmic Game Theory - CMU School of Computer Science*. cs.cmu.edu.
<https://www.cs.cmu.edu/~sandholm/cs15-892F13/algorithmic-game-theory.pdf>
- Pallini, K. (2022, October 25). *6 Principles of Persuasive Marketing: How to Influence People*. InSites Consulting.
<https://www.insites-consulting.com/blog/6-principles-of-persuasive-marketing-how-to-influence-people/>
- Roughgarden, T. (2016). *Twenty Lectures on Algorithmic Game Theory*. Cambridge University Press.
- Smith, C. (2013, December 17). *Social Psychology*. Medium.
[https://medium.com/@carolinekyungae/social-psychology-10bb0f48e649#:~:text=The%20%20ingredients%20%20of%20%20persuasion%3A%201,is%20communicated%204\)%20the%20audience.](https://medium.com/@carolinekyungae/social-psychology-10bb0f48e649#:~:text=The%20%20ingredients%20%20of%20%20persuasion%3A%201,is%20communicated%204)%20the%20audience.)
- Unknown, U. (2022, November 25). *Persuasion Marketing*. Marketing Schools.
<https://www.marketing-schools.org/types-of-marketing/persuasion-marketing/#section-3>

Effect of Social Media on Neuroplasticity in Adolescents By Vanshi Garodia

Abstract

This literature review investigates the impact of social media and smartphone use on the adolescent brain, summarizing current research in a straightforward manner. It explores how these technologies affect brains in the short term and considers possible long-term consequences. The review reveals that social media and smartphone usage can influence brain reward systems, possibly leading to addictive behaviors and impulsive online actions. Adolescents, whose brains are still developing, seem particularly vulnerable, experiencing heightened emotional responses, social anxiety, and potential changes in their ability to make decisions. Some studies suggest structural alterations in specific brain regions, raising questions about the potential cognitive effects. However, it is important to note that while some research points to negative mental health outcomes, there are also positive aspects like identity exploration, social connection, and digital skills development associated with technology use. This review emphasizes the need for future research to look into long-term effects, consider different social media platforms, and develop tailored interventions. By simplifying the complex relationship between technology and the brain, this paper aims to inform individuals, parents, and policymakers about the intricate connection between modern technology and changes in the adolescent brain.

Introduction

Social media is changing our brains. Are we ready for the consequences?

“Social media is about sociology and psychology more than technology”, stated Brian Solis, a globally recognized independent digital analyst and anthropologist (Allton). Social media has become an integral part of our lives, and has formed the basis around which our societies revolve, from the ways users view themselves, others, their thoughts, actions and lifestyle, especially more so for Generation Z and millennials. Social media has a number of positive and negative impacts on society. On the one hand, it can help people to connect with others and form new relationships, which can lead to increased social well-being. On the other hand, it can result in scrolling addiction which can be more addictive than drugs, gambling or alcohol (Woolley and Sharif 19). Alarmingly, research has shown that Facebook, Snapchat, and Instagram leverage the very same neural circuitry used by slot machines and cocaine to keep us using their products as much as possible (Alter).

Most social media users do not realize the health consequences that they face while social media. The majority of social media users do not consider the short and long-term effects it has on their emotions, cognitive function and productivity. However, users often feel that they have failed to absorb any content and are left feeling like they have wasted their time over social media (Baughan et al. 2).

Social media has a vast positive potential. In the domain of health, social media can be used by healthcare professionals to improve health outcomes, build a professional network, spread news and findings, encourage patients, and provide health information to the public

(Gupta 294). In the field of education, social media can help students to share information and learn from diverse sources; and also to stay up-to-date on trends and prepare for their future (Gupta 295). In one's personal life, social media can help people to connect with others and form new relationships, which can lead to increased social well-being (Wasim and Kumar 347). In the area of mental health, social media use has a positive indirect impact on psychological well-being, mainly due to the positive effect of bonding and bridging social capital (Ostic et.al. 9).

While social media can be used to connect with others and learn new things, it can also have a negative impact on our brains if users overuse it. The rapid growth of social media has led to concerns about the potential negative effects that it can have on our brains. It can capture and scatter our attention, making it difficult to focus on anything for long periods of time. Heavy social media use can lead to addiction, as it provides immediate rewards in the form of dopamine releases. Social media can also lead to memory deficits, as it makes us rely on external sources for information instead of our own memories. The constant use of social media can shrink parts of the brain associated with attention and cognitive function (Fotuhi). Thus, it is important to be mindful of our social media use and to take breaks from it regularly in order to protect our brains.

The use of social media is extremely widespread today. As of 2023, there are over 4.48 billion people using social media worldwide, representing over 56% of the world's population. The most popular social media platforms are Facebook, Instagram, YouTube, and WhatsApp. Social media has evolved rapidly over the past few decades with a shift towards video-based platforms like YouTube and TikTok (Dean).

As social media use has become increasingly prevalent, researchers have begun to study the impact of this technology on the adolescent brain. Adolescence (13-20 years of age) is a time of rapid brain development, and social media use can have a significant impact on this development. Studies have shown that social media use can activate the reward circuit of the brain, which is involved in processing rewards such as food and drugs leading to increased dopamine levels, which can make social media use addictive. Social media use can also affect the development of brain regions involved in social cognition, such as the medial prefrontal cortex. This region is involved in self-regulation and processes information about oneself and others. Increased social media use has been linked to decreased activity in the prefrontal cortex, which can lead to problems with self-esteem and social skills (Åström). Overall, the research on the effects of social media use on the adolescent brain is still in its early stages. However, the available evidence strongly suggests that social media use can have a drastic negative impact on adolescent brain development.

This paper will focus on the cognitive and psychological changes that can occur in adolescents who use social media excessively. This will be done by examining the effects of social media on attention spans, knowledge, transactive memory, social cognition, delayed gratification, reward, memory, multimedia tasking, neural connectivity, neuroplasticity and pathways in the brain. Additionally we will discuss the relationship between social media use

and disorders such as ADHD, depression, anxiety, and addiction. Finally, we will explore some therapies and treatments that can help to mitigate the negative effects of social media on the developing adolescent mind.

Cognitive/Psychological Changes in Adolescent Brain

1. Measuring cognitive function

Social media and smartphone dependency dramatically affect cognitive functions during adolescent years. Cognitive function includes the mental processes involved in the acquisition of knowledge, manipulation of information, and reasoning. It includes the domains of perception, memory, learning, attention, decision making, and language abilities. Cognitive functions rapidly develop during adolescent years while increasing the ability to reason and think abstractly. Between the age of 12-18, adolescents start thinking about their future, become more self-conscious and fantasize ways to deal with hypothetical situations.

Measuring cognitive function in adolescents who use social media is a complex task as there are many different factors that can affect cognitive performance. The increasing use of social media among adolescents has raised concerns about the potential negative effects of its use on cognitive function.

2. Attention spans

Attention spans are a key component of higher cortical function and are often decreased due to excessive social media usage. The ability to organize, plan, analyze, and compare past and present actions and to control impulses are described as executive functions. A study found that multitasking with social media and school work leads to difficulty in paying attention and retaining information (Kokoç 493). Another study found that attention control moderated the relationship between social media use and psychological distress in adolescents. This means that adolescents with lower levels of attention control were more likely to experience anxiety and other negative emotions if they used social media heavily (Mahalingham, Howell and Clarke 536)

Our cognitive resources are limited, and individuals are already permanently surrounded by distractions through the constant availability of a smartphone. Research found when an individual is near a smartphone it activates the involuntary attention system, searching for a source of dopamine, which actively listens to your smartphone even when it's not ringing. Without consciously realizing, users might pick up their phone and spend hours scrolling through Instagram, and feel tired and exhausted even when users did not completely exercise their brains (Eriş et al. 433).

Research also shows that smartphone presence influences working memory capacity and fluid intelligence, which leads to lower attention on complex tasks. One study found that the effect of smartphone presence on attentional functioning is only present in high-level tasks that require meta-cognitive decisions and not in low-level attentional tasks. The allocation of attentional resources are divided between the central task and the additional cognitive load seen

through smartphones (Skowronek et.al.) Another found that our brain circuits involved in ‘wanting’ are much more powerful than the brain circuits involved in ‘satisfaction’. Social media capitalizes on this satisfaction, providing endless possibilities for seeking attention in the form of features like infinite scroll, recommendations, and swiping to new content (Richard et.al.).

3. Knowledge capacity

There is a growing body of research on the impact of social media use on adolescents' knowledge capacity. Some studies have found that social media use can be beneficial for knowledge acquisition, as it can provide adolescents with access to a wide range of information and resources. One study found that adolescents who used social media were likely to be knowledgeable about current events and pop culture. However, near-constant access to information might disrupt memory abilities or the utilization of effortful thinking and have difficulty in recalling information. It is important to note that the effects of social media use on knowledge capacity varies depending on a number of factors, such as the type of social media platform used, the amount of time spent on social media, and the individual's cognitive abilities (Sivakumar, Sudarsan and Shaik 5) More research is needed to better understand the complex relationship between social media use and knowledge capacity in adolescents.

4. Social cognition

Social media has had a profound impact on the way adolescents communicate and interact with each other. Adolescents who use social media often are more likely to be hypersensitive to social feedback from their peers. This may be due to the constant stream of social rewards and punishments that social media platforms provide. A study found that young adults who use their mobile devices more often prefer to interact via social media rather than in person. This is because they feel more comfortable and in control when communicating through social media (Jimenez and Morreale). This shows how social media has reduced the amount of face-to-face interaction between people. A recent study also found that adolescents who check social media more often may change their brains to become more sensitive to social rewards and punishments. This could have long-standing effects, as well as leading to adolescents going on to compulsive social media use (Waugh). The findings of these studies suggest that social media use can have a significant impact on the way adolescents communicate and interact with each other.

Social media use can impact the way adolescents communicate and interact with each other, as well as on their cognitive biases and emotional responses. Cognitive bias is the tendency of the human brain to filter out information and simplify it based on past experiences and preferences, it is a form of systemic thinking and error. A recent article suggests that social media can exploit adolescents' cognitive biases to keep them engaged and clicking. For example, social media notifications trigger their salience network, which helps them determine what is important. This can lead them to check their social media accounts more often than they need to. Additionally, social media algorithms are designed to show them content that they are likely to engage with, which can reinforce their existing beliefs and opinions. This can lead to

confirmation bias, where users only see information that confirms what they already believe. Social media can also make adolescents more susceptible to negative emotions, such as fear and anxiety. This is because social media is often filled with negative content, such as news about violence or disasters. When users see this kind of content, their brains are wired to pay more attention to it, which can make them feel more stressed and anxious ("Social Media and the Brain").

5. Effects of delayed gratification and reward

While the ability to delay gratification is a key cognitive skill essential for success in many areas of life, there is evidence that social media may decrease capacity to resist instant gratification. Recently, researchers examined the disruptive effects of constant positive and negative stimuli from social media on the ability to delay gratification and pursue long-term goals. It emphasized how excessive social media use led individuals to prioritize short-term, immediate goals over meaningful achievements. The author noted the challenges faced by younger generations in practicing delayed gratification due to their exposure to social media. The article concluded by advocating for moderation and awareness in social media consumption to reinstate delayed gratification (Karthic and Sudhakaran 1792).

6. Effects on Memory

Social media use has been linked to a number of negative effects on memory in adolescents, including decreased working memory, impaired long-term memory consolidation, and increased forgetfulness. Transactive memory, the collective knowledge shared within a group, was examined in several studies. One study explored how the internet might affect memory and attention, suggesting that the abundance of online information could lead to attention shifts and multitasking, potentially challenging sustained attention and individual memory consolidation (Firth et.al. 121). A subsequent study extended this, noting that the online social world blended with offline cognitive processes, possibly altering traditional transactive memory dynamics. The integration of digital and real-life experiences could disrupt conventional patterns of knowledge sharing. This indicates how the extensive use of online platforms might influence how transactive memory operates (Firth, Torous and Firth 2)

There is a growing body of research that suggests that adolescent social media use can have a negative impact on memory retention. In a study, fifty teenagers were randomly assigned to either a control group or an experimental group. The control group was asked to sit in silence for five minutes, while the experimental group was asked to use social media for five minutes. After five minutes, both groups were given a memory test. The results showed that the teenagers who had used social media for five minutes recalled significantly fewer words than the teenagers who had sat in silence. This study suggests that social media use can interfere with short-term memory. This study found that daily social media use is linked to weakened everyday memory functioning in adults (Sharifian and Laura 71).

7. Effect of multimedia tasking on high frequency effects

Multimedia multitasking, which is the practice of doing multiple tasks at the same time, can have a negative impact on cognitive function, especially in adolescents.

Adolescents who frequently engage in media multitasking may experience a number of immediate and chronic effects. In the short-term, media multitasking can lead to poorer performance on cognitive tasks, such as working memory and attention (Cain et.al. 1938). This is because media multitasking requires the brain to switch between multiple tasks, which can overload cognitive resources and lead to errors. In the long-term, chronic media multitasking has been linked to poorer academic achievement, increased attention problems and decreased well-being (Van der Schuur et. al. 149). Media multitasking can disrupt the development of executive functions, and interfere with attention span and working memory, therefore negatively affecting development.

Neural connectivity and pathways in the brain

The constant use of smartphones and social media can have a significant impact on the neural connectivity and pathways in the adolescent brain, as reflected in the following sections.

1. Structural changes in the brain (Memory and Processing changes)

The human brain is not completely developed until the age of 25, making it particularly vulnerable to the effects of social media in adolescence.

Adolescents who habitually check social media have lower neural sensitivity to social anticipation in the amygdala, a region of the brain that is involved in processing emotions such as fear and anxiety (Maza, et al. E6). This suggests that habitual social media checking may lead to a hypersensitive amygdala, which is important for processing social feedback, such as receiving likes or comments on social media posts. When the amygdala is hypersensitive, it can lead to increased anxiety, stress, and a decrease in self-esteem. This is because the amygdala is constantly scanning for threats, even when there are none. Adolescents who use social media more frequently are likely to experience anxiety and depression in a correlative manner rather than causative.

Adolescents who frequently check social media also show an increasing sensitivity to peer feedback and greater activation in the prefrontal cortex, a region of the brain that is involved in decision-making and self-regulation. A study showed that when participants received negative peer feedback, they showed increased activity in the ventrolateral prefrontal cortex (VLPFC) and anterior insula, as well as the medial prefrontal cortex (MPFC). These regions are involved in emotion regulation, pain processing, and salience detection. The increased activity in these regions suggests that negative peer feedback can lead to a negative emotional state, as well as the need for emotion regulation. In contrast, when participants received positive peer feedback, they showed increased activity in the superior frontal gyrus (SFG), posterior cingulate cortex (PCC), precuneus, and temporoparietal junction (TPJ). These regions are involved in impulse control,

self-referential processing, theorizing about other minds, and social cognition. Positive peer feedback leads to increased self-awareness and understanding of others (Wikman et.al. 10).

One study found that the nucleus accumbens (NAcc) and the prefrontal cortex (PFC) were active when participants viewed photos that had been endorsed by their peers. The NAcc is a region of the brain that is involved in reward processing and addiction. The PFC is involved in social cognition, which is the ability to understand and respond to the thoughts and feelings of others. The PFC was less active when participants viewed photos of risky behaviors. This suggests that social media can lead to decreased cognitive control, which can make it more difficult to resist peer pressure and make healthy decisions. These findings suggest that social media can have a significant impact on the brain, and that it is important to be aware of the potential risks of social media use. The anterior cingulate cortex (ACC) was not activated in response to negative peer feedback. This is surprising, as the ACC is typically involved in social and emotional evaluation and affective distress. This is because the negative feedback was not as unexpected or aversive (Wikman et.al.)

Another study found that the nucleus accumbens (NAcc) is activated when people process gains in reputation for themselves in comparison to those of others. The NAcc was not activated when people processed gains in monetary reward. Suggesting that the NAcc processes rewards that are relevant to the individual's self-concept. The NAcc may play a role in how people use social media to manage their reputation and the ways it can alter the brain's reward system. These changes could make people more sensitive to social feedback and more likely to use social media to manage their reputation (Meshi, Carmen and Heekeren 9).

The brain's dynamic response to positive and negative peer feedback suggests that it can adapt its neural pathways based on input and conditioning. Heightened activity through positive feedback indicates that the brain is engaging in self-reflection, empathy, and understanding others. Over time, these areas may become more adept at processing social cues and regulating emotions, enhancing an individual's social cognition skills. As these regions become more active, the brain may strengthen its ability to cope with negative emotions and regulate distress. This suggests that repeated exposure to negative feedback could lead to improved emotional resilience and coping mechanisms over time. The brain might rewire its responses based on the familiarity and intensity of the feedback, demonstrating how plasticity allows for context-specific adjustments.

A longitudinal study revealed that the associations between structural brain development, social media use, and mental well-being in adolescents follow distinct pathways of brain development. Adolescents are characterized by significant neuromaturation, including cortical thinning particularly in frontal regions. Adolescents who used social media more than their peers had higher baseline cortical thickness, or “width of gray matter”, in the lateral prefrontal cortex (PFC) and medial PFC. Higher baseline cortical thickness is associated positively with increased intelligence and higher performance on delayed visual memory tasks. However, they also experienced more pronounced decreases in the lateral PFC and temporal parietal junction over

time. In contrast, adolescents with lower mental well-being had lower baseline surface area in the medial PFC and posterior superior temporal sulcus (Achterberg et al. 10).

With regards to cortical remapping, the study mentioned earlier found that adolescents who engage in habitual social media checking behaviors show a distinct neurodevelopmental trajectory, ‘maturation of behavior’, within regions of the brain comprising the affective salience, motivational, and cognitive control networks in response to anticipating social rewards and punishments. Adolescents with habitual social media checking behaviors showed lower neural sensitivity to social anticipation at age 12 years in the left amygdala, posterior insula (PI), and ventral striatum (VS). However, among those with habitual checking behaviors, there were longitudinal (sustained) increases in the amygdala during social anticipation, whereas among those with non habitual checking behaviors, longitudinal decreases were seen in the amygdala (Maza, et al. E6). These findings suggest that habitual social media checking in early adolescence may be steadily associated with changes in neural sensitivity to anticipation of social rewards and punishments.

2. Neuroplasticity

Neuroplasticity is the ability of the brain to change and adapt over time. This means that the brain can be shaped by experiences, including the use of social media.

The digitized world is altering our minds, while cultivating certain cognitive capabilities at the expense of other important ones due to our heavy dependence on social media. A study explored the “use-it-or-lose-it” phenomenon of neuroplasticity and depicts how the process of learning and memory stimulates new neural connections and stabilizes synaptic connections in areas of the brain frequently used. While it permanently eliminates connections through the loss of a brain cell function when dormant (Rechnitz and Derdikman).

The parts of the brain that control emotions and social interactions are still developing, making adolescents more sensitive to emotional media. This can lead to heightened reactivity to social media platforms, where adolescents may experience acceptance or rejection. A study found that adolescents are more responsive to emotion-arousing media than adults. This is because their brains are still developing and they are not as good at controlling their emotions. Myelination takes place through the rapid transition of neurons and synaptic pruning through the removal of excess synapse in the prefrontal cortex, improving the efficiency of information processing (Crone and Konijn 6). Myelination strengthens neural connections between the prefrontal cortex and other regions of the brain. The limbic system also develops around this time, playing a crucial role in determining rewards and punishments and processing emotional experience and social information. Although the limbic system is not mature until adulthood, social media plays a role in detecting these constant inputs of awards making us impatient and anxious until users have that next wave of dopamine rush (Rechnitz and Derdikman).

Yet another study found that young adults with smartphone dependence (SPD) had significantly lower white matter integrity than controls in the superior longitudinal fasciculus (SLF), superior corona radiata (SCR), internal capsule, external capsule, sagittal stratum,

fornix/stria terminalis and midbrain structures. White matter integrity is a measure of the quality and connectivity of the network of nerve fibers that connect different parts of the brain. The findings of this study suggest that SPD may be associated with structural changes in the brain that could lead to impaired cognitive and behavioral functioning. Smartphone use can lead to chronic stress and inflammation, which can damage the brain's white matter. Additionally, smartphone use can disrupt sleep, which is also known to have a negative impact on white matter integrity (Yuanming et.al. 533).

One more study discussed that regular engagement with social media, which involves activating the dorsolateral prefrontal cortex (DLPFC), medial prefrontal cortex (MPFC), and ventral striatum (VS), could lead to the strengthening of neural pathways associated with these brain networks. Just as practicing a skill can enhance the corresponding neural connections, the repeated use of social media could reinforce the pathways related to understanding others' thoughts and motivations (DLPFC), self-reflection (MPFC), and processing social rewards. Positive feedback and social rewards, such as likes and comments, trigger the ventral striatum (VS). Over time, the brain might become more attuned to seeking out and valuing these online social rewards. This could potentially impact real-world behavior, as individuals might seek validation and positive feedback more actively, both online and offline. This selective reinforcement might shape individuals' behaviors and attitudes in their social interactions. The engagement of the medial prefrontal cortex (MPFC) when posting about oneself suggests that social media use could influence how individuals perceive themselves. Over time, the way people construct their online personas might have a reciprocal effect on their self-concept and offline interactions. Similarly, the activation of the dorsolateral prefrontal cortex (DLPFC) when trying to understand others' mental states could influence empathy and perspective-taking skills. While moderate engagement with social media might offer cognitive and emotional benefits, excessive or compulsive use could lead to maladaptive changes in neural pathways. For instance, over-reliance on online social rewards might affect real-world relationships and social interactions. Similarly, constant self-presentation on social media might impact individuals' genuine self-perception (Meshi, Tamir and Heekeren 778).

3. Altered connectivity and communication between networks of the brain

Social media use can lead to altered connectivity and communication between networks of the brain, which can have a negative impact on cognitive function and emotional regulation.

Dopamine is referred to as a 'feel good' chemical rather than giving us pleasure; it motivates us to do things that bring us pleasure. Scientists use dopamine to measure the addictive potential of an experience. A dopamine loop is a self-reinforcing cycle that is driven by the way a certain activity interacts with our brain's reward system. There are four major dopamine pathways or connections found in our brain that act as a "highway" for neurotransmitters. Each pathway has its own associated motor or cognitive process. Three of these pathways, the mesocortical, mesolimbic, and nigrostriatal pathways, are considered our reward pathways and are malconditioned in addiction. All three of these distinct anatomical pathways create an

association between the stimulus and a reward which induces positive feelings that become stronger each time through a process termed 'long-term potentiation'. This potential strengthens synapses over time, increasing the intensity at which they respond to a particular stimulus. Dopamine neurons in the midbrain signal the brain's reward system by releasing dopamine in response to unexpected or rewarding events. This process, known as long-term potentiation, strengthens the connections between neurons and helps the brain learn to repeat behaviors that are associated with rewards.

A readily graspable analogy can be drawn to elucidate this phenomenon: the human brain exhibits a proclivity for seeking the path of least resistance when allocating cognitive effort. It tends to veer away from tasks involving problem-solving, such as tackling complex challenges, crafting academic assignments, or engaging in formal educational activities. Instead, it inclines toward the allure of facile dopamine release, facilitated by the act of perusing social media platforms and assimilating copious volumes of information that foster cognitive associations, thereby evoking ephemeral sensations of contentment.

Subsequently, following this indulgence, the brain enters a state of deactivation, signaling the need for bodily repose. Upon reactivation, the brain instinctively seeks to recreate the pleasurable dopamine surge, driving individuals back toward their smartphones or online entertainment venues. Notably, individuals derive reduced satisfaction when abstaining from social media use, primarily attributable to the depletion of dopamine levels below their baseline.

This research underscores the pivotal significance of comprehending the intricate dynamics governing the interaction between social media consumption and the dopaminergic systems of adolescents. Furthermore, it underscores the imperative need to scrutinize the potential ramifications of this interaction on the overall well-being of this specific demographic group.

The impact of social media on cognitive function and brain development has gained attention recently, with TikTok being a prime example. Called "TikTok brain," the short video platform has been linked to negative effects on memory, attention span, and overall mental functioning. Studies have shown that addiction to online platforms can lead to atrophy in brain regions like the frontal cortex. Moreover, the visual-centric nature of TikTok hinders the development of reading skills, making it harder for younger users to engage with text-based information. Similarly, Instagram's effect on the brain's dopamine system may contribute to "age-related cognitive decline" due to the atrophy of brain networks caused by insufficient engagement in deep thinking and cognitive exercise.

An analogy that can help illuminate this concept is to liken the brain to a creature of habit, one that tends to gravitate towards the path of least resistance. In this context, it appears that the brain is not particularly inclined to engage in the mentally taxing activities of problem-solving, completing assignments, or participating in educational pursuits such as attending classes. Instead, it exhibits a distinct preference for seeking immediate gratification by indulging in the constant stream of dopamine made available through the habitual scrolling of social media feeds and the consumption of copious amounts of information that trigger feelings of temporary joy

and satisfaction. Consequently, following such indulgence, the brain seems to enter a state of temporary inactivity, signaling the body to take a respite. When the brain eventually rouses from this restful state, it seeks out the same source of dopamine, which is often found in the form of a smartphone or online entertainment (Meshi, Tamir and Heekeren 778).

Disorders

The use of smartphones and social media has been linked to a number of disorders in adolescents, including anxiety, depression, and addiction.

1. ADHD and Brain changes

There is growing concern that social media use may be contributing to the development of ADHD in adolescents, as well as exacerbating symptoms in those who already have the disorder.

One study found that college students with ADHD (Attention-deficit/hyperactivity disorder) were more likely to be internet addicts than those without ADHD. Attention deficiency was the most associated symptom of internet addiction among college students with ADHD. This correlation shows that ADHD may not be completely related to internet addiction and it doesn't establish a causation.. This framework encompasses neurobiological mechanisms (i.e., imbalance theory of brain development and dual pathway model of ADHD) and social mechanisms, including influences from peers and parents. The findings of these studies suggest that adolescents with ADHD may be more likely to use social media problematically than their peers without ADHD. This is likely due to a combination of factors, including neurobiological vulnerabilities, social influences, and the nature of social media itself (Aleksandra).

A set of researchers found that the relationship between social media use and executive functioning may be best, with moderate levels of usage and decreased at low levels of usage. Moderate-intensity social media use, which emphasizes social interaction, helps individuals filter distractions and focus on valuable information, leading to improved executive functions, especially in the shifting function. Additionally, this type of usage provides social rewards and emotional support, mitigating the negative impact of excessive social media use on cognitive function. Social media use can be associated with impaired executive functioning, particularly in the areas of impulsivity and inhibitory control (Yen et al). Another set of researchers found that social media addiction was associated with impaired executive functioning in young adults. Emotional disturbance and poor sleep quality were mediators in the relationship between social media addiction and executive function. Problematic social media use is associated with executive functioning impairment such as inhibitory control, decision-making, and working memory (Zhang et al. 1916).

2. Depression/Anxiety

Depression and anxiety are two of the most common mental disorders in adolescents, and there is growing evidence that social media use can contribute to the development of these disorders.

The Surgeon General's advisory, released in 2023, highlighted the escalating rates of teenage depression, sadness, and hopelessness, particularly among girls. The review's findings found that females using social media were more prone to experiencing depression symptoms than males. While the combined evidence underlines the need for strong safety standards and collaboration between policymakers, it also underscores the necessity for further research to comprehend the intricacies of this connection and to develop standardized assessment methods (Edwards and Hallie).

Recent research indicates that young adults who use social media excessively are three times more prone to experiencing depression, which could potentially lead to thoughts of self-harm or suicidal behaviors. This concern is especially pertinent given the rising suicide rates among teenagers, particularly in the 15-24 age group.

There is a strong connection between how much teenagers use social media and the impact it has on their sleep quality and mental health. A strong link between the amount of time adolescents spend on electronic devices, their engagement with social media, and various sleep and mental health aspects can be seen. Factors such as genetic predispositions, environmental influences, and the interplay between brain regions contribute to this vulnerability. The hormonal changes during adolescence, specifically in melatonin levels, lead to natural shifts in sleep-wake cycles, making teens more likely to stay up late and struggle with early mornings. Adolescents who spent more time on their devices experienced poorer sleep quality and shorter sleep durations, potentially due to the stimulating blue light emitted by screens.

Recent research has shed light on the relationship between social media use and social anxiety among adolescents. One study investigated the connection between life goals for using social media and Fear of Missing Out (FoMO) in Turkish adolescents. They found that extrinsic goals, such as seeking popularity and conveying a positive self-image, were associated with FoMO, which in turn led to lower academic performance due to distractions and study interference (Tanrikulu and Mouratidis 28695). Other researchers found that passive social media use was positively correlated with social anxiety, while active social media use had a negative correlation. The study also revealed that communication capacity mediated the relationship between social media use and social anxiety, suggesting that active social media use could reduce anxiety through improved communication (Lai et al. 9). These studies collectively emphasize the intricate relationship between social media usage patterns, communication capacity, and social anxiety in adolescents, underscoring the need for further exploration and potential interventions to mitigate the adverse effects of excessive social media engagement on their mental well-being.

FOMO is attributed to the pressure of staying connected and the fear of missing out (FOMO), leading teenagers to frequently check social media even near bedtime. The pervasive influence of social media platforms, exemplified by Instagram, has kindled a pertinent concern regarding FOMO among adolescents. This apprehension revolves around the incessant need to stay abreast of peers' activities, thereby driving teenagers to incessantly peruse their social media feeds and juxtapose their own lives with the seemingly idealized portrayals online. The

ramifications of this unremitting self-comparison are manifold, encompassing negative self-appraisals and often engendering perilous undertakings to capture captivating photographs or utilizing filters to enhance one's appearance. Notably, despite the inaccuracy inherent in numerous images circulating on Instagram, teenagers persistently gauge their self-worth against these visuals, potentially fostering detrimental self-perceptions. This study also suggests a complex connection between FOMO and well-being, highlighting how FOMO can both increase social media use and potentially strengthen social bonds, affecting overall well-being. Collectively, these seminal studies shed light on the intricate entwinement of social media, the pervasive FOMO, incessant self-comparison, and the intricate interplay they have on the psychological well-being of adolescents (Oberst et al. 53)

Recent investigations have delved into the intricate relationship between social media and the phenomenon known as nomophobia, which refers to the fear of being without a mobile phone. A group of researchers conducted a longitudinal study to uncover the temporal dynamics linking nomophobia, addictive use of social media, and insomnia among Iranian adolescents. This study of 1098 adolescents revealed that increased engagement in addictive use of social media was associated with a rise in nomophobia and insomnia over a three-month period. Notably, the interaction between nomophobia, addictive use of social media, and time was found to significantly impact insomnia. This suggests that both nomophobia and addictive social media usage could potentially serve as risk factors for adolescent insomnia. The intricate interplay unveiled in this study emphasizes the need for intervention strategies that target the reduction of nomophobia and addictive social media behavior to enhance the sleep quality of adolescents (Lin et al.).

Moreover, social media usage can lead to impulsive posting, unrealistic standards, and body image issues. Recent research delving into the realm of social media has shed light on its connection with the internalization of beauty ideals among individuals. Researchers conducted a study exploring the intricate interplay between appearance-related photo activity on social media, internalization of the ideal body, appearance comparison, and body satisfaction. Their investigation, centered around young South Korean female social media users, found that engaging in appearance-related photo activity on social media was associated with heightened levels of internalization and appearance comparison. Subsequently, these factors contributed to diminished body satisfaction. Notably, the study also highlighted the moderating influence of photo-editing behavior on the relationship between appearance-related photo activity, internalization, and appearance comparison (Lee and Hyun-Hwa). On a broader scale, another set of researchers delved into the association between the use of social networking sites and the internalization of a thin ideal among females. Their meta-analysis revealed a positive correlation between social networking site use and the internalization of a thin ideal. Additionally, their findings indicated that engagement with appearance-related features within social networking site had a stronger relationship with body image disturbance than broader social networking site use. This suggests that the interactive nature of online platforms, particularly concerning

appearance-related content, may contribute to the amplification of body image concerns. Together, these studies underscore the complex dynamics between social media, internalization of beauty standards, and its subsequent impact on body satisfaction and well-being (Mingioia et al. 7).

To address these concerns, parents are urged to monitor their teenagers' online activity and look for signs of excessive social media use that may contribute to mental health problems. Both parents and teenagers should prioritize healthier social media habits and recognize the importance of balancing screen time for overall mental well-being.

3. Digital Interactions

The adolescent brain is remarkably adaptable to the rapid changes in technology and is shaped by the experiences gained through digital interactions. While this adaptability is beneficial in a constantly evolving digital world, it also poses challenges as adolescents grapple with navigating the digital landscape responsibly and managing potential negative impacts on their mental well-being.

The striatum is a part of the brain that is responsible for processing rewards and reinforcing behaviors. Recently, researchers investigated the long-term effects of different types of digital media use, such as watching television, using social media, and playing video games, on brain development in children aged 9-11 years. The researchers found that digital media usage, especially playing video games, did not have a significant effect on the change in the cortical surface or volume of the cerebellum. However, a notable finding was that girls who spent more time playing video games exhibited a smaller change in the volumes of the striatum, a region associated with cognitive functions. This effect was not observed in boys. While the effect sizes were generally small, these findings underscore the need for further research with longer follow-ups to better understand the nuanced impact of digital media on brain development in children.

The diffusion of information and sharing is occurring at a massive scale across geographic proximity. Despite ample research, it still remains unclear as to how people determine which information to share. However, a person's psychological and social factors shape an individual's perception of information value, especially in a networked environment like social media. Before people share information they assess the outcome of doing so based on whether it would be beneficial, what others would think of it or is it worth sharing. This means that sharing information is a conscious act based on personal and social consequences. It can be said that the ways in which users choose what to share depends on (1) attributes of the information, (2) the perceiver's domain knowledge, and (3) the social relationships given.

When users share information on social media, it gives them the false impression that they know more about a certain topic than they actually do. Doing this can create a rise in self-esteem and confidence because by putting information online the share commits then he/she knows a substantial amount about the topic. Just after reading the headline of a trending post people post it to show that they 'care' about it or have a substantial knowledge base on the topic. Social

media takes control of who users are and what they really believe. While it can be used wisely, it mostly warps our perceptions and reasoning, which can lead to errors in judgment.

Social media use can lead to a condition called phantom vibration syndrome (PVS). PVS is the feeling that your phone is vibrating or ringing when it is not. It is thought to be caused by excessive checking of social media, which can lead to a heightened sensitivity to vibrations (Nivins et al.).

Social media algorithms are designed to show users content that is most likely to keep them engaged. This can lead to users being exposed to a narrow range of information, which can in turn alter their perception of reality. For example, an algorithm that favors PRIME information (content that is popular, engaging, or controversial) may show users more polarizing or politically extreme content than they would otherwise see. This can lead to users becoming more divided and less tolerant of other viewpoints (Brady et al. 950).

4. Benefits of Technology Usage

While there are some potential negative consequences of using technology, there are also many benefits that can be gained from its use.

Social media in adolescence offers a range of benefits that contribute to their overall development and well-being. Research highlights the positive impacts of social media on teenagers. Adolescents utilize social media platforms as tools for increasing communication skills, obtaining information, and developing technical proficiencies. These platforms provide a space for identity exploration, enabling them to express themselves, practice self-presentation, and gain self-concept clarity. Moreover, social media offers avenues for connection and social support, aiding in friendships and relationships. Through digital interactions, teenagers find like-minded peers, potentially reducing feelings of loneliness and boosting confidence (Tartari). Social media also promotes digital literacy and skill development. However, it is vital to acknowledge the potential downsides, as excessive social media use can lead to issues such as cyberbullying, exposure to inappropriate content, and sleep disturbances. Hence, while social media holds promise in facilitating positive growth, a balanced and mindful approach is necessary to harness its benefits while minimizing its risks (Uhls, Ellison and Subrahmanyam S69).

Therapies/Treatments/Recommendations

1. Dopamine fasting and Medication

Treatment for social media addiction requires a comprehensive approach that addresses the biological, psychological, and social aspects of the disorder.

A leading expert in addiction, emphasizes that addiction is a complex biopsychosocial disease that demands a range of interventions. One effective method is the concept of a "dopamine fast," which entails resetting reward pathways and allowing individuals to recognize cause-and-effect relationships. This technique encompasses biological changes by giving reward

pathways time to reset and psychological changes by altering the environment and social interactions. It's essential to recognize addiction as a spectrum disorder, varying in severity from mild to severe. While a dopamine fast might help those with mild to moderate addiction, individuals with more severe addiction may require professional intervention, such as rehab or medication-assisted treatment. This approach caters to individuals across the addiction spectrum, providing a nuanced and comprehensive strategy to address social media addiction (McKay and McKay).

2. Limiting social media addiction

A recent study highlights the positive impact of limiting social media use on psychological well-being. Involving 230 undergraduate students, the experiment assigned participants to either restrict social media use to 30 minutes daily or continue regular usage. After two weeks, those who self-monitored their usage experienced reduced anxiety, depression, loneliness, and negative affect, along with increased positive affect. The study stands out for focusing on self-monitoring as an alternative to complete abstinence, suggesting that moderation can enhance well-being. While some limitations exist, such as a lack of detailed content analysis, the findings urge further exploration of the nuanced link between social media and mental well-being (Faulhaber, Lee and Gentile 16). An article in ReachOut, an Australian website dedicated to the mental wellbeing of the youth, offers insightful strategies for curbing excessive social media usage through gradual changes. Platform features such as Instagram's personal activity and screen time settings on iPhones allow users to monitor their usage and establish limits. Engaging in screen-free hobbies for at least an hour per week can provide a refreshing break from digital devices and contribute to improved mental well-being. During such activities, employing the 'Do Not Disturb' mode minimizes distractions. Additionally, practicing phone-free dinners and leaving devices outside the bedroom to promote genuine interactions and better sleep quality. Recognizing that transformation takes time, the approach advises starting with small, manageable adjustments to gradually reduce screen time. By following these strategies, individuals can take control of their digital habits and achieve a healthier balance in their lives (ReachOut "5 ways to tame your social media use").

Conclusions

The present discussion revolved around the limitations in studies related to the impact of social media and smartphone usage on brain changes. Many studies focused on groups of people who already used or didn't use social media, potentially overlooking hidden variables that might have affected brain changes. One important concern was that observed brain changes might have been correlated with technology usage rather than caused by it. For instance, Dr. Clifford Sussman, MD, a child and adolescent psychiatrist specializing in internet and video game addiction has suggested that social media platforms are designed to keep users engaged, which can be difficult for people with ADHD to resist. In an interview with the CHADD organization, she suggested that social media use is a high-dopamine activity, which means it can be addictive.

Children and teens with ADHD are more likely to be drawn to high-dopamine activities and may have difficulty regulating their time spent on social media. Spending too much time on social media can lead to a child developing a higher tolerance for dopamine, making it more difficult for them to enjoy other activities. This can lead to boredom and irritability, as well as problems with attention and focus (Sussman et al. 320).

A study highlighted that most studies investigated instantaneous usage effects, rather than sustained or long-term usage impacts. The study discussed the potential negative impact of habitual smartphone use on cognitive functions like attention, memory, and delay of gratification (Wilmer, Sherman and Chein 10). While the research field was evolving, there was a need to address limitations in empirical methodology and interpretation for a more comprehensive understanding of the relationship between technology habits and cognitive functioning.

When reviewing the literature on the impact of social media use on adolescents, several challenges and factors come into play. One primary constraint is the limited scope of measurements used in studying the brain, including size and signals. Utilizing techniques like MRI scans for brain analysis can be expensive and may pose difficulties, potentially impeding comprehensive exploration.

The sensitivity of the subject itself can also present challenges. Some studies delve into the negative effects of social media, addressing topics like mental health, self-esteem, and well-being. For instance, one study posed sensitive questions to participants, probing into their personal experiences related to depression and sleep deprivation with potential adverse effects of social media use (Mohammadbeigi et al. 78).

One problem researchers face when studying how social media affects teenagers is finding a suitable group of teenagers to study. Nowadays, almost all teenagers use social media, so it's hard to find teenagers who have never used it. This makes it tough to compare the differences in the brains of teenagers who use social media a lot versus those who use it less. Because of this, it's not easy to say for sure that any changes in the brain are only because of how much they use social media.

Other factors, such as individual differences and lifestyle choices, could also be contributing to the observed changes in the brains of adolescents who use social media heavily. For example, these individuals may also differ in their physical activity levels, sleep habits, and social interactions. These lifestyle factors could also be impacting brain development, making it difficult to isolate the effects of social media use.

Therefore, researchers must be cautious when interpreting the results of studies on the effects of social media on the brain. They must use rigorous methods to control for other factors that could be contributing to the observed changes. Additionally, they may need to conduct longitudinal studies that follow adolescents over time to obtain a more complete understanding of the long-term effects of social media use.

In conclusion, the research on the effects of social media on the brain is still in its early stages, but it is clear that there are both positive and negative impacts. On the one hand, social

media can help to connect people, share information, and learn new things. It can also be a source of entertainment and support. On the other hand, social media can also be addictive and contribute to mental health problems such as anxiety and depression.

Future research should continue to explore the long-term effects of social media on the brain, which is an area that scientists still don't have concrete evidence for. Moreover, researchers should look into easy and affordable ways to have standardized measurement techniques to see the effect of social media on brain activity to draw long term conclusions. It is also important to consider the ways in which social media use is changing, as more and more people are using smartphones and other mobile devices. The ways in which smartphones may be harming us through constantly bombarding us with notifications and distractions.

In conclusion, the research on the effects of social media and smartphone usage on brain changes is still in its early stages. There are many limitations to the existing studies, such as the use of cross-sectional designs, the lack of control groups, and the difficulty of controlling for confounding variables. However, the available evidence suggests that there is a potential for both positive and negative effects of social media and smartphone usage on the brain. Future research should use more rigorous designs and methods to better understand these effects.

Future Directions

Studies have mostly focused on the short-term effects of social media use, but it is also important to understand the long-term impacts on people of all ages and the different ways that different types of social media platforms affect our brains. For example, how does social media use affect our cognitive abilities, mental health, and relationships over time? How does social media use affect the cognitive development of children? How does it affect the mental health of older adults? How do the different types of social media platforms, such as those that focus on sharing photos and videos versus those that focus on text-based communication, affect our brains differently? More research is needed to better understand the long term effects.

Works Cited

- Achterberg, Michelle et al. "Longitudinal Associations Between Social Media Use, Mental Well-being, and Structural Brain Development Across Adolescence." *Developmental Cognitive Neuroscience*, vol. 54, 2022, pp. 1-17.
- Akram, Wasim, and R. Kumar. "A Study on Positive and Negative Effects of Social Media on Society." *International Journal of Computer Sciences and Engineering*, vol. 5, no. 10, 2017, pp. 347-354. www.ijcseonline.org, E-ISSN: 2347-2693.
- Aleksandra. "TikTok is Killing Your Brain, One Short-Form Video at a Time." *Social Media Psychology*, 18 Aug. 2022, <https://socialmediapsychology.eu/2022/08/18/tiktok-is-killing-your-brain-right-now/>
- Allton, Mike. "Social Media is about sociology and psychology more than technology." – Brian Solis. *The Social Media Hat*, 18 Sept. 2012, <https://www.thesocialmediahat.com/blog/social-media-is-about-sociology-and-psychology-more-than-technology-brian-solis/>
- Alter, Adam. *Irresistible: The Rise of Addictive Technology and the Business of Keeping Us Hooked*. Penguin Press, 2017.
- Åström, Michaela. "The Adolescent Brain on Social Media: A Systematic Review." Bachelor's thesis, University of Skövde, School of Bioscience, 2021.
- Baughan, Amanda, et al. "I Don't Even Remember What I Read: How Design Influences Dissociation on Social Media." *CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, vol. 18, 2022, pp. 1-13.
- Brady, William, et al. "Algorithm-mediated social learning in online social networks." *Trends in Cognitive Sciences*, 2023, pp. 947-960.
- Cain, Matthew S., et al. "Media multitasking in adolescence." *Psychonomic Bulletin & Review*, vol. 23, no. 5, 2016, pp. 1932-1941.
- Center for Humane Technology. "Social Media and the Brain." Center for Humane Technology, 17 Aug. 2021, <https://www.humanetech.com/youth/social-media-and-the-brain>
- Crone, Eveline A., and Elly A. Konijn. "Media Use and Brain Development During Adolescence." *Nature Communications*, vol. 9, 2018, pp. 1-10, doi:10.1038/s41467-018-03126-x. www.nature.com/naturecommunications.
- Dean, Brian. "Social Network Usage & Growth Statistics: How Many People Use Social Media in 2023?" Backlinko, 27 March 2023, <https://backlinko.com/social-media-users>.
- Edwards, Erika, and Jackson, Hallie. (2023, May 23). Social media is driving teen mental health crisis, surgeon general warns. NBC News. <https://www.nbcnews.com/health/mental-health/social-media-driving-teen-mental-health-crisis-surgeon-general-warns-rcna35766>
- Eriş, Hüseyin et al. "The Effect of Mobile Phone Use of University Students on Their Academic Success." *European Journal of Science and Technology*, no. 27, 2021, pp. 433-438. EJOSAT, www.ejosat.com, doi: 10.31590/ejosat.942753.

- Faulhaber, Manuela Ellen, Jeong Eun Lee, and Douglas A. Gentile. "The Effect of Self-Monitoring Limited Social Media Use on Psychological Well-Being." *Technology, Mind, and Behavior*, vol. 4, no. 2, 2023, pp. 1-23.
- Firth, Joseph et al. "The 'Online Brain': How the Internet May Be Changing Our Cognition." *World Psychiatry*, vol. 18, no. 2, 2019, pp. 119-129.
- Firth, Josh A., Torous, John and Firth, Joseph. "Exploring the Impact of Internet Use on Memory and Attention Processes." *International Journal of Environmental Research and Public Health*, vol. 17, no. 24, 2020, pp. 1-12.
- Fotuhi, Majid. "What Social Media Does to Your Brain." NeuroGrow, 21 Sept. 2020, neurogrow.com/what-social-media-does-to-your-brain/
- Gupta, Sachin. "A Review on the Impacts of Social Media." *International Journal of Innovative Research in Engineering & Management (IJIREM)*, vol. 8, no. 6, November 2021, ISSN: 2350-0557, pp. 294-297, <https://doi.org/10.55524/ijirem.2021.8.6.60>. Accessed from www.ijirem.org.
- Hu, Yuanming et al. "Alterations in White Matter Integrity in Young Adults with Smartphone Dependence." *Frontiers in Human Neuroscience*, vol. 11, 2017, pp. 532-537.
- Jimenez, Yerika and Morreale, Patricia. "Social Media Use and Impact on Interpersonal Communication." *HCI International 2015 - Posters' Extended Abstracts*, vol. 529, Springer, 2015, pp. 91-96.
- Karthic, Ananthi and Sudhakaran M.V. "Perceived Self-Esteem and Assertiveness among Adolescents." *Journal of Interdisciplinary Cycle Research*, vol. XIII, no. III, March 2021, pp. 1785-1794. ISSN 0022-1945.
- Kokoç, Mehmat. "The Mediating Role of Attention Control in the Link Between Multitasking with Social Media and Academic Performances Among Adolescents." *Scandinavian Journal of Psychology*, vol. 62, no. 4, 2021, pp. 493-501.
- Lai, Fengxia et al. "Relationship between Social Media Use and Social Anxiety in College Students: Mediation Effect of Communication Capacity." *International Journal of Environmental Research and Public Health*, vol. 20, no. 4, 2023, pp. 1-14.
- Mahalingham, T., Howell, J., & Clarke, P. J. F. "Attention control moderates the relationship between social media use and psychological distress." *Journal of Affective Disorders*, 297, 536-541 (2022)
- Maza, Maria T. et al. "Association of Habitual Checking Behaviors on Social Media with Longitudinal Functional Brain Development." *JAMA Pediatrics*, 2023, pp. E1-E8 doi:10.1001/jamapediatrics.2022.4924.
- McKay, Brett, and Kate McKay. "How to Do a Dopamine Reset." *Art of Manliness*, <https://www.artofmanliness.com/character/habits/how-to-do-a-dopamine-reset/#:~:text=One%20of%20the%20first%20things>, 10 Sep. 2023.
- Meshi, Dar, Tamir, Diana I. and Heekeren, Hauke R.. "The Emerging Neuroscience of Social Media." *Trends in Cognitive Sciences*, vol. 19, no. 12, 2015, pp. 771-782.

- Meshi, Dar, Carmen Morawetz, and Hauke R. Heekeren. "Nucleus accumbens response to gains in reputation for the self relative to gains for others predicts social media use." *Frontiers in Human Neuroscience*, vol. 7, 2013, pp. 1-11.
- Mingoia, John et al. "The Relationship between Social Networking Site Use and the Internalization of a Thin Ideal in Females: A Meta-Analytic Review." *Frontiers in Psychology*, vol. 8, 2017, pp. 1-10
- Mohammadbeigi, Abolfazl et al. "The Prevalence of Phantom Vibration/Ringing Syndromes and Their Related Factors in Iranian Students of Medical Sciences." *Asian Journal of Psychiatry*, vol. 27, 2017, pp. 76-80.
- Lee, Minsun, and Hyun-Hwa. "Social media photo activity, internalization, appearance comparison, and body satisfaction: The moderating role of photo-editing behavior." *Computers in Human Behavior*, vol. 114, 2021.
- Lin, C. Y., et al. "Longitudinal Relationships between Nomophobia, Addictive Use of Social Media, and Insomnia in Adolescents." *Healthcare*, vol. 9, no. 9, 2021.
- Nivins, Simon et al. "The Long-Term Impact of Digital Media on Brain Development in Children." *MedRxiv*, 2022.
- Oberst, Ursula et al. "Negative consequences from heavy social networking in adolescents: The mediating role of fear of missing out." *Journal of Adolescence*, 55, 2017, pp. 51-60.
- Ostic, Dragana et al. "Effects of Social Media Use on Psychological Well-Being: A Mediated Model." *Frontiers in Psychology*, vol. 12, 2021, article 678766, pp. 1-13. www.frontiersin.org, doi:10.3389/fpsyg.2021.678766.
- ReachOut. "5 Ways to Tame Your Social Media Use." ReachOut, <https://au.reachout.com/articles/5-ways-to-tame-your-social-media-use>.
- Richard, Jocelyn M., et al. "Mapping brain circuits of reward and motivation: In the footsteps of Ann Kelley." *Neuroscience and Biobehavioral Reviews*, vol. 37, no. 0, 2013.
- Sharifian, Neika, and Zahodne, Laura B. "Daily associations between social media use and memory failures: the mediating role of negative affect." *Journal of General Psychology*, vol. 148, no. 1, 2021, pp. 67-83.
- Sivakumar, Arunkumar, Sudarsan Jayasingh, and Shahenaz Shaik. "Social Media Influence on Students' Knowledge Sharing and Learning: An Empirical Study." *Educ. Sci.*, vol. 13, no. 7, 745, pp. 1-16.
- Skowronek, Jeanette, et al. "The mere presence of a smartphone reduces basal attentional performance." *Scientific Reports*, vol. 13, 2023, <https://doi.org/10.1038/s41598-023-36256-4>.
- Sussman, Clifford J. et al. "Internet and Video Game Addictions: Diagnosis, Epidemiology, and Neurobiology." *Child Adolesc Psychiatr Clin N Am*, vol. 27, no. 2, 2018, pp. 307-326.
- Tanrikulu, Gulfem, and Athanasios Mouratidis. "Life aspirations, school engagement, social anxiety, social media use and fear of missing out among adolescents." *Current Psychology* 42 (2023): 28689–28699. <https://doi.org/10.1007/s12144-022-03917-y>.

- Tartari, Elda. "Benefits and Risks of Children and Adolescents Using Social Media." *European Scientific Journal*, vol. 11, no. 13, 2015.
- Uhls, Yalda T., Ellison, Nicole B., & Subrahmanyam, Kaveri. "Benefits and Costs of Social Media in Adolescence." *Pediatrics*, vol. 140, no. Supplement_2, 2017, pp. S67–S70.
- Van der Schuur, Winneke A. et. al. "Exploring the Long-term Relationship Between Academic-Media Multitasking and Adolescents' Academic Achievement." *Journal of Youth and Adolescence*, vol. 49, no. 1, 2020, pp. 139-158.
- Waugh, Rob. "Teenagers who use social media 15 times a day become 'hypersensitive'." *Yahoo News*, 4 Jan. 2023,
https://uk.news.yahoo.com/teenagers-who-use-social-media-15-times-a-day-become-hypersensitive-142329108.html?soc_src=social-sh&soc_trk=ma.
- Wikman, Patrik et al. "Brain Responses to Peer Feedback in Social Media Are Modulated by Valence in Late Adolescence." *Frontiers in Behavioral Neuroscience*, vol. 16, 2022, pp. 1-14.
- Wilmer, Henry H., Sherman, Lauren E, and Chein, Jason M. "Smartphones and Cognition: A Review of Research Exploring the Links between Mobile Technology Habits and Cognitive Functioning." *Frontiers in Psychology*, vol. 8, 2017, pp. 1-16
- Woolley, Kaitlin, and Marissa A. Sharif. "Down a Rabbit Hole: How Prior Media Consumption Shapes Subsequent Media Consumption." *Journal of Marketing Research*, vol. 59, no. 3, 2021, pp. 11-25. <https://doi.org/10.1177/00222437211055403>
- Yen, Ju-Yu et al. "The association between adult ADHD symptoms and Internet addiction among college students: The gender difference." *Cyberpsychology & Behavior*, vol. 12, no. 2, 2009, pp. 203-207.
- Zhang, Kuo et al. "Effect of social media addiction on executive functioning among young adults: The mediating roles of emotional disturbance and sleep quality." *Psychological Research and Behavior Management*, vol. 16, 2023, pp. 1911-1920.

Overlooked Predispositions to Anorexia and Improving Treatment By Lainey Bradley

Abstract

Every year, more people are diagnosed with anorexia nervosa (anorexia), yet there is a lack of public awareness for the factors behind this psychological disorder (Holcombe; “What Is Anorexia: Symptoms, Complications and Causes.”). Neurobiological-based studies on patients with anorexia depict the chemical imbalances and cognitive functioning abnormalities contributing to the emergence of the eating disorder (Frank et al.; Lozano-Serra et al.). This review paper analyzes the psychopathology of anorexia relating to the formation and progression of the illness as well as discussing current treatment approaches. The brain-based anatomy of anorexia will be exemplified through a review of controlled experiments using brain scans, notably positron emission tomography (PET) technology. Qualitative data referring to the symptoms of anorexia and the thoughts and emotions that may accompany the disorder will be illustrated through the inclusion of podcasts hosted by young women who have had anorexia and peer-reviewed articles that include psychiatric evaluations for anorexia, interviews with patients, and longitudinal studies of patients throughout recovery. Treatment for anorexia is also reviewed. The unpredictability surrounding the intensity of the eating disorder through its development makes anorexia difficult to treat. Anorexia manifests differently in each patient, so a universal approach to treatment is not effective in curing a patient’s disorder (Kenny et al.). Furthermore, lack of understanding from clinicians and professionals weakens a patient’s chances for full recovery because staff do not know how to balance restoring physical health with psychotherapy (Hernberg; Paintain; Roloff). We present that if treatment is individualized to address the struggles of each patient’s case of anorexia, then patients are able to heal the causes that fueled the eating disorder, increasing the likelihood of staying healthy in recovery.

Keywords: anorexia nervosa, neurobiological, qualitative data, psychotherapy, recovery

Introduction:

Anorexia nervosa has the highest mortality rate among all psychiatric disorders (“What Is Anorexia: Symptoms, Complications and Causes.”). In the United States, ten thousand people die from eating disorders each year, equating to one person every hour (“Anorexia Statistics: Gender, Race & Socioeconomics”). From spring of 2020 to spring of 2021, the number of patients admitted to inpatient care for eating disorders more than doubled (Holcombe). Thus, there is a need for a greater understanding of anorexia and effective treatment for the disorder. However, when considering effective treatments, there is not a singular cause for an eating disorder (Klump et al.; Paintain; Roloff). A multitude of factors can lead to the development of an eating disorder (Frank et al., 2019; Holtkamp et al.; Kenny et al.).

The *Diagnostic and Statistical Manual for Mental Disorders, Fifth Edition (DSM-5-TR*; American Psychiatric Association, 2022) categorizes an individual as having anorexia nervosa if the following requirements are met: a significant restriction of energy intake resulting in a low

body weight in the context of age, sex, developmental trajectory, and physical health; an intense fear of gaining weight or being fat despite being underweight; current disturbance with weight or shape which influences self-evaluation and denial of being underweight (“Anorexia Nervosa”).

Research in the field of anorexia has found commonalities stretching across individuals diagnosed with this disorder. Starvation does not allow the brain to function properly, and areas of the brain are at risk of losing their proper ability to function (Frank et al., 2013). Individuals experiencing anorexia have a decreased ability to exhibit higher level cognitive functioning, for patients with anorexia record lower than average scores on tests measuring visual-spatial ability, visual memory, and executive functioning. This lack of ability then makes it more difficult for the individual to combat anorexia’s influence (Lozano-Serra et al.). The body does not have enough energy to put towards fighting self-destructive coping strategies (e.g., eating disorder behaviors), and certain trait tendencies become more prevalent as the disorder develops. Specifically, perfectionism, cognitive rigidity, overcontrolled emotions, and obsessional behaviors are heightened with anorexia (Holtkamp et al.). Fortunately, through treatment, weight restoration takes the body out of survival mode. Energy can be distributed to areas of the body that are not deemed as essential for keeping oneself alive, so some areas of the brain normalize during recovery and weight restoration; nevertheless, trait tendencies persist even once the physiological functions of the body normalize (Bezance & Holliday; Frank et al., 2013). This review explores the theory that the neurobiological and psychological symptoms that remain present through recovery pre-existed the diagnosis of anorexia nervosa and were intensified by the mental disorder.

Characteristics of Anorexia

Dopamine

Dopamine pathways of the brain become retrained through the development of anorexia nervosa (Frank et al., 2013). Dopamine is a neurotransmitter responsible for feelings of pleasure and motivation (“2 Minute Neuroscience: Dopamine”). This chemical messenger is also linked to reinforcement and reward learning, memory, sleep regulation, and arousal (“Dopamine: What It Is, Function & Symptoms.”). There is a positive correlation between the activation of the dopamine neurotransmitter and anorexia nervosa (Frank et al., 2013; Frank et al., 2019). When people with anorexia lose weight, the body wants to remain homeostatic, so the body increases its dopamine signaling, communicating the body’s craving for caloric energy (Frank et al., 2019). The brain is then in a motivational state, where a reward is sought out, and an aversive stimuli is avoided (“2 Minute Neuroscience: Dopamine”). For the brain, the reward is food, and the aversive stimulus is hunger and a non-homeostatic state. The food is considered a reward to the point where there is a motivation to feed oneself and experience an associated pleasurable feeling. However, when a person has anorexia, the brain’s dopamine pathways are modified; individuals do not see food as the rewarding stimulus, and starvation is the new reward (Frank, 2013).

People with anorexia nervosa are likely to have increased prediction error signaling (Frank, 2013). An unexpected reward will trigger a spike in the firing of dopamine neurons, but the more the unexpected reward occurs, the brain begins to expect the stimulus (Humphries). Rewards are sought out to avoid an aversive stimuli, an unpleasant event that decreases the future action of the behavior causing the aversive stimuli, and a reward is a desirable object or action that reinforces the behavior resulting in the reward (Pritchard & Chong). On that account, an unexpected reward resulting from a certain action or behavior increases the likelihood of the individual engaging in the behavior again, and the reward is no longer unexpected (Garrison et al.). After habituating to the rewarding stimulus, more dopamine is released when the brain suspects the reward is coming, rather than releasing a greater amount of dopamine when the reward is actually delivered. After a predicted reward is received, dopamine levels return to baseline. However, if a reward is expected and no reward is received, dopamine activity falls below baseline (Humphries). In a study performed by Frank (2013), adolescent girls were tested for prediction error signaling using a monetary task, where the participants learned to associate images with certain outcomes, so they could predict the consequences, as opposed to guessing what the effect of their associations would be. Participants were noted to have a higher rate of correctly pairing images with outcomes in relation to the control group, suggesting a higher sensitivity to rewards in patients with anorexia due to the quicker connections of causes and effects in the brains of patients with anorexia. This study also suggests that anorexia corresponds with more harm avoidance. Patients with anorexia tend to refrain from “normal” activities or behaviors in an attempt to avoid punishment. Those with anorexia when compared to a control group of those without the disorder tend to learn the associations between images and outcomes in the monetary reward activity faster because they are fearful of punishment. Dopamine levels in this experiment conclude that the participants with anorexia are more likely to accurately predict the reward, suggesting a higher cognitive ability to acquire information on the connections between a stimulus and its effect.

Patients with anorexia nervosa are apt to release more dopamine due to their conditioned understanding of reinforced behaviors (Frank et al., 2019). For those with anorexia, weight loss is predictable considering the low food intake and there is excitement surrounding weight loss (Coniglio et al.). Dopamine neurons habituate to the reward of losing weight, so as weight loss is expected by someone with anorexia, the excitement deriving from the activation of dopamine is occurring before seeing the weight loss (Humphries). The process of restricting food, exercise, or engaging in other eating disorder behaviors is what then triggers the release of dopamine (Coniglio et al.). As the antecedent stimulus becomes exciting, and as the latter stimulus continues to please the person due to eating disorder symptomatology, the cycle grows more powerful to encapsulate more of the eating disorder into the individual’s daily life (Frank et al., 2019). The observable behaviors of an individual with anorexia derive from rewired motivational brain circuits, so their actions, while causing harm to themselves, are self-gratifying. Hence, dopamine’s influence on a person with anorexia takes over primitive human functioning leading to significant changes in a person’s behaviors (Frank 2013; Frank et al., 2019).

Exercise

Typically, exercise is important for the body to stay healthy, but for someone with an eating disorder, exercise can be dangerous (Archer; Coniglio et al.; Dittmer et al.). From a physiological perspective, blood flow to the brain is increased through movement, arousing areas of the brain. In particular, exercise is helpful for encoding memories and the process of learning. Brain cells receive more nutrients through increased blood flow, which promotes neuroplasticity. On top of that, physical activity releases more dopamine and endorphins while decreasing the amount of stress receptors in the brain, creating pleasurable experiences (McGregor). Yet with anorexia, a liking for exercise goes deeper than the physiological benefits. Exercise can be a way to alter shape, weight, and appearance, thus becoming a coping mechanism for someone with anorexia (Archer).

For people with anorexia, the primary reason individuals initially fall into a pattern of increasing exercise is to heighten positive affect (Coniglio et al.). Anorexia, aligning with symptoms of depression, takes away joy that was once found in certain activities, as seen through dopamine's influence on the brain (Frank, 2013). At a time when individuals are not finding pleasure in other aspects of their life, exercise elicits the desired positive emotions.

The positive correlation between eating disorders, cognitive rigidity, and easily becoming addicted to certain behaviors is seen through pathological exercise. Pathological exercise, although there is no universal definition, is reported by Dittmer as "excessive exercise that a patient feels driven to perform in response to an obsession or according to rules that must be applied rigidly, and exercise that is aimed at preventing or reducing distress or at preventing some dreaded consequences." Exercise becomes pathological through building up a tolerance, meaning more exercise is needed to have the same effects on mood with beta-endorphins (Coniglio et al.). Anorexia drives patients to exercise because they feel the need to "burn off the meal," or they may exercise in advance to "earn the meal." Patients find a sense of comfort in exercise since it relieves the guilt surrounding food that is difficult for those with an eating disorder to get rid of (Dittmer et al.).

Along these lines, this definition of pathological exercise supports that people with anorexia are compelled to exercise in order to decrease negative affect. Experiments with animals depict that there is lowered availability of beta-endorphins in animals who exercise regularly, lowering the capability to produce the desired positive affect that comes from movement (Coniglio et al.). In Coniglio's study, participants in the control group only mentioned exercise as heightening positive affect while patients with anorexia also reported exercise as removing aversive stimuli. Most compulsive exercisers continued to adhere to their routine even though they no longer found happiness in the activity, illustrating that as the eating disorder develops, the motivation of patients with anorexia to exercise changes from doing so to better their mood to then serving as a coping mechanism for emotions stemming from the eating disorder, such as shame and guilt (Dittmer et al.).

Anorexia escalates levels of cognitive rigidity and perfectionism, which are theorized to have existed prior to the eating disorder, notably observed in relation to exercise (Kaye et al.;; Lozano-Serra et al.; Nahman & Holland). A compulsive exerciser usually has a strict routine as to when the appropriate time is to exercise, for how long, the number of exercises done, and the order of the exercises. Individuals with anorexia take all of these factors into great consideration and some have voiced that they plan their day around exercise (Cresswell et al.). Perfectionists feel pressured to keep up with their exercise routine and continue to add to it as a way to maintain self-control and reach the feeling of achievement and reward, even though this is at the cost of the patient's standards for exercise being unrelentingly high (Dittmer et al.). When any part of the exercise routine is in danger of not being at the "normal" standard for the person with an eating disorder, there is distress. Instead of sitting with that distress, patients note that their mind instantly directs its thoughts to be about how to compensate for not exercising to the full amount they would have liked to, whether that be through eating less or working out more the next day (Nahman & Holland; Roloff). Individuals with eating disorders have a hard time adjusting their thinking to fit the flow and uncertainty of life. Patients are constantly planning out their next moves in relation to the eating disorder that they become fully consumed by its influence (Dittmer et al.). Exercise impedes on daily life, taking away both mental and physical energy from a person with anorexia that used to be spent on more satisfactory activities and experiences.

On a neurobiological level, for patients with anorexia, exercise progresses into a compulsion that becomes independent from dopamine reward processes and, instead, is more dependent on corticostriatal connections in the brain, which are linked to habitual and addictive behaviors (Archer; Cresswell et al.). Archer conducted a study to decrease the synthesis and transmission of dopamine by applying acute phenylalanine and tyrosine depletion in patients with anorexia and the control group. Patients with anorexia showed a higher motivation to exercise than controls, despite exercise being reinforced for both groups. In patients with anorexia, reduced levels of dopamine do not decrease an individual's motivation to exercise, but for healthy controls, their motivation to exercise did decrease. Although dopamine plays a crucial role in eating disorder behaviors being viewed as rewarding, the data indicated that anorexia continued to develop .

Trait and Cognitive Tendencies

Across people with anorexia, many similarities exist when observing cognitive abilities and character traits. Black and white thinking, cognitive rigidity, and perfectionism are the most common cognitive distortions, faulty inner mental thinking, perceptions or beliefs that increase misery and fuel anxiety, for individuals with anorexia to exhibit (Coniglio et al.; Lozano-Serra et al.). The question arises of whether or not default thinking patterns emerge as an effect of anorexia or if they predispose an individual to anorexia. Due to research showing that irrational traits and cognitive tendencies persist throughout recovery, it is suggested that these propensities are temperamental, existing before the eating disorder (Lozano-Serra et al.).

Although there are not any studies of individuals with anorexia before they developed an eating disorder, evidence from an individual's past can suggest the presence of certain cognitive distortions. Also prevalent in those with anorexia, patients indicate a history of trying harder than peers in school and exhausting themselves by putting more effort into academics, believing that if they did not, they would be seen as a failure. In her podcast, *Best of Bettina*, Bettina, a young woman in recovery from anorexia, discloses that growing up, she did well in school. She began to demand more from herself, putting in more effort than her classmates because it was the only way for her to feel a sense of pride and achievement. These unrealistic expectations Bettina set for herself are instances where, as an innocent child, years before the development of an eating disorder, her mind naturally thought in absolutes. Her unreasonable standards began to infiltrate her life as Bettina recalls the initiation of anorexia while growing up arrived when the voice in her head told her that since she was doing so well in school and succeeding in other areas of her life, she needed to turn that back to her body. Bettina began to strive for a perfect body as a result of having automatic, reactive thoughts in extreme measures. As a child, nobody was punishing her for not being the best student or solving arithmetic problems the fastest. Bettina's motivation to thrive academically was internal, and the motivation to meet a certain beauty standard was intrinsically motivated, as well.

In patients with anorexia, many cognitive distortions exist simultaneously, contributing to the deathliness of the illness, for 95% of people with anorexia suffer from another mental disorder ("What Is Anorexia: Symptoms, Complications and Causes."). Black and white thinking compels viewing either side of an argument, event, or action as extreme, which is convergent with perfectionism (Bezance & Holliday, 2013). Perfectionists are internally driven to avoid disappointment and seek out accomplishment. Perfectionism causes individuals to have a negative outlook on things on one end of the extreme ("*Perfectionism*"). With anorexia, perfectionism manifests as only eating foods when they meet a certain criteria. Sorting food labels also reflects black or white thinking, arguing whether foods are "healthy" or not (Frank et al., 2019). Brooke, an adolescent recovering from anorexia, remarks that while she was anorexic, her mind communicated, "If I eat, it's going to be worth it. Why am I eating this many calories if it's not going to be worth it. It's got to be perfect to me" (Paintain 33:45). Forming judgments about food, and then acting on that judgment enhances an unhealthy relationship with food (Frank et al., 2019). The preceding worries and daily stressful thoughts that cognitive distortions communicate to individuals prior to the appearance of an eating disorder are continuously triggered through the eating disorder's progression, verifying why the layers to this mental health disorder are so dense (Cogiblio et al.). When multiple negative thought patterns occur continuously, they reinforce the eating disorder (Frank et al., 2019).

In addition to cognitive distortions, people with anorexia display depressive, anxious, and obsessive-compulsive behaviors. Recorded in a study conducted by Holtkamp et al., even after ten years following discharge from an inpatient setting, the majority of patients with anorexia continue to display these behaviors. Further, compared to the control group, former patients with anorexia recorded a higher level of interpersonal sensitivity, including feelings of inadequacy,

inferiority, and lower self-image. It is important to note that none of the patients from this study met the full criteria for any other mental health disorders other than anorexia, despite embodying some behaviors and symptoms that align with depression, anxiety, and obsessive-compulsive disorder. One possible explanation for the enduring behaviors is that starvation and weight loss, primary symptoms of anorexia, triggered these traits, and they remained throughout recovery as a result of the eating disorder's influence.

Various behavioral tendencies portrayed through research experiments in those with anorexia validate the theory that when the brain is deprived of energy, behaviors and motivations seen in the average person also decline. For example, Klump et al. measured personality characteristics of women during a period where they met criteria for an eating disorder and then through recovery. Data from this study shows that women with anorexia tend to be more nervous in new situations and refrain from typical behaviors as a means of avoiding judgment from others. Coined "harm avoidance," feelings of fear and shyness triumph over the standard emotions of relaxation and optimism found in participants without anorexia (Markett et al.). Likewise, the concept of novelty seeking – behavioral activation to pursue rewards – negatively correlates with the development of a restrictive eating disorder where internal motivations, seen in the brain's production of dopamine, are lessened; for patients with bulimia nervosa, levels of novelty seeking were closer to the levels of novelty seeking in the control group than the levels of novelty seeking in the group of patients with anorexia, which were noticeably lower. The lack of care for goal or reward in patients with anorexia highlights the decreased activation of the amygdala, the part of the brain responsible for regulating and releasing emotions. Yet when levels of novelty seeking are measured again during recovery, the scores of patients who previously had anorexia increased, presenting scores more relative to those of the control group (Klump et al.). It can be postulated that when the brain is not receiving enough caloric energy as a result of significantly lowering food intake, areas of the brain begin to shut down as a means of saving energy for essential functions that keep the body alive, and once the intake of food escalates with recovery, previously "dead" regions of the brain come back to life as a result of the body having an adequate amount of energy.

From a neurobiological perspective, brain scans emphasize a more specific idea pertaining to the state of the amygdala during starvation. Anxiety is commonly tied with anorexia due to deeply rooted fears surrounding food, body image, and other eating disorder idealizations, including the intense motivation to exercise, eat in public, or discuss food ("Anorexia Nervosa"; Seeger et al.). Positron emission tomography (PET) scans of women with anorexia reveal anxiety as it is fueled when patients are shown target images related to the body or food (Fig. 1). The prefrontal lobe and anterior cingulate cortex are more activated with the photographic models, proving that the brain spends more effort processing visual food stimuli (Frank et al., 2004). The prefrontal lobe and anterior cingulate are responsible for interpreting visual stimuli and pairing those visuals with complex cognitive functions, ranging from the experience of feelings of empathy to impulse control and decision-making (Marusak et al.). The ability of the anterior cingulate to trigger higher-level thoughts enables the amygdala to be more

sensitive to food and body-related images (Frank et al., 2004). The minimal supply of energy that the brain can utilize is fully absorbed and manipulated by the anorexia-related symptomatology, so any area that is not related to the eating disorder essentially does lose the extent to which it used to be activated prior to the illness.

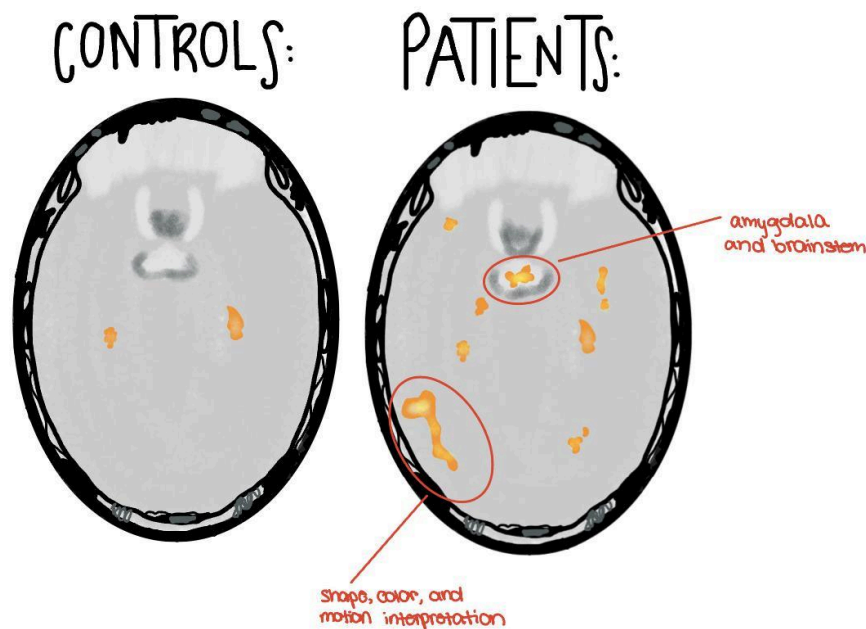


Fig. 1: PET scans portray that target stimuli (visuals relating to food and body image) elicits activity in the amygdala, brainstem, and visual processing areas of the brain in patients with anorexia but not the controls.

A cycle emerges between different process centers of the brain, beginning with the visual input being transcribed in the prefrontal lobe and anterior cingulate. Once the image is interpreted, emotions stemming from the eating disorder relay the amount of fear and degree to which the disorder is activated, and the negative emotions stemming from the eating disorder lead to an individual giving in to the judgments and processing of the images when shown to the patient again (Frank et al., 2013). Because patients with anorexia who were underweight and ill received low scores on a psychological evaluation specifying eating attitudes and symptoms, the results illustrate that their levels of cognitive processing are rigid (Lozano-Serra et al.). Inflexibility when it comes to thought increases the likelihood that negative beliefs surrounding food and weight appear more frequently, which fuels the cycle of disordered psychological symptoms that patients with anorexia experience (Marusak et al.).

Accordingly, it might be a result of the state of starvation that deprives an individual with anorexia from the ability to counter intrusive eating disorder thoughts and feelings (Frank et al.,

2019; Lozano-Serra et al.). As starvation continues, the connection of white matter from the frontal cortex to the amygdala is weakening, which can eventually lead the prefrontal brain region to be incapable of controlling and countering the hypersensitized amygdala that has been affected by the eating disorder (i.e., calming oneself down) (Frank et al., 2013). The reinforcing mechanism is to avoid the anxiety and fear stemming from the eating disorder, so the individual engages in behaviors that will make them feel better on the inside, lessening the emotional pain that is experienced (Paintain). The cycle is fed through the desire to alleviate the uneasiness that a patient with anorexia faces, yet the eating disorder does not allow the patient to see the danger in their actions, presuming that the damage done to the brain is worth the risk to satisfy the goals of the eating disorder (Frank, 2013; Frank et al., 2019).

Treatment Flaws

Overview of Treatment

Despite the prevalence of anorexia nervosa within society, with 1.2% of Americans over the age of 15 meeting the criteria for anorexia, in comparison to other mental disorders (ie. depression, anxiety, obsessive-compulsive disorder), little research on anorexia has been performed (“Anorexia Statistics: Gender, Race & Socioeconomics”). From studies that have been conducted, most are only composed of white females, making the results of the experiments difficult to generalize to other patients (Frank et al., 2004; Holtkamp et al.; Kenny et al.; Seeger et al.). Further, data across different studies is not always the same, illustrating that there is no “one size fits all” policy when it comes to assessing and examining anorexia. Experiments regarding principle symptoms and areas of neural activation when examining the brains of patients with anorexia can show different results as to the severity of symptoms or regions of the brain where the most activation occurs (Frank et al., 2004; Kaye et al.; Seeger et al.). The ambiguity and large range of personality characteristics, cognitions, and actions endured by individuals with this specific eating disorder make it hard to identify what recovery is and how it is achieved (Bezance & Holliday; Kenny et al.; Paintain).

Across all studies included in this review that include personal perspectives from patients with anorexia, the patients voiced concerns in regards to the definition of anorexia focusing on weight and food intake while excluding what they believe to be crucial aspects of the disorder. Treatment is based upon targeting the clinically agreed upon definition from the *DSM-5-TR*, and when that definition leaves out important factors, medical professionals may not address every aspect of the psychological illness (Bezance & Holliday; Cook et al.; Holtkamp et al.; Kenny et al.; Klump et al.; Paintain; Roloff). The foundation from which anorexia emerges, negative thought patterns and behaviors, continue to eat away at the individual’s healthy self (Hernberg). It is the responsibility of the specialists in the eating disorder field to care for each of the illness’s facets as symptoms that remain post-treatment are prone to inflict pain in the future (Bezance & Holliday; Kenny et al.). The second part of this review discusses the shortcomings within the current treatment system for anorexia nervosa, emphasizing the need for individualized treatment plans in order to meet the definition of recovery that each individual is striving to reach.

Absence of Universal Definitions

Although many note anorexia recovery as including weight restoration, the absence of eating disorder behaviors, healthy organ functioning, proper food intake, and the ability to exercise in a healthy way, there is no set definition that the medical community provides for recovery from anorexia (Herberg; Roloff) . Because there is no operationalized definition of recovery, it is left up for interpretation. Treatment is not comprehensive since some centers focus more on quantitative data (e.g., Body Mass Index (BMI), weight, heart rate, blood pressure) as opposed to psychological recovery (Holtkamp, et al.). Without a standard for what recovery looks like, patients with anorexia are dismissed from treatment centers at different stages in recovery (Paintain; Roloff). Furthermore, it is hard to compare scientific data through research experiments that included recovered patients because across multiple studies, all patients were not necessarily at the same level of recovery whereas in one study, all fit the provided definition of recovery. A medically appropriate specification for defining recovery from anorexia would benefit all individuals struggling with this psychological disorder.

Clinical studies range from providing explanations for long-term recovery versus short-term recovery, incorporating psychological testing into the analysis of recovery, or focusing mainly on weight (Bezance & Holliday; Kenny et al.). Holtkamp defines long-term recovery for females as a time period of at least three years when there are normal menstrual cycles, a stable weight, no “weight phobia” (i.e., anxiety around weight and body image when it comes to food and eating), and does not meet the criteria for any eating disorders. The criteria here for long-term recovery includes both physical and psychological measures, but there is vagueness as to the concept of “weight phobia.” Like most definitions of recovery, this one is all-or-nothing, where failing to meet one of these qualifications within the three year period results in that individual not being categorized as being in long-term recovery. However, that raises questions as to where the individual stands. For example, if a patient who previously had anorexia has been in recovery for two years, has maintained a stable weight and normal menstrual cycles, but has a slip where compulsive exercise occurs, there is a question as to whether or not that individual is no longer in long-term recovery. The all-or-nothing emphasis of the requirements to fit this definition of recovery separates individuals into two groups, ignoring those who do not clearly fit into either category. Many definitions of recovery do not take into account that recovery is not linear, which devalues the effort that an individual has put towards recovery.

One of the most popularly accepted descriptions of recovery from anorexia is the Bardone-Cone’s definition. Under this standard, full recovery occurs once the patient has a BMI greater than 18.5; has not engaged in fasting, purging, or binge eating in the past three months; and obtains a score on the Eating Disorder Examination – Questionnaire (EDE-Q) less than one standard deviation from the norm (Kenny et al.). This definition is more extensive and complete than Holtkamp’s definition of long-term recovery described above due to the specified behaviors listed. It could be implied to the public that the three noted behaviors in the Bardone-Cone definition are the most essential behaviors that distinguish an eating disorder, and that the other

behaviors may not appear without partaking in the behaviors of fasting, purging, or binge eating. Nonetheless, considering that behaviors related to compulsive exercise, body checking, counting calories, and negotiating food preferences are excluded from the Bardone-Cone definition falsifies the criteria for recovery. There is room for someone who previously had anorexia to exhibit the less commonly known eating disorder symptoms while being told that they are progressing through recovery at a successful rate (Paintain). Hence, patients may be misguided throughout their treatment, for they are told that the more secretive behaviors are essentially fine to engage in.

Additionally, Bardone-Cone created a definition for partial recovery. Partial recovery is when someone who previously had anorexia has a BMI greater than 18.5, and has not engaged in fasting, binge eating, or purging in the past three months, but the individual scores higher than one standard deviation above the norm on the EDE-Q. The patient demonstrates growth in recovery, and a psychological breakthrough regarding eating disorder thoughts, voices, emotions, and goals (Kenny et al.). With two definitions pertaining to particular stages of recovery, there is an opportunity for patients and their treatment teams to see where the eating disorder is present and what is holding the individual back from full recovery. Plus, the larger idea of recovery is broken down into smaller steps, making goals seem more attainable.

The Bardone-Cone definition has received critical opinions from patients with anorexia who do not believe the provisions for recovery are realistic. For example, participants in Kenny's study expressed that a time frame for recovery is impractical. Depending on the amount of time that someone had an eating disorder and the severity of the illness, reaching goals in recovery looks different for each individual. A time frame can inaccurately categorize individuals as being recovered or not. Furthering this idea, recovery is more than the absence of eating disorder symptoms, but recovery is also the presence of normal aspects of life that were shadowed during the time of the eating disorder. This looks different for everyone based on what life looked like prior to the development of the eating disorder, and this is another complexity to the definition of recovery that needs to be considered (Hernberg; Paintain). In order for a patient, in their eyes, to truly recover, they must gain control back over their life, and that comes along with engaging in activities and experiences that they used to enjoy, without the effects of the eating disorder determining the thoughts, feelings, and behaviors during these times (Bezance & Holliday). However, the Bardone-Cone definition strictly fixates on the absence of eating disorder behaviors without mentioning the resurgence of positive life activities (Kenny et al.).

Considering the information highlighted above, there is discussion around having two definitions of recovery: one that can be used for research purposes and another that is individualized for treatment purposes (Kenny et al.). The medical model's definition of recovery mainly focuses on the behavioral aspect of the disorder, whereas the recovery model takes into account personal experiences that individuals with eating disorders face, which tend to dig deeper into the cognitions fueling the eating disorder (Bezance & Holliday; Klump et al.). We propose that the research definition of recovery should broaden, so more patients with anorexia can participate in certain studies. This way, the criteria does not exclude or attempt to categorize

patients into groups. Everyone's eating disorder is unique, with one behavior coming up the most for one patient and never occurring for the other (Hernberg; Roloff). Therefore, patients should not be categorized into groups, for there must be one broader definition of recovery that encompasses the general idea of progress, moving towards a normal life that is free from the eating disorder (Paintain). With a specific definition of recovery for research and medical purposes, the goal is not to hone in on precise, elaborate standards, but rather for the definition to embrace the idea that recovery is not all-or-nothing (Hernberg); rather, recovery is an ongoing, never-ending process that is portrayed through the patient making steps in the right direction, slowly taking control away from anorexia (Kenny et al.; Klump et al.).

We put forward the idea that for treatment to be more successful, however, the definition of recovery must be individualized for each patient. With the help of professionals, patients immensely benefit from establishing goals to work towards, breaking down the long-term goal of full recovery into more attainable steps (Roloff). All patients who suffer from anorexia have their own primary urges, behaviors, and thoughts that cannot be categorized into short definitions explaining recovery (Crow et al.). The definition of recovery for research purposes has the intent of being used for finding participants for certain medical studies, but applying this definition universally across treatment centers for anorexia is not useful; it would lead to patients being discharged without having the roots of their eating disorder addressed because a model of treatment that works for a few people is not guaranteed to work for everyone. In her podcast, *We're All Insane*, Devorah notes the importance of examining the initiating causes of the eating disorder. If an individual develops anorexia out of a desire for control, in recovery, that sense of control needs to be manifested in another area, not disregarded. Without this focus in treatment, with the internal need for control not acknowledged and addressed, it is likely that when the individual has the drive for control again, anorexia may reappear, for that is all the individual knows of how to take control (Frank et al., 2013).

Recovery looks different for all depending on how long the eating disorder has been present and the specific behaviors that were engaged in, as well as how much and what types of foods were eaten (Kenny et al.). Patients with anorexia even voice that with some eating disorder behaviors, other people who interact with the patient daily are not even aware that certain behaviors connect to the anorexia. Bettina voices that there is a difference between looking at a menu, picking the lower calorie option, and claiming that it is her taste preference versus ordering the meal she would rather eat. (Hernberg). Yet, the emergence of the eating disorder in more subtle ways, similar to this example, can be overlooked by family members, friends, and even treatment teams. As a result, patients are the ones who need to construct their own goals for recovery based on their experiences. Doctors and clinicians may be unaware of the thoughts and behaviors that contributed to the eating disorder. The return to a life free from the eating disorder cannot occur without the uprooting and weakening of these factors, which holds a huge role in recovery and is unique for every patient (Klump et al.).

Compliance

Most patients with anorexia put into treatment were done so against their will. Out of concern, family members or friends bring the individual struggling with anorexia to a doctor, and the doctor is the one referring them to a treatment facility (Kenny et al.). The truth for many people with eating disorders is that they will never believe that they are sick enough or require medical care (Roloff). As the eating disorder continues to accumulate power, the brain loses its ability for rational thought, and individuals do not nearly see the problem as intensely as the reality of it is (Lozano-Serra et al.). Given that many patients are forced into treatment, they struggle to adhere to the goals of treatment and to progress through the challenges their anorexia poses (Bezance & Holliday; Hernberg). Compliance rates with treatment are on the lower end. Research suggests ½ of patients recover with treatment, ⅓ of patients improve, and the remaining stay ill (“What Is Anorexia: Symptoms, Complications and Causes.”). In order to advance through recovery, patients must abide by the goals of their treatment (Hernberg; Paintain).

The aspect of treatment where there is usually the most resistance is with medications (Crow et al.). The eating disorder influences how patients see themselves: as tied to the eating disorder, only acting in a way that aligns with the personality of the eating disorder (Klump et al.). Even if a patient wants to recover, there is a high tendency that they will struggle to agree to take a medication that ultimately reduces the symptoms of the eating disorder because it goes against the factors fueling the anorexia. Anorexia is prone to be ego-syntonic, so patients are more likely to exhibit behaviors, values, and feelings that are in congruence with their ideal self-image. People with anorexia are in constant conflict over whether or not to support or abandon the eating disorder. Agreeing to take a medication that weakens the symptoms of the eating disorder is completely against the rules set for themselves, and the individual may feel the need to punish themselves or continue to struggle in another area to compensate for feeling a lack of control (Crow et al.).

Relating to this idea, when one harmful coping mechanism is taken away, another replacement behavior, which is often a sabotaging one, will emerge. Even after recovering from anorexia, one study illustrated that most patients engage in obsessional behaviors, have inflexible thinking, and over-control their emotions and impulses (Holtkamp et al.). Additionally, patients in recovery have “increased perfectionism, and their most common obsessional target symptoms were the need for symmetry and ordering/arranging” (Kaye et al.). Likewise, Devorah mentions that “the beginning of recovery was even worse than just having the eating disorder” (Roloff). Hearing that taking away one coping mechanism only leads to the development of others can be demotivating in recovery; once a patient has initiated the beginnings of recovery, that does not mean that things will just start to improve (Bezance & Holliday). After one barrier is broken, and an unhealthy habit is destroyed, there are still more obstacles to work through, which may reinforce the idea that recovery is a never-ending journey (Holtkamp et al.). The harsh reality of treatment, where there are always battles to fight through, can greatly discourage an individual from committing to recovery because it is an arduous, exhaustive process, and if a patient begins

to give up halfway through, they run the risk of sinking back to where they started or worse (Bezance & Holliday).

Another factor that influences willingness for engagement in treatment is the atmosphere that the patients are in (Roloff). Unfortunately, a large number of patients describe the treatment setting to be negative, and combining that with a patient's low amounts of motivation to recover, the individual is not set up to succeed (Paintain). One of the biggest critiques from patients who previously had anorexia revolves around conformity. When one motivated patient is surrounded by five patients who are refusing to cooperate in treatment and are not making any efforts towards recovery, it is very likely that the motivated individual will lose their desire to succeed in treatment (Bezance & Holliday). It can be seen as unfair to have to meet certain expectations when nobody else in the room is meeting them or appears motivated. Similarly, when patients spend the majority or all of their days in a treatment center, the idea of a normal life becomes more imaginative and obscure. Patients must put a pause on the development of social relationships, family connections, and fun activities, arguably all of which are aspects that make an individual who they are. Without the sprinkling of normal experiences into treatment, motivation to participate in treatment can decrease, for there are not any moments of joy that make the patient believe that recovery is worth the effort (Hernberg). Patients struggle to engage in treatment when there are no reinforcers and punishments (Bezance & Holliday).

New Relationship with Exercise

The majority of anorexia treatment facilities do not include care or concerns for exercise reincorporation in the patient's life. Institutions tend to focus on the more vital aspects of recovery that can determine between life and death (Hernberg; Nahman & Holland;; Paintain). Correspondingly, the primary concern is weight gain and getting the patient to eat enough solid food (Kenny et al.). Treatment centers tend to view exercise as a secondary concern when it comes to initial recovery, leaving it up to other professionals and treatment team members who help the patient progress through later levels of recovery (Roloff). However, even at the most intensive levels of care, it is crucial to acknowledge unhealthy exercise when present in a patient's eating disorder. At a time when alternative, healthy coping strategies are being learned to replace destructive eating disorder behaviors, leaving exercise out of the equation can lead to more harm when the individual is out of the treatment facility (Nahman & Holland). Thus, it is important to intervene on unhealthy exercise behaviors and patterns while in treatment as a way to establish a healthier relationship with the body and allow for personalized treatment.

One argument against addressing exercise in treatment is that it "increases staff workload" (Dittmer et al.). However, patients' needs should not be neglected due to the lack of knowledge that staff may have around exercise or because it requires more effort from the staff. The benefits of integrating exercise in treatment include preventing relapse (Nahman & Holland). Consequently, the system as a whole is responsible for this issue, and the patients should not have to receive inadequate treatment due to an institutional flaw.

The standard way for treatment centers to tackle the challenge of exercise as it relates to anorexia is by banning the patient from exercising and allowing them to return to movement once they are ready to discharge from the program (Cook et al.). Typically, patients are in distress when unable to exercise, for it has been a compulsive habit that their life had revolved around prior to treatment (Dittmer et al.). Nonetheless, it is not as though the patient is stress free while in treatment, and it is probable that distress is already experienced since patients are not allowed to engage in other eating disorder behaviors, such as restriction, purging, and bingeing (Holtkamp et al.). The negative emotions that stem from not being able to exercise derive from the eating disorder's control of the brain and rational mind (Frank et al.). In order for the eating disorder to be treated completely, it has to lose all control. The individual has to sit with the discomfort of not having the ability to exercise, or else, the patient runs the risk of the eating disorder maintaining control over that area of the brain pushing for compulsive exercise despite the eating disorder giving up other forms of control (Nahman & Holland).

Effective treatment includes learning healthy coping mechanisms, meaning there does not need to be one clear-cut answer as to what coping mechanisms should be turned to instead of compulsively exercising (Kenny et al.). There is a middle path for treatment centers to follow that does not require a specialized field of exercise treatment and does not have to cover all of the details surrounding compulsive exercise, but the simple task of discussing the aspect of exercise in a patient's life is a step in the right direction towards learning how to combat compulsions (Nahman & Holland). Discussions with a professional regarding thoughts and emotions around exercise provides an opportunity for the patient to realize when exercise becomes compulsive, driven by the eating disorder, and how to have control over urges to exercise (Cook et al.).

Fortunately, it might be easier than treatment centers realize to address compulsive exercise. Fostering discussions is the first step, and next, the question usually arises as to what patients are to do when they have the urge to compulsively exercise. The useful tool for treatment centers is that there are not specific solutions that relate to exercise urges specifically, but any healthy coping mechanism that professionals at a treatment center teach to patients can be used to replace the actions under the umbrella of unhealthy exercise (Hernberg; Nahman & Holland). All patients in treatment for eating disorders have unhealthy coping mechanisms (e.g., restricting as a way to control emotions), so throughout treatment, patients are taught many coping mechanisms that combat any eating disorder behaviors, including excessive exercise (Kenny et al.). For that reason, it does not require much additional effort by eating disorder professionals to address exercise within higher levels of care that gravitate towards leaving this aspect of the eating disorder unaddressed. In the end, for patients who engage in compulsive exercise, relapse is not as worrisome since this aspect of the eating disorder would be dealt with during treatment (Dittmer et al.).

On top of securing a higher success rate through recovery for patients where exercise is dealt with throughout treatment, there is a positive correlation between individualized treatment and compliance from the patient (Bezance & Holliday; Crow et al.). When unhealthy exercise

patterns are addressed by clinical professionals, treatment becomes individualized. Patients with eating disorders are more likely to comply with psychotherapy when the patient's feelings and thoughts behind the eating disorder are heard and valued (Kenny et al.). For a patient who is banned from movement while in treatment, having a therapist to talk to and create goals surrounding reincorporating movement in the patient's life can be internally motivating, increasing the likelihood of the patient actively participating in recovery (Dittmer et al.).

Conclusion

Information from this review highlights the neurobiological and psychological complexities of anorexia nervosa, portraying the multitude of symptoms and effects this mental disorder has on individuals ("Anorexia Nervosa"). When the brain is starved, neural pathways are altered, which reinforces the eating disorder and triggers certain trait tendencies (Frank et al., 2013; Klump et al.). Every case of anorexia presents itself differently, and attempting to treat this mental illness using the same standards and techniques on all patients does not work (Bezance & Holiday; "What Is Anorexia: Symptoms, Complications and Causes."). The modern day treatment system does not cover every layer that comes along with the eating disorder as it develops, failing patients and their families as the eating disorder is not always weakened through treatment (Kenny et al.; Nahman & Holland). In the future, treatment needs to be individualized in order to support better results throughout recovery. Individualized treatment looks different for everyone and requires staff that is both knowledgeable and willing to listen to each patient (Hernberg; Nahman & Holland). When core aspects of the eating disorder are left unaddressed, and the more common symptoms are the ones corrected, the remaining behaviors and thought patterns continue to impede on an individual's quality of life (Frank et al., 2013; Kaye et al.). Additional research and training for clinicians specializing in anorexia is necessary in order to improve recovery outcomes.

Works Cited

- “Anorexia Nervosa.” *National Eating Disorders Association*, 28 Feb. 2018, www.nationaleatingdisorders.org/learn/by-eating-disorder/anorexia.
- “Anorexia Statistics: Gender, Race & Socioeconomics.” *Bulimia.Com*, <https://bulimia.com/anorexia/statistics/>
- Archer, Trevor. ‘Exercise in Anorexia Nervosa: Complexity of Pathology and Health’. *Clinical and Experimental Psychology*, vol. 2, no. 3, OMICS Publishing Group, 2016, <https://doi.org/10.4172/2471-2701.1000e107>
- Bezance, Jessica, and Joanna Holliday. ‘Adolescents with Anorexia Nervosa Have Their Say: A Review of Qualitative Studies on Treatment and Recovery from Anorexia Nervosa’. *European Eating Disorders Review: The Journal of the Eating Disorders Association*, vol. 21, no. 5, Wiley, Sept. 2013, pp. 352–360, <https://doi.org/10.1002/erv.2239>
- Coniglio, Kathryn A., et al. ‘Behavioral Reinforcement of Pathological Exercise in Anorexia Nervosa’. *The International Journal of Eating Disorders*, vol. 55, no. 2, Wiley, Feb. 2022, pp. 184–192, <https://doi.org/10.1002/eat.23626>
- Cook, Brian J., et al. ‘Exercise in Eating Disorders Treatment: Systematic Review and Proposal of Guidelines’. *Medicine and Science in Sports and Exercise*, vol. 48, no. 7, Ovid Technologies (Wolters Kluwer Health), July 2016, pp. 1408–1414, <https://doi.org/10.1249/mss.0000000000000912>
- Cresswell, Camilla, et al. ‘The Role of Compulsive Exercise in the Relationship between Perfectionism and Eating Disorder Pathology in Underweight Adolescents with Eating Disorders’. *Eating Behaviors*, vol. 47, no. 101683, Elsevier BV, Dec. 2022, p. 101683, <https://doi.org/10.1016/j.eatbeh.2022.101683>
- Crow, Scott J., et al. ‘What Potential Role Is There for Medication Treatment in Anorexia Nervosa?’ *The International Journal of Eating Disorders*, vol. 42, no. 1, Wiley, Jan. 2009, pp. 1–8, <https://doi.org/10.1002/eat.20576>
- Dittmer, Nina, et al. ‘Compulsive Exercise in Eating Disorders: Proposal for a Definition and a Clinical Assessment’. *Journal of Eating Disorders*, vol. 6, no. 1, Springer Science and Business Media LLC, Nov. 2018, p. 42, <https://doi.org/10.1186/s40337-018-0219-x>
- “Dopamine: What It Is, Function & Symptoms.” *Cleveland Clinic*, <https://my.clevelandclinic.org/health/articles/22581-dopamine>.
- Frank, Guido K. W. ‘Altered Brain Reward Circuits in Eating Disorders: Chicken or Egg?’ *Current Psychiatry Reports*, vol. 15, no. 10, Springer Science and Business Media LLC, Oct. 2013, p. 396, <https://doi.org/10.1007/s11920-013-0396-x>
- Frank, Guido, et al. ‘Motivation to Eat and Not to Eat - The Psycho-Biological Conflict in Anorexia Nervosa’. *Physiology & Behavior*, vol. 206, Elsevier BV, July 2019, pp. 185–190, <https://doi.org/10.1016/j.physbeh.2019.04.007>

- Garrison, Jane, et al. ‘Prediction Error in Reinforcement Learning: A Meta-Analysis of Neuroimaging Studies’. *Neuroscience and Biobehavioral Reviews*, vol. 37, no. 7, Elsevier BV, Aug. 2013, pp. 1297–1310, <https://doi.org/10.1016/j.neubiorev.2013.03.023>
- Hernberg, Bettina, host. “My eating disorder story: Being diagnosed with anorexia at 15.” *Best of Bettina*, April 2020. *Spotify*, <https://open.spotify.com/show/2CSVbEFxw14s9kV7XCwmkb>
- Holcombe, Madeline. “Hospitalizations for Eating Disorders Grew in the Pandemic. the Problem Isn’t over, Experts Say.” *CNN*, Cable News Network, 7 Nov. 2022, www.cnn.com/2022/11/07/health/eating-disorders-pandemic-wellness/index.html.
- Holtkamp, K., et al. ‘Depression, Anxiety, and Obsessionality in Long-Term Recovered Patients with Adolescent-Onset Anorexia Nervosa’. *European Child & Adolescent Psychiatry*, vol. 14, no. 2, Springer Science and Business Media LLC, Mar. 2005, pp. 106–110, <https://doi.org/10.1007/s00787-005-0431-5>
- Humphries, Mark. ‘Why Does the Brain Have a Reward Prediction Error? - The Spike - Medium’. *The Spike*, 11 Feb. 2019 <https://medium.com/the-spike/why-does-the-brain-have-a-reward-prediction-error-6d52773bd9e7>.
- Kaye, W., et al. ‘Serotonin Neuronal Function and Selective Serotonin Reuptake Inhibitor Treatment in Anorexia and Bulimia Nervosa’. *Biological Psychiatry*, vol. 44, no. 9, Elsevier BV, Nov. 1998, pp. 825–838, [https://doi.org/10.1016/s0006-3223\(98\)00195-4](https://doi.org/10.1016/s0006-3223(98)00195-4)
- Kenny, Therese E., et al. ‘Lived Experience Perspectives on a Definition of Eating Disorder Recovery in a Sample of Predominantly White Women: A Mixed Method Study’. *Journal of Eating Disorders*, vol. 10, no. 1, Oct. 2022, p. 149, <https://doi.org/10.1186/s40337-022-00670-2>
- Lozano-Serra, Estefanía, et al. ‘Adolescent Anorexia Nervosa: Cognitive Performance after Weight Recovery’. *Journal of Psychosomatic Research*, vol. 76, no. 1, Elsevier BV, Jan. 2014, pp. 6–1, <https://doi.org/10.1016/j.jpsychores.2013.10.009>
- Markett, Sebastian, et al. ‘Anxiety and Harm Avoidance’. *Neuroimaging Personality, Social Cognition, and Character*, Elsevier, 2016, pp. 91–112, <https://doi.org/10.1016/b978-0-12-800935-2.00005-1>
- Marusak, H. A., et al. ‘You Say “Prefrontal Cortex” and I Say “Anterior Cingulate”’: Meta-Analysis of Spatial Overlap in Amygdala-to-Prefrontal Connectivity and Internalizing Symptomology’. *Translational Psychiatry*, vol. 6, no. 11, Springer Science and Business Media LLC, Nov. 2016, pp. e944–e944, <https://doi.org/10.1038/tp.2016.218>
- McGregor, Grace. “How Exercise Affects the Brain.” *Life Sciences*, Life Sciences, 28 Jan. 2022, <https://lifesciences.byu.edu/how-exercise-affects-your-brain>
- Paintain, Brooke, host. “My anorexia story: tw.” *Positively Me*, 31 Jan. 2023. *Spotify*, <https://open.spotify.com/episode/3I6etgOdIF6GyA7PhKeXnM>
- “Perfectionism.” *Psychology Today*, www.psychologytoday.com/us/basics/perfectionism.
- Seeger, Gert, et al. ‘Body Image Distortion Reveals Amygdala Activation in Patients with Anorexia Nervosa -- a Functional Magnetic Resonance Imaging Study’. *Neuroscience*

Letters, vol. 326, no. 1, Elsevier BV, June 2002, pp. 25–28, [https://doi.org/10.1016/S0304-3940\(02\)00312-9](https://doi.org/10.1016/S0304-3940(02)00312-9)

“2-Minute Neuroscience: Dopamine.” Youtube, uploaded by Neuroscientifically Challenged, 27 Apr. 2018, https://www.youtube.com/watch?v=Wa8_nLwQIpg.

Roloff, Devorah, host. “Surviving Anorexia.” *We’re All Insane*, 27 Mar. 2023. *Spotify* <https://open.spotify.com/episode/0dfz1mw2ydjTZtYIuia7b>

“What Is Anorexia: Symptoms, Complications and Causes.” *Eating Disorder Hope*, 29 July 2022, www.eatingdisorderhope.com/information/anorexia.

The Efficacy of HPV Vaccines and a Healthy Microbiome on the Development of Cervical Cancer By Sarah Wang

Abstract

Specific strains of HPV frequently result in the development of cancer, such as cervical cancer, vaginal cancer, oropharyngeal cancer, and anal cancer. Despite technological advances, however, cervical cancer is still one of the leading causes of cancer-related deaths for women worldwide, particularly in developing countries. Here, we review different studies that discuss the efficacy of the bivalent, quadrivalent, and nonavalent HPV vaccines considering both sexes and how successfully they protect against the development of cervical cancer. We also discuss different factors that may significantly impact the efficacy of HPV vaccines and cancer screening success, such as the vaginal microbiome, pre-existing health conditions, and geographical location. Although we hypothesize that all three of the most used vaccines are effective against HPV in both males and females, we found that only the quadrivalent and nonavalent vaccines prevent HPV and its subsequent cancers in both sexes. At the same time, the bivalent is only effective in females. In addition, we found that after HPV infection, a significant factor contributing to whether cervical cancer develops is the presence of healthy versus unhealthy vaginal bacteria.

Introduction

Cervical cancer is the second most common cancer in women worldwide, the third most deadly cancer in women, and the most common malignancy in developing countries.¹⁻³ One of the major causes of cervical cancer is the Human Papillomavirus (HPV). HPV will affect over 90% of sexually active men and 80% of sexually active women in their lifetime (Minnesota Department of Health). Consequently, because of the prevalence of HPV, an effective treatment must be developed to prevent the spread. In this section, we look at the bivalent, quadrivalent, and nonavalent vaccines, which are the three most widely used. We will also go over their respective efficacy in preventing the spread of HPV and the subsequent development of cervical malignancies. In addition, other factors may aid vaccines in preventing the development of HPV and cervical cancer, such as vaginal microbiomes, the use of proper protection during intercourse, and avoiding harmful habits like smoking (CDC). It is hypothesized that all current HPV vaccines will effectively prevent the development of cervical cancer in females and other HPV-related cancers in both sexes. We also wanted to know if the vaginal microbiome contributed to the development of cervical cancer, an area not yet commonly explored.

However, despite having such dire consequences, current HPV vaccines only protect against specific types of HPV, which are the most common ones that cause cancer (Muñoz et al.). A vaccine like the one developed by the GlaxoSmithKline (GSK) clinical trials that specifically and successfully protects against HPV-16 and HPV 18, the two most common cancer-causing strains of HPV, could prevent 71% of cervical cancer cases worldwide. More importantly,

HPV-16 and HPV-18 are found to cause different types of cervical cancer; HPV-16 is responsible for squamous cell carcinoma, while HPV-18 is primarily responsible for adenocarcinoma. It is also important to point out that HPV 16 causes cervical cancer at a much higher percentage than HPV 18, having a 57.4% and 16.6% cancer rate respectively.

HPV and Cervical Cancer Identification

Cervical cancer can be fatal, and therefore it is essential that there is an efficient screening method. Current screening methods require much effort and cost from both the patient and the hospital, which makes it difficult to access in developing countries (Brown et al.). In fact, 85% of all cervical cancer cases occur in developing countries (Parkin). Because screening programs can work to reduce cervical cancer rates by 80%, other HPV cancers are more prevalent in developed countries, and a higher ratio of these cancers subsequently occur in men (Peto et al.). The most common type of screening involves a Papanicolaou smear (Pap smear) and subsequent testing of HPV DNA and cytology (the exam of a single cell type). Scientists have been working to develop new ways of screening for cervical cancer, including more efficient tests that can catch cancer growth early, which could contribute to better patient outcomes. One method of early detection that scientists have been working on is the discovery of potential biomarkers for more effective cervical cancer diagnosis (Huo et al.). To do this, they obtained and analyzed differentially expressed prognosis-related genes and eventually came up with a six-gene signature that can be used to predict the overall survival of a cervical cancer patient. This may be helpful for decision-making and individual treatment, but not enough research has been done for this to be put into practice. It is also important to note that, unless this technology can become much more accessible, developing countries, particularly in Africa, arguably the most impacted by cervical cancer, will still struggle to obtain affordable screening technology (Singh). Not only do these countries lack the necessary screening technology, but they also have limited HPV vaccination programs, which makes HPV and cervical cancer rates relatively high in these regions (Amponsah-Dacosta et al.).

Pap smears and current screening options are not feasible in developing countries due to their high cost, low sensitivity, and subsequent need for laboratory and necessary expertise (Cronjé). Thus, it is important for screening in developing countries to be affordable, accessible, and usable in remote locations. Thus, because of the shortcomings of the Pap smear, alternative screening methods have been considered, such as cervicography, acetic acid test, speculoscopy, HPV DNA identification, and colonoscopy. Of these methods, cervicography, HPV DNA identification, and colonoscopy all have high costs and little accessibility in developing countries due to their need for experienced medical personnel and underdeveloped technology. While a colonoscopy does provide immediate results, it would be used to detect other cancers caused by HPV, such as anal cancer but not cervical cancer itself (Forbes et al.). The acetic acid test and speculoscopy are similar in function and result, and advantages to both include their simplicity, immediate results, and affordability. However, one disadvantage to these screenings is that they have low specificity and often produce false positives. As such, no current option is accurate,

developed, cost-effective, or plausible in developing countries. This highlights the need for more research in implementing screening methods for cervical cancer and HPV in developing countries. At this point, HPV DNA identification used in conjunction with other methods seems to be the most promising if the tools are available.

Effectiveness of HPV Vaccine in Preventing the Transmission of HPV

HPV is a viral disease that often clears in most infected sexually active individuals after two years with limited clinical disease (Stanley). Currently, three types of HPV vaccines are widely used, each targeting specific strains of HPV (St. Laurent et al.). The three vaccines are bivalent (16, 18), quadrivalent (6, 11, 16, 18), and nonavalent (6, 11, 16, 18, 31, 33, 45, 52, 58). HPV vaccines are highly effective in preventing HPV and HPV-related diseases, especially when administered to prepubertal females and males (Kechagias et al.). However, the efficacy of the HPV virus substantially decreases after age 26. Despite using the vaccine, some populations are still more at risk for HPV than others. Among these include women undergoing surgical procedures to treat cervical intraepithelial neoplasia (CIN) or abnormally growing cells in the cervix. These women would benefit from additional vaccination applied after treatment to reduce the risk of cervical cancer.

CIN consists of different grades, with CIN 1 being the least severe and CIN 3 being the most severe. Thus, women who develop an exceptionally high grade of CIN become more susceptible to HPV infections and can quickly reacquire infections after surgical procedures. This also means that these women have an exceptionally high risk of repeatedly developing CIN and other HPV-related malignancies.

Previous studies have also shown that the HPV vaccine produces a much larger immune response than the infection itself. The natural immunity against HPV wanes over time, so the HPV vaccines have been seen to protect against reactivation and reinfection against HPV and HPV-related malignancies for individuals who have previously tested positive for a currently cleared infection. Thus, the HPV vaccines have benefits for both new infections and reinfections of their respective HPV subtypes.

Effectiveness of HPV Vaccine in Preventing the Transmission of Cervical Cancer

It has become clear that HPV is linked to at least five different types of malignancies concerning epithelial cells: vulvar, vaginal, anal, penile, oropharyngeal, and cervical cancer (St. Laurent et al.). Among these, cervical cancer is the most severe in women. Of the three widely accepted vaccines currently available (bivalent, quadrivalent, and nonavalent), the bivalent and quadrivalent vaccines have been available for the longest and, thus, have undergone the most rigorous testing.

The bivalent vaccine was tested on 18,000 women using a placebo vaccine as a control (St. Laurent et al.). This vaccine was 92.9% effective in protecting 15–25-year-old women with either no history of malignancies or normal cervical cytology from CIN 2. It was 52.8% effective in preventing CIN 2 lesions often associated with HPV 16 and HPV 18 in women with a history

of previous HPV infection. The same vaccine is only 33.6% effective in preventing CIN 3 lesions in patients with a history of infection. Thus, the bivalent vaccine effectively prevents less harmful/life-threatening conditions, but efficacy decreases with severity and recurrence.

The same study tested the quadrivalent vaccine in a placebo-controlled experiment with over 12,000 women aged 15-26 with abnormal Pap smear results and less than four sexual partners in their lifetime. They found the vaccine was 98% effective against a mixture of CIN 2/3, HPV 16/18, and adenocarcinoma in situ (AIS). Specifically, the vaccine was 42% effective against HPV 16 and 79% effective against HPV 18. However, the quadrivalent vaccine only prevents 17% of all CIN 2 lesions and does not demonstrate high levels of cross-protection (32.5% against CIN2/3 and AIS).

Finally, the newest vaccine, the nonavalent vaccine, which protects against nine HPV strains, was evaluated in a double-blind, non-inferiority trial (St. Laurent et al.). A non-inferiority trial ensures that a new treatment is at least as effective as a currently used treatment, such as the quadrivalent vaccine. In this study, they looked at 14,215 low-risk women aged 16-26 that were either given: the quadrivalent vaccine or the new experimental nonavalent vaccine. Study participants were given three randomized doses of either the nonavalent or quadrivalent vaccine. It turns out that the new nonavalent vaccine is 96% effective at preventing high-grade malignancies such as CIN and AIS. The nonavalent vaccine is also 96% effective when preventing recurring infections related to the same HPV types. There was, however, no difference between the quadrivalent and nonavalent vaccines when it came to those with prior HPV infections. This shows that the nonavalent vaccine is not inferior to the existing vaccines and therefore is an effective new treatment.

Based on the trials above, it can be seen that the three current vaccines are efficient in protecting their specific subtypes. However, individuals must be vaccinated before having their first sexual experience to maximize the effects of the vaccine and limit the possibility that they have already been infected with HPV.

Impact of HPV Vaccine in People with Current Malignancies

The above only discussed prevention of cervical cancer through vaccine treatments. However, while prevention is a necessary precaution, treating cervical cancer once contracted is just as important. Thus, it can be helpful to look at the usage and efficacy of the HPV vaccine in individuals currently diagnosed with cervical cancer. It is also essential to consider immunocompromised individuals or those with current malignancies because they are at a higher risk of being infected by HPV and developing subsequent HPV-related cancers, including cervical cancer (Palefsky et al.).

Individuals currently affected with HIV have a notably higher chance of developing cervical cancer from HPV (Palefsky et al.). This is because HPV infection is not only incited from sexual transmission but also a reactivation of previously acquired infections. A study done on HIV-infected women shows that HPV-16, the strain responsible for over half of HPV-induced cancers, in particular, is the least responsive to any HIV medication taken that may help suppress

other HPV strains. HIV is not the only pre-existing infection that increases the development of HPV and cervical cancer. In fact, it is also known that individuals who have recently undergone organ transplants or invasive surgery are at a higher risk for HPV-induced cancers and CIN when compared to the general population. CIN is also more prevalent in chronic, taxing conditions such as renal failure and lupus erythematosus.

However, HPV vaccines may perform worse on immunocompromised individuals, as their body does not respond normally to the vaccine. Nevertheless, there is still good reason to believe the HPV vaccine will work on immunocompromised patients because HPV is highly immunogenic, HPV-positive women can mount a humoral immune response to HPV antigens. However, HIV is a highly complex autoimmune disorder, and more testing would be needed to see if there is sufficient response to HIV-positive patients, which is difficult to test because there is a thin, invisible line between the vaccine functioning properly and overtaking an individual's immune system. Because of this, it is important to find a way to test the efficacy of the HPV vaccine on individuals with pre-existing conditions.

Here, we look at a study where 29 patients 18 years or older with either confirmed stage Ib or IIa cervical cancer were given two vaccinations with TA-HPV four to twelve weeks apart (Kaufmann et al.). The first vaccine injection was administered at least two weeks before the scheduled radical hysterectomy surgery, while the second was given four to eight weeks after surgery. The test subjects also must have had a life expectancy of over three months. It has also been noticed that oncoproteins E6 and E7 can always be expressed in tumor cells because they are fundamentally involved in the process of carcinogenesis by inhibiting apoptosis in cells. These proteins may provide a tumor-specific target for vaccines or treatment of cervical cancer in patients.

After vaccinating patients with progressive cervical cancer with the TA-HPV vaccine, it can be surmised that a measurable immune response could be produced (Kaufmann et al.). In fact, a single vaccination produced responses in 12 of the 29 tested individuals. This study has confirmed that vaccination with the TA-HPV vaccine outside of the hospital is ultimately a viable option for treating women with active cervical cancer.

Vaginal Microbiomes and Their Relationship With HPV

Many factors contribute to the development of cervical cancer. One such factor includes the vaginal microbiomes uniquely present in each individual (Xu et al.). Previous studies have shown that the *Prevotella* species is often associated with inflammation of the cervix and vagina, causing an imbalance in the microbiota. The *Lactobacillus gasseri* species tend to correlate with a healthy vaginal microbiome with decreased rates of HPV infection and cervical cancer (Shannon et al.). Thus, it is interesting to see the relationship specific vaginal microbiomes have with an individual's health.

It is important to note that the state of the vaginal microbiome is constantly changing, but a balance must remain to maintain vaginal health. An abundance of the *Lactobacillus* species is most often associated with good gynecological health because it secretes lactic acid to compete

with pathogens for space on the epithelial wall (Shannon et al.). The makeup of the vaginal microbiome may vary depending on various external factors such as race, diet, and geological location, as well as internal factors such as the host's immune system or hormone levels. Further, not all types of *Lactobacillus* may be beneficial. *L. crispatus* seems to be the most beneficial and protective, while *L. iners* is often associated with diseases such as HPV and HIV.

Suppose the balance of the cervical and vaginal microbiota is disrupted. In that case, increased amounts of pro-inflammatory cytokines may be produced as part of the immune response, resulting in increased inflammation in the cervix (Shannon et al.). This also disrupts the immune regulation in these areas of the female reproductive tract, thus creating an appropriate environment for a tumor to develop more easily. Microbial infection can increase the rate of HPV reproduction while also increasing inflammation and damaging epithelial cells and eventually resulting in CIN. These can both lead to cervical cancer.

Interestingly, this study mentions that HPV alone is not enough to lead to the development of cervical cancer (Shannon et al.). Instead, HPV, vaginal bacteriosis and microbiota dysregulation all influence each other in ways that can eventually lead to the development of cervical cancer. For example, inflammation caused by an unbalanced microbiota can lead to increased inflammation and increased vulnerability to HPV. Thus, although important, other factors impact the health of the female reproductive tract that goes beyond HPV.

Variable Effects of HPV Vaccines

Infection of HPV and other diseases caused by HPV in males are also common. Contrary to my hypothesis, of the three main HPV vaccines, only the quadrivalent and nonavalent vaccines are recommended for both males and females (Handler et al.). The quadrivalent vaccine, an HPV vaccine active against HPV types 6, 11, 16, and 18, was tested in a study on males (Giuliano et al.). It is important to pay attention and vaccinate against HPV in males because it also causes cancers related explicitly to males, such as cancers of the penis and cancers of the anus. As most other HPV vaccine experiments discussed, this experiment was also done with a group of 2,033 males who received a placebo vaccine and a group of 2,032 who received the proper quadrivalent vaccine. These groups were carefully balanced regarding race, ethnicity, sexual history, smoking history, and age.

Eventually, it was proven that the quadrivalent is effective in males (Giuliano et al.). Within one month after the third dose of the quadrivalent vaccine, antibodies targeting HPV strains 6, 11, 16, and 18 were present in at least 97.4% of vaccinated subjects. Only 0.9% of the vaccinated group did not have seroconversion to any of the targeted HPV strains, while 15% only had seroconversion to three of the four strains. The vaccine was also effective against internal and external genital lesions that resulted due to HPV 6, 11, 16, and 18 while significantly decreasing their persistent infection rates. Thus, from this study, it can be concluded that the quadrivalent vaccine is effective in men.

Another study found the vaccine to be moderately effective against genital HPV infections, and higher efficacy rates were noted in males who were naïve for their respective

tested HPV strains (Harder et al.). It is important to note, however, that compared to females, substantially less research has been done on HPV in males and more is needed to draw an indisputable conclusion.

Conclusion

In summary, the HPV vaccine, in general, is an effective method in preventing the occurrence of cervical cancer. Of course, not all of the over 200 HPV strains can cause cancer, but the ones that do cause cancer are responsible for 90% of all anal and cervical cancer cases (CDC). The three most common HPV vaccines all protect against HPV 16 and HPV 18, the two HPV strains responsible for the most cervical cancer cases (Muñoz et al.). For optimal effect, the first dose of these vaccines is generally recommended at around 12 years of age, and the second dose should be given no later than 26 years of age (CDC). Also, studies have shown that the bivalent, quadrivalent, and nonavalent vaccines are all relatively effective in protecting against cervical cancer. However, of the three main HPV vaccines mentioned, only the quadrivalent and nonavalent vaccines effectively prevent HPV in both females and males. This is possible because males were not originally intended to be vaccinated against HPV, so older vaccines were not designed to protect them. With that being said, there are many factors beyond sexual history and lifestyle that can contribute to the development of HPV and subsequent cervical cancer.

One such factor is the vaginal microbiome, where different microbiomes in different geographical regions are often linked with positive or negative results. In fact, it has been seen that around 15-20% of all cancers, including cervical cancer, may be caused by microbial pathogens (Bhatt et al.). Generally, most *Lactobacillus* bacteria species are known to affect vaginal health and prevent infections positively. One or two of the four *Lactobacillus* species, *L. crispatus*, *L. gasseri*, *L. jensenii*, and *L. iners*, predominantly dominate the vaginal microbiome. Of these, *L. crispatus*, *L. gasseri*, and *L. jensenii* are often associated with maintaining a healthy, non-inflamed environment, while *L. iners* is responsible for causing inflammation that can eventually lead to cancer. Ultimately, a healthy vaginal microbiome is also important to prevent cervical cancer development after HPV infection.

A healthy vaginal microbiome can provide substantial health benefits, including its impact on cervical cancer prevention. To prevent the development or disturbance of bacteria in the first place, things like safe sex, regulation of hormones like estrogen, and decreased stress and smoking can be practiced (Barrientos-Durán et al.). That being said, the many contributing factors to a healthy vaginal microbiome are still being investigated, and more research needs to be done (Reid). In cases where the balance of the vaginal microbiome has been disrupted, probiotics containing beneficial strains of *Lactobacillus* can be taken to help restore the balance (Joseph et al.).

Works Cited

- Amponsah-Dacosta, Edina, et al. "Human Papillomavirus Vaccination in South Africa: Programmatic Challenges and Opportunities for Integration With Other Adolescent Health Services?" *Frontiers in Public Health*, vol. 10, Jan. 2022, p. 799984. DOI.org (Crossref), <https://doi.org/10.3389/fpubh.2022.799984>.
- Barrientos-Durán, Antonio, et al. "Reviewing the Composition of Vaginal Microbiota: Inclusion of Nutrition and Probiotic Factors in the Maintenance of Eubiosis." *Nutrients*, vol. 12, no. 2, Feb. 2020, p. 419. DOI.org (Crossref), <https://doi.org/10.3390/nu12020419>.
- Bhatt, Aadra P., et al. "The Role of the Microbiome in Cancer Development and Therapy." *CA: A Cancer Journal for Clinicians*, vol. 67, no. 4, July 2017, pp. 326–44. DOI.org (Crossref), <https://doi.org/10.3322/caac.21398>.
- Brown, Alaina J., and Cornelia L. Trimble. "New Technologies for Cervical Cancer Screening." *Best Practice & Research Clinical Obstetrics & Gynaecology*, vol. 26, no. 2, Apr. 2012, pp. 233–42. DOI.org (Crossref), <https://doi.org/10.1016/j.bpobgyn.2011.11.001>.
- CDC. "What Can I Do to Reduce My Risk of Cervical Cancer?" *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 21 Aug. 2023, www.cdc.gov/cancer/cervical/basic_info/prevention.htm.
- Cronjé, H. S. "Screening for Cervical Cancer in Developing Countries." *International Journal of Gynecology & Obstetrics*, vol. 84, no. 2, Feb. 2004, pp. 101–08. DOI.org (Crossref), <https://doi.org/10.1016/j.ijgo.2003.09.009>.
- Forbes, Clara M., et al. "Opportunistic Detection of Anal Intraepithelial Neoplasia at Colonoscopy." *JGH Open*, vol. 4, no. 6, Dec. 2020, pp. 1207–10. DOI.org (Crossref), <https://doi.org/10.1002/jgh3.12424>.
- Giuliano, Anna R., et al. "Efficacy of Quadrivalent HPV Vaccine against HPV Infection and Disease in Males." *New England Journal of Medicine*, vol. 364, no. 5, Feb. 2011, pp. 401–11. DOI.org (Crossref), <https://doi.org/10.1056/NEJMoa0909537>.
- Handler, Nancy S., et al. "Human Papillomavirus Vaccine Trials and Tribulations." *Journal of the American Academy of Dermatology*, vol. 73, no. 5, Nov. 2015, pp. 759–67. DOI.org (Crossref), <https://doi.org/10.1016/j.jaad.2015.05.041>.
- Harder, Thomas, et al. "Efficacy, Effectiveness and Safety of Vaccination against Human Papillomavirus in Males: A Systematic Review." *BMC Medicine*, vol. 16, no. 1, Dec. 2018, p. 110. DOI.org (Crossref), <https://doi.org/10.1186/s12916-018-1098-3>.
- Huo, Xiao, et al. "Identification of a Six-Gene Signature for Predicting the Overall Survival of Cervical Cancer Patients." *OncoTargets and Therapy*, vol. Volume 14, Feb. 2021, pp. 809–22. DOI.org (Crossref), <https://doi.org/10.2147/OTT.S276553>.
- Joseph, Rebecca Jane, et al. "Finding a Balance in the Vaginal Microbiome: How Do We Treat and Prevent the Occurrence of Bacterial Vaginosis?" *Antibiotics*, vol. 10, no. 6, June 2021, p. 719. DOI.org (Crossref), <https://doi.org/10.3390/antibiotics10060719>.

- Kaufmann, Andreas M., et al. *Safety and Immunogenicity of TA-HPV, a Recombinant Vaccinia Virus Expressing Modified Human Papillomavirus (HPV)-16 and HPV-18 E6 and E7 Genes, in Women with Progressive Cervical Cancer.*
- Kechagias, Konstantinos S., et al. “Role of Human Papillomavirus (HPV) Vaccination on HPV Infection and Recurrence of HPV Related Disease after Local Surgical Treatment: Systematic Review and Meta-Analysis.” *BMJ*, Aug. 2022, p. e070135. *DOI.org (Crossref)*, <https://doi.org/10.1136/bmj-2022-070135>.
- Minnesota Department of Health. “Quick Facts: HPV-Associated Cancer.” *Quick Facts: HPV-Associated Cancer - MN Dept. of Health*, 3 Oct. 2022, [www.health.state.mn.us/data/mcrs/data/qfhpv.html#:~:text=Human%20Papillomavirus%20\(HPV\)%20is%20a,with%20HPV%20in%20their%20lifetime](http://www.health.state.mn.us/data/mcrs/data/qfhpv.html#:~:text=Human%20Papillomavirus%20(HPV)%20is%20a,with%20HPV%20in%20their%20lifetime).
- Muñoz, Nubia, et al. “Against Which Human Papillomavirus Types Shall We Vaccinate and Screen? The International Perspective: HPV Types for Vaccines and Screening Tests.” *International Journal of Cancer*, vol. 111, no. 2, Aug. 2004, pp. 278–85. *DOI.org (Crossref)*, <https://doi.org/10.1002/ijc.20244>.
- Palefsky, Joel M., et al. “Chapter 16: HPV Vaccines in Immunocompromised Women and Men.” *Vaccine*, vol. 24, Aug. 2006, pp. S140–46. *DOI.org (Crossref)*, <https://doi.org/10.1016/j.vaccine.2006.05.120>.
- Parkin, Donald Maxwell. “The Global Health Burden of Infection-Associated Cancers in the Year 2002.” *International Journal of Cancer*, vol. 118, no. 12, June 2006, pp. 3030–44. *DOI.org (Crossref)*, <https://doi.org/10.1002/ijc.21731>.
- Peto, Julian, et al. *The Cervical Cancer Epidemic That Screening Has Prevented in the UK*. 2004.
- Reid, Gregor. “Therapeutic Opportunities in the Vaginal Microbiome.” *Microbiology Spectrum*, edited by Robert Allen Britton and Patrice D. Cani, vol. 5, no. 3, May 2017, p. 5.3.06. *DOI.org (Crossref)*, <https://doi.org/10.1128/microbiolspec.BAD-0001-2016>.
- Shannon, B., et al. “Association of HPV Infection and Clearance with Cervicovaginal Immunology and the Vaginal Microbiota.” *Mucosal Immunology*, vol. 10, no. 5, Sept. 2017, pp. 1310–19. *DOI.org (Crossref)*, <https://doi.org/10.1038/mi.2016.129>.
- Singh, Gopal. “Global Inequalities in Cervical Cancer Incidence and Mortality Are Linked to Deprivation, Low Socioeconomic Status, and Human Development.” *International Journal of MCH and AIDS (IJMA)*, vol. 1, no. 1, 2012. *DOI.org (Crossref)*, <https://doi.org/10.21106/ijma.12>.
- Stanley, Margaret. “HPV - Immune Response to Infection and Vaccination.” *Infectious Agents and Cancer*, vol. 5, no. 1, Dec. 2010, p. 19. *DOI.org (Crossref)*, <https://doi.org/10.1186/1750-9378-5-19>.
- St Laurent, Jessica et al. “HPV vaccination and the effects on rates of HPV-related cancers.” *Current problems in cancer* vol. 42,5 (2018): 493-506.
doi:10.1016/j.currproblcancer.2018.06.004
- Xu, Jinyun, et al. *Vaginal Microbiomes and Ovarian Cancer: A Review*.

The Chronicle of Social Movements: How Museums Influence and Reflect Social Movements in the 21st Century By Laura Wylie

Abstract

Museums may seem like they are timeless, unchanging presentations of history, but this is not so. They are not stagnant; they change as societies change. But how do museums change over time? How do significant social changes and movements affect how museums present historical information to the public, and how do museums influence the societies around them? In this paper, I will explore the intricate relationship between museums and societies—the role that museums play in society, and in turn, what role society’s evolution plays in museums. Museums are affected by the ideologies of their visitors and curators. When the ideologies of their visitors and curators change through generations, the museums must also evolve with them. Ideologies can be changed by influential events in society and evolving lifestyles. Thus, social changes and movements are reflected in the changes in history, science, and art museums. In this paper, I will address what museums are, and how social movements relate to museums. Then, I will explore how the American Museum of Natural History, California Academy of Sciences, Museum of Us, and Getty Center—institutions in some of the largest cultural centers in New York and California—relate to this topic. Through these explorations, I will evaluate how social movements and museums in America influence and reflect each other.

What are Museums?

In this first section, I will evaluate what museums are. Museums are educational displays of the human experience and knowledge that play a large role in the cultural expression of societies; museums are essentially a condensed reflection of our societies. The author Mackenzie Laminack describes in the *Journal of Museum Studies* that museums are “places of conversation with diverse dialogue, places that celebrate the richness of individual and collective experience, and places that are active and visible players in the civic life of their communities” (Laminack, 2015, 20). In “From Taking in to Reaching Out: How Collections and Collections Staff Are Being Used to Create a Community-Centered Museum,” she explores the pattern of how museums shift from being simply spaces for education to being places that also interact with and reflect society by “encourag[ing] critical thinking, inspir[ing] discussion, and chang[ing] paradigms, as well as educa[ing] about the outside world,” to create a meaningful impact on the communities around them. (Laminack, 2015, 20). However, to truly understand the value of museums that closely interact with society, we must first understand that most were not always so. Antonis Chaliakopoulos explores in “History of Museums: A Look at The Learning Institutions Through Time” how the settings and purposes of museums have evolved through history. Some of the first museums, in Rome, were collections kept in private residences that were open to the public. These museums were made in the eyes of the wealthy; they were not made to be representative of all (Chaliakopoulos, 2020). Thus, they were not interactive or reflective of the communities around them.

Museums get information about their exhibits from anthropologists, scientists, art historians, and researchers. Then, curators decide how to present the information to the public by utilizing psychology and organizational techniques (Laminack, 2015, 22). It is very important to know what museums decide to show in their exhibits. Recently, many museums are moving towards showing not only new discoveries and valuable histories but also exhibits that directly affect visitors; curators interact with visitors to learn what is most relevant to the community that learns from the museum (Laminack, 2015, 25). As museums directly interact with and affect the communities that surround them, they have a great social significance in society.

How Do Social Movements Relate to Museums?

One of the reasons that museums change over time is due to the influence of social movements. When you think of significant social movements in our time, a couple of examples may come to mind: Black Lives Matter and environmental activism. These social movements are brought to the attention of museums when the movements relate to the museum's exhibits. Social movements connect to museums because the exhibits in the museums themselves reflect social topics. For example, science museums have exhibits on climate change, natural history museums have exhibits on the colonization of cultures, and art museums have exhibitions on artwork from artists of various backgrounds. These exhibits reflect environmental, racial, and cultural movements.

One of the most influential social movements in the twenty-first century is the movement for environmental activism. Amid climate change and mass extinction, many organizations have been formed to protest against the large corporations that play the largest part in climate change. Not only have organizations been formed, but environmental activism has become a part of daily life for many Americans. From consciously separating between recycling and landfill waste to remembering to turn off the faucet when not in use, being conscious of our human effects on the environment is a large part of how we live today. Another influential social movement today is the Black Lives Matter movement. Leslie R. Crutchfield writes in "Black Lives Matter: From Protests to Lasting Change" that the BLM movement has gone beyond a momentary spark in American history, but has instead made a lasting mark by going "beyond protest and policy change" to "[dismantle] deeply rooted social and cultural norms" (Crutchfield, 2020). By bringing to mainstream attention the systemic oppression of Black Americans and inspiring countless peaceful protests in support of Black rights, the BLM movement has made an irreversible change in the social atmosphere; it has opened doors for not only Black Americans but Americans of various minorities to fight for representation in society. In the following sections, I will continue to explore how racial and cultural representation of minorities and environmental activism are reflected in museums and how museums contribute to representation and conservation.

The Influence of the American Museum of Natural History

Social movements can impact museums by influencing which topics are most important to visitors. Museums are some of the most prevalent educational sources in our societies. Thus, the way that museums present information and the way that they emphasize those topics changes the public perception of those topics. I think that social movements affect museums because I have seen the phenomenon in real time. As a child growing up in New York City, I frequented the American Museum of Natural History. Some of the exhibits that I saw as a child, such as a diorama depicting the meeting of Dutch and Lenape people, have since changed in reflection of the evolving societal environment. When I first saw the original version of the diorama as a child, I thought that it truthfully depicted the reality of colonization. It showed “Dutch leader Peter Stuyvesant receiving a delegation of Lenape, including Oratamin, a sachem (leader) of the Munsee branch” (American Museum of Natural History, 2019). However, the exhibit portrayed the Native American people with inaccurate stereotypes: they were portrayed as wearing little clothing, which would not have been what they would have worn to the important event, and Lenape women were depicted as being subservient and uninvolved with the leadership responsibilities, but in reality, the women held positions of authority and would have been involved with the Dutch interaction (American Museum of Natural History, 2019). These erroneous portrayals made the exhibit non-educational and harmful to the Native American people today whose history had been misrepresented. In 2018, the museum made necessary changes to the exhibit by adding labels that gave accurate information and educated visitors on “how complex and violent colonization was for Native People” (American Museum of Natural History, 2018). This edit with more accurate information instated the museum as a place for education and cultural connection with Native American history.

The change that the museum made was impactful to me. Once it was updated, I learned not only a more accurate interpretation of American history but also how much museum exhibits can affect and be affected by its visitors. Prior to my post-2018 visit, my only knowledge of the Lenape people’s history was, admittedly, through the American Museum of Natural History’s exhibit. In my education in the American school system, there is not enough focus on Native American history, and in popular media such as movies, music, and fashion, there is very little focus on representing Native people. The change in the American Museum of Natural History’s diorama singularly changed my perception of that moment in history. Because of the lack of education and representation of minorities and marginalized cultures, museums become one of the only sources of knowledge on these topics that are widely accessible and commonly visited. Thus, they have an obligation to educate and represent history. The New York museum’s diorama was updated largely because the history that it shows directly affects people in New York. Bradley Pecore, a visual historian, explains that “Lenape people still live in Manhattan,” and “there needs to be a conversation and I that hope that this just begins a conversation that we can have”; righting the exhibit is essential because the history is relevant to the community that the museum serves (American Natural History Museum, 2018). Lauri Halderman, the Vice President of Exhibitions in the museum, explains that “our eyes are different today than they were in the 1930s” (The American Museum of Natural History, 2018). Thus, the museum has a

responsibility to its visitors to evolve history to make an accurate and all-inclusive view of history that represents the culture and society of the people who rely on museums.

Spotlight on Museums: The California Academy of Sciences, The Museum of Us, and the Getty Center

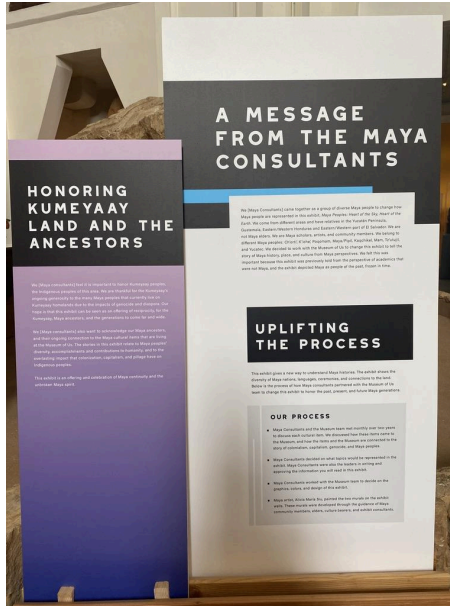
For this section, I will focus on the Getty Museum, the Museum of Man, and the California Academy of Sciences. By investigating the common patterns between these three museums, I will explore social movements that connect to the museums and show how each museum evolved.

The Museum of Us in San Diego has an exhibit: “Maya Peoples: Heart of Sky, Heart of Earth,” dedicated to representing and appreciating Mayan cultures that have survived despite the negative effects of colonialism. It shows neck-craning statues with Mayan symbols, traditional clothes with intricate designs, and pottery made by hands of generations past.

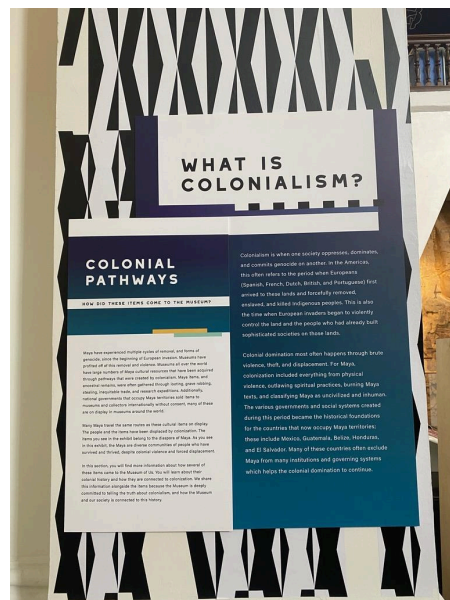


Mayan Statue, The Museum of Us, 2023

It also displays plaques that explain the harms of colonialism and how the Mayan people and culture persisted. They are displayed in the largest room of the museum, the first room visitors enter, so as not to erase the horrors of colonialism but to educate visitors and respect the survivors.



“A Message from the Maya Consultants”, The Museum of Us, 2023



“What is Colonialism?”, The Museum of Us, 2023

The upfront display also centralizes the South American history in the museum. Rather than just dedicating a small room to the display—for it not being a part of Western history—the Museum of Us emphasizes the importance of educating visitors on the Mayan history and culture, disallowing it to fade into the background. The exhibit was designed by a half-Mayan, half-Cantonese artist, Alicia María Siu. Mayan culture is dear to her heart and is inspired by the stories of her mother and ancestors (Museum of Us, 2023). Her heritage is a story that is not often illuminated in America, so this exhibit is a significant form of representation for people of

Mayan ancestry. Though a museum exhibit cannot undo or erase the damage done by colonialism, perhaps the representation and respect that the exhibit offers can be a step towards preventing the regrettable parts of history from repeating.

The Museum of Us has another exhibit dedicated to the appreciation of culture; entitled “Kumeyaay Nation,” the exhibit appreciates and respects the native Kumeyaay Nation—a people that still exist in their native homeland in California. The exhibit extends gratitude to the Kumeyaay people and preserves their culture (Museum of Us, 2023). It shows traditional handmade baskets and ingenious hunting tools and explains their usage and significance in Kumeyaay culture. The exhibit also displays an immersive overhead mural depicting the culture’s cosmological belief system.



Woven baskets, The Museum of Us, 2023

Both of these exhibits represent cultures that have been neglected in mainstream society. They give spaces for diverse artists to exhibit their history, and for museum visitors to learn about things that they could not have learned about anywhere else because Mayan and Kumeyaay nations are not often represented in other places. The exhibit follows the social movement to encourage more diversity in public spaces such as media and museums.

With this exhibit, the Museum of Us makes impactful steps towards reducing the lack of diverse cultural representation in American society and creating an inspirational space for cultures to mingle. However, the impact of one exhibit cannot be enough to fully reform the social status and deeply rooted systemic oppression of Native Americans in the United States. However, for the Kumeyaay Nation, the Museum of Us is one of the only museums to dedicate an exhibit to their culture, and it is the first to spotlight their cosmological belief system (Museum of Us, 2023). Thus, though one museum exhibit’s impact may seem small, to the Kumeyaay Nation, it is an important step in representation. The positive example of representation bodes well for the future of inclusive curations in American museums.

The Getty Center has an exhibit: “Woshaa’axre Yaang’aro,” that presents the cultural values of the Tovaangar people, who are native to California. The artist, Mercedes Dorame, combines her Native American cultural heritage and history with the ecological values of the area. The piece features large seashells that hang over the guests and reflect sunlight around the circular room, which exemplifies the message of preserving the importance of the ocean.



Mercedes Dorame, The Getty Center, 2023

The artist displays how her culture intertwines with the California ecosystem. Thus, this exhibit is a beautiful combination of the movements for environmental restoration and cultural representation. Her vision and artistic talent are displayed in the main room of the museum, which gives her the kind of spotlight representation that many Native American artists in America cannot have. Her work creates new steps for Native Americans in America.

The California Academy of Sciences is a natural sciences museum in California. The California Academy of Sciences has a mission to “repair our damaged ecosystems and climate” (California Academy of Sciences, 2023). They have programs that breed coral reefs, increase biodiversity through islands, and strengthen forests against wildfires. These programs respond to the widespread awareness of the negative man-made impacts on the environment and create a new positive impact. The building of the museum itself is the most environmentally friendly museum in the world. The current facility opened in 2008 and has a state-of-the-art design that incorporates structure, lighting, heating, and landscaping designs that minimize energy usage and waste (Arup, 2023). This environmentally friendly design aligns with the museum’s mission to regenerate our ecosystems and also society’s consciousness of environmental issues. The museum actively impacts society because every time a visitor enters the museum, they are

transported to a world where environmental consciousness is the most important value and the visitor immediately becomes a part of the solution for the climate crisis.

Conclusions

In this paper, I have discussed how museums are related to social movements. Museums are spaces for education and interaction with society. They are connected to social movements through their interaction with society. Major museums in America are connected to the social movements of racial and cultural representation, and environmental consciousness. As a museum visitor, museums have a large impact on my perception of history, culture, and the society around me. They educate visitors on the events in their world and expand their understanding of the people and cultures around them. Through this cultural education, they become not only places of knowledge, but also places of social connection.

Thus, as I have explored in this paper, museums have an innate relationship with social movements because museums reflect the events and values of the communities that surround them. Not only do they simply reflect the events that have already happened, but many museums actively take steps forward to improve the lives of the communities around them and nurture new forms of connection and education by presenting untold histories of minority cultures and taking part in conserving the local and global environment. In this way, museums have evolved from the scribes of history to the facilitators of revolutionary creation. However, the steps that museums take cannot entirely influence society. It is up to the visitors of museums to interact with the messages of the exhibits, engage with the communities represented in museums, and participate in social movements in order to take steps toward making real positive changes for minorities and our environment.

Works Cited

- American Museum of Natural History. 2018. "Behind the Updates to Old New York Diorama." *American Museum of Natural History*. Video. Accessed on September 17, 2023.
- American Museum of Natural History. 2019. *Teaching the Old New York Diorama*. Accessed on September 17, 2023.
<https://www.amnh.org/content/download/261787/4452028/file/teaching-the-old-new-york-diorama-grade-3-5.pdf>
- Arup. 2023. "California Academy of Sciences is the most sustainable museum in the world." *Arup*. Accessed on September 17, 2023.
<https://www.arup.com/projects/california-academy-of-sciences#>
- California Academy of Sciences. 2023. "Mission: Regeneration." *California Academy of Sciences*. Accessed on September 17, 2023.
<https://www.calacademy.org/about-us/regenerating-the-natural-world>
- Chaliakopoulos, Antonis. 2020. "History of Museums: A Look at The Learning Institutions Through Time." *The Collector*. Accessed on September 17, 2023.
<https://www.thecollector.com/history-of-museums/>
- Crutchfield, Leslie R. 2020. "Black Lives Matter: From Protests to Lasting Change." *The Chronicle of Philanthropy*. Accessed on September 17, 2023.
https://www.philanthropy.com/article/Black-Lives-Matter-From/249017?cid=cdfd_home
- Laminack, Mackenzie. 2015. "From Taking in to Reaching Out: How Collections and Collections Staff Are Being Used to Create a Community-Centered Museum." *Journal of Museum Studies* 9 (1): 20-28. Accessed on September 17, 2023.
- Museum of Us. 2023. "Kumeyaay: Native Californians/Iipai-Tipai." *Museum of Us*. Exhibit. Accessed on September 17, 2023.
<https://museumofus.org/exhibits/kumeyaay-native-californians-iipai-tipai>
- Museum of Us. 2023. "Maya Peoples: Heart Of Sky, Heart Of Earth." *Museum of Us*. Exhibit. Accessed on September 17, 2023.
<https://museumofus.org/exhibits/maya-peoples-heart-of-sky-heart-of-earth>

Analysis of Common Injuries in Tennis By Sunay Iyer

Abstract

Tennis is a very physically demanding, full body sport that requires intense transfer of power and quick movement for a long period of time, which allows for many injuries. Some of the injuries in the sport have been studied frequently, while others have not been covered extensively. The goal of this review is to analyze the prevalence of common injuries in tennis and compare them to each other as well as explain the etiology and biomechanics behind the injuries.

Introduction

Tennis is an open-skilled sport, where every point is unique and demands a different movement from the body each time. It requires twisting and sudden changes of direction, and as the level of the player increases, so does the frequency of injury. The contortion of the joints in modern techniques, maximization of power through the kinetic chain, and strenuous, repetitive motions combine to cause a great degree of musculoskeletal injuries. The aim of this literature review is to summarize data on common injuries in various locations of the body and provide a comparison in prevalence with other joints and sites of injury, present causes of injury through biomechanics and anatomy, and explore prevention and treatment protocols. This study will primarily use data collected from The Championships, Wimbledon (1) and USTA Boys Championships (2) for comparison of injury rates.

Table 1

	Male (% of male injuries)	Female (% of female injuries)	Total (% of total injuries)
Upper Extremities	28	28	28
Lower Extremities	47	49	48
Trunk	25	23	24

*Data from “Tennis injury data from The Championships, Wimbledon, from 2003 to 2012” (1)

Table 2

	Prevalence (per 100) athletes	Percentage of total injuries (derived from prevalence)
Upper Extremities	5.6	26.5
Lower Extremities	10.3	48.8
Trunk	5.2	24.6

*Data from “Injury surveillance at the USTA Boys' Tennis Championships: a 6-yr study” (2)

**Prevalence was defined as the number of new and recurrent injuries that required medical evaluation or treatment.

Analysis of Overall Injury Rates

Both tables provided similar data and injury rates. Lower body injuries constituted almost half of injuries in both studies. These were primarily due to the number of acute injuries for each location. Although tennis is an open skilled sport overall, the upper body is relatively closed skill, as the swing motions of players do not vary as much; groundstrokes generally follow a similar, repeatable technique. This leads to mostly chronic injuries resulting from overuse. The exception in the upper body was the wrist, which suffered from the second most acute injuries (3). The ankle suffered the most acute injuries (3), therefore elevating the total percentage of lower body injuries. This could be explained by the quick, reactive nature of footwork and sliding on court, placing a heavier strain on the ankle and all lower body muscles. The trunk performs the least varied movements, primarily rotation as part of the kinetic chain (4) but also extension and flexion during the service motion. This limited movement causes the least amount of acute injuries, and injuries that do arise are usually the product of a repeated, strenuous service motion. Muscle imbalances can also cause various musculoskeletal injuries, and these are present with the knees and hips, the lumbar and abdominals, and the shoulders and elbows (5). Imbalances are particularly common in the hip, knee, and shoulder, contributing to the data summarized in Tables 1 and 2.

Table 3

	Male (% of all injuries)	Female (% of all injuries)	Combined (% of all injuries)	Combined (% of upper extremity injuries)
Shoulder	14	10	12	42.9
Elbow	6	6	6	21.4
Wrist	8	12	10	35.7

*Data extrapolated from “Tennis injury data from The Championships, Wimbledon, from 2003 to 2012” (1)

Shoulder Injury

Within the upper extremities, the shoulder was most commonly injured. Most upper extremity injuries are likely to stem from the serve, as the most intense stages of the kinetic chain in the serve place the most duress on distal locations like the shoulder, elbow, and wrist (6). However, the serve places the shoulder in a more extreme range of motion during the early and late cocking stages (see **Figure 1**) of the serve, causing many more injuries (7). Additionally, the shoulder contains many intricate structures, being the most flexible joint in the body (8), which increases injury rates for it, especially relative to the restricted mobility of the elbow as a hinge joint.

Tendinopathy in the rotator cuff is a common byproduct of the repetitive use of the kinetic chain in the serve, and can lead to other injuries such as scapular dyskinesia(9). The serve can account for 45-60 percent of strokes performed in a match (9), and as a result the tendinitis is a primary factor in shoulder injury.

Scapular dyskinesia is another common injury, causing a protruding shoulder blade and a loss of range of motion. It may be asymptomatic, but left untreated, it could cause a greater risk of injury. It is caused mainly by instability or inflexibility of the surrounding structures. The cocking position of the serve places the scapula and the surrounding muscles at risk, and the injured muscle or structure may cause a different type of scapular dyskinesia to develop (10). Scapular dyskinesia is classified into three types. Type 1 is seen in the cocking position and is commonly paired with inflexibility of the pectoralis major/minor and weakness of the lower trapezius and serratus anterior. Type 2 is seen even at rest, and is associated with weak stabilizers such as the trapezius and rhomboideus. Type 3 is associated with a correlation to rotator cuff injury (7).

Lastly, SLAP tears occur in tennis, usually in the chronic form, since the acute form is caused by falls and fast movement of heavy weights. Chronic slap tears in tennis are a result of repetitive overhead motions, such as the serve. A theory that provides a potential explanation is, in overhead motions, extreme external rotations (cocked phase of service motion) cause the ends of greater tuberosity to contact the posterior and superior labrum, leading to SLAP tears (11).

Figure 1



Elbow Injury

Tennis elbow is one of the most well-known athletic injuries, however, it constitutes a smaller percentage of upper extremity injury than both the wrist and the shoulder. Elbow injuries in tennis are generally limited to tendinopathy, caused by repetitive motions and the overuse of the forearm muscles (12). There are two primary tendinopathies in the elbow: medial and lateral. Medial tendinopathy, specifically medial epicondylitis is known as golfer's elbow and lateral epicondylitis is known as tennis elbow, although both can arise from repetitive motions in tennis. The condyles are rounded ends of bone and the tendons that connect them to specific muscles may become inflamed due to overuse of the respective muscle groups. Medial epicondylitis is more common in golf and pitching in baseball, as the acceleration phase (see **Figure 2**) places a great load through flexor pronator muscles that leads to tendinopathy in the medial epicondyle (13). However, this acceleration motion can also be seen in serve and forehand motions, leading to the same tendinopathy (compare **Figure 2** and **Figure 3**).

Figure 2



Figure 3



Lateral epicondylitis is more unique to tennis, as it is caused by repetitive stress on the extensor muscles in the wrist such as the extensor carpi radialis brevis (ECBR), extensor carpi radialis (ECR), and extensor carpi ulnaris longus (ECUL). It was noted by Jobe and Cicatti (13) that the extensor brevis received greatest activation in the forehand groundstroke. This is due to these muscles having to repeatedly stabilize in an extended position for both the forehand and backhand (see **Figure 4 and 5**). Increased reliance on wrist lag for topspin and power in modern tennis technique also contribute to these issues.

Figure 4



*The wrist lag in the forehand is the last step in maximizing power in the kinetic chain of the forehand

Figure 5



*The left hand during the take back for the two handed position activates the extensor muscles.

Wrist Injury

The biomechanical causes of wrist injury are similar to those of the elbow, as the same forearm muscles and tendons are connected to the elbow and wrist. For example, the previously mentioned extensor carpi ulnaris connects at the lateral epicondyle and the fifth metacarpal base (14), therefore causing tendinopathies through overuse in both the wrist and elbow. However, wrist injuries are more prevalent because the joint is more mobile, and injury in the triangular fibrocartilage complex (TFCC) and tears in the ligaments connecting carpal bones increase the risk of upper extremity injury.

Ulnar-sided wrist tendinitis is often linked to the same repetitive overuse of the ECU and the instability of the ECU tendon in the forehand and two-handed backhand groundstroke (see **Figure 4** and **Figure 5**). In the low forehand specifically, the supination required to get under and over the ball can cause rupture in the ECU tendon’s subsheath, causing instability (15). The instability due to rupture is known as subluxation, and can lead to “attenuation in both the tendon and subsheath” (16). Extreme cases involve full tears which usually do not recover fully.

Radial-sided tendinitis (De Quervain's) is far less common, and is associated with repetitive forehand motions while using the Eastern grip (15). In the Eastern grip, the abductor pollicis longus (APL) muscle is activated for stability (17), and repeated stress can cause De Quervain’s to develop.

The triangular fibrocartilage complex (TFCC) is located on the ulnar side of the wrist, and composed of fibrous tissue and cartilage. It provides joint stability and “load transmission across the wrist” (15). The most common injury is a slow degradation of cartilage through overuse that can lead to a tear, and it is caused by excessive rotation or “axial loading of an ulnar deviated wrist” (18). Both axial loading (loading with racket weight) and rotation are seen in the groundstrokes through dropping the racket head below the ball and brushing over it.

Table 4

	Male (% of all injuries)	Female (% of all injuries)	Combined (% of all injuries)	Combined (% of trunk injuries)
Abdominal	6	4	5	20.8
Lumbar	13	12	12.5	52.1
Chest	3	4	3.5	14.6
Head/Neck	3	3	3	12.5

*Data extrapolated from “Tennis injury data from The Championships, Wimbledon, from 2003 to 2012” (1)

**Injury to chest, head, and neck will not be reviewed in this article

Abdominal injuries

Strain of the rectus abdominis is a common tennis-related injury, generally caused by repeated, powerful trunk flexion from a hyperextended position due to incorrect technique or certain types of serves (19). Since the rectus abdominis is a primary muscle for trunk flexion, it suffers from strains contralateral to the dominant arm (19). The serve itself places the back in a hyperextended state to store energy, and there is more extension on the nondominant side because the nondominant arm is extended upwards (see **Figure 6**). Electromyography (EMG) activity values in the abdominals were higher for kick (topspin) serves according to a study by

Chow, Shim, et al. (19). This is consistent with the findings that increased hyperextension and resulting flexion more severely strain abdominal muscles (20) because the kick serve motion requires more arching of the back (see **Figure 7**). More infrequently, the rotational movements of tennis groundstrokes do also cause internal oblique strains (21).

Figure 6



*In the early cocking position of the serve, the extended left arm causes increased activity (19) in the abdominals contralateral to the dominant side.

Figure 7



*A similar early cocking position, but the back is in a more extreme hyperextended position for a topspin serve

Lumbar injuries

Lumbar strains are generally acute injuries of tendons that occur due to the sudden twisting and stretching that tennis requires for moving and performing certain strokes. These are fairly common and can be caused by a single event or repetitive microtrauma (22).

The lumbar spine's role in shock absorption and transmission of force between the upper and lower body makes it "essential...as the center of rotation" (23). The transmission of force is seen at its maximum in the serve, as the kinetic chain transfers power from the legs (see **Figure 7**) to the late cocking position (see **Figure 1**). However, the repetitive trauma and overuse of the lumbar spine in deceleration, change of direction, and the kinetic chain leads to disc degeneration. Change of direction is facilitated by a low center of gravity, requiring players to keep their hamstring in a contracted state for an extended period of time, which can lead to disc degeneration (23).

Table 5

	Male (% of all injuries)	Female (% of all injuries)	Combined (% of all injuries)	Combined (% of trunk injuries)
Hip/groin	10	7	8.5	17.7
Thigh	7	8	7.5	15.6
Knee	12	13	12.5	26.0
Lower leg	4	3	3.5	7.3
Ankle	8	8	8	16.7
Foot	6	10	8	16.7

*Data extrapolated from "Tennis injury data from The Championships, Wimbledon, from 2003 to 2012" (1)

**Injury to thigh and lower leg will not be reviewed in this article

Hip and Groin Injury

The primary cause of groin injury is an adductor strain, which often occurs with changing direction and exaggerated stress on the tendon" (24). Modern tennis requires quick, explosive change in direction, making the adductors susceptible to injury. In addition to changing of direction, excessive pronation of the leg can contribute to adductor strains (24). In high-level tennis, players often slide to balls to change direction quicker, causing overpronation (see **Figure 8**). Muscular imbalances can also play a role in injuries. For example, asymmetrical hypertrophy of the iliopsoas and gluteal muscles can also cause groin injury, and this muscle imbalance is common in the non-dominant leg (27) Lastly, the rectus abdominis and the adductor longus are

connected by the same tendons, therefore weakness and injury to the rectus abdominis as previously discussed could make a player prone to adductor strain. (25)

Figure 8



*Overpronation (inside turning of the foot) while sliding to change direction

The hip is a complex joint, and different types of injuries can occur due to overuse as it plays a role in transferring power in the kinetic chain from the lower body to the upper extremities (26). . A study by Martin et al. investigated the correlation between the open stance forehand and likelihood of hip injury, and found that it resulted in more extreme hip angles which could increase risk of overuse injuries and hip impingement (26). In modern tennis, the open stance forehand is used to keep up with the faster pace and defend better.

Knee Injury

The knee is the most commonly injured structure in the lower leg, as it is one of the most anatomically complex joints (28) making it more susceptible to injury than other joints. Chronic injury and repeated microtrauma are common in tennis, whereas certain acute injuries like ACL tears are more prevalent in contact sports like American football and rugby (29). Overuse related knee injuries are common due to the stress from low load, high repetitions in the countless number of small steps a player takes on the court, and from high load, relatively lower repetition from extremely sudden change of direction and jumping (30).

Jumper's knee, or patellar tendinitis, is most common in basketball and volleyball. It is caused by the repeated extending of the knee and great force produced by the quadriceps in this motion. However, patellar tendinitis is common in tennis with rapid knee extension playing a role in the service motion and the change of direction (**Figure 8**). The movements are extremely explosive, and occur repeatedly, during almost every point, causing repeated microtrauma. This results in a "manifestation of microtrauma" including tendinitis, specifically jumper's knee (30).

Another knee injury is a torn meniscus, and it has two types: degenerative and acute. The acute variety is more common in sports, often due to pivoting and changing direction, both of which are common in modern tennis. It is most commonly injured in a "semiflexed position and

when a torsional stress is applied to the knee” (31). An example of this position is in the closed stance forehand, where the back leg is bent and rotated inwards (see **Figure 9**). The knee is in a vulnerable position, and the frequency of the forehand causes meniscus injury.

Figure 9



*The back leg rotates inwards as the body rotates to hit a forehand to maximize power in the kinetic chain

Ankle Injury

The ankle has a relatively lower injury rate, and a possible cause for this is the decreased involvement in the kinetic chain. Thus, it is not involved in as many repetitive, powerful motions that cause overuse injury. A main contributor to ankle injury is an ankle sprain due to inversion, damaging the lateral ligaments of the ankle (31).

Although there have been a few studies correlating the surface of play and lower limb injuries (32, 33), none directly address its relation to ankle sprains, partially due to the wide variety of factors that are difficult to control. The ankle injuries in table 5 could be higher from the slippery courts at Wimbledon but also counteracted by the shorter matches on grass. Clay courts promote sliding, matches are longer and there is a risk of tripping on the plastic lines, all contributing to ankle injuries. Hard courts vary greatly by composition, with corresponding differences in the grip and duration of matches. However, it can be assumed that sliding on hard courts with more grip may cause inversion and related ankle injuries.

Conclusion

Injuries in tennis were most commonly a product of repeated microtrauma and overuse.. The kinetic chain was a primary cause of injury, as the repetitive transfer of power places heavy

stress on the joints. Of all the strokes, the serve was the most injury-provoking movement, so correct serve technique should be an emphasis for coaches and players seeking to minimize injury. Additionally, many of the injuries occur at the ends of each joint's range of motion, such as the hyperextended risk in the forehand and hyperextended lumbar in the serve. This suggests that as the sport is becoming more fast-paced, frequency and severity of injuries are also changing. Further research into the effects of the modern forehand and kick serve are needed. The relationship between the court surface and injuries has also not been studied adequately, and could be an avenue for further exploration in future research.

Works Cited

1. McCurdie, I., et al. "Tennis injury data from The Championships, Wimbledon, from 2003 to 2012." *British journal of sports medicine* 51.7 (2017): 607-611.
2. Hutchinson, MARK R., et al. "Injury surveillance at the USTA Boys' Tennis Championships: a 6-yr study." *Medicine and science in sports and exercise* 27.6 (1995): 826-831.
3. Sell, Katie, et al. "Injury trend analysis from the US Open Tennis Championships between 1994 and 2009." *British journal of sports medicine* 48.7 (2014): 546-551.
4. Elliott, Bruce. "Biomechanics and tennis." *British journal of sports medicine* 40.5 (2006): 392-396.
5. Croisier, Jean-Louis. "Muscular imbalance and acute lower extremity muscle injuries in sport." *International SportMed Journal* 5.3 (2004): 169-176.
6. Van der Hoeven, H., and W. Ben Kibler. "Shoulder injuries in tennis players." *British journal of sports medicine* 40.5 (2006): 435-440.
7. Lee, Helen WM. "Mechanisms of neck and shoulder injuries in tennis players." *Journal of Orthopaedic & Sports Physical Therapy* 21.1 (1995): 28-37.
8. InformedHealth.org. Cologne, Germany: Institute for Quality and Efficiency in Health Care (IQWiG); 2006-. How does the shoulder work? 2020 Feb 13. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK554696/>
9. Alrabaa, Rami G., Mario H. Lobao, and William N. Levine. "Rotator cuff injuries in tennis players." *Current reviews in musculoskeletal medicine* 13 (2020): 734-747.
10. Kibler, W. Ben, and Aaron Sciascia. "Current concepts: scapular dyskinesis." *British journal of sports medicine* 44.5 (2010): 300-305.
11. Familiari, Filippo, et al. "SLAP lesions: current controversies." *EFORT open reviews* 4.1 (2019): 25-32.
12. Flatt, Adrian E. "Tennis elbow." *Baylor University Medical Center Proceedings*. Vol. 21. No. 4. Taylor & Francis, 2008.
13. Jobe, Frank W., and Michael G. Ciccotti. "Lateral and medial epicondylitis of the elbow." *JAAOS-Journal of the American Academy of Orthopaedic Surgeons* 2.1 (1994): 1-8.
14. Campbell, Doug, et al. "Sports-related extensor carpi ulnaris pathology: a review of functional anatomy, sports injury and management." *British journal of sports medicine* 47.17 (2013): 1105-1111.
15. Gil, Joseph A., and Sanjeev Kakar. "Hand and wrist injuries in tennis players." *Current reviews in musculoskeletal medicine* 12 (2019): 87-97.
16. Byrd, Jacqueline N., Sarah E. Sasor, and Kevin C. Chung. "Extensor carpi ulnaris subluxation." *Hand clinics* 37.4 (2021): 487-491.
17. Van Oudenaarde, E., and R. A. Oostendorp. "Functional relationship between the abductor pollicis longus and abductor pollicis brevis muscles: an EMG analysis." *Journal of anatomy* 186.Pt 3 (1995): 509.
18. Jawed, Akram, Mohammed Tahir Ansari, and Vikas Gupta. "TFCC injuries: how we treat?." *Journal of Clinical Orthopaedics and Trauma* 11.4 (2020): 570-579.

19. Maquirriain, Javier, Juan P. Ghisi, and Antonio M. Kokalj. "Rectus abdominis muscle strains in tennis players." *British journal of sports medicine* 41.11 (2007): 842-848.
20. Chow, John W., Soo-An Park, and Mark D. Tillman. "Lower trunk kinematics and muscle activity during different types of tennis serves." *BMC Sports Science, Medicine and Rehabilitation* 1 (2009): 1-14.
21. Maquirriain, J., and J. P. Ghisi. "Uncommon abdominal muscle injury in a tennis player: internal oblique strain." *British journal of sports medicine* 40.5 (2006): 462-463.
22. Lawrence, James P., Hunter S. Greene, and Jonathan N. Grauer. "Back pain in athletes." *JAAOS-Journal of the American Academy of Orthopaedic Surgeons* 14.13 (2006): 726-735.
23. Fiani, Brian, et al. "Repetitive traumatic discopathy in the modern-era tennis player." *Cureus* 12.8 (2020).
24. Kiel, John, and Kimberly Kaiser. "Adductor strain." (2018).
25. Koulouris, George. "Imaging review of groin pain in elite athletes: an anatomic approach to imaging findings." *American Journal of Roentgenology* 191.4 (2008): 962-972.
26. Martin, Caroline, et al. "Can the open stance forehand increase the risk of hip injuries in tennis players?." *Orthopaedic journal of sports medicine* 8.12 (2020): 2325967120966297.
27. Abrams, Geoffrey D., Per A. Renstrom, and Marc R. Safran. "Epidemiology of musculoskeletal injury in the tennis player." *British journal of sports medicine* 46.7 (2012): 492-498.
28. Majewski, Martin, Habelt Susanne, and Steinbrück Klaus. "Epidemiology of athletic knee injuries: A 10-year study." *The knee* 13.3 (2006): 184-188.
29. Maquirriain, J., and P. J. Megey. "Tennis specific limitations in players with an ACL deficient knee." *British journal of sports medicine* 40.5 (2006): 451-453.
30. Davies, George J., Lynn A. Wallace, and Terry Malone. "Mechanisms of selected knee injuries." *Physical Therapy* 60.12 (1980): 1590-1595.
31. Perkins, Robert H., and Denise Davis. "Musculoskeletal injuries in tennis." *Physical Medicine and Rehabilitation Clinics* 17.3 (2006): 609-631.
32. Girard, Olivier, et al. "Effects of the playing surface on plantar pressures and potential injuries in tennis." *British journal of sports medicine* 41.11 (2007): 733-738.
33. Kim, Shinyang. "The intrinsic and extrinsic risk factors for injury in professional tennis players on clay and grass court: a systematic." (2022).

Images and Figures Works Cited

1. El Attrache, Neal, et al. "The Shoulder in Athletes." *Musculoskeletal Key*, 8 Sept. 2016, musculoskeletalkey.com/the-shoulder-in-athletes/. Accessed 13 Sept. 2023.
2. Rosengren, Phil. "Should Pitchers Stretch to Increase External Rotation? – BetterPitching.com." *Betterpitching.com*, betterpitching.com/should-pitchers-stretch-to-increase-external-rotation/. Accessed 8 Sept. 2023.
3. "Roger Federer Forehand in Super Slow Motion 9 - Indian Wells 2013 - BNP Paribas Open." *Www.youtube.com*, 28 July 2013, www.youtube.com/watch?v=N4VxccAUPto. Accessed 22 Sept. 2023.
4. "Roger Federer Forehand Slow Motion Court Level View - ATP Modern Tennis Forehand Technique." *Www.youtube.com*, 31 Mar. 2020, www.youtube.com/watch?v=EFY460oquXw. Accessed 3 Oct. 2023.
5. Paloma, Marylyn. "Novak Djokovic Forehand and Backhand in Super Slow Motion - 2013 Cincinnati Open -." *Essential Tennis*, 8 June 2016, www.essentialtennis.com/novak-djokovic-forehand-backhand-super-slow-motion-2013-cincinnati-open/. Accessed 14 Oct. 2023.
6. "Andy Roddick Serve Analysis - the Supersonic Roddick Serve Technique." *Tennis Instruction*, 13 Jan. 2021, tennisinstruction.com/andy-roddick-serve/. Accessed 1 Nov. 2023.
7. Vincent, Rick. "Tennis Kick Serve." *Www.youtube.com*, 19 Feb. 2019, www.youtube.com/watch?v=nsr-dxCzyp0. Accessed 10 Nov. 2023.
8. Wolfond, Joe. "VIDEO: Djokovic Slides into Brilliant, Split-Legged Backhand Pass." *TheScore.com*, 31 Jan. 2016, www.thescore.com/news/949156. Accessed 12 Nov. 2023.
9. "Kyle Edmund Forehand Slow Motion - ATP Modern Tennis Forehand Technique (NextGen Forehand in Tennis)." *Www.youtube.com*, 17 June 2020, www.youtube.com/watch?v=5g3PjMKxVgs. Accessed 20 Nov. 2023.

Panopticon Dilemma: Surveillance as Freedom and Captivity in Law By Zhenze Yu

Introduction

Thomas Hobbes once stated that in the absence of laws, human existence would be "solitary, poor, nasty, brutish, and short" (Hobbes, 2018). In any civilized society, the rule of law serves as the bedrock upon which social cohesion and stability are established. Laws are not merely an arbitrary set of rules; they often emerge from significant philosophical doctrines. One such doctrine is utilitarianism, an ethical theory that aims to maximize overall happiness and minimize suffering, serving as a guiding philosophy in legislative processes. Within utilitarian principles, one key concept is that of the Panopticon, which was initially conceptualized by Jeremy Bentham and later elaborated upon by Michel Foucault. The Panopticon is a philosophical construct that has evolved from its initial design as a prison structure to impact multiple facets of modern society. The original idea was proposed by the utilitarian philosopher Jeremy Bentham in the late 18th century. Bentham envisioned a circular prison layout where a single watchman could observe all inmates without being seen, thereby creating a sense of constant surveillance.⁴⁰ Inmates would act as though they were always being watched, thus self-regulating their behavior, even when no one was watching them. This method of behavioral regulation replaces the most popular method back at the time, which is binding the inmates to a heavy object and placing them into darkness to prevent escape attempts; despite its less aggressive means, the Panopticon is in reality

Today, much of contemporary lawmaking can be closely related to the philosophical constructs of utilitarianism and the Panopticon effect, as it relates to criminal justice. Specifically, this paper will argue that the Panopticon effect is indispensable to rehabilitative justice because it instills a psychological "gaze"⁴¹ that serves as a deterrent for former prisoners. This "gaze" creates a self-regulating behavior, discouraging reoffending by placing the prisoners under a false sense of being under constant scrutiny, even when they are not. Through the lens of the IRAA law and its 'second look amendment,' this paper aims to provide a detailed analysis of how the Panopticon effect, when integrated with utilitarian principles, can inform the development of more effective rehabilitative justice policies and contribute to the prevention of crimes in modern legislative acts.

I. The Panopticon Effect and Its Influence on Law

The Panopticon effect, initially a philosophical concept, has far-reaching implications for the legal system. This psychological mechanism of surveillance has been integrated into the very fabric of modern legal systems, serving as a tool for control and a subject of ongoing ethical discussion.

While some critics argue that the Panopticon effect is ethically questionable due to its emphasis on surveillance, it is essential to understand that the true power of the Panopticon lies

⁴⁰ Bentham, Jeremy, *The Panopticon Writings*.

⁴¹ A concept put-forth by Jean-Paul Sartre in his existentialism play *No Exit*, later elaborated in the text

in the fear of being caught while committing a crime. This fear serves as a deterrent, guiding individuals towards societal norms and expectations, and in extreme cases, preventing violence. Specifically, this fear acts as a form of socialization, compelling individuals to internalize societal values and norms as their own, thereby fostering a sense of community and shared morality. In this way, the Panopticon effect serves not merely as a punitive measure, but as an instrument for shaping social behavior and cohesion.

Aside from the ethical implications of the Panopticon, it is worth examining the philosophical underpinnings of this theory using existentialism. In Jean-Paul Sartre's existentialist play *No Exit*, he introduces the concept of "the gaze," positing that the awareness of being monitored by others shapes human behavior and identity. Sartre's concept of the gaze broadly aligns with the Panopticon effect, as both emphasize the psychological impact of perceived surveillance. Sartre's "gaze" serves to reinforce the Panopticon's efficacy by highlighting that the feeling of being observed is an inherent aspect of human existence, thereby making the Panopticon's surveillance mechanisms all the more potent.

In summary, the Panopticon effect transcends the conventional understanding of surveillance and policing to serve as a new approach in contemporary legal systems. Rooted in the philosophical works of Bentham and Foucault, it operates as a psychological mechanism that instills a sense of perpetual oversight, thereby shaping human behavior and social norms. This effect is not only a tool for post-crime punishment, but also acts proactively to deter criminal activity, serving as an instrument for socialization. Its principles are deeply interwoven into the fabric of modern law, influencing not only penal institutions, but also broader legislative measures aimed at crime prevention. Furthermore, its relevance is underscored by existential concepts such as Sartre's "gaze," which highlights the inherent human sensitivity to perceived observation. Thus, the Panopticon effect serves as both a tool for societal control and a subject of ongoing ethical and philosophical discourse, making it a foundational element in the modern legal landscape.

II. The Panopticon's Broad Reach: Influences on Societal Behavior and Structure

The Panopticon essentially achieves its goal of regulating human behavior through imposing a psychological regulation. While its physical manifestation may be limited to specific institutions like prisons, its psychological reach extends far beyond. Foucault's interpretation shifted the Panopticon from a design for efficient control within a prison to a model of societal power dynamics. The notion is that the perception of constant surveillance results in self-regulation, thus maintaining order and compliance within a community. The Panopticon operates on the principle of "disciplinary power," which has a profound impact on how individuals behave in society.

One of the key psychological effects of the Panopticon is the fear of punishment. According to Hsin-Wen Lee's *Taking Deterrence Seriously*⁴², the fear of external sanctions, such

⁴² Hsin-Wen Lee, "Taking Deterrence Seriously: The Wide-Scope Deterrence Theory of Punishment," *Journal of Research in Crime and Delinquency*, 2017.

as legal repercussions, serves as a significant deterrent against criminal behavior. This fear is not just about the immediate consequences, but also includes internal sanctions like guilt or social shame. The Panopticon effect amplifies this fear by creating a sense of constant surveillance, making individuals think twice before engaging in unlawful activities.

Moreover, the Panopticon's influence isn't limited to deterring criminal behavior. It also shapes social norms and expectations. The mere perception of being watched can influence people to conform to societal norms, even when no one is actually watching. This self-regulation, induced by the Panopticon effect, contributes to social cohesion and stability.

The Panopticon's transition from philosophy to societal control showcases how deeply embedded philosophical ideas can be in shaping societal norms and systems. Its legacy serves as a case study for how complex philosophical concepts can significantly influence, not only specific sectors like the justice system, but also the broader context of how power is exercised and maintained in society.

III. IRAA Law and the 'Second Look Amendment'

The Incarceration Reduction Amendment Act (IRAA)⁴³ is legislation aimed at criminal justice reform. Passed in Washington D.C., its central goal is to provide inmates, particularly those who have been incarcerated since they were minors, the chance to petition for resentencing after a specific period of time. The intent is to acknowledge growth and change in individuals, providing them a "second look" and possibly reducing their time in prison.

The 'Second Look Amendment' is a crucial part of this act. Philosophically, it draws from ideas of rehabilitation and humanism, contending that humans are not static beings, but rather individuals who can change and evolve over time. This aligns with broader moral philosophies that advocate for the potential for human change and growth, contrasting with the more punitive or deterministic philosophies that see individuals as immutable.

However, the implications of this amendment are complex. While it seeks to give a second chance to those who have shown significant rehabilitation, it also raises questions about justice and societal protection. The possibility of releasing individuals back into society after they have served only a part of their sentence can be seen as disrupting societal norms and possibly even weakening the control mechanisms that the Panopticon represents.

In broader legal framework, the 'Second Look Amendment' under the IRAA represents a shift towards restorative justice from a more traditional, retributive model. This is in line with a growing movement in modern lawmaking that focus on rehabilitation over punishment. However, it stands in contrast to philosophical concepts like the Panopticon⁴⁴, which stress the importance of surveillance and disciplinary power as methods of societal control and protection.

IV. The Contradiction Between IRAA and the Panopticon

⁴³ For the specific statute, see “§ 24–403.03. Modification of an imposed term of imprisonment for violations of law committed before 25 years of age.”

⁴⁴ Impacts are elaborated in Section V

The Incarceration Reduction Amendment Act (IRAA) and its 'Second Look Amendment' represent a significant departure from the Panopticon's principles of surveillance and control. While the Panopticon operates on the fear of constant surveillance to deter criminal behavior, the 'Second Look Amendment' under IRAA focuses on the potential for human growth and rehabilitation. This shift towards a more humanistic approach contradicts the Panopticon's emphasis on disciplinary power as a method of societal control.

Data from the Indiana Department of Correction reveals a startling picture. In 2018, of the 532 juveniles released, 24.8% were re-incarcerated within three years. Even more shocking, those discharged from incarceration had a recidivism rate of 88.2%, compared to only 9.8% for those released on parole⁴⁵. The implication here is clear: modern prison systems that focus on reform are failing to achieve their intended outcomes; the idea is further supported by Markus Dubber's article, which states that modern penal codes rely largely on treatment theory, which has been largely discredited⁴⁶. The high recidivism rates indicate that the focus on rehabilitation over surveillance and control is not yielding the desired outcomes.

This failure raises serious concerns about the IRAA's 'Second Look Amendment.' Releasing inmates back into society, especially those who were incarcerated as minors, the group that tends to have the highest recidivism rates seems not to be the most rational decision, suggesting that the IRAA's approach may be fundamentally flawed.

Moreover, if the IRAA law becomes more widespread and gains public recognition, it could have the unintended consequence of encouraging criminal behavior among teenagers. The false perception that crimes committed before the age of 18 will result in lighter sentences or early release could incentivize unlawful activities, thereby increasing crime rates.

As stated previously, the Panopticon's principle of surveillance indicates the importance of a psychological deterrent that minimizes the need for physical force or violence. This is particularly important in the context of the juvenile justice system, where the aim should be not just to punish, but to prevent future offenses. The Panopticon, therefore, offers a more effective model for reducing recidivism rates compared to the humanistic approach of the 'Second Look Amendment.' which simply evades the problem of high recidivism rate within the federal justice system instead of proposing a long-term solution.

V. Conclusion

The Incarceration Reduction Amendment Act (IRAA) and its 'Second Look Amendment' undoubtedly come from a place of good intentions. The idea that people, especially those who have made mistakes at a young age, deserve a second chance is a noble one. However, the road to hell is often paved with good intentions. While the prison system is far from perfect in reforming inmates, releasing them back into society without adequate preparation poses risks that cannot be ignored.

⁴⁵ Indiana Department of Correction Juvenile Recidivism Rates, 2021.

⁴⁶ Dubber, Markus. "Theories of Crime and Punishment in German Criminal Law." *American Journal of Comparative Law*, vol. 53, no. 3, 2005, pp. 679-757.

The potential for the IRAA to encourage criminal behavior among teenagers, based on the false perception that they will receive lighter sentences or early release, is a serious issue. This could lead to an increase in crime rates, further destabilizing society and possibly contributing to a potential increase in recidivism rates.

While the IRAA may be well-intentioned, it highlights the urgent need for a more comprehensive approach to prison reform. This should involve interdisciplinary research and the inclusion of experts from various fields such as psychology, philosophy, law, and law enforcement. Only through a multi-faceted approach can we hope to develop a more effective and ethical system for rehabilitating inmates.

In closing, through examining the potential impacts of IRAA through the Panopticon effect put forth by Foucault, this paper comes to the conclusion that the IRAA essentially do not propose a solution to the problem the law is aimed to solve, but instead would possibly contribute to the opposite of its original purpose. This conclusion further emphasizes the idea that ethical philosophy principals could, and should, receive more recognition in contemporary lawmaking to mitigate the potential risks that was not considered in the course of proposing laws.

Works Cited

- "§ 24–403.03. Modification of an Imposed Term of Imprisonment for Violations of Law Committed before 25 Years of Age." *D.C. Law Library*, Council of the District of Columbia, code.dccouncil.gov/us/dc/council/code/sections/24-403.03. Accessed 26 Aug. 2023.
- Bentham, Jeremy, 1748-1832. *The Panopticon Writings*. London ; New York :Verso, 1995.
- "Biannual Reports on Juvenile Arrests." *Metropolitan Police Department of the District of Columbia*, mpdc.dc.gov/page/biannual-reports-juvenile-arrests.
- Britannica, The Editors of Encyclopaedia. "deontological ethics." *Encyclopedia Britannica*, 17 Aug. 2023, <https://www.britannica.com/topic/deontological-ethics>. Accessed 10 September 2023.
- Dubber, Markus Dirk. "Penal Panopticon: The Idea of a Modern Model Penal Code." *Buffalo Criminal Law Review*, Vol. 4, No. 1, April 2000, pp. 53-100.
- Felluga, Dino. "Modules on Foucault: On Panoptic and Carceral Society." *Introductory Guide to Critical Theory*, 2015.
- Foucault, Michel, 1926-1984. *Discipline and Punish : the Birth of the Prison*. New York :Pantheon Books, 1977.
- Gutting, Gary and Johanna Oksala, "Michel Foucault", *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = [<https://plato.stanford.edu/archives/fall2022/entries/foucault/>](https://plato.stanford.edu/archives/fall2022/entries/foucault/).
- Hobbes, Thomas. *Leviathan*. First Avenue Editions TM, 2018.
- Hsin-Wen Lee, "Taking Deterrence Seriously: The Wide-Scope Deterrence Theory of Punishment," *Journal of Research in Crime and Delinquency*, 2017.
- Indiana Department of Correction. "JUVENILE RECIDIVISM 2021." Indiana Department of Correction, 2021, <https://www.in.gov/idoc/files/data-and-statistics/2021.JuvRecidivismRpt-1.pdf>.
- Michael Fiddler, "Phantom architecture: Jeremy Bentham's haunted and haunting panopticon," *Theoretical Criminology*, 2022.
- Sylvia Elias, "Foucault's panoptic disciplinary power and torture: the power of the gaze in Jean-Paul Sartre's No Exit (1944)," *Human Studies*, 2020.

The Potential of Using Neural Networks in the Diagnosis of Cardiovascular Disease By Salil Belgal

Abstract

Cardiovascular diseases are the leading cause of death globally. Early detection and prognosis of these diseases is vital, which is why scientists have turned over to new technologies such as neural networks to speed up the long manual process of identifying key features from both cardiac imaging data and cardiac time-series data that can help doctors predict and diagnose cardiovascular risk. This paper specifically looked at the effectiveness of Convolutional neural networks (CNNs) in analyzing echocardiograms and the effectiveness of Recurrent neural networks (RNNs) in analyzing electrocardiograms and patient records. This paper also explains the basics on how both types of networks are structured, trained, and evaluated. These basics help to understand the existing studies that this paper analyzes to evaluate the effectiveness of both CNNs and RNNs in the cardiovascular healthcare field. The key statistics from the studies suggest that both RNNs and CNNs have great capabilities in helping doctors find key features and predict cardiovascular risk from both cardiac imaging and time-series data.

Introduction

Cardiovascular diseases (CVDs) are the leading cause of death globally, killing an estimated 17.9 million people every year¹. And since early detection and treatment can be the difference between life and death for a cardiovascular patient, it is necessary for scientists to implement ways to detect and treat CVDs more efficiently. This is why neural networks have emerged as an innovative technology in the diagnosis and prognosis of many diseases, including cardiovascular diseases.

Neural networks are inspired by the human brain, mimicking the way that neurons of the human brain signal to one another. Once these networks are trained for accuracy, they can classify data at very high velocity². More specifically, in tasks such as image recognition, neural networks can be way more efficient than manual identification done by human experts². Traditional risk predictors such as the American College of Cardiology/American Heart Association (ACC/AHA) risk model and the Framingham risk score are developed based on clinical data through medical records and patient interviews³. However, these models tend to be inaccurate and inefficient due to the large manpower and time required to properly analyze all of the clinical data provided by patients³. This again is where the fast data classification abilities of neural networks can help healthcare workers. All in all, in the realm of cardiovascular disease, these networks can help with efficiently identifying and understanding the features of cardiac data such as medical imaging data, electrocardiographs, and patient records.

Although neural networks offer a promising path in the treatment of cardiovascular disease, it is necessary to understand how they work, their capabilities, and their limitations in regards to their integration in the fight against cardiovascular disease. The purpose of this paper is to do just that by first explaining the basics of how neural networks work and how they can be

trained to do pattern recognition tasks. Then, this paper will analyze existing literature and studies to determine the effectiveness of neural networks in the interpretation of electrocardiograms(ECGs) and cardiac imaging such as echocardiograms(echos). This paper will also examine existing literature to discuss the ability of neural networks to assist in cardiovascular risk prediction by analyzing patient records and also any limitations or problems involved in using these networks.

Neural Network types

There are two main types of neural networks that I will discuss in this paper, recurrent neural networks (RNNs) and convolutional neural networks(CNNs). A RNN is a type of neural network that uses time series data, which is data that is recorded over consistent intervals of time, which can include a collection of observations over time⁴. A RNN analyzes the characteristics of this data and recognizes patterns to predict the next likely scenario⁵. This is why RNNs can be effective at analyzing both patient records for risk prediction and ECG data for arrhythmia detection. On the other hand, a CNN is a type of neural network that is suited for computer vision tasks involving the processing of pixel data⁶. This makes CNNs the best neural networks for object classification and pattern recognition from images⁶. This means that CNNs can effectively be used to interpret cardiac imaging such as echocardiograms(echos). Before I examine the effectiveness of the application of RNNs and CNNs in the cardiovascular healthcare field, it is first necessary to explain the basics of how both of them work, how they are trained, and how they are tested.

Basics of CNN structure and training

The basic concept of a CNN is that it learns filters of increasing complexity layer by layer. Each feature that we want the model to detect usually has a filter, such as a filter for seeing noses which would give an indication of how strongly a nose appears in an image and also where it appears in the image⁷. A typical CNN has three layers, including the convolution layer, the pooling layer, and the fully connected layer⁷. In the Convolutional layer, the filters, also known as kernels, are moved across the pixels of an image from the top left to the bottom right. The values in the filters are also known as weights⁸. The filters perform a dot product with the source pixel, resulting in a value being stored at each point in the image⁷. The resulting map of values is called a feature map, which exists for each filter. The values on the feature maps are then taken through an activation function, which adds non-linearity to combine the feature maps produced by all the filters and also to decide whether a certain feature exists at a location⁷. The most common activation function used in the convolutional layers is known as ReLu. Typically, the pooling layer is applied after the convolutional layer. The pooling layer takes the feature maps as an input and usually selects the largest values on the feature maps, also known as max pooling which helps the network find outliers⁷. These largest values from the pooling layers can be used as inputs to subsequent layers in the network, which can be more convolutional or pooling layers depending on the task at hand. Usually, the final layer placed before the classification output of a

CNN is the fully connected layer. This layer uses an activation function, most commonly Softmax, to combine elements of the final feature maps to ultimately create the final classification output⁷.

This processing of the image from the input to the output is known as the forward pass of the neural network⁹. However, the CNNs are trained backwards through a process called backpropagation. I will explain the basic concept of how backpropagation works with steps below

Loss function-Backpropagation works by first calculating the loss function at the end of a forward pass, which is essentially a function that shows the error between the predictions shown by the output of the model and the actual truth⁹.

Gradient calculations- The loss function is sent backwards from the output layers to the previous layers of the CNN. Then, the chain rule of calculus is used to find the gradient of the error with respect to each of the weights⁹.

Gradient descent- Optimization algorithms use the gradients to update the weights in order to minimize the loss⁹.

The above backpropagation process is iterative for many forward passes, meaning that backpropagation occurs every time after multiple forward passes of the same dataset in order to tune the model for accuracy¹⁰.

Basics of RNN structure and training

One main attribute of RNNs that differentiates them from regular neural networks is their ability to memorize parts of the inputs in order to make predictions. A typical RNN has three layers: the input layer referred to as $x(t)$, the hidden state referred to as $h(t)$, and the output layer referred to as $o(t)$ ¹¹. The t is the time step, which ranges from 1 to the number of iterations required to process a whole sequence of data. RNNs also have weights that parameterize the connections between the three layers and are also shared across time¹¹. Also, RNNs use activation functions, typically at the hidden state, for the same reason as CNNs- to introduce nonlinearity. The hidden state $h(t)$ serves as the “memory” of the RNN, which is why $h(t)$ depends on both the current input $x(t)$ and also the previous hidden state $h(t-1)$ ¹¹. This memorizing ability is what makes RNNs suitable in analyzing time-series data.

RNNs are trained with a type of backpropagation called backpropagation through time (BPTT). BPTT is very similar to regular backpropagation performed by CNNs, as it still involves a loss function and the calculation of gradients¹². However, in BPTT gradients are calculated backwards at each time step before the gradient descent algorithm is used to update weights¹². Just like for CNN backpropagation, BPTT is iterative.

One of the main issues with the training of regular RNNs is the vanishing gradient problem. This problem can occur because since gradients are multiplied together, as you go backpropagate to previous time steps, the gradients can become vanishingly small¹³. When the gradients become too small, they can't be used to update the weights and hence the model cannot improve its performance¹³. This is why for long-range dependencies, regular RNNs are replaced

by types of RNNs that are structured to diminish the vanishing gradient, including LSTM(Long Short-Term Memory) and GRU (Gated Recurrent Unit)¹³.

Testing of Neural Networks

To understand how classification models like neural networks are tested and evaluated, it is first important to understand the definitions of the following terms: True Positive, True Negative, False Positive, and False Negative. True Positive means that the model correctly identified that a feature is present, True Negative means that the model correctly identified that a feature is not present, False Positive means that the model incorrectly claimed that a feature is present, and False Negative means that the model incorrectly claimed that a feature is not present¹⁴. For example, let's say a model has to predict whether someone is sick or healthy. True positive would mean that the model correctly identified sick people as sick, True Negative would mean that the model correctly identified healthy people as not sick or healthy, False Positive would mean that the model incorrectly identified healthy people as sick, and False Negative would mean that the model incorrectly identified sick people as healthy. In this case, the positive class would be sick people while the negative class would be healthy people. Almost all methods of neural network evaluation use these values. Below is a numbered list of many of the most commonly used evaluation methods and how they are calculated.

Accuracy: Measures overall accuracy of all predictions made by the model. Calculated by doing $(\text{True Positive} + \text{True Negative}) / (\text{Total number of predictions made by model})$ ¹⁵.

Precision: Measures the accuracy of all positive predictions. Calculated by doing $\text{True Positive} / (\text{True Positive} + \text{False Positive})$ ¹⁵.

AUC(Area under the curve): AUC measures the area under the curve with the True Positive rate on the y-axis and the False Positive rate on the x-axis. This test demonstrates how well the model can distinguish between a feature being present or not, with an area of 1 showing a perfect distinguishing ability¹⁵.

Recall: This metric measures the amount of values that are actually predicted positive out of all actually positive values. Calculated by doing $\text{True Positive} / (\text{True Positive} + \text{False Negative})$ ¹⁵.

R-squared value: This evaluation method measures the goodness of fit of a regression-like model by measuring the variance in the predictions explained by the dataset. A value of 1 means that the model perfectly fits the data¹⁶.

Effectiveness of RNNs for Arrhythmia detection and risk prediction

Electrocardiograms(ECGs) record heart rate and heart rhythm, giving doctors important information such as the possibility of arrhythmia¹⁷. In order to determine the effectiveness of RNNs in Arrhythmia detection, I will examine and explain a study that was a part of the International Conference on Computational Intelligence and Data Science issue and is published on ScienceDirect. In their simulation, the authors trained three types of RNNs- a regular RNN, an LSTM, and a GRU- to detect arrhythmia from ECG recordings¹⁸. The data set they used to

train and test the RNNs was the MIT-BIH arrhythmia dataset, which contains 47 ECG records. They performed a 70:30 split on the data, meaning that 70% of the data was used for training while the remaining 30% was used for testing¹⁸. The main evaluation method used by the authors for the RNN models was accuracy. The results showed that for arrhythmia detection the regular RNN had an accuracy of 85.4%, the GRU had an accuracy of 82.5%, and the LSTM had an accuracy of 88.1%¹⁸.

Along with interpreting ECGs for arrhythmia detection, RNNs, especially LSTM models, also have the ability to make cardiovascular risk predictions based on patient records. In order to evaluate RNN effectiveness in making cardiovascular risk predictions based on patient records, I will be reviewing a study published in the BioMed Central Journal where the authors used an LSTM model to predict cardiovascular health trajectories from patient data. The dataset used by the authors is the The Guideline Advantage (TGA) dataset, which contains electronic health records (EHRs) that provided the authors with 14-year longitudinal cardiovascular health measures¹⁹. The authors used this data to train the LSTM model with an 80:20 split, meaning 80 percent of the data was used to train the model and 20 percent was used to test the model¹⁹. The LSTM model was employed to predict cardiovascular health measure categories in five different submetrics, which included smoking status (SMK), A1C levels (A1C), blood pressure (BP), body mass index (BMI), and LDL cholesterol (LDL)¹⁹. The main evaluation method the authors used was AUC. The authors split the patients from the dataset into either the poor, intermediate, or ideal category for each submetric. The authors tested the model and found an AUC value for each category for each of the submetrics. They found out that the LSTM model proved to be accurate in predicting all five submetrics: the micro-average AUC was 0.99 for SMK prediction, 0.97 for BMI, 0.84 for BP, 0.91 for A1C, and 0.93 for LDL prediction¹⁹.

Effectiveness of CNNs in analyzing cardiac imaging

One of the most common cardiac imaging techniques used by doctors is the use of echocardiograms (echos), which checks the structure of the heart, the surrounding blood vessels, blood flow, and the pumping chambers of the heart²⁰. In order to examine the effectiveness of CNN's in analyzing echos, I will review a study published in the NPJ Digital Medicine journal. In this study, the authors trained a CNN model called EchoNet to see if the model could identify cardiac structures, estimate cardiac function, and predict phenotypes that affect cardiovascular risk²¹. They trained the model on a data set of over 2.6 million echo images all from an accredited echocardiography lab²¹. The main evaluation methods used by the authors for EchoNet were AUC and R-squared. EchoNet was able to accurately to identify the presence of many features: pacemaker leads with an AUC of 0.89, enlarged left atrium with an AUC of 0.86, left ventricular hypertrophy with an AUC of 0.75, left ventricular end systolic volume with an R-squared value of 0.74, left ventricular end diastolic volume with an R-squared value of 0.70, and ejection fraction with an R-squared value of 0.50²¹. Lastly, EchoNet was able to also predict phenotypes that affect cardiovascular risk with moderate accuracy: age with an R-squared value

of 0.46, sex with an AUC of 0.88, weight with an R-squared value of 0.56, and height with an R-squared value of 0.33²¹.

The Black Box problem of neural networks

The inability to see how deep learning systems such as neural networks make their decisions is known as the black box problem²². For example, if even after training, a neural network model continues to produce unwanted outcomes, the trainers won't be able to see why exactly the neural network is making bad decisions. In the context of neural network application in healthcare, the black box problem can serve as a big obstacle to the concept of patient trust, where patients want transparency in medical recommendations. The lack of explainability of neural networks can make it very hard for both patients and doctors to understand exactly why the model is predicting or recommending something. There is still ongoing research to try and look into this "box" to find out why neural networks make certain decisions. This ongoing research field is known as "explainable AI", where computer scientists are trying to develop algorithms to make deep learning models like neural networks more transparent and accountable²².

Conclusion

Through the review of the existing studies, it is clear that CNNs have great capabilities in analyzing echocardiograms and RNNs have great capabilities in detecting arrhythmia from electrocardiograms and predicting submetrics that can help doctors assess cardiovascular risk. However, the integration of these neural networks in the cardiovascular healthcare field still has its limitations because the black box problem can make it hard for patients to understand why certain predictions are being made by the model.

Works Cited

- World Health Organization. (2021). Cardiovascular diseases.
https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1
- IBM. (n.d.). What are neural networks? <https://www.ibm.com/topics/neural-networks>
- Kim, J. O., Jeong, Y.-S., Kim, J. H., Lee, J.-W., Park, D., & Kim, H.-S. (2021). Machine Learning-Based Cardiovascular Disease Prediction Model: A Cohort Study on the Korean National Health Insurance Service Health Screening Database. *Diagnostics*, *11*(6), 943.
- IBM. (n.d.). What are recurrent neural networks?
<https://www.ibm.com/topics/recurrent-neural-networks>
- Laskowski, N. (2021). What are recurrent neural networks? TechTarget.
<https://www.techtarget.com/searchenterpriseai/definition/recurrent-neural-networks#:~:text=A%20recurrent%20neural%20network%20is,predict%20the%20next%20likely%20scenario>
- Awati, R. (2023). Convolutional neural network (CNN). TechTarget.
https://www.techtarget.com/searchenterpriseai/definition/convolutional-neural-network?Offer=abMeterCharCount_var1#
- Stewart, M. (2019). Simple Introduction to Convolutional Neural Networks. Towards Data Science.
<https://towardsdatascience.com/simple-introduction-to-convolutional-neural-networks-cdf8d3077bac>
- Shah, S. (2022). Convolutional Neural Network: An Overview. Analytics Vidhya.
<https://www.analyticsvidhya.com/blog/2022/01/convolutional-neural-network-an-overview/#:~:text=The%20filters%20are%20learned%20during,called%20the%20weights%20of%20CNN.&text=A%20feature%20map%20is%20a,inputs%20with%20the%20same%20weights>
- Roy, R. (2022). Neural Networks: Forward Pass and Backpropagation. Towards Data Science.
<https://towardsdatascience.com/neural-networks-forward-pass-and-backpropagation-be3b75a1cfcc>
- Arie, L. G. (2021). Neural Network Backpropagation made easy. Towards Data Science.
<https://towardsdatascience.com/neural-networks-backpropagation-by-dr-lihiur-arie-27be67d8fdce>
- Nabi, J. (2019). Recurrent Neural Networks (RNNs). Towards Data Science.
<https://towardsdatascience.com/recurrent-neural-networks-rnns-3f06d7653a85>
- Brownlee, J. (2020). A Gentle Introduction to Backpropagation Through Time. Machine Learning Mastery.
<https://machinelearningmastery.com/gentle-introduction-backpropagation-time/>
- Patil, S. (2023). Vanishing Gradient Problem in RNNs. Medium.
<https://medium.com/@sagarpatiler/vanishing-gradient-problem-in-rnns-d362235005c>

- Dilmegani, C. (2022). Machine Learning Accuracy: True-False Positive/Negative. AIMultiple. <https://research.aimultiple.com/machine-learning-accuracy/>
- Mankad, S. (2020). A Tour of Evaluation Metrics for Machine Learning. Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2020/11/a-tour-of-evaluation-metrics-for-machine-learning/>
- Kharwal, A. (2021). R2 Score in Machine Learning. The Clever Programmer. <https://thecleverprogrammer.com/2021/06/22/r2-score-in-machine-learning/>
- Mayo Clinic Staff. (2022). Electrocardiogram (ECG or EKG). Mayo Clinic. <https://www.mayoclinic.org/tests-procedures/ekg/about/pac-20384983>
- Singh, S., Pandey, S. K., Pawar, U., & Janghel, R. R. (2018). Classification of ECG Arrhythmia using Recurrent Neural Networks. *Procedia Computer Science*, 132, 1290-1297.
- Guo, A., Beheshti, R., Khan, Y. M., Langabeer II, J. R., Foraker, R. E. (2021). Predicting cardiovascular health trajectories in time-series electronic health records with LSTM models. *BMC Medical Informatics and Decision Making*, 21(1), 5.
- NHS. (2022). Echocardiogram. <https://www.nhs.uk/conditions/echocardiogram/>
- Ghorbani, A., Ouyang, D., Abid, A., He, B., Chen, J. H., Harrington, R. A., ... Zou, J. Y. (2020). Deep Learning interpretation of echocardiograms. *npj Digital Medicine*, 3, 10.
- Blouin, L. (2023). AI's mysterious "black box" problem, explained. University of Michigan-Dearborn. <https://umdearborn.edu/news/ais-mysterious-black-box-problem-explained>

Art of Healing – A Comprehensive Review of Visual Art Interventions in Dementia Care and their Influence on Memory, Quality of Life and Well-Being By Yejie Choi

Abstract

Dementia is a neurocognitive disease that impairs memory function and drastically reduces quality of life and well-being (Hoe et al., 2009). This systematic review aims to examine the effect of art therapy on memory, quality of life, and well-being in patients with dementia. We collected articles reporting on the impact of art therapy for people with dementia from ResearchGate, Google Scholar, Scopus, PubMed and ScienceDirect with the keywords “dementia”, “art therapy”, “well-being”, “quality of life”, and “memory” guiding the search. We found that art therapy interventions improve memory with a small average effect size of 0.32 ($n=5$), quality of life with a small average effect size of 0.25, and well-being with a small average effect size of 0.37 in people with dementia. We also examined the difference between interventions focused on viewing compared to making art and discovered that while viewing art has a greater impact on memory (t-test = 0.000575893695, $p < 0.05$), making art improves quality of life (t-test = 0.08829026805, $p < 0.05$) and well-being (t-test = 0.03780487276, $p < 0.05$) more than viewing art based on this review. Additionally, only participating in viewing art or only participating in making art has a greater impact than simultaneously participating in viewing and making in not only memory (t-test = 0.001240643861, $p < 0.05$) but also quality of life (t-test = 0.01061624383, $p < 0.05$) and well-being (t-test = 0.0004410984931, $p < 0.05$). Future studies should include a randomized control group and explicitly examine the difference between viewing and making art for people with dementia. Based on current evidence, we recommend further development of art therapy interventions focused on making art.

Introduction

Dementia refers to an impaired ability to think, remember, and reason that interferes with everyday life. This is a widespread and debilitating condition that significantly impacts the lives of many individuals. One in 10 older Americans have dementia, the most common types being Alzheimer’s disease, vascular dementia, and fronto-temporal dementia (U.S. Department of Health and Human Services [HHS], n.d.). Comparing males and females, 13 out of 1,000 developed Alzheimer’s each year compared to only seven men (Niu, 2016). The underlying cause of this neurodegenerative disease is attributed to structural changes in the brain, related to the abnormal accumulation of proteins known as amyloid plaques and tau tangles (Mayo Foundation for Medical Education and Research, 2023).

While a definitive cure remains elusive for most types of dementia, there are commonly prescribed medications, such as cholinesterase inhibitors and memantine, designed to temporarily provide relief from the associated symptoms. However, these pharmacological approaches often come with side effects, such as nausea, vomiting, and, in some serious cases,

even death (Ali, 2015). Medication is sometimes paired with occupational therapy for those struggling with severe symptoms and behavior problems (O'Brien, 2017). Occupational therapy involves one-on-one sessions with a therapist to develop strategies for improving day-to-day function in patients with dementia through adaptation of their environment. Both medication and occupational therapy can be effective, but may be expensive and difficult to access. Many patients continue to suffer from severe memory impairment, reduced quality of life, and poor well-being (World Health Organization, n.d.).

Ongoing research explores various non-pharmacological approaches to enhance symptom management and the overall quality of life for those affected by dementia. These strategies typically use group class formats to offer activities that are mentally stimulating, require sustained attention, and are enjoyable.

Art-based interventions represent a relatively recent development that holds promise as a potential therapeutic approach for addressing dementia. Different research groups have developed art interventions centered around museum visits, creative art therapy, touch-screen or computer-based art, sculpture-based classes, and more. In small-scale studies, preliminary results suggest that these interventions have positive effects on enhancing cognition, quality of life, and well-being. Current literature examines how art interventions may improve engagement of the patients and their relationship with caregivers. This review offers novel insight into how specific types of art interventions that focus on making vs. viewing art may offer different benefits in terms of memory, wellbeing, and quality of life.

Methodology

The current literature review was conducted through an extensive search of online databases such as PubMed and ResearchGate, applying a wide range of search terms like “dementia”, “art interventions”, “memory”, “quality of life”, and “well-being”. The inclusion criteria for experiments in this review was:

- 1) used standardized questionnaires to assess memory and/or quality of life, and/or well-being,;
- 2) included a pre-post testing strategy to evaluate the results;
- 3) examined effects on people with dementia (sometimes including caretakers of the patients);
- 4) focused on visual art, including both the making of art and the experience of viewing artworks; and
- 5) had interventions in a group setting. Given the scarcity of research done around this topic, control groups were not required. Notably, individuals with substantial psychiatric disorders and chronic illnesses, such as but not limited to schizophrenia, bipolar disorder, and clinically significant depression disorder or anxiety, were not included in the review. This review included 20 experiments that shed light on the effects of visual art interventions on memory, quality of life, and wellness for people with dementia

Population

With an average age of 79.57 and an age range spanning from 51 to 102 years, the studies included in this review have a diverse spectrum of elderly individuals diagnosed with dementia. The data reflects memory, quality of life, and well-being in people with dementia across different stages of aging. Most of the participants in these experiments are ones with mild to moderate levels of dementia, suggesting that these individuals are better suited for active participation in the interventions compared to ones with severe dementia. Females account for greater than 60% of the study population, which closely reflects the distribution of people with dementia in the population by sex. Notably, there was an absence of specification of race in 11 out of the 20 studies. Within the population of the nine studies that did specify the race of participants, 92.78% were Caucasian, 2.09% Asian, 2.93% African-American, and 0.84% were from other racial backgrounds.

Study Designs

In examining the study designs across the 20 experiments reviewed in this paper, several key findings emerged. The average sample size was approximately 29 participants, though there was substantial variability with the largest study encompassing 215 individuals with dementia and the smallest involving a mere 6 individuals. The majority of these studies did not incorporate control groups, with only 6 studies out of the 20 implementing a control group to compare the intervention group with. Most of the experiments were within-person designs rather than between-person designs. Additionally, it was observed that 11 of the studies incorporated caregivers, likely due to the necessity of their involvement in supervising and supporting patients during the span of the interventions. The presence of caregivers can enhance the comfort of patients participating in these experiments. All of the studies featured in this paper pertain to art interventions, covering various forms of art such as art-museum based interventions, art gallery interventions, visual arts programs, art training programs, sculpture-based interventions, art-making activities (including painting, coloring, and drawing), and computer-based interventions (digital art). Furthermore, the average total intervention duration across the studies was approximately 28 hours, with a significant number of interventions being a minimum of six weeks, conducted at least once per week. Finally, the observation that all the studies included were published within the last 20 years suggests that this is a very new and relatively undeveloped field that requires further funding and attention.

Outcomes

This review focused on the examination of three core aspects of patients with dementia: memory, quality of life, and well-being. Memory-related assessments were conducted using various questionnaire

including the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog), Mini-Addenbrooke's Cognitive Examine (M-ACE), Montreal Cognitive Assessment (MoCA) Delayed Recall, Mini-Mental State Examination (MMSE), and the Quality of Life Questionnaire for People with Dementia (DEMQOL), which also measures aspects related to memory. In cases where the quantitative memory assessment was not applied, observational techniques such as Frequency of Code Occurrence and facilitator observations were used. Quality of life was comprehensively measured via the Quality of Life in Alzheimer's Disease Scale (QoL-AD) and DEMQOL. The measure of well-being encompassed a diverse array of questionnaires, such as the Geriatric Depression Scale (GDS), General Well-being Questionnaire (GWQ), the Greater Cincinnati Well-being Scales (CWS), Bristol Activities of Daily Living Scales (BADLS), Visual Analog Scale for Well-being (VAS-well), and, when quantitative data was not applicable, custom-designed questionnaires were also used. These comprehensive assessments enabled detailed results as to how visual art interventions affect memory, quality of life, and well-being in individuals with dementia.

While the majority of experiments showed improvement in the outcomes of interest, they did not have a big enough effect size to be considered significant. This is why they are categorized as "no effect" in Table 1. Memory improvements were observed in 40% of the studies (with small to medium effect sizes), yet it is crucial to recognize that the majority of studies did not focus on memory assessment. In terms of quality of life, only 25% of the studies demonstrated results indicative of enhanced quality of life after the interventions. Conversely, 45% of the studies reviewed found improvements in wellbeing measures, suggesting that nearly half of the experiments recorded better well-being among the patients post-intervention in comparison to the baseline data. However, it is important to note that these percentages may not fully capture the outcomes or the effect of art interventions on patients with dementia accurately due to variations in the measurement tool across studies.

Discussion

Art interventions have the potential to improve the memory, quality of life, and well-being of dementia patients. Analysis of the effect sizes from the various experiments reviewed revealed a discernible pattern: engagement in viewing visual art exhibits had a positive impact on memory, while active art creation had benefits in improvement of quality of life and overall well-being. Moreover, the findings suggest that while viewing art has a greater impact on memory (t-test = 0.000575893695, $p < 0.05$), making art improves quality of life (t-test = 0.08829026805, $p < 0.05$) and well-being (t-test = 0.03780487276, $p < 0.05$) more than viewing art. In addition to the difference between viewing and making art, doing both viewing and making was compared to only doing one. This comparison suggests that only participating i

viewing art or only participating in making art has a greater impact than simultaneously participating in viewing and making in not only memory (t-test = 0.001240643861, $p < 0.05$) but also quality of life (t-test = 0.01061624383, $p < 0.05$) and well-being (t-test = 0.0004410984931, $p < 0.05$).

Another interesting observation arising from the experimental outcomes pertains to the substantial predominance of female patients, as greater than 60% of the study population were women. Although this may seemingly necessitate gender-specific considerations, it is essential to understand this gender disparity within the broader epidemiological context. Women are known to face a higher susceptibility to developing dementia, particularly Alzheimer's disease, over the course of their lifetimes (Alzheimer's Society, n.d.). Consequently, the “overrepresentation” of females in these studies may solely be a reflection of this demographic reality.

A vitally important limitation regarding this topic of research is the fact that patients with severe dementia are not able to participate in these interventions, because this level of cognitive impairment presents obstacles to active participation in visual art-based interventions, particularly those involving creative engagement, although viewing interventions are still important and interventions should aim to include them. Patients diagnosed with severe dementia may encounter challenges in comprehending and executing artistic and creative tasks due to the advanced nature of their condition. This limitation takes on a central role given the heightened urgency in meeting the unique requirements of individuals with severe dementia, for whom the imperative for effective interventions becomes even more pronounced. This limitation displays the underlying complexity of conducting research in this population, necessitating the development of methodologies and accommodations to ensure inclusion and enable the exploration of tailored interventions that resonate with the patients' specific challenges and requirements.

Additionally, another limitation among the studies reviewed in this paper was the absence of specification of race in 11 out of the 20 studies. This observation may indicate that studies are underpowered to examine differences in outcomes by race, as the population of these experiments are rather small to accurately figure out whether there are various outcomes for various races or not. As mentioned in the outcomes, throughout all the studies, 92.78% of the population were Caucasian, 2.09% Asian, 2.93% African-American, and 0.84% were from other racial backgrounds. This evident lack of representation of individuals of color in dementia research emphasizes the urgent need for more diverse and inclusive studies in regions where the majority of the population are not White. Future studies should include larger samples and examine race-specific dynamics

The review of the 20 experiments in the realm of visual art interventions for individuals with dementia reveals the power of viewing art to improve memory. When individuals with dementia engage in viewing art, they often experience sensory and emotional stimulation, which can lead to enhanced memory retention (Deshmukh, 2018). Furthermore, the aesthetic and emotional qualities of art may evoke personal memories and emotional connections, contributing to the observed memory-enhancing effects. Overall, these findings underscore the therapeutic potential of art as a non-pharmacological intervention for individuals with dementia, especially its ability to enhance memory function. Art intervention based treatments are a considerable alternative to medications and occupational therapy, which can be expensive and hard to access. Although there is some evidence showing the positive effect of visual art interventions on patients with dementia, further research involving large sample sizes and randomized control groups should be conducted in order to gain deeper insights into the favorable results as an outcome of these interventions.

Art interventions are innovative, holistic strategies for dementia care. The enhancement of memory, quality of life, and well-being in patients with dementia is of growing interest should be of top priority, which is why this is a field in need of more research and exploration.

Conclusion

This comprehensive literature review of visual art interventions for dementia patients highlights consistent patterns in the literature. Art appreciation appears to have a positive impact on memory, while active art creation is associated with improved quality of life and well-being for individuals with dementia. However, the review also underscores the challenge of including patients with severe dementia in art-based interventions, emphasizing the need for tailored approaches. Additionally, the lack of racial diversity in study populations calls for more inclusive research efforts. Overall, while promising, further research with larger sample sizes and randomized control groups is essential to deepen our understanding of the potential benefits of visual art interventions in dementia care.

Works Cited

- Camic, P., Baker, E., & Tischler, V. (2015). Theorizing How Art Gallery Interventions Impact People With Dementia and Their Caregivers. *The Gerontologist*, 56(6), 10.1093/geront/gnv063.
- Camic, P., Tischler, V., & Pearman, C. (2013). Viewing and making art together: A multi-session art-gallery-based intervention for people with dementia and their carers. *Aging & mental health*, 18. 10.1080/13607863.2013.818101.
- Centers for Disease Control and Prevention. (2019, April 5). What is dementia?. Centers for Disease Control and Prevention. <https://www.cdc.gov/aging/dementia/index.html#:~:text=Dementia%20is%20not%20a%20specific,most%20common%20type%20of%20dementia>. D'Cunha, N. M., McKune, A. J., Isbel, S., Kellett, J., Georgousopoulou, E. N., & Naumovski, N. (2019).
- Psychophysiological Responses in People Living with Dementia after an Art Gallery Intervention: An Exploratory Study. *Journal of Alzheimer's disease : JAD*, 72(2), 549–562. <https://doi.org/10.3233/JAD-190784>.
- Eekelaar, C., Camic, P. M., & Springham, N. (2012). Art galleries, episodic memory and verbal fluency in dementia: An exploratory study. *Psychology of Aesthetics, Creativity, and the Arts*, 6(3), 262-272. doi:<https://doi.org/10.1037/a0027499>.
- Flatt, J. D., Liptak, A., Oakley, M. A., Gogan, J., Varner, T., & Lingler, J. H. (2015). Subjective experiences of an art museum engagement activity for persons with early-stage Alzheimer' disease and their family caregivers. *American journal of Alzheimer's disease and other dementias*, 30(4), 380–389. <https://doi.org/10.1177/1533317514549953>.
- Graham, M. & Fabricius, A. (2017). Painting in situ : the benefits of live mural painting for residents on two secure long-term care dementia units. *Arts & Health*. 10. 1-17. 10.1080/17533015.2017.1370717.
- Hattori, H., Hattori, C., Hokao, C., Mizushima, K., & Mase, T. (2011). Controlled study on the cognitive and psychological effect of coloring and drawing in mild Alzheimer's disease patients. *Geriatrics & gerontology international*, 11(4), 431–437. <https://doi.org/10.1111/j.1447-0594.2011.00698>.
- Hendriks, I., Meiland, F. J. M., Slotwinska, K., Kroeze, R., Weinstein, H., Gerritsen, D. L., & Dröes, R. M. (2019). How do people with dementia respond to different types of art? An explorative study into interactive museum programs. *International psychogeriatrics*, 31(6), 857–868. <https://doi.org/10.1017/S1041610218001266>.
- Hoe, J., Hancock, G., Livingston, G., Woods, B., Challis, D., & Orrell, M. (2009, July 23). Changes in the quality of life of people with dementia living in Care Homes. *Alzheimer*

- disease and associated disorders.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2759656/>.
- Johnson, K., LaPlume, A., & Wiseheart, M. (2020). Art Training in Dementia: A Randomized Controlled Trial. *Frontiers in Psychology*. 11. 585508. 10.3389/fpsyg.2020.585508.
- Kinney, J. M., & Rentz, C. A. (2005). Observed well-being among individuals with dementia: Memories in the Making, an art program, versus other structured activity. *American journal of Alzheimer's disease and other dementias*, 20(4), 220–227.
<https://doi.org/10.1177/153331750502000406>.
- MacPherson, S., Bird, M., Anderson, K., Davis, T., & Blair, A. (2009). An art gallery access programme for people with dementia: 'you do it for the moment'. *Aging & mental health*, 13(5), 744–752. <https://doi.org/10.1080/13607860902918207>
- Mayo Foundation for Medical Education and Research. (2023, August 30). Dementia. Mayo Clinic.
<https://www.mayoclinic.org/diseases-conditions/dementia/diagnosis-treatment/drc-20352019>.
- McGuigan, K., Legget, J., & Horsburgh, M. (2015). Visiting the museum together: Evaluating a programme at Auckland Museum for people living with dementia and their carers. *Arts & Health*. 7. 1-10. 10.1080/17533015.2015.1045531.
- O'Brien, J. T., Holmes, C., Jones, M., Jones, R., Livingston, G., McKeith, I., Mittler, P., Passmore, P., Ritchie, C., Robinson, L., Sampson, E. L., Taylor, J. P., Thomas, A., & Burns, A. (2017). Clinical practice with anti-dementia drugs: A revised (third) consensus statement from the British Association for Psychopharmacology. *Journal of psychopharmacology (Oxford, England)*, 31(2), 147–168.
<https://doi.org/10.1177/0269881116680924>.
- Savazzi, F., Isernia, S., Farina, E., Fioravanti, R., D'Amico, A., Saibene, F., Rabuffetti, M., Gilli, G., Alberoni, M., Nemni, R., & Baglio, F. (2020). “Art, Colors, and Emotions” Treatment (ACE-t): A Pilot Study on the Efficacy of an Art-Based Intervention for People With Alzheimer’s Disease. *Frontiers in Psychology*. 11. 10.3389/fpsyg.2020.01467.
- Schall, A., Tesky, V. A., Adams, A. K., & Pantel, J. (2018). Art museum-based intervention to promote emotional well-being and improve quality of life in people with dementia: The ARTEMIS project. *Dementia (London, England)*, 17(6), 728–743.
<https://doi.org/10.1177/1471301217730451>.
- Seifert, K., Spottke, A., & Fliessbach, K. (2017). Effects of sculpture based art therapy in dementia patients—A pilot study. *Heliyon*. 3. e00460. 10.1016/j.heliyon.2017.e00460.
- Strohmaier, S., Homans, K.M., Hulbert, S. et al. Arts-based interventions for people living with dementia: Measuring ‘in the moment’ wellbeing with the Canterbury Wellbeing Scales [version 3; peer review: 2 approved]. *Wellcome Open Res* 2021, 6:59 (<https://doi.org/10.12688/wellcomeopenres.16596.3>)

- Tyack, C., Camic, P. M., Heron, M. J., & Hulbert, S. (2017). Viewing Art on a Tablet Computer: A Well-Being Intervention for People With Dementia and Their Caregivers. *Journal of applied gerontology : the official journal of the Southern Gerontological Society*, 36(7), 864–894. <https://doi.org/10.1177/0733464815617287>
- U.S. Department of Health and Human Services. (n.d.). What is dementia? symptoms, types, and diagnosis. National Institute on Aging. <https://www.nia.nih.gov/health/what-is-dementia>.
- Why is dementia different for women?. Alzheimer's Society. (n.d.). <https://www.alzheimers.org.uk/blog/why-dementia-different-women#:~:text=Women%20have%20a%20greater%20risk%20factor%20for%20this%20disease>.
- Windle, G., Joling, K. J., Howson-Griffiths, T., Woods, B., Jones, C. H., van de Ven, P. M., Newman, A., & Parkinson, C. (2018). The impact of a visual arts program on quality of life, communication, and well-being of people living with dementia: a mixed-methods longitudinal investigation. *International psychogeriatrics*, 30(3), 409–423. <https://doi.org/10.1017/S1041610217002162>.
- World Health Organization. (n.d.). Dementia. World Health Organization. <https://www.who.int/news-room/fact-sheets/detail/dementia#:~:text=People%20with%20dementia%20may%20not,well%20as%20those%20around%20them>.
- Young, R., Tischler, V., Hulbert, S., & Camic, P. (2015). The Impact of Viewing and Making Art on Verbal Fluency and Memory in People With Dementia in an Art Gallery Setting. *Psychology of Aesthetics Creativity and the Arts*. 9. 10.1037/aca0000030.

Applications of Atom Manipulation in Scanning Tunneling Microscopy By Nikita Chanda

Abstract

Scanning Tunneling Microscopy (STM) stands at the forefront of nanotechnology, enabling the precise probing and manipulation of surfaces at the atomic and molecular levels. Originating in the 1980s, STM has catalyzed breakthroughs in materials science, quantum mechanics, and molecular chemistry. This literature review explores the current techniques in STM atom manipulation, delving into lateral manipulation, vertical manipulation, electric field manipulation, inelastic tunneling manipulation, and electron-induced manipulation. Through an examination of these techniques, the review highlights the broad applications in constructing quantum structures, creating quantum dots, inducing controlled chemical reactions, and manipulating near-field properties. The study encompasses the latest developments in each category, emphasizing their impact on nanoscience and nanotechnology. The comprehensive analysis provided herein aims to inspire further innovations and advancements in the exploration of atomic-scale phenomena.

Introduction

Standing at the forefront of nanotechnology, scanning tunneling microscopy (STM) enables researchers to probe and manipulate surfaces at both the atomic and molecular level. This concept first emerged during the early 1980s, when it was invented by IBM scientists Gerd Binnig and Heinrich Rohrer. In 1986, this groundbreaking discovery led them to earn the Nobel Prize in Physics, opening up the world to breakthroughs in materials science, molecular chemistry, quantum mechanics, and more(Oxford Instruments). Scanning tunneling microscopy has made a significant impact on the world, from the creation of nanostructures to the world's smallest movie, a film produced by IBM researchers by manipulating carbon monoxide molecules and magnifying them 100 million times(IBM).

This literature review seeks to investigate the current techniques used in STM atom manipulation, explore the applications of STM manipulation in various fields, and assess the impact of these advances on the understanding of atomic-scale phenomena. In the following sections, we discuss the different types of atom manipulation techniques, as well as their unique applications, focusing on 5 main areas: lateral manipulation, vertical manipulation, electric field manipulation, inelastic tunneling manipulation, and electron-induced manipulation. By examining these fields of STM manipulation, this review aims to bring to light the impact these techniques have on nanoscience and nanotechnology, hoping to inspire further innovations in the science of the very, very small.

Lateral Manipulation

Lateral manipulation is perhaps the most popular method of STM manipulation. This technique involves the creation of a temporary atom-tip attractive force between the probe tip atoms and the adatom. The atom is then moved across the surface to a new position(Celotta et

al). Finally, the STM tip is withdrawn to a point where the force between the adatom and the tip is negligible, leaving the atom bound to the final location on the surface(Eigler & Schweizer).

The first instance of lateral manipulation was used to position individual xenon atoms on a single-crystal nickel surface, building the “IBM” company logo(Eigler & Schweizer). Since then, lateral manipulation has proven to be a powerful tool with various applications that extend beyond mere surface manipulation. The following sections explore the significance of lateral manipulation through its applications in constructing quantum corrals, quantum dots, and inducing chemical reactions.

A. Quantum Corrals

Lateral manipulation has played a pivotal role in the creation of quantum corrals, two-dimensional structures built atom by atom on a clean metallic surface. The seminal work by Crommie et al. in 1993 marked the first instance of building quantum corrals with iron atoms on a copper surface. These corrals confine electrons in the surface states of noble metals, leading to standing-wave patterns and discrete resonances. Quantum corrals serve as a unique platform for studying confined electrons' properties, interaction with adsorbates, and size quantization effects. Furthermore, recent studies by Jolie et al. (2022) extended the exploration of quantum corrals to unique surface alloys, introducing the concept of Rashba surface alloy, offering new possibilities for creating artificial lattices with distinct properties.

B. Quantum Dots

Lateral manipulation has been instrumental in creating quantum dots with confined electronic states. By removing a single CO molecule from the lattice on a Cu(111) surface, interacting quantum dots are formed. Each cell bounded by four CO molecules acts as a confining barrier, and the removal of a single CO molecule enlarges the confinement area, resulting in stronger intensity in confined quantum states. This controlled manipulation of quantum dots demonstrates the potential for efficient single atom and molecule manipulation. The interaction between multiple quantum dots, showcasing bonding and anti-bonding states, offers possibilities for applications in quantum information processing and nanophotonics.

C. Chemical Reactions

The soft lateral manipulation technique has been applied to induce chemical reactions with the STM tip. This technique relies on tip-adsorbate forces for controlled movement, allowing the STM tip to interact with adsorbates on the surface and influence their positions. This approach was demonstrated in a study by Hla et al. (2000), where a chemical reaction was induced by manipulating phenyl molecules along a straight step edge on a Cu(111) surface. The ability to precisely control the movement of molecules during chemical reactions opens new avenues for single-molecule engineering.

Conclusion

In summary, lateral manipulation in STM is a versatile technique with applications ranging from constructing quantum structures to creating quantum dots and inducing controlled chemical reactions. Its significance lies in its ability to provide unprecedented control over atomic-scale manipulation, offering insights into quantum phenomena and inspiring advancements in nanoscience and nanotechnology.

I. Vertical Manipulation

Vertical manipulation is characterized by the transfer of individual atoms or molecules between the scanning tunneling microscope tip and the sample surface. While lateral manipulation orchestrates horizontal movements, VM involves the transfer of individual atoms or molecules between the scanning tunneling microscope tip and the sample surface. This nuanced technique distinguishes itself by engaging diverse mechanisms, including the application of electric fields, inelastic tunneling electrons, and direct tip-surface contact. Unlike lateral manipulation, which relies on creating temporary forces between the STM probe tip and adatoms for lateral movements, VM takes a step further by venturing into the controlled transfer of atoms in a vertical direction. This distinction establishes VM as a versatile tool for creating structures and exploring phenomena that may prove challenging for lateral manipulation. Applications of VM abound, from the realization of atomic switches and bistable switches to the assembly of nanostructures and the construction of atomic-scale quantum qubit platforms. Pioneering studies by Hla et al. (2005), Eigler et al. (1991), Celotta et al. (2014), and Wang et al. (2023) showcase the potential of VM across diverse domains, laying the groundwork for future explorations in nanoscience and technology. As researchers explore the potentials of vertical manipulation, it is poised to play a pivotal role in advancing our understanding and applications at the atomic and molecular scales.

A. Atomic Switches and Bistable Switches

Hla et al. (2005) introduced an "atomic switch" realized through repeated transfer of a xenon (Xe) atom between the STM tip and a nickel substrate. This mode utilized electric fields, inelastic tunneling electrons, and tip-atom/molecule mechanical contact to transfer atoms between the tip and the surface. The reversible transfer of atoms between the tip and substrate demonstrated the feasibility of creating a novel class of electronic devices known as "atom switches." Eigler et al. (1991) pioneered vertical manipulation by creating a bistable switch using xenon atoms. This bistable switch operated by moving a single xenon atom reversibly between stable positions on two stationary conducting 'leads,' corresponding to the STM tip and nickel surface. This innovative approach laid the foundation for atom switches, offering potential applications in atomic-scale electronic devices.

B. Nanostructure Assembly

Celotta et al. (2014) explored the assembly of atomically perfect nanostructures using

VM. This involved strengthening the probe-adatom tip bond to pick up an adatom from the surface, repositioning it solely with the probe tip, and placing it back onto the surface. While less reliable than lateral manipulation, this method facilitated the construction of simple nanostructures and enabled experiments on quantum magnetic interactions.

C. Quantum Qubit Platforms

Wang et al. (2023) utilized VM for the construction of an atomic-scale multi-qubit system. By arranging titanium atoms on a magnesium oxide surface and employing the STM tip for vertical manipulation, the researchers demonstrated coherent operations and readout of electron-spin qubits. This approach showcased the potential for practical applications in quantum information processing, emphasizing the versatility of VM in creating atomic-scale quantum architectures.

Conclusion

Vertical manipulation within STM transcends the limitations of lateral manipulation by venturing into the vertical dimension to unlock new possibilities in atom and molecule manipulation. From the creation of atomic switches to the assembly of nanostructures and the realization of quantum qubit platforms, VM contributes to the expanding landscape of nanoscience and technology. As researchers delve deeper into the potentials of vertical manipulation, it is poised to play a pivotal role in advancing our understanding and applications at the atomic and molecular scales.

IV. Electric Field Induced Manipulation

Electric Field Induced Manipulation (EFIM) represents a cutting-edge paradigm in the realm of nanoscience and technology, harnessing the power of electric fields to precisely control the position and behavior of individual atoms and molecules. This innovative technique, pioneered by Saw-Wai Hla and further explored by subsequent researchers, leverages the voltage applied between the scanning tunneling microscope (STM) tip and the sample to induce a host of controlled movements and reactions at the atomic and molecular scales.

At its core, EFIM relies on the manipulation of single atoms and molecules through the modulation of the electric field between the tip and the sample. The key mechanism involves changing the bias polarity—alternating between positive and negative charges—resulting in attractive or repulsive forces that influence the interaction between the STM tip and the atom or molecule on the surface. This dynamic interplay allows scientists to navigate and reposition these entities with unprecedented precision.

The impact of EFIM extends beyond mere repositioning, encompassing intricate processes such as bond breaking and complex reactions. By applying high electric fields, bond dissociation becomes feasible, opening avenues for controlled displacement, dissociation, and fragmentation of molecules. Additionally, EFIM enables structural rearrangement on surfaces, offering a powerful tool for creating new atomic and molecular arrangements.

A. Precision Movement

Saw-Wai Hla's seminal work in 2005 laid the groundwork for precision movements of individual atoms and molecules through EFIM. By toggling bias polarity, attractive or repulsive forces induced by the electric field between the STM tip and sample allowed for controlled movements. The study demonstrated the versatility of EFIM in facilitating lateral and vertical manipulation, showcasing its potential for atomically precise relocation on surfaces.

B. Bond Breaking and Complex Reactions

EFIM's impact extends beyond mere movement, as evidenced by the high electric fields (>3 V) demonstrated by Saw-Wai Hla, enabling bond breaking of molecules. The intricate study by G. Dujardin et al. in 1992 delved into the dissociation of individual molecules with electrons from the STM tip. This highlighted EFIM's role in inducing complex reactions, such as displacement, dissociation, and fragmentation of molecules. The ability to control these processes through carefully tuned bias voltages adds a layer of sophistication to the manipulation toolkit.

C. Structural Rearrangement and Surface Modification

Devel et al.'s theoretical exploration in 1995 emphasized the self-consistent aspects of EFIM, revealing that the last atoms of the STM tip apex create a strong electric field. This insight underlines EFIM's role in structural rearrangement, offering a means to create new atomic or molecular arrangements on surfaces. The capacity to induce modifications at the molecular scale opens avenues for tailoring surface structures and properties with unprecedented precision.

D. Coherent Control over Electrons at Atomic Scales

M. Garg and K. Kern's work in 2020 introduced EFIM as a pivotal tool for coherent control over electrons at atomic length and time scales. Through precise tuning of the carrier-envelope phase of optical pulses, the study explored photon-driven and field-driven tunneling regimes. This highlights EFIM's potential for manipulating electron behavior with sub-femtosecond resolution. The demonstrated capability to control electron tunneling at extremely low time scales positions EFIM as a cornerstone for advancing nanoelectronics and microscopy.

Conclusion

The applications discussed reveal the multifaceted capabilities inherent in Electric Field Induced Manipulation (EFIM). From orchestrating precision movements and effecting bond dissociation to enabling intricate structural rearrangements and facilitating coherent control over electrons, EFIM emerges as a transformative technique within scanning tunneling microscopy. This categorization not only delineates the diverse roles EFIM assumes but also underscores its profound impact. The collective body of evidence presented herein attests to EFIM's

instrumental role in reshaping our understanding and application of STM techniques, facilitating significant progress in nanoscience and technology.

V. Inelastic Tunnelling Induced Manipulation

In the dynamic field of scanning tunneling microscopy (STM), the use of tunneling electrons has evolved significantly, playing a crucial role in both identifying and manipulating molecules. A particularly intriguing avenue of progress is the exploration of inelastic tunneling induced manipulation (ITIM) and its multifaceted applications. Inelastic tunneling involves a process where electrons experience energy changes during tunneling events, thereby influencing the manipulation and analysis of molecules.

This section seeks to offer a thorough understanding of the groundbreaking research and wide-ranging applications of ITIM by delving into key literature. It illuminates the strides made in the manipulation and identification of molecules through tunneling electrons, specifically highlighting the intricate processes involved in inelastic tunneling that contribute to the advancement of scanning tunneling microscopy.

A. Desorption Induced by Tunneling Current Injection

Examining the execution of electron-induced changes in adsorbates using tunneling electrons with STM, highlighting the benefits of atomic-scale resolution and the limited number of molecules required for investigation (Devel et al., 1995). The research team meticulously controlled the tunneling current to induce changes in adsorbates, showcasing the benefits of atomic-scale resolution. The limited number of molecules required for investigation emphasized the precision achievable through this method.

B. Vibration-Induced Hopping and Chemical Reaction

Exploring surface phenomena induced by vibration-excitation, including hopping of adsorbates on the surface and rotation (Komeda et al., 2005). Additionally, the discussion extends to the long-standing efforts to control chemical reactions with non-thermal processes, focusing on vibration as the reaction coordinate. The researchers meticulously investigated surface phenomena induced by vibration-excitation, shedding light on hopping of adsorbates and rotational motion. The study delved into the historical context of efforts to control chemical reactions, highlighting the role of vibration as a crucial reaction coordinate.

C. Study Focus on C60 Molecules on Cu(111)

A detailed investigation into the properties of C60 molecules adsorbed on Cu(111) using low-temperature scanning tunneling microscopy (LTSTM) and spectroscopy (Stróżecka et al., 2007). The study demonstrates controlled manipulation of single C60 molecules on Cu(111) and analyzes the electronic spectrum and vibrational properties. The researchers employed LTSTM to meticulously study the properties of C60 molecules on Cu(111), showcasing the controlled manipulation of single molecules. The electronic spectrum and vibrational properties were

systematically analyzed, providing valuable insights into the behavior of C60 on the surface.

D. Manipulation of Near-Field Properties in Plasmonic Nanocavities

A study showcasing the manipulation of spectral features in a plasmonic STM junction by nanofabricating gold tips (Nano Lett., 2019). The paper emphasizes the potential for inelastic tunneling induced manipulation to influence and control plasmonic properties in nanoscale cavities. The research team employed nanofabrication techniques to create gold tips and manipulate spectral features in a plasmonic STM junction. The study's significance lies in the ability to influence and control plasmonic properties in nanoscale cavities, paving the way for advancements in near-field imaging and spectroscopy accuracy.

E. Coherent Manipulation of Electrons

An exploration of the need for coherent manipulation and control over electrons at atomic length and time scales in nanoelectronic devices (Garg and Kern, 2020). The study demonstrates coherent control over electrons in an STM tunnel junction through precise tuning of the carrier-envelope phase of optical pulses. The researchers focused on achieving coherent manipulation and control over electrons, addressing challenges associated with atomic length and time scales. Precise tuning of the carrier-envelope phase of optical pulses was a key methodological aspect, enabling the demonstration of coherent control over electrons in an STM tunnel junction.

Conclusion

In conclusion, the diverse applications of inelastic tunneling induced manipulation underscore its multifaceted nature, from precision movements and bond breaking to structural rearrangement and coherent control over electrons. This categorization elucidates the diverse roles ITIM plays in pushing the boundaries of STM techniques across various scientific domains. By bridging precision with innovation, ITIM emerges as a pivotal tool, reshaping our understanding and capabilities in nanoscience and technology.

VI. Outlook

As we progress through the exploration of atom manipulation in scanning tunneling microscopy (STM), it's crucial to scrutinize methodologies for future advancements. The journey through lateral manipulation, vertical manipulation, electric field-induced manipulation (EFIM), and inelastic tunneling-induced manipulation (ITIM) has unearthed a wealth of possibilities and challenges. This section critically assesses the nuances of our methods, addresses limitations, and outlines avenues for refinement and innovation.

In lateral manipulation, refining temporal and spatial control over atom movements is an ongoing quest for enhanced precision. Innovations in probe-tip engineering, feedback mechanisms, and computational modeling hold the key to achieving subatomic precision. A deeper exploration of material-specific interactions is warranted, ensuring the universality of

lateral manipulation techniques across diverse surfaces.

Vertical manipulation promises groundbreaking applications, yet challenges persist in the controlled transfer of atoms and molecules. Heightened scrutiny of the role of substrate materials, tip design, and the influence of environmental factors is crucial. Pioneering studies hint at the potential for creating atomic switches and quantum qubit platforms, necessitating a systematic investigation into scalability, reliability, and the broader implications for quantum information processing.

Electric Field-Induced Manipulation (EFIM), with its prowess in precision movement, bond breaking, and coherent electron control, sets a high standard for complexity. Future research endeavors should focus on elucidating the intricate dynamics of electric field interactions at the atomic scale. Advancements in computational simulations, coupled with experimental verifications, will deepen our understanding and aid in optimizing EFIM for diverse molecular structures and surfaces.

Inelastic Tunneling-Induced Manipulation (ITIM) opens new frontiers in spectroscopy and manipulation but requires an interdisciplinary approach. Methodological innovations in spectroscopy, combined with advancements in tunneling current control, will unlock new dimensions of manipulation and analysis. Future studies should address the scalability of ITIM techniques and their integration into practical applications, bridging the gap between theoretical potential and experimental feasibility.

The interdisciplinary nature of atom manipulation in STM demands collaborative efforts and shared methodologies. Bridging the expertise of physicists, chemists, materials scientists, and engineers will foster a holistic approach to innovation. Standardization of experimental protocols, data sharing, and cross-disciplinary training will be pivotal in maximizing the impact of atom manipulation techniques. Critically assessing the literature echoes a call for rigor in future research. Researchers must focus on the reproducibility and reliability of reported findings, ensuring attention to experimental variables, robust statistical analyses, and transparent reporting to fortify the foundations of atom manipulation studies.

Our outlook extends beyond the horizon of current achievements, beckoning us to refine our methods, overcome challenges, and collectively propel the field of atom manipulation in STM towards new frontiers. The integration of diverse perspectives and a commitment to methodological excellence will be the catalysts for transformative breakthroughs in the exciting realm of nanoscience and nanotechnology.

VII. Conclusion

This literature review has introduced the diverse techniques employed in STM atom manipulation, unraveling their applications across various scientific domains. Lateral Manipulation has emerged as a powerful tool, enabling the construction of quantum corrals, quantum dots, and the induction of controlled chemical reactions. The ability to position individual atoms with subatomic precision has opened avenues for unprecedented insights into quantum phenomena, offering a platform for advancements in materials science and molecular

engineering. Vertical Manipulation explores the controlled transfer of atoms and molecules between the STM tip and sample surface. This technique, showcased through the creation of atomic switches, nanostructure assembly, and quantum qubit platforms, demonstrates the versatility of STM in manipulating matter not only laterally but also in the vertical dimension, leading to groundbreaking applications in quantum information processing. Electric Field-Induced Manipulation (EFIM) harnesses the power of electric fields to orchestrate precision movements, induce bond breaking, and achieve structural rearrangement at the atomic scale. Its multifaceted capabilities position EFIM as a transformative technique, offering unparalleled control over electrons and opening new frontiers in nanoelectronics and microscopy. Inelastic Tunneling-Induced Manipulation (ITIM), driven by tunneling electrons, emerges as a pivotal tool in reshaping STM techniques and advancing our understanding of molecular behavior with sub-femtosecond resolution with applications from precision movements to bond breaking, structural rearrangement, and coherent control over electrons

In conclusion, the convergence of lateral manipulation, vertical manipulation, EFIM, and ITIM invites the manipulation of individual atoms to become a canvas for scientific exploration and technological innovation. As we stand at the intersection of current achievements and future possibilities, the commitment to methodological excellence, collaborative interdisciplinary efforts, and a relentless pursuit of knowledge will steer the field of atom manipulation in STM towards uncharted frontiers, promising transformative breakthroughs in the captivating domain of nanoscience and nanotechnology.

Works Cited

- Bartels, L., et al. "Basic Steps of Lateral Manipulation of Single Atoms and Diatomic Clusters with a Scanning Tunneling Microscope Tip." *Physical Review Letters*, American Physical Society, 28 July 1997, doi.org/10.1103/PhysRevLett.79.697.
- Celotta, Robert J., et al. "Autonomous Assembly of Atomically Perfect Nanostructures using a Scanning Tunneling Microscope." *Review of Scientific Instruments*, vol. 85, no. 12, 2014, 121301, <https://doi.org/10.1063/1.4902536>.
- Crommie, M. F., et al. "Confinement of Electrons to Quantum Corrals on a Metal Surface." *Science*, vol. 262, no. 5131, 1993, pp. 218-220, DOI: <https://doi.org/10.1126/science.262.5131.218>.
- Devel, M., et al. "Field Induced Manipulation of Fullerene Molecules with the STM: A Self-consistent Theoretical Study." *Applied Surface Science*, vol. 87–88, 1995, pp. 390-397, ScienceDirect, [https://doi.org/10.1016/0169-4332\(94\)00557-5](https://doi.org/10.1016/0169-4332(94)00557-5).
- Dujardin, G., et al. "Vertical Manipulation of Individual Atoms by a Direct STM Tip-Surface Contact on Ge(111)." *Physical Review Letters*, vol. 80, no. 14, 1998, pp. 3085-3088, doi: 10.1103/PhysRevLett.80.3085.
- Eigler, D., Lutz, C., & Rudge, W. "An Atomic Switch Realized with the Scanning Tunneling Microscope." *Nature*, vol. 352, 1991, pp. 600–603, <https://doi.org/10.1038/352600a0>.
- Eigler, D., Schweizer, E. "Positioning Single Atoms with a Scanning Tunneling Microscope." *Nature*, vol. 344, 1990, pp. 524–526, <https://doi.org/10.1038/344524a0>.
- Fiete, Gregory A., and Eric J. Heller. "Colloquium: Theory of Quantum Corrals and Quantum Mirages." *Reviews of Modern Physics*, vol. 75, no. 3, 2003, pp. 933-948, American Physical Society, doi: 10.1103/RevModPhys.75.933, URL: <https://link.aps.org/doi/10.1103/RevModPhys.75.933>.
- Garg, M., Kern, K. "Attosecond Coherent Manipulation of Electrons in Tunneling Microscopy." *Science*, vol. 367, 2020, pp. 411-415, DOI: <https://doi.org/10.1126/science.aaz1098>.
- Hla, Saw-Wai, et al. "Inducing All Steps of a Chemical Reaction with the Scanning Tunneling Microscope Tip: Towards Single Molecule Engineering." *Physical Review Letters*, vol. 85, no. 13, 2000, pp. 2777-2780, American Physical Society, doi: <https://doi.org/10.1103/PhysRevLett.85.2777>.
- Hla, Saw-Wai. "Scanning Tunneling Microscopy Single Atom/Molecule Manipulation and Its Application to Nanoscience and Technology." *Journal of Vacuum Science and Technology B*, vol. 23, no. 4, 2005, pp. 1351–1360, <https://doi.org/10.1116/1.1990161>.
- IBM. "A Boy and His Atom: The World's Smallest Movie." YouTube, YouTube, 30 Apr. 2013, www.youtube.com/watch?v=oSCX78-8-q0.
- Jolie, Wouter, et al. "ACS Nano 2022 16 (3), 4876-4883." *ACS Nano*, vol. 16, no. 3, 2022, pp. 4876-4883, DOI: 10.1021/acsnano.2c00467.
- Komeda, Tadahiro. "Chemical Identification and Manipulation of Molecules by Vibrational Excitation via Inelastic Tunneling Process with Scanning Tunneling Microscopy." *Progress in Surface Science*, vol. 78, no. 2, 2005, pp. 41-85, ISSN: 0079-6816,

- <https://doi.org/10.1016/j.progsurf.2005.05.001>.
- Othman, Ahmed M., et al. "Nature of Confining Potentials in Adatom-Based Quantum Corrals and Superlattices." *Physical Review B*, vol. 108, no. 20, 2023, p. 205409, American Physical Society, Nov. 2023, doi: 10.1103/PhysRevB.108.205409, URL: <https://link.aps.org/doi/10.1103/PhysRevB.108.205409>.
- Oxford Instruments. "Scanning Tunneling Microscopy (STM): An Overview." Oxford Instruments, afm.oxinst.com/modes/scanning-tunneling-microscopy-stm.
- Robert J. Celotta, Stephen B. Balakirsky, Aaron P. Fein, Frank M. Hess, Gregory M. Rutter, Joseph A. Stroscio. "Invited Article: Autonomous Assembly of Atomically Perfect Nanostructures using a Scanning Tunneling Microscope." *Review of Scientific Instruments*, vol. 85, no. 12, 2014, 121301, <https://doi.org/10.1063/1.4902536>.
- Saw-Wai Hla, et al. "Scanning Tunneling Spectroscopy and Manipulation of C60 on Cu(111)." *Applied Physics A*, vol. 87, 2007, pp. 475–478, <https://doi.org/10.1007/s00339-007-3914-z>.
- Saw-Wai Hla, et al. "Dissociation of Individual Molecules with Electrons from the Tip of a Scanning Tunneling Microscope." *Science*, vol. 255, 1992, pp. 1232-1235, DOI: 10.1126/science.255.5049.1232.
- Stróżecka, A., Mysliveček, J. & Voigtländer, B. "Scanning Tunneling Spectroscopy and Manipulation of C60 on Cu(111)." *Applied Physics A*, vol. 87, 2007, pp. 475–478, <https://doi.org/10.1007/s00339-007-3914-z>.
- Wang, Yu, et al. "An Atomic-Scale Multi-Qubit Platform." *Science*, vol. 382, 2023, pp. 87-92, DOI: <https://doi.org/10.1126/science.ade5050>.

Color Psychology and Pediatrician Attire By Katherine Pyasik

Abstract

Pediatric healthcare-induced anxiety arises in children starting from three years of age and stems from a perceived loss of bodily autonomy during clinical encounters. Since children often have a prolonged relationship with their pediatrician, it is vital that pediatric physicians take steps to address verbal and nonverbal cues that have the potential to ameliorate their patients' anxiety. One nonverbal cue that has been previously researched across a range of medical specialties is physician attire. Physician attire has been shown to have an impact on patient perceptions and color has been shown to influence human emotions starting from age 5. The present study investigated the impact of physician attire color on elementary school students' perceptions of physicians. Using the Round Robin comparison method to compare images of physicians in different colored scrubs, 18 students ages 5-11 were interviewed. The results did not show any significant preferences, although a slight aversion to dark colors was discovered. In addition, analysis of color preference based on participant age, gender, and healthcare-induced anxiety were consistent with previous literature. The results also showed that the genders of the people in the images used may have skewed the results in favor of the colors that the female physicians wore. However, the results of this study warrant further research because some aspects of the hypothesis were supported. It is recommended that this study is repeated with a larger sample size to provide narrow confidence intervals and determine if physician attire color preferences exist.

Introduction

Literature Review: The Doctor-Patient Relationship

Research from the World Journal of Clinical Pediatrics identifies that healthcare-induced anxiety frequently arises in children during routine encounters with their primary care physician (PCP), such as preventative clinic visits and medical procedures (Lerwick 143-144). A University of Michigan survey reported that half of parents in the US state that their preschooler (aged three to five) fears doctor's visits ("Half of parents"). This anxiety stems from a perceived loss of bodily autonomy and control as children put their well-being in the hands of their PCP. Additionally, children often exhibit aggression, withdrawal, and lack of cooperation with necessary medical procedures, like immunizations (Lerwick 144). These symptoms are manifestations of physiological and behavioral distress, which are symptoms of pediatric healthcare-induced anxiety (Smith, 1-2). If not mitigated, constant anxiety like this may cause long-term emotional, behavioral, and cognitive disorders (Lerwick 144). The American Academy of Pediatrics recommends that children visit their PCP 31 times between birth and age 21, so it is vital that they feel comfortable at PCP visits to prevent anxiety (Smith 1-2).

It is essential to establish trust early in the pediatrician-patient relationship, as this allows time for the relationship to grow stronger and reduces the likelihood of broken trust. Broken trust in the pediatrician-patient relationship leads to healthcare-induced anxiety, as well as both

parents and children second-guessing the pediatrician's competence (Sisk). The maintenance of trust between a doctor and patient of any age often leads to increased adherence to treatment plans, which improves health outcomes (Ha et al. 39). The CARE (choices, agenda, resilience, emotions) principles is a framework created to combat children's perceived loss of control and instill trust in the pediatrician-patient relationship (Lerwick 143). However, despite their potential usefulness, integration of this and similar strategies into medical practices is limited and a lack of trust in pediatrician-patient relationships, and healthcare-induced anxiety in the US persists (Lerwick 144; "Half of parents"; Feldscher 1).

As early as the first interaction a child has with their doctor, the child makes assumptions about that doctor's competence based on their appearance, gestures, and verbal and nonverbal cues. Therefore, pediatricians must control these aspects to ensure patient comfort and prevent the adverse effects of a child's loss of trust in their pediatrician (Sisk; Asokan et al. 30). One nonverbal cue that has been extensively researched is the doctor's attire.

An article from Johns Hopkins Medicine reports that pediatricians have recently neglected the centuries-old tradition of wearing a white coat because they report seeing expressions of anxiety and intimidation on their patients' faces while wearing it ("White Jacket"). One study, however, showed the opposite: when children aged 4-8 were asked to choose which doctor they would prefer to see from images of doctors who wore or did not wear white coats, 69% chose images of doctors in white coats. Moreover, the study found that white coats make a doctor seem confident and safe ("White Jacket").

There have also been many studies that assess children's preferences for their dentist's attire, as pediatric healthcare-induced anxiety is a prevalent issue in dentistry as well (Asokan et al. 30).⁴⁷ One study assessed whether pediatric attire (bright-colored scrubs with iron-ons of familiar movie characters or objects) or traditional white attire is preferred by children ages 2-15. The findings were consistent across age groups and indicated that dentally anxious children prefer pediatric attire (Elmore et al. 105-106). A similar study published in the *International Journal of Pediatric Dentistry* was done with children ages 9-13 and identified that they regard formally-dressed doctors in white coats as more competent than friendly. Nevertheless, more children preferred white coat attire, although child friendly attire was shown to be important for anxious children (Aranha 15-16). One study looked solely at dentally anxious children ages 9-12 and found that they prefer bright, colored attire without a white coat (Asokan et al. 32-33).

The importance of the physician's clothing has also been shown to have an impact on adults' perceptions of their physicians. Although a study published in *Biomed Central's Family Practice Journal* found that patient satisfaction is not related to physician attire in primary care (Matsuhisa et al. 4), a multitude of studies argues the opposite point. One study even found that adult patients are more likely to share physical, social, and psychological problems with a doctor

⁴⁷ It is important to note that although dentists are doctors and healthcare professionals, they are not physicians. This research focuses solely on pediatric primary care physicians, but literature on pediatric dentistry is included here because pediatric healthcare-induced anxiety is similar in pediatric primary care and pediatric dentistry. Usually, no acute medical procedures are done, but children experience an unprecedented perceived loss of control (Elmore et al., 2015).

who is professionally dressed (Rehman). According to studies from *The American Journal of Medicine* and *BMJ Open*, formal attire with a white coat is most widely preferred by adults and physicians who wear formal attire with a white coat are perceived as more knowledgeable, trustworthy, caring, approachable, and comfortable (Rehman; Petrelli 5). Furthermore, a study from the *Scandinavian Journal of Primary Health Care* assessed patient attitudes towards white coats and found that the preference for seeing a physician wearing a white coat increases with age (Anvik 91). Finally, a study from *JAMA Internal Medicine* that assessed patient family preferences for physician attire in the intensive care unit found that scrubs and a white coat are most strongly preferred by families because physicians wearing this attire are perceived to be the most competent to perform a life-saving procedure (Au et al.).

Another approach used in an attempt to combat healthcare-induced anxiety in children and adults has been to examine the use of color in medical offices. One investigation aimed to find preferences for the wall color in medical offices. Participants were aged 18-55 and the results showed that, among warm, cool, bright, and muted, cool was preferred by 49% of respondents (Babin 13). Another study, which was published by the *Journal of the Indian Society of Pedodontics and Preventive Dentistry*, aimed to improve children's attitudes toward dental medical settings and found that children ages 6-12 have the most positive color-emotion association with yellow and blue, while red and black were associated with fear and boredom. Accordingly, the recommendation of this study was to assure that blue and yellow are present in pediatric medical offices (Umamaheshwari 30).

Literature Review: Color Theory

There is a vast body of literature on color psychology and emotion-color associations in both adults and children. According to a study from the journal *Frontiers in Psychology*, Johann Wolfgang von Goethe pioneered the idea of color-emotion associations in his 1810 book *Theory of Colors*. One of the most important aspects of modern research in color theory is that colors are perceived in context. For example, blue on a ribbon is perceived as positive because of a presumed victory, whereas blue on a piece of meat communicates that it is spoiled (Elliot). Next, one study from the journal *Evolutionary Psychology* identified that college students received higher attractiveness ratings when wearing red than any other color, likely because red is innately associated with fertility. On the contrary, it has been shown that sports teams with red uniforms are viewed as aggressive and dominant by their opposing team (Roberts et al.). Another study found that customers are more likely to maintain long-term trust in brands with blue logos than red logos. Blue is the color of natural space, like the sky and ocean, so humans may have a natural evolutionary predisposition to favoring blue over red (Su et al. 270-271). Environments with blue in them have been linked to higher-order attentional processing, increased focus, and calmness (Elliot).

In children, color-emotion and color-object associations arise from socialization and enculturation, so they are predominantly present in older children (Miles et al. 3-4). In one study, researchers showed children aged 5-6.5 different colored pieces of construction paper and asked

them to name emotions that they associate with each color. The study found that positive emotions were mostly linked with bright colors, including pink, yellow, and blue, and negative emotions are linked with black, brown, and gray, which are dark colors (Boyatzis & Varghese). In a study published in the Journal of Modern Psychological Studies, preschoolers (ages three to four) were read a happy or sad story and then provided a corresponding picture to color. The happiness or sadness of the story didn't make the children use certain colors (Miles et al. 9). Therefore, for the purposes of this study, it is reasonable to conclude that color-emotion associations become defined in children at around age 5.

Gap in the Literature

The style of attire that doctors wear is an important nonverbal cue that can help establish trust in the doctor-patient relationship. Patient perceptions of physician attire have been researched across various medical specialties such as pediatrics, internal medicine, and pediatric dentistry. There is also a vast body of literature on color-emotion associations in adults and children. Color theory has been tangentially looked at in relation to dentist's scrubs, but it has never been looked at in relation to pediatric physicians' scrubs. Therefore, this research aims to find whether current knowledge on color psychology can be applied to physician attire color in pediatric primary care. This is an important endeavor in pediatric primary care specifically because of the prevalence of pediatric healthcare-induced anxiety, as previously discussed. The question being researched is: To what extent do preferences for physician attire color exist in pediatric primary care in children ages 5-11?

The age range begins at 5 because color-emotion associations arise at this age. Doing this study with participants younger than 5 likely would not have provided meaningful results, as color-emotion associations in this less age group are less defined. The upper bound for the range is 11 because elementary school students are aged 5-11, making it more feasible to find a facility to complete this research where all ages in this range are present.

Hypothesis

The hypothesis is that black scrubs will be preferred the least and blue, yellow, and red scrubs will be preferred the most. This is based on previous literature on children's color-emotion associations. The secondary hypothesis is that the participant population's assessment of white scrubs will likely depend on the mean anxiety level, as white attire has been shown to increase anxiety in dentally anxious children.

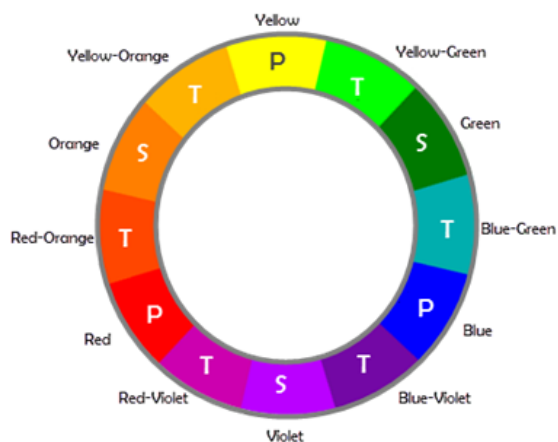
Methodology

Method

To fill the proposed gap in the literature, I obtained consent from two facilities to conduct interviews with their students: one after-school program for elementary school students and one preschool. Because preschoolers are ages three to five, only students who were five at the time that the interviews were conducted participated. Two weeks prior to beginning the interviews at

the aftercare program, consent forms (*Appendix A*) were emailed to parents. This facility required an email confirmation from each parent to allow their child to participate in this study. One week before beginning interviews at the preschool, the director of the preschool asked parents to read the consent form and obtained verbal consent from each parent for their child to participate. For both locations, a student’s consent alone was not enough, as children aren’t capable of giving informed consent. If a student had stated that they did not want to participate before or during the interviews, they would have been excluded from the study, but this did not occur.

A set of five images (*Appendix B*) was used in this study. These were images of adults wearing different colored standard medical scrubs, which addressed the main part of the research question on physician attire color. The colors I accessed the most research on were used to ensure that the findings could be adequately analyzed in conversation with other literature. These colors were blue, yellow, white, red, and black. To ensure that this study used similar colors to other studies in the field of color psychology, the red, blue, and yellow scrubs were matched up with the standard 3-primary-color wheel (“Color Basics”), shown below. This color wheel was obtained from usability.gov, a website that is devoted to making websites aesthetically pleasing and easy to use. This does not apply to black or white scrubs because black and white are not on the standard color wheel. It was assumed that previous experiments used very similar colors to those on the standard wheel. This was necessary because some of the literature on color psychology used for this study did not include images of the colors used in their experiments (Umamaheshwari 226; Miles 6). This allowed the findings of this study to be applied to a discussion on whether the results of this study support previous research with more confidence.



The individuals in the images are not wearing white coats to ensure that the white coats would not distract the participants from the color of their scrubs, as scrub color was the aspect being researched. The images were obtained from online stores and scrub catalogs like Amazon, Scrubs and Beyond, and Walmart. These platforms were the only ones accessed that provided full-body images of individuals who were standing up, facing the camera, and smiling. Images of five different people were used because the students interviewed likely would have been able to

decipher what was being researched. When this occurs, participants often develop research biases and select answers that are different from their actual preferences (McCambridge et al.).

The individuals in all the images are Caucasian because, unfortunately, racial biases are present in children from a young age, so including images of individuals of different races would have likely skewed the results so that they identified pediatrician race preferences instead of attire color preferences (Elliot). The individuals are white, opposed to another race, because this research was done in a majority white community, so the physicians that the students are used to seeing are likely white. Additionally, the people in the images all have pleasant facial expressions and are smiling. They also all have a similar posture to one another, although it is not exactly the same. All of the individuals are standing up, facing forward, and some have one or both hands in their pockets. Using images of physicians who were the same race and had similar facial expressions and stances ensured that participants' choices were based solely on scrub color.

Despite the fact that many aspects of the images were controlled, physician gender was not a controlled aspect in these images because no literature was accessed that stated that defined preferences for physician gender exist in pediatric patients. Research published in the *Ochsner Journal* identified that children prefer to see gender-congruent physicians (those who appear traditionally male or traditionally female) (Turow & Sterling 340). All of the images are of gender-congruent individuals. Therefore, it was assumed that participants' attitudes toward male and female physicians would be similar and that the differences in gender would not affect the results.

During the main portion of the interviews, each student was presented with ten sets of two images each and asked which doctor they would rather see for their yearly check-up. They were asked about yearly check-ups because children generally see their primary care pediatrician for this purpose and because the research question focuses on primary care (Lerwick 144). If simply asked which doctor they would rather see, the participants, especially those who see subspecialists or surgeons, may have identified their preferences based on which doctor they would rather have as their subspecialist perform their surgery.

Ten sets of two images were presented to each student and each image was compared to each of the others exactly one time using the round robin comparison method. This allowed preferences for certain attire colors to be determined, as this method makes it easy to gauge how many times certain images were preferred over others. The formula $T(n)=n*[(n-1)/2]$ was used to find how many sets of two images would be needed to complete this task. This formula is used to determine the number of different pairs, $T(n)$, that can be made from n objects. Only five images were used to ensure that interviews were relatively short so that the participants could stay focused throughout their duration.

Using this method aligns well with the research question. By asking students which image they prefer and collecting the data, I was able to gauge which images students preferred the least and the most. The results also showed how strong the preferences were, which addresses the extent to which attire color preferences exist.

Each participant was presented with the same sets in the same order to ensure that, if the order did make a difference, this difference would be present among all the participants. The color sets and the order they were presented in can be found in **Figure 3** in the results section.

Students were interviewed one at a time. If multiple students had been interviewed at once, it is possible that some students would have exhibited conformity bias by conforming to their peers' choices rather than stating their own preferences (Sun & Yu 4). At both childcare facilities, I sat across from each participant and began the interviews by introducing myself and the notetaker, who recorded interview results using the chart (*Appendix C*). I explained that I would be asking a few questions, that there are no right or wrong answers to those questions, and that the interviews would only take a few minutes. First, I asked each child their age. Second, I asked them how nervous they are during their yearly check-up on a 1-4 scale. I clarified that 1 would indicate none or minimal nervousness, while 4 would signify extreme nervousness. Each student's gender was assumed based on their name and appearance. Finally, I explained that I would show them images of two doctors⁴⁸ side by side, and that they should point to whichever one they would rather see for their yearly check-up. I also stated that, for each set, they should make their choice as fast as possible. This ensured that the participants' choices were based solely on their first impression of the individual. Because the scrubs that each individual wore took up the majority of the physical space on the image, I assumed that this would be one of the first things each participant would consciously or subconsciously notice. Prolonged examination of the images likely would have caused the students to notice other factors, such as posture or stance, that may have also contributed to their decision.

Studies similar to this discussed how their results differed with age and healthcare-induced anxiety level (Elmore et al. 106). A study on color-emotion associations in children stated that these associations differed by gender (Boyatzis & Varghese). Using this data, this study was able to support the conclusions of these studies, as will be discussed later. Participants were asked about anxiety during check-ups on a 1-4 scale because there is no middle value on this scale for participants to select if they are unsure. On a scale with an odd number of choices, like a 1-5 scale, if a student wasn't sure whether their anxiety was minimal or significant, they could have chosen 3.

A quantitative method of data collection was utilized in this study because, if students had been asked to describe their perception of the individuals in the images as physicians (like other studies have done (Roberts et al.)), their perceptions would have likely been affected by more than just attire color.

Results

Primary Results

Altogether, 18 students were interviewed. No results were omitted, as all students took no more than 10-15 seconds to make their decisions for each image set. All students were within the

⁴⁸Although it is unlikely that the individuals in the images were actually pediatric primary care physicians, each participant was told that these were images of doctors to fulfill the goal of the research question.

5-11 age range, but, by chance, no students aged 10 were interviewed. The age distributions for this study can be found in **Figure 1** below. The ages present within the sample population are listed on the vertical axis and there is a percentage scale on the horizontal axis. The mean age for this study was 6.83 years. Data on self-reported pediatric healthcare-induced anxiety on a 1-4 scale can be found in **Figure 2** and is organized like **Figure 1**. The mean self-reported anxiety level was 1.78.

Figure 1:

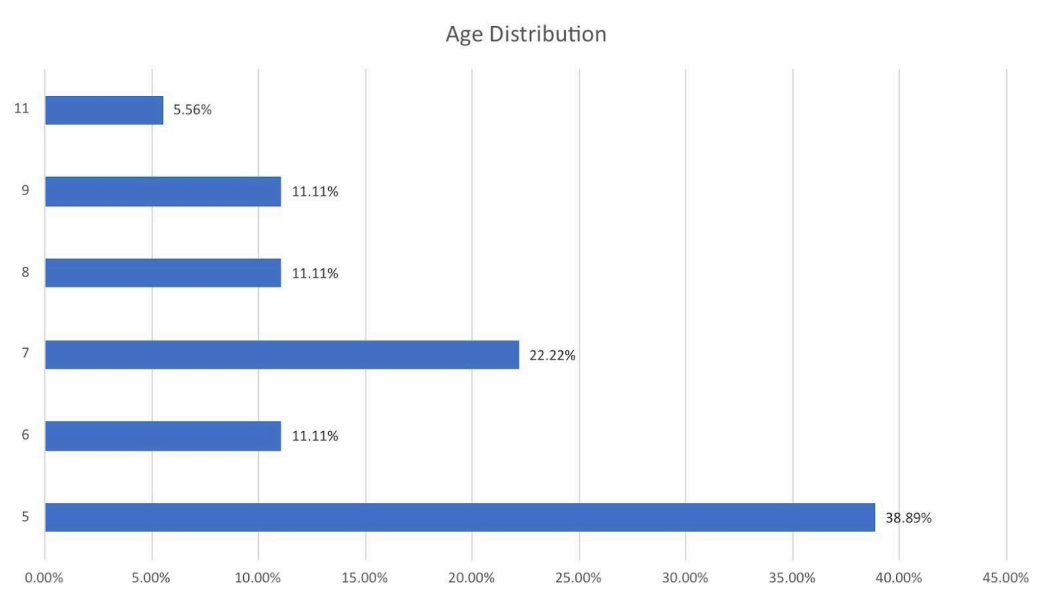
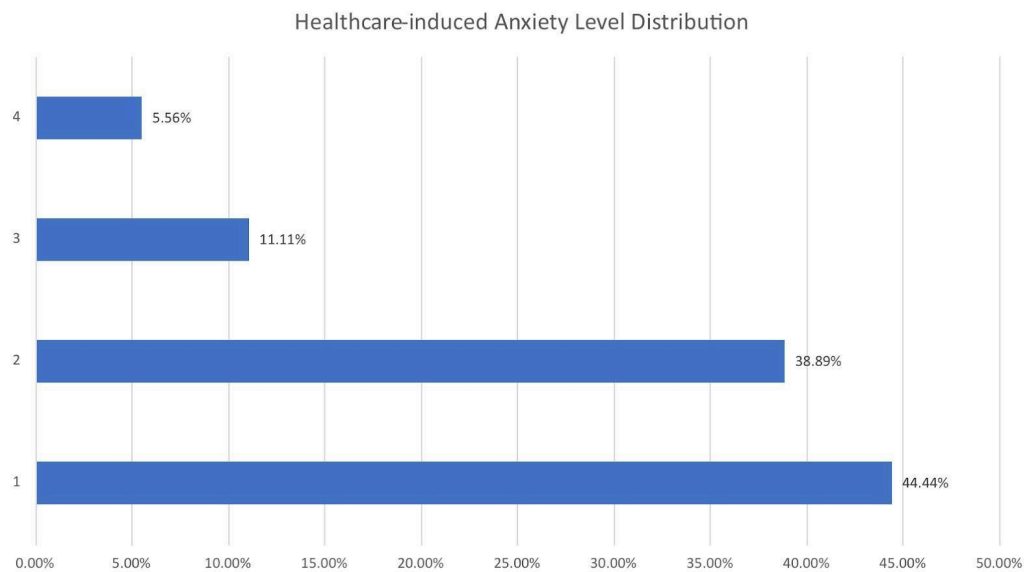


Figure 2:



As previously stated, all the participants were shown the same image sets in the same order. The results for the number of students that chose each image from each set can be found in **Figure 3**. Sets 1-10 are numbered in the order that they were shown to each student. The scrub

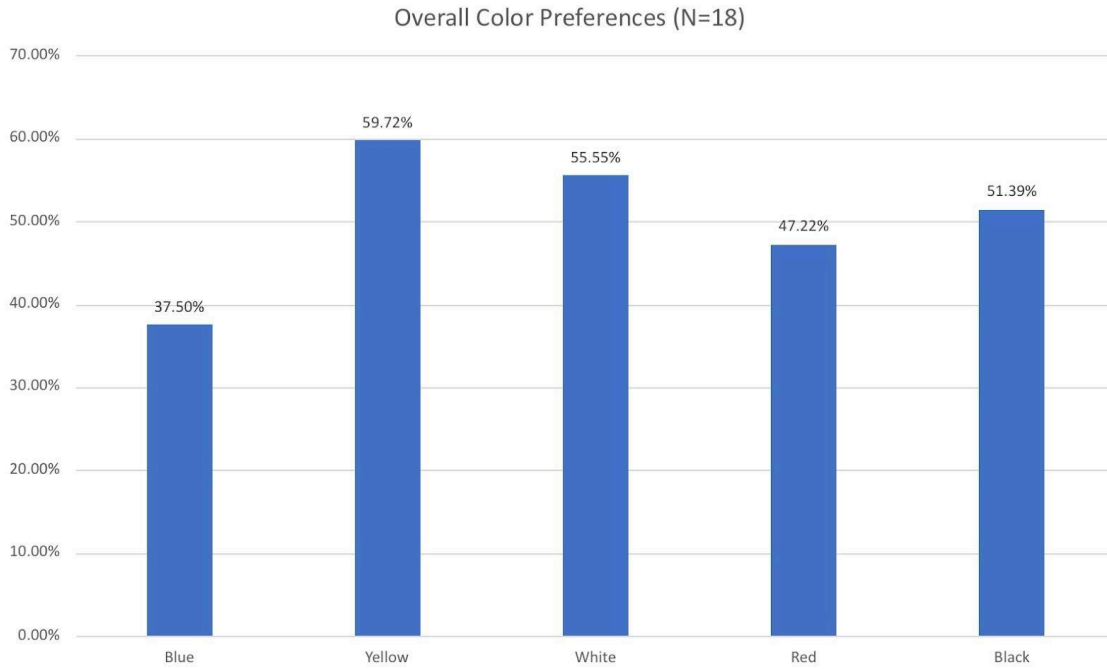
color listed first in the column labeled “Color Combinations” is color 1 and, therefore, corresponds to the column labeled “Frequency 1 (%)”. This column identifies the number of students that chose each color from that set, and the corresponding percentage out of the entire participant population. The same is true for the second color listed and the column labeled “Frequency 2 (%)”. As an example, from set 1, seven children indicated that they would like to see the individual in blue scrubs over the one in yellow scrubs, and 11 indicated the opposite.

Figure 3:

#	Color Combinations	Frequency 1 (%)	Frequency 2 (%)
1	Blue, Yellow	7 (38.89)	11 (61.11)
2	Blue, White	5 (27.78)	13 (72.22)
3	Blue, Red	8 (44.44)	10 (55.56)
4	Blue, Black	7 (38.89)	11 (61.11)
5	Black, Yellow	6 (33.33)	12 (66.67)
6	Black, White	10 (55.56)	8 (44.44)
7	Black, Red	9 (50.00)	9 (50.00)
8	Red, Yellow	9 (50.00)	9 (50.00)
9	White, Red	12 (66.67)	6 (33.33)
10	White, Yellow	7 (38.89)	11 (61.11)

Using the data above, **Figure 4** was created. It shows the percentage of the time that each image was chosen over another image. Because the sample size was 18 and each of the five images was compared with the four others one time, each image could have been chosen four times by 18 students. Therefore, each image could have been chosen 72 (18*4) times total over other images. The percentage at the top of each bar identifies the percent of the time each image was chosen over another. The N=18 in the title means that the total sample size was 18.

Figure 4:



Secondary Results

Figure 5 presents the results exactly like **Figure 4** does, but it presents the results with respect to the gender of the participants. The percentages were calculated the same way, but separately for male and female students. **Figure 6** and **Figure 7** present the data with respect to age and self-reported healthcare-induced anxiety level respectively. These tables are the same as the others as well. The age range was collapsed into younger students (ages 5-7) and older students (ages 8,9, and 11 because no students of age 10 were interviewed). The anxiety levels were collapsed into the categories 1-2 and 3-4, indicating low and high healthcare-induced anxiety respectively. Not providing a middle value on the scale allowed the results to be presented like this. The N-values on both charts indicate the number of students in each category.

Figure 5:

Color Preferences by Gender

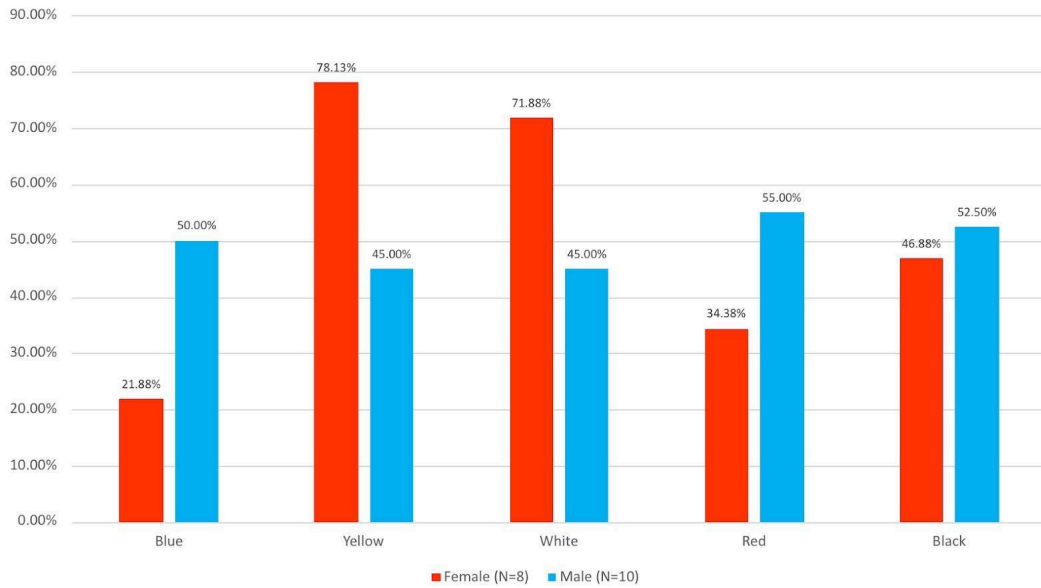


Figure 6:

Color Preferences by Age

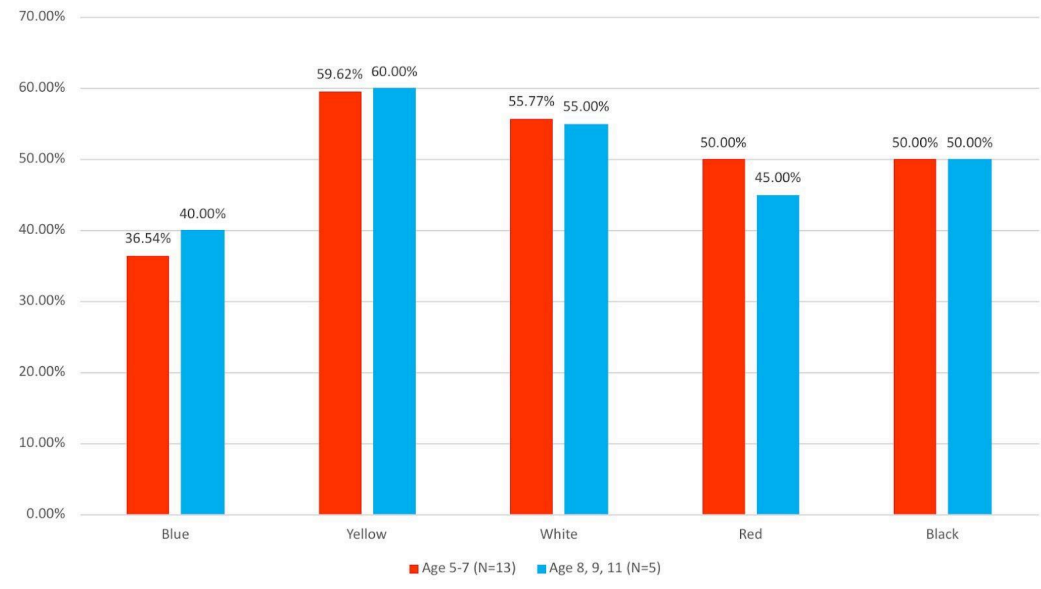
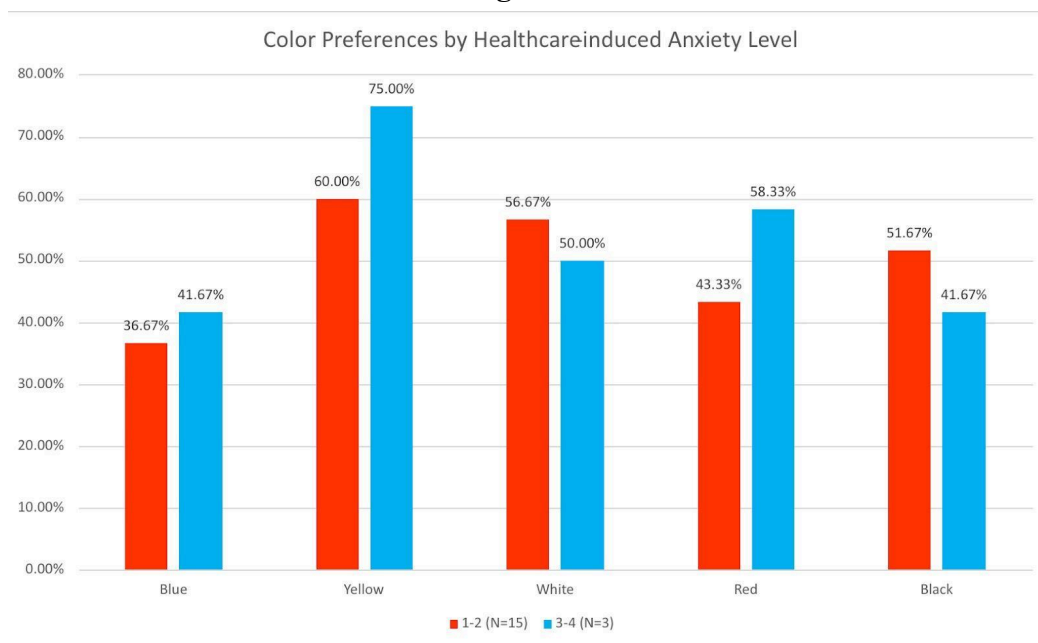


Figure 7:



Discussion

Statistical Analysis of the Data

90% confidence intervals were calculated for the overall data (**Figure 4**) and can be found on the graph in *Appendix D*. Because of the low sample size, the intervals were relatively large (about 40% for each) and did not show that the data is statistically significant. Confidence intervals that don't overlap would have indicated statistical differences between physician attire color preferences, but all of the intervals for this study's data overlapped. To determine whether these differences exist, this study would need to be repeated with a larger sample size so that narrower intervals can be calculated. Despite this caveat, much of this study's data did align with previous literature.

Descriptive Analysis of the Data

As expected, no statistically significant preference was found for physician attire color. There are a multitude of other factors that contribute to patient comfort in pediatrics that are as important as physician attire. These include the physician's demeanor, their ability to explain difficult concepts to the patient, and general cleanliness and other nonverbal cues (Sisk, 2019; Rehman, 2005). However, colors are closely tied with emotional responses, and consistent, albeit slight, color preferences were discovered. **Table 2** shows that there were no extreme preferences, as all percentages for the number of times each image was chosen out of all the times it could have been chosen fall between 37.50% and 59.72%.

Looking at **Table 2**, the image of the physician wearing yellow scrubs was most preferred, as it was chosen 59.72% of the time over other images. The literature states that pediatric dentistry patients feel more at ease when yellow is present in a medical office and that they have positive color-emotion associations with this color, which explains this finding

(Umamaheshwari, 2013; Boyatzis & Varghese, 2010). The least preferred color was blue, as it was only chosen 37.5% of the time over other images, which contradicts the study that supports the finding that yellow and blue are the most preferred colors in a pediatric dental office (Umamaheshwari 30). Furthermore, this finding doesn't align with the current understanding in color psychology that states that blue is associated with calmness and trust in both adults and children (Umamaheshwari 30; Elliot; Boyatzis & Varghese). This might be explained by one study that found that children are slightly more likely to share sensitive health information with female physicians in white coats, and the individual in the image used in this study was male (Hochberg 310). However, this may not apply to those who aren't wearing white coats, like those in the images used in this study. This was the only study accessed that addressed physician gender preferences; there is popular literature on the subject, but it is based purely on speculation rather than experimental or empirical evidence. There is popular literature suggesting that children prefer to see a pediatrician who is the same gender as they are, but, again, no scholarly literature supporting this conjecture was accessed.

The image of an individual in red scrubs was the only other color that was chosen less than 50% of the time. This indicates an aversion towards this image and aligns with the conclusion that red evokes feelings of untrustworthiness and aggression in adults and older children (Roberts et al.; Elliot), although red has been shown to be correlated with more positive emotions in children because it is classified as a bright color (Boyatzis & Varghese). It is possible that, if a general preference for female physicians existed within this study's sample, this could have affected students' choices and skewed the results against this image, as a higher evaluation of red was expected.

The image of an individual in black scrubs was chosen 51.39% of the time, which is more than expected because previous studies show that black is associated with negative emotions, like fear and boredom, in children because it is a darker color (Umamaheshwari 30; Boyatzis & Varghese). However, as stated above, an innate preference for female physicians within the sample population may have skewed this. There is literature that states that female physicians may be seen as more maternal and nurturing, so the study participants may have been predisposed to favoring female physicians over male physicians (McCambridge et al.). However, black was still preferred less than white and yellow, which were the other colors that the female physicians wore. This supports the hypothesis and previous literature because it shows that black, a dark color, is preferred less than yellow and white. Still, it is important to note that the images that were most preferred were images of females.

The percent of the time that the image of the physician in white scrubs was chosen falls between that of the physician in black scrubs and the physician in yellow scrubs at 55.55%, which supports the conclusion that dark colors evoke negative emotions in people across all ages (Umamaheshwari 30; Elliot; Boyatzis & Varghese). Based on this, it is also reasonable to conclude that non-anxious children prefer to see doctors dressed in white, as they associate this color with increased physician competence ("White Jacket").

When analyzing color preference with respect to participant gender, it was determined that girls showed far more significant preferences than boys. For female students, the least preferred color was blue, which was only 21.88% of the time over other images. The highest was a 78.13% preference for yellow. For boys, the lowest percentage was 45.00% for yellow and white and the highest was 55.00% for red. The percentage range was much larger for female participants. This aligns with research that found that positive and negative color-emotion associations are significantly more defined in girls than in boys (Boyaztis & Varghese). Additionally, based on **Figure 6**, color preferences did not differ between younger and older students, as the percentages for each color don't differ by more than 5.00% between age groups. This supports a previous study that found that preferences for dentist attire color are consistent in the age range 5-11. Looking at **Figure 7**, it is evident that preferences for brighter colors (red, yellow, and blue) and a dispreference for black were more defined in children who reported a high anxiety level (3-4). This supports the findings of the same study, as it identified that dentally anxious children prefer brighter colored attire (Elmore et al. 105-106).

Limitations

One possible limitation for this study is that physician gender was not a controlled variable and may have skewed the results, as discussed in more detail above. Additionally, the individuals in the images have slightly different stances, so certain images may have been chosen less because that individual's stance was viewed as threatening or intimidating compared to the others.

Another limitation is that the sample did not have equal numbers of children of each age that was included in the age range, or equal numbers of students who reported minimal and extreme anxiety. 38.89% of the participants were five years old and 83.33% reported an anxiety level of 1-2, so the results were likely not representative of each of these groups. No students were 10 years old, so these findings cannot be applied to any children who are 10 years old. Additionally, the sample size was only 18, so the findings cannot be applied with full confidence to all children within the age range before this study is repeated with a higher sample size. It is not reasonable to conclude that the findings of this study would be similar with another group of students within the age range, especially if the students' mean and anxiety level was different. A previous study showed that anxiety level may affect this factor (Elmore et al. 106). Still, because the findings were mostly consistent with previous research, preferences for bright colored physician attire would likely be present in a different participant population within the same age range.

Conclusion

Fulfillment of the Gap in the Literature

Altogether, the results show that slight preferences for physician attire color existed within the participant population. This study addresses the main purpose of the research question, as some knowledge now exists on physician attire color preferences in pediatric

primary care. Previous studies addressed adults' preferences for physician attire, as well as that of pediatric dental patients. Before this, no literature existed on pediatric primary care patients' preferences for physician attire, other than literature addressing whether pediatricians should wear white coats.

Implications

The findings of this study have shown that, within the participant population, current knowledge on color-emotion associations in children do apply to physician attire color preferences in some cases. The main implication is that pediatricians should refrain from wearing dark colors. This is the only recommendation being offered because this is the only finding in the study that lines up perfectly with current literature on color psychology. As stated earlier, among the female physicians, students ranked the black scrubs the lowest, and ranked yellow and white higher. Wearing brighter colors is not being recommended because the results were inconclusive on what bright colors were preferred, as red and blue were ranked last and yellow was ranked first. Therefore, repetition of this study is necessary to provide more specific recommendations.

The results of this study warrant repetition because they have supported some aspects of the hypothesis. Color is closely associated with human emotion and has been shown to influence decision-making and perceptions of others. Therefore, it is important that literature, such as this, exists on the applications of color theory in pediatric medical offices. This study provides evidence that slight physician attire color preferences may exist and, if repeated, can provide more specific recommendations on the right attire to wear to keep pediatric healthcare-induced anxiety at a minimum.

Avenues for Future Research

As previously discussed, it is likely that physician gender affected this study's results. Therefore, this study should be repeated with images of all female or all male physicians. It is possible that, if this is done, preferences would be different for female and male physicians. If they aren't however, this research could still confirm or deny the findings of this study based on the colors that are most preferred. Secondly, physician attire color preferences can be researched in different medical specialties in pediatric and adult medicine. Just as preferences for ICU physicians and internal medicine physicians have been researched, physician attire color preferences can be researched in surgical patients.

Finally, because there is limited scholarly literature on children's preferences for their pediatrician's gender, this should be further researched. This could be done by explicitly asking children this or by using image sets similar to those in this study. In the latter method, the researcher would present numerous image sets (all containing images of different people) of male and female physicians two at a time, and each participant would select which doctor they would rather see. This may also play an important role in mitigating pediatric healthcare-induced anxiety.

Works Cited

- Lerwick, Julie. "Minimizing pediatric healthcare-induced anxiety and trauma." *World Journal of Clinical Pediatrics*, vol. 5, no. 2, 2016, pp. 143-150, DOI: 10.5409/JCPenney.v5.i2.143.
- "Half of parents say their preschooler fears doctor's visits." *University of Michigan Health*, vol. 32, no. 5, 2018,
<https://www.uofmhealth.org/news/archive/201810/half-parents-say-their-preschooler-fears-doctor%E2%80%99s-visits>.
- Smith, Michelle. "Interventions to Minimize Distress During Pediatric Primary Care Visits: A Systematic Literature Review." *BYU ScholarsArchive*. 2014,
<https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=1003&context=studentpub>.
- Feldscher, Talia. "Medical Trust in Pediatric Care in the United States." *Independent Study Project (ISP) Collection*. 2020,
https://digitalcollections.sit.edu/cgi/viewcontent.cgi?article=4316&context=isp_collection.
- Sisk, Bryan, and Justin Baker. "A Model of Interpersonal Trust, Credibility, and Relationship Maintenance." *American Academy of Pediatrics*, vol. 144, no. 6, 2019,
<https://publications.aap.org/pediatrics/article/144/6/e20191319/37945/A-Model-of-Interpersonal-Trust-Credibility-and?autologincheck=redirected>.
- Aranha, Andreza, M. et al. "Do Children and Adolescents Prefer Pediatric Attire over White Attire during Dental Appointments? A Meta-analysis of Prevalence Data." *International Journal of Pediatric Dentistry*, vol. 14, no. 1, 2012, pp. 14-29, DOI: 10.5005/up-journals-10005-1861.
- Hochberg, Mark S. "The Doctor's White Coat: An Historical Perspective." *AMA Journal of Ethics*, vol. 9, no. 4, 2007, pp. 310-314, DOI: 10.1001/virtualmentor.2007.9.4.mhst1-0704.
- "White Jacket On or Off?." *Johns Hopkins Medicine*. 2014,
<https://www.hopkinsmedicine.org/news/articles/white-jacket-on-or-off?scrlbrkr=7758b8f>.
- Gunjalli, Gunjalli et al. "Children and Parent's Attitude and Preferences of Dentist's Attire in Pediatric Dental Practice." *International Journal of Clinical Pediatric Dentistry*, vol. 8, no. 2, 2015, pp. 102-107, DOI: 10.5005/jp-journals-10005-1203.
- Asokan, Alexander et al. "A survey of the dentist attire and gender preferences in dentally anxious children." *Journal of Indian Society of Pedodontics and Preventive Dentistry*, vol. 34, no. 1, 2016, pp. 30-35, DOI: 10.4103/0970-4388.175507.
- Matsuhisa, Takaharu et al. "Effect of physician attire on patient perceptions of empathy in Japan: a quasi-randomized controlled trial in primary care." *BMC Family Practice*. 2021,
<https://doi.org/10.1186/s12875-021-01416-w>.
- Rehman, Shakaib U. et al. (2005, November). "What to wear today? Effect of doctor's attire on the trust and confidence of patients." *The American Journal of Medicine*, vol. 118, no. 11, 2005, pp. 1279-86. DOI: 10.1016/j.amjmed.2005.04.026.

- Petrilli, Chris M. et al.(2018, April 19). “Understanding patient preference for physician attire: a cross-sectional observational study of 10 academic medical centers in the USA.” *BMJ Open*, vol. 8, no. 5, 2018. DOI: 10.1136/bmjopen-2017-021239.
- Anvik, T. (2009, July 12). “Doctors in a White Coat-what do Patients think and what do Doctors do? 3739 patients, 137 general practitioners, and 150 staff members give their answers.” *Scandinavian Journal of Primary Health Care*, vol. 8, no. 2, 2009, pp. 91-94. DOI: 10.3109/02813439008994937.
- Au, Selena et al. “Physician attire in the intensive care unit and patient family perceptions of physician professional characteristics.” *JAMA Internal Medicine*, vol. 173, no. 6, 2013, pp. 465-476. DOI: 10.1001/jamainternmed.2013.2732.
- Babin, Sarah E. “Colors theory-the effects of color in medical environments.” *The Aquila Digital Community*. 2013, https://aquila.usm.edu/cgi/viewcontent.cgi?article=1173&context=honors_theses.
- Asokan, Sharath et al. “Child-friendly colors in a pediatric dental practice.” *Journal of Indian Society of Pedodontics and Preventive Dentistry*, vol. 31, no. 4, 2013, pp. 225-228. DOI: 10.4103/0970-4388.121817.
- Miles, Emily et al. “The relationship between color shade and emotion association in pre-school aged children.” *Modern Psychological Studies*. 2021, <https://scholar.utc.edu/cgi/viewcontent.cgi?article=1537&context=mps>.
- Roberts, S. Craig et al. “Distinguishing between perceived and wearer effects in clothing color-associated attributions.” *Evolutionary Psychology*. 2010, DOI: 10.1177/147470491000800304.
- Elliot, Andrew J. “Color and psychological functioning: a review of theoretical and empirical work.” *Frontiers in Psychology*. 2015, DOI: 10.3389/fpsyg.2015.00368.
- Su, Lixu et al. “Trustworthy Blue or Untrustworthy Red: The Influence of Colors on Trust.” *The Journal of Marketing Theory and Practice*, vol. 27, no. 3, 2019, pp. 269-281. DOI: 10.1080/10696679.2019.1616560.
- Boyatzis, Chris J., and Reenu Varghese. “Children's Emotional Associations with Colors.” *The Journal of Genetic Psychology*, vol. 155, no. 1, 2010, pp. 77-85, DOI: 10.1080/00221325.1994.9914760.
- Ha, Jennifer F. et al. “Doctor-Patient Communication: A Review.” *The Ochsner Journal*, vol. 10, no. 1, 2010, pp. 38-43, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3096184/>.
- Turow, Judy A., and Robert C. Sterling. “The role and impact of gender and age on children's preferences for pediatricians.” *Ambulatory Pediatrics*, vol. 4, no. 4, 2004, pp. 340-343, DOI: 10.1367/A03-110R1.1.
- McCambridge, Jim et al. “The Effects of Demand Characteristics on Research Participant Behaviours in Non-Laboratory Settings: A Systematic Review.” *PLoS ONE*, vol. 7, no. 6, 2012, DOI: 10.1371/journal.pone.0039116.
- Usability.gov staff. “Color Basics.” *Usability.gov*, n.d., <https://www.usability.gov/how-to-and-tools/methods/color-basics.html>.

Financing Modeling of Renewable Energy By Xavier Ma

Abstract

As the world increases its reliance on energy, the practice of adopting innovative, renewable energy sources has grown in significance. This study focuses on the effects of regulations and incentives on renewable energy financing in the U.S. The goal is to examine the connection between factors such as GDP and renewable energy financing. Using data from 2016 to 2019, a regression model is used to analyze how factors like GDP affect renewable energy consumption. Findings suggest that adopting certain renewable energy sources are more correlated with GDP, including solar, geothermal, biomass, and total renewable energy. In addition, economic factors were examined to determine which affected renewable energy consumption the most. The analysis suggested that the most effective energy source in the U.S. was solar, due to the high GDP in the U.S. as well as the rate of technological advancement in solar energy.

Introduction

The modeling of renewable energy finance is essential as the world begins to follow the guidelines of numerous climate agreements. The Paris Agreement focuses on reducing emissions in high-emission sectors by implementing zero-carbon solutions and considering the impact of policies on renewable energy construction (UN, “The Paris Agreement”).

Following the onset of the Russian invasion of Ukraine, global nations have been made aware of their energy supplies. Energy prices have surged at least 63% and up to 113% in some areas (Guan et al., 304). In addition, government spending worldwide has increased by \$500 billion since the invasion (Ellerbeck, “Government spending on clean energy...”). The global energy economy’s heavy reliance on Russian oil and gas has resulted in the consequence of this heavy reliance affecting certain regions significantly. A few years before the war started, two-thirds of European imports from Russia comprised energy-related products (Politico, “The delayed impact of the EU’s...”).

Even without the onset of the ‘special military operation’ over the past few years, renewable energy investment has also increased in many countries worldwide including China and India. In these two countries, the lack of a consistent energy supply in rural villages has become increasingly apparent. China still harbours 30 million people who do not have access to electricity (Liming, 1097). In addition, India accounts for a third of the world’s population who still do not have access to electricity (Bhattacharyya, “Energy access problem of the poor...”). Thus the need for energy becomes ever apparent, yet the method by which that electricity is generated has still leaned towards fossil fuels for centuries. Investment in fossil fuel energies remains as high as ever as over 1 trillion USD continues to go towards gas, coal, and oil. However, evidence points to the surge of renewable energy finance, in 2023, 1.7 trillion is set to be invested into clean technologies. Furthermore, for the first time, solar power is expected to

exceed oil production (IEA, “Overview and Key Findings...”). This indicates the growing investment in renewable energy financing.

The importance of the way we finance clean technologies remains essential to effectively shift the world away from fossil fuels. Specifically, financiers tend to focus on obtaining finance instead of managing the financial assets effectively, this leads to skewed distribution of funds and leads to certain sectors of renewable energy research needing to be more balanced. (Mazzucato, Semieniuk, 2018, p.8) On the other hand, many factors come into consideration when building renewable energy infrastructure such as location, demand, capacity, and resource availability. Wind and Solar for instance require careful planning when erecting power plants. Wind in particular requires at least 9 mph for smaller wind turbines and 13 mph for utility-scale turbines, and not all locations are optimal (EIA, “Where Wind Power Is Harnessed...”).

Consequently, in the U.S. alone, five states comprised 57% of wind energy generation in 2022 (EIA, “Where Wind Power Is Harnessed...”). In addition, Solar requires areas with high concentrations of solar radiation further highlighting the importance of location as California and Texas alone comprised 42% of utility-scale solar electricity “(EIA, Where Solar Is Found...)”. Therefore, the financial planning of investment into renewable energies needs to be controlled and managed for further efficacy.

Our results find that GDP affects different types of renewable energy, however, the benefits of financing each renewable energy source differ, with some sources being better for public investors than others. In our data, we found that solar is currently the most prominent renewable energy source being developed and is a renewable energy source worth financing. This project will aim to explain why that is the case through regression analysis of energy consumption data from the IEA and research from outside sources.

Data

This research aims to find the most effective way to finance renewable energy. Using the data below, we analyze the relationships between different variables. In this project, data from 42 U.S. states is used and the data taken for the tables is from 2016-2019. The data is compiled from different sources and relisted as graphs and tables to help better understand the relationship between different variables. We were not able to acquire data for all U.S. states and variables; however, the data is still taken from a majority of U.S. states meaning that the results are still significant. The data is in a panel structure because we have time series and cross-sectional components. Some of the data samples are taken from the IEA, which lists detailed information on energy consumption and price. This data was directly implemented into our compiled data, which was then run through code and Excel to find the listed values in the tables below.

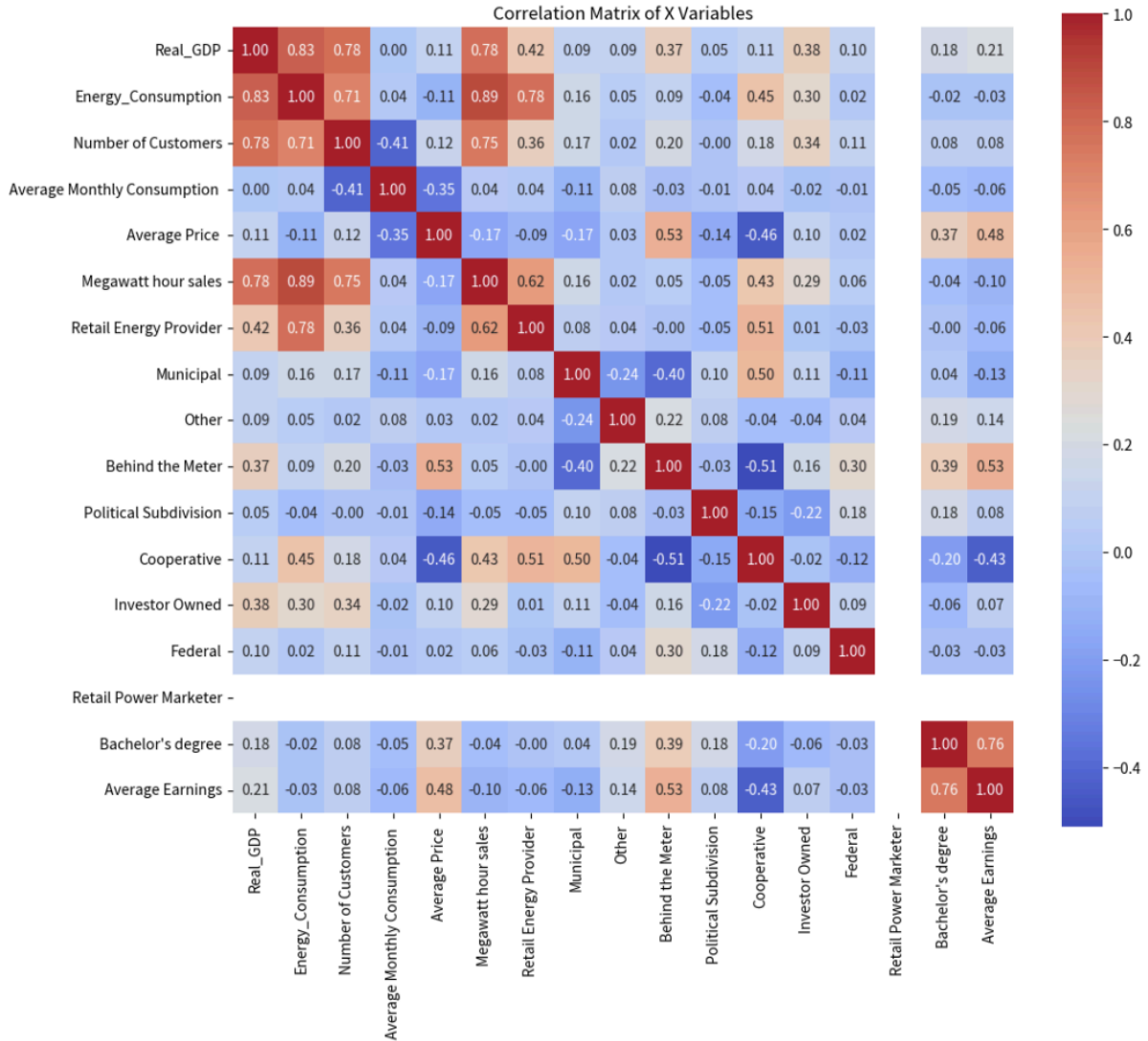
Results

We noticed many signs of collinearity throughout the results using the correlation matrix below. The closer to the red, which means 1, the more correlated they were. However, we

excluded certain variables that displayed signs of multicollinearity to avoid biased estimates. We used a cutoff point of 0.8 to exclude certain variables such as energy consumption.

Table 1

Full data set on correlation between all variables used in the research paper.



Source: Data from *Electric Power Monthly - U.S. Energy Information Administration (EIA)*, EIA, 22 Nov. 2023, www.eia.gov/electricity/monthly/.

Table 2

Total Renewable Energy Consumption Compared with other Independent Variables and finding the variation the data can account for.

PanelOLS Estimation Summary						
Dep. Variable:	Total Renewable	R-squared:	0.7851			
Estimator:	PanelOLS	R-squared (Between):	-2.1278			
No. Observations:	163	R-squared (Within):	0.7851			
Date:	Sun, Nov 12 2023	R-squared (Overall):	-2.1103			
Time:	16:29:19	Log-likelihood	-1767.0			
Cov. Estimator:	Unadjusted					
		F-statistic:	27.923			
Entities:	42	P-value	0.0000			
Avg Obs:	3.8810	Distribution:	F(14,107)			
Min Obs:	2.0000					
Max Obs:	4.0000	F-statistic (robust):	27.923			
		P-value	0.0000			
Time periods:	4	Distribution:	F(14,107)			
Avg Obs:	40.750					
Min Obs:	38.000					
Max Obs:	42.000					
Parameter Estimates						
	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
const	-1.451e+05	9.669e+04	-1.5006	0.1364	-3.368e+05	4.659e+04
Real_GDP	0.3809	0.0828	4.6001	0.0000	0.2167	0.5450
Energy_Consumption	0.0970	0.0243	4.0016	0.0001	0.0490	0.1451
Number of Customers	-0.0089	0.0013	-6.9936	0.0000	-0.0114	-0.0063
Average Monthly Consumption	-0.1998	1.2064	-0.1656	0.8688	-2.5914	2.1918
Average Price	2695.8	1746.3	1.5438	0.1256	-765.93	6157.6
Megawatt hour sales	0.0051	0.0017	2.9750	0.0036	0.0017	0.0086
Municipal	475.48	115.60	4.1130	0.0001	246.31	704.65
Behind the Meter	4800.3	2512.9	1.9102	0.0588	-181.31	9781.8
Political Subdivision	-3.785e+04	4411.4	-8.5795	0.0000	-4.659e+04	-2.91e+04
Cooperative	-3090.6	2047.0	-1.5098	0.1341	-7148.6	967.47
Investor Owned	-5474.8	1.057e+04	-0.5178	0.6057	-2.644e+04	1.549e+04
Federal	1.781e+05	2.002e+04	8.8975	0.0000	1.384e+05	2.178e+05
Bachelor's degree	3841.2	5033.3	0.7631	0.4471	-6136.8	1.382e+04
Average Earnings	-2.8287	1.3749	-2.0574	0.0421	-5.5542	-0.1031
F-test for Poolability: 212.47						
P-value: 0.0000						
Distribution: F(41,107)						
Included effects: Entity						

Source: Data from *Electric Power Monthly - U.S. Energy Information Administration (EIA)*, EIA, 22 Nov. 2023, www.eia.gov/electricity/monthly/.

Note: Table 2 shows the fixed effects model results using total renewable energy consumption as the dependent variable. Time-fixed effects and state-fixed effects are used to account for time-invariant characteristics and deal with the limitations of omitted variable bias. The R squared is 0.7851, indicating that 78.51% of the variation in the dependent variable—total renewable energy consumption—is explained by our fixed effects model. Using the significance

level of 5%, we can see that only the following variables are not significant: average monthly consumption, average price, behind the meter, cooperative, investor-owner, and bachelor's degree. This is a surprising result since I thought that education would affect the given dependent variable due to the increased knowledge that people would have on environmental topics. Yet this is not the case as shown with the P value. The coefficient with the Real GDP shows that an increase of one thousand dollars in Real GDP is partially correlated with an increase of 0.38 million BTU in total renewable energy consumption.

Table 3

Total Solar Energy Consumption Compared with other Independent Variables and finding the variation the data can account for.

PanelOLS Estimation Summary						
Dep. Variable:	Solar_MBTu	R-squared:	0.8536			
Estimator:	PanelOLS	R-squared (Between):	-5.9986			
No. Observations:	163	R-squared (Within):	0.8536			
Date:	Sun, Nov 12 2023	R-squared (Overall):	-5.7610			
Time:	16:34:12	Log-likelihood	-1565.7			
Cov. Estimator:	Unadjusted					
		F-statistic:	44.577			
Entities:	42	P-value	0.0000			
Avg Obs:	3.8810	Distribution:	F(14,107)			
Min Obs:	2.0000					
Max Obs:	4.0000	F-statistic (robust):	44.577			
		P-value	0.0000			
Time periods:	4	Distribution:	F(14,107)			
Avg Obs:	40.750					
Min Obs:	38.000					
Max Obs:	42.000					
Parameter Estimates						
	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
const	-1.171e+05	2.812e+04	-4.1633	0.0001	-1.728e+05	-6.133e+04
Real_GDP	0.4244	0.0241	17.625	0.0000	0.3767	0.4722
Energy_Consumption	-0.0258	0.0071	-3.6570	0.0004	-0.0398	-0.0118
Number of Customers	-0.0009	0.0004	-2.5764	0.0113	-0.0017	-0.0002
Average Monthly Consumption	0.2265	0.3509	0.6457	0.5199	-0.4690	0.9221
Average Price	703.46	507.88	1.3851	0.1689	-303.35	1710.3
Megawatt hour sales	0.0011	0.0005	2.1593	0.0331	8.895e-05	0.0021
Municipal	-24.945	33.622	-0.7419	0.4598	-91.597	41.707
Behind the Meter	463.04	730.85	0.6336	0.5277	-985.78	1911.9
Political Subdivision	-2441.1	1283.0	-1.9027	0.0598	-4984.5	102.24
Cooperative	-454.64	595.36	-0.7637	0.4468	-1634.9	725.58
Investor Owned	342.20	3075.1	0.1113	0.9116	-5753.9	6438.3
Federal	1.05e+04	5822.8	1.8027	0.0743	-1046.4	2.204e+04
Bachelor's degree	366.00	1463.9	0.2500	0.8030	-2536.0	3268.0
Average Earnings	-0.5317	0.3999	-1.3297	0.1864	-1.3244	0.2610
F-test for Poolability: 124.44						
P-value: 0.0000						
Distribution: F(41,107)						
Included effects: Entity						

Source: Data from *Electric Power Monthly - U.S. Energy Information Administration (EIA)*, EIA, 22 Nov. 2023, www.eia.gov/electricity/monthly/.

Note: Table 3 shows the fixed effects model results using total solar energy consumption as the dependent variable. The data uses time- and state-fixed effects to reduce omitted variable bias and account for variables that change due to time factors. The R squared is 0.8536, indicating that 85.36% of the variation in the dependent variable can be explained through the model. Keeping in mind the significance level of 5%, we can see that the variables average monthly consumption, behind the meter, investor-owned, average price, municipal, political subdivision, cooperative, federal, average earnings, and bachelor's degree are **not** significant. This is surprising because I thought that education would affect renewable energy consumption as a more educated populace would be able to identify the effects of climate change and attempt to reduce their carbon footprint. The coefficient with the Real GDP, shows that an increase of one thousand dollars in Real GDP leads to an increase of 0.42 million BTU in solar energy consumption.

Table 4

Total Solar Energy Consumption Compared with other Independent Variables and finding the variation the data can account for.

PanelOLS Estimation Summary						
Dep. Variable:	Wind_MBTu	R-squared:	0.7816			
Estimator:	PanelOLS	R-squared (Between):	-2.5768			
No. Observations:	163	R-squared (Within):	0.7816			
Date:	Sun, Nov 12 2023	R-squared (Overall):	-2.5205			
Time:	16:34:44	Log-likelihood	-1668.7			
Cov. Estimator:	Unadjusted	F-statistic:	27.349			
Entities:	42	P-value	0.0000			
Avg Obs:	3.8810	Distribution:	F(14,107)			
Min Obs:	2.0000					
Max Obs:	4.0000	F-statistic (robust):	27.349			
		P-value	0.0000			
Time periods:	4	Distribution:	F(14,107)			
Avg Obs:	40.750					
Min Obs:	38.000					
Max Obs:	42.000					
Parameter Estimates						
	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
const	-1.445e+05	5.291e+04	-2.7317	0.0074	-2.494e+05	-3.964e+04
Real_GDP	-0.0782	0.0453	-1.7261	0.0872	-0.1680	0.0116
Energy_Consumption	0.1625	0.0133	12.250	0.0000	0.1362	0.1888
Number of Customers	-0.0010	0.0007	-1.4809	0.1416	-0.0024	0.0003
Average Monthly Consumption	-0.3177	0.6601	-0.4813	0.6313	-1.6263	0.9909
Average Price	-828.47	955.52	-0.8670	0.3879	-2722.7	1065.7
Megawatt hour sales	-0.0025	0.0009	-2.6463	0.0094	-0.0044	-0.0006
Municipal	365.07	63.256	5.7712	0.0000	239.67	490.47
Behind the Meter	2989.7	1375.0	2.1743	0.0319	263.90	5715.5
Political Subdivision	-2282.6	2413.8	-0.9456	0.3465	-7067.6	2502.5
Cooperative	-1310.7	1120.1	-1.1702	0.2445	-3531.2	909.72
Investor Owned	1235.6	5785.5	0.2136	0.8313	-1.023e+04	1.27e+04
Federal	9225.4	1.095e+04	0.8421	0.4016	-1.249e+04	3.094e+04
Bachelor's degree	-3208.3	2754.1	-1.1649	0.2466	-8668.0	2251.4
Average Earnings	0.5368	0.7523	0.7136	0.4770	-0.9545	2.0282
F-test for Poolability: 166.44						
P-value: 0.0000						
Distribution: F(41,107)						
Included effects: Entity						

Source: Data from *Electric Power Monthly - U.S. Energy Information Administration (EIA)*, EIA, 22 Nov. 2023, www.eia.gov/electricity/monthly/.

Note: With total wind energy consumption as the dependent variable, the findings of the fixed effects model are displayed in Table 4. Temporal- and state-fixed effects are used in the data to account for variables that vary as a result of temporal factors and minimize the bias caused by omitted variables. With an R squared of 0.7816, the model can account for 78.16% of the variation observed in the dependent variable. The average monthly consumption, investor-owned, number of customers, average price, political subdivision, cooperative, Real GDP, federal, bachelor's degree, and average earnings variables are **not** significant, according to the 5% significance level. This is unexpected since, as Table 1 explains, we expected education to have an impact on the dependent variable; however, this is not the case because the effect is not statistically significant. Federal is also not significant, which is also surprising considering the government often invests in wind energy.

Table 5
Total Hydropower Energy Consumption Compared with other Independent Variables and finding the variation the data can account for.

PanelOLS Estimation Summary						
Dep. Variable:	Hydropower_MBTu	R-squared:	0.5822			
Estimator:	PanelOLS	R-squared (Between):	-3.9599			
No. Observations:	163	R-squared (Within):	0.5822			
Date:	Sun, Nov 12 2023	R-squared (Overall):	-3.9118			
Time:	16:35:13	Log-likelihood	-1732.5			
Cov. Estimator:	Unadjusted	F-statistic:	10.649			
Entities:	42	P-value	0.0000			
Avg Obs:	3.8810	Distribution:	F(14,107)			
Min Obs:	2.0000	F-statistic (robust):	10.649			
Max Obs:	4.0000	P-value	0.0000			
Time periods:	4	Distribution:	F(14,107)			
Avg Obs:	40.750					
Min Obs:	38.000					
Max Obs:	42.000					
Parameter Estimates						
	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
const	-2.252e+04	7.823e+04	-0.2878	0.7740	-1.776e+05	1.326e+05
Real_GDP	-0.0293	0.0670	-0.4380	0.6623	-0.1621	0.1035
Energy_Consumption	-0.0414	0.0196	-2.1114	0.0371	-0.0803	-0.0025
Number of Customers	-0.0066	0.0010	-6.4256	0.0000	-0.0086	-0.0046
Average Monthly Consumption	0.1766	0.9761	0.1809	0.8568	-1.7584	2.1115
Average Price	2965.5	1412.8	2.0990	0.0382	164.73	5766.2
Megawatt hour sales	0.0067	0.0014	4.7896	0.0000	0.0039	0.0095
Municipal	123.73	93.531	1.3229	0.1887	-61.684	309.14
Behind the Meter	2501.7	2033.1	1.2305	0.2212	-1528.7	6532.0
Political Subdivision	-3.202e+04	3569.0	-8.9714	0.0000	-3.909e+04	-2.494e+04
Cooperative	-881.01	1656.2	-0.5320	0.5959	-4164.2	2402.2
Investor Owned	-7627.8	8554.4	-0.8917	0.3746	-2.459e+04	9330.4
Federal	1.39e+05	1.62e+04	8.5834	0.0000	1.069e+05	1.711e+05
Bachelor's degree	7978.1	4072.2	1.9591	0.0527	-94.631	1.605e+04
Average Earnings	-2.6620	1.1123	-2.3931	0.0184	-4.8671	-0.4569
F-test for Poolability: 194.63						
P-value: 0.0000						
Distribution: F(41,107)						
Included effects: Entity						

Source: Data from *Electric Power Monthly - U.S. Energy Information Administration (EIA)*, EIA, 22 Nov. 2023, www.eia.gov/electricity/monthly/.

Note: With total hydroenergy consumption as the dependent variable, the fixed effects model's findings are shown in Table 5. Temporal- and state-fixed effects are added to the data to reduce the bias arising from omitted variables and account for variables that fluctuate owing to temporal conditions. With an R squared of 0.5822, the model accounts for 58.22% of the variance observed in the dependent variable. At the 5% significance level, the energy consumption, the number of customers, megawatt-hour sales, average price, political subdivision, federal, and average earnings variables are all significant. This is expected, as the government is the main financier of hydropower due to the scale of the construction projects for hydro dams. Real GDP is not a significant factor, which is important to consider.

Table 6

Total Geothermal Energy Consumption Compared with other Independent Variables and finding the variation the data can account for.

PanelOLS Estimation Summary						
Dep. Variable:	Geothermal_MBTu	R-squared:	0.6155			
Estimator:	PanelOLS	R-squared (Between):	-0.3197			
No. Observations:	163	R-squared (Within):	0.6155			
Date:	Sun, Nov 12 2023	R-squared (Overall):	-0.3187			
Time:	16:35:38	Log-likelihood	-1222.0			
Cov. Estimator:	Unadjusted					
		F-statistic:	12.233			
Entities:	42	P-value	0.0000			
Avg Obs:	3.8810	Distribution:	F(14,107)			
Min Obs:	2.0000					
Max Obs:	4.0000	F-statistic (robust):	12.233			
		P-value	0.0000			
Time periods:	4	Distribution:	F(14,107)			
Avg Obs:	40.750					
Min Obs:	38.000					
Max Obs:	42.000					
Parameter Estimates						
	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
const	4142.9	3413.3	1.2137	0.2275	-2623.7	1.091e+04
Real_GDP	-0.0101	0.0029	-3.4674	0.0008	-0.0159	-0.0043
Energy_Consumption	0.0008	0.0009	0.9040	0.3680	-0.0009	0.0025
Number of Customers	-5.451e-05	4.469e-05	-1.2197	0.2252	-0.0001	3.408e-05
Average Monthly Consumption	-0.0259	0.0426	-0.6084	0.5442	-0.1103	0.0585
Average Price	-59.629	61.644	-0.9673	0.3356	-181.83	62.573
Megawatt hour sales	7.149e-05	6.103e-05	1.1713	0.2441	-4.95e-05	0.0002
Municipal	2.2633	4.0809	0.5546	0.5803	-5.8266	10.353
Behind the Meter	-47.927	88.707	-0.5403	0.5901	-223.78	127.92
Political Subdivision	970.97	155.72	6.2352	0.0000	662.26	1279.7
Cooperative	46.908	72.261	0.6491	0.5176	-96.342	190.16
Investor Owned	33.541	373.24	0.0899	0.9286	-706.37	773.45
Federal	-6247.7	706.74	-8.8402	0.0000	-7648.8	-4846.7
Bachelor's degree	-80.509	177.68	-0.4531	0.6514	-432.74	271.72
Average Earnings	0.0221	0.0485	0.4549	0.6501	-0.0741	0.1183
F-test for Poolability: 1078.0						
P-value: 0.0000						
Distribution: F(41,107)						
Included effects: Entity						

Source: Data from *Electric Power Monthly - U.S. Energy Information Administration (EIA)*, EIA, 22 Nov. 2023, www.eia.gov/electricity/monthly/.

Note: Table 6 displays the results of the fixed effects model using total geothermal consumption as the dependent variable. To account for variables that change due to temporal circumstances and to lessen the bias resulting from omitted variables, temporal- and state-fixed effects are added to the data. The model explains 61.55% of the variation seen in the dependent variable, with an R squared of 0.6155. The only significant variables are Real GDP, political subdivision, and federal. This is expected, as geothermal is less commonly invested in and financed by the public sector; wind and solar are often more popular choices. The correlation between geothermal energy consumption and Real GDP indicates that an increase in Real GDP of \$1,000 is associated with a corresponding decrease in geothermal energy consumption of 0.0101 million BTU.

Table 7

Total Biomass Energy Consumption Compared with other Independent Variables and finding the variation the data can account for.

```

=====
PanelOLS Estimation Summary
=====
Dep. Variable:      Biomass_MBTu  R-squared:          0.3989
Estimator:         PanelOLS      R-squared (Between): 0.2880
No. Observations:  163          R-squared (Within):  0.3989
Date:              Sun, Nov 12 2023  R-squared (Overall): 0.2926
Time:              16:35:42      Log-likelihood       -1570.4
Cov. Estimator:    Unadjusted

Entities:          42          F-statistic:        5.0712
Avg Obs:           3.8810      P-value             0.0000
Min Obs:           2.0000      Distribution:        F(14,107)
Max Obs:           4.0000      F-statistic (robust): 5.0712
Time periods:     4          P-value             0.0000
Avg Obs:           40.750     Distribution:        F(14,107)
Min Obs:           38.000
Max Obs:           42.000

=====
Parameter Estimates
=====
Parameter  Std. Err.  T-stat  P-value  Lower CI  Upper CI
-----
const      1.349e+05  2.895e+04  4.6597  0.0000  7.75e+04  1.923e+05
Real_GDP   0.0741     0.0248    2.9903  0.0035  0.0250    0.1233
Energy_Consumption 0.0009     0.0073    0.1292  0.8975  -0.0135   0.0153
Number of Customers -0.0002     0.0004   -0.6421  0.5222  -0.0010   0.0005
Average Monthly Consumption -0.2593     0.3612   -0.7180  0.4743  -0.9753   0.4566
Average Price -85.000     522.78    -0.1626  0.8711  -1121.3    951.35
Megawatt hour sales -0.0002     0.0005   -0.4055  0.6859  -0.0012   0.0008
Municipal   9.3626     34.609    0.2705  0.7873  -59.245    77.970
Behind the Meter -1106.2     752.29    -1.4705  0.1444  -2597.6    385.09
Political Subdivision -2075.5     1320.6    -1.5716  0.1190  -4693.4    542.52
Cooperative -491.07     612.82    -0.8013  0.4247  -1705.9    723.77
Investor Owned 541.61     3165.3    0.1711  0.8645  -5733.3    6816.5
Federal     2.563e+04  5993.6    4.2760  0.0000  1.375e+04  3.751e+04
Bachelor's degree -1214.1     1506.8    -0.8057  0.4222  -4201.2    1773.0
Average Earnings -0.1939     0.4116   -0.4711  0.6385  -1.0099    0.6220
=====

F-test for Poolability: 354.42
P-value: 0.0000
Distribution: F(41,107)

Included effects: Entity

```

Source: Data from *Electric Power Monthly - U.S. Energy Information Administration (EIA)*, EIA, 22 Nov. 2023, www.eia.gov/electricity/monthly/.

Note: With total biomass use as the dependent variable, the fixed effects model's findings are shown in Table 7. State- and temporal-fixed effects are added to the data to reduce the bias caused by missing variables and account for factors that vary with temporal conditions. With an

R squared of 0.3989, the model accounts for 39.89% of the variance observed in the dependent variable. Only two variables—Real GDP and Federal—are statistically significant at the 5% significance level. These results are somewhat expected, as the government mainly funds biomass operations and biomass investment opportunities are not as easily accessible to public investors. Thus, it is not as easily financeable as other renewable energy sources like solar and wind. In addition, biomass energy consumption increases by 0.0741 million BTU for every \$1,000 rise in Real GDP, according to the link between biomass energy consumption and Real GDP.

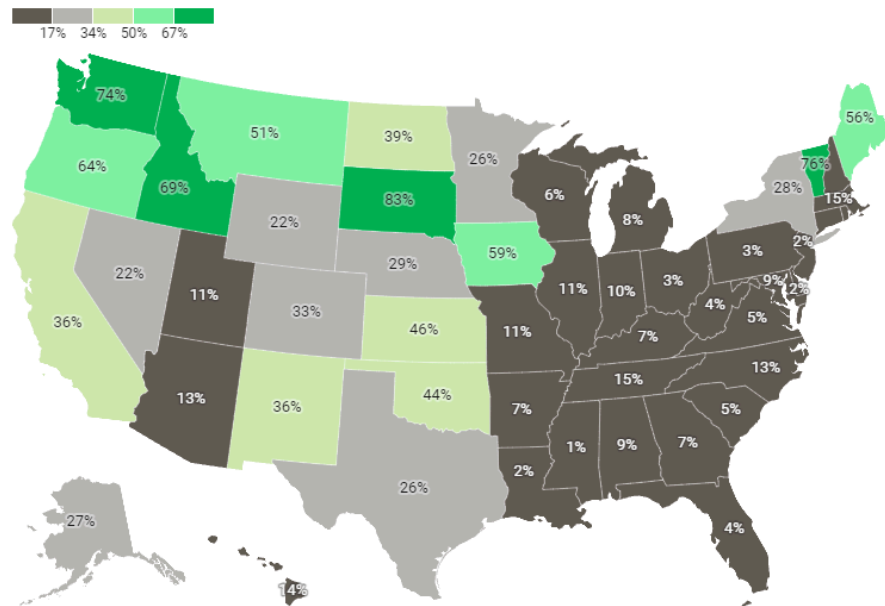


Fig. 1. A map showing the percentage of total renewable energy production in each state from Kirk, Karin. “Which State Is Winning at Renewable Energy Production? ” Yale Climate Connections.” Yale Climate Connections, 1 May 2023, yaleclimateconnections.org/2023/02/us-state-with-most-renewable-energy-production/.

Note: This map shows the raw percentage of energy demand fulfilled by renewable energy in each U.S. state with the range being between 1% and 83%. In this case, brighter colors mean more renewable energy production and darker colors mean less renewable energy production. As shown in the map, Eastern states have a significantly lower level of renewable energy in their energy portfolios when compared with Western states like California, Oregon, and Washington.

U.S. primary energy consumption by energy source, 2022

total = 100.41 quadrillion
British thermal units (Btu)

total = 13.18 quadrillion Btu

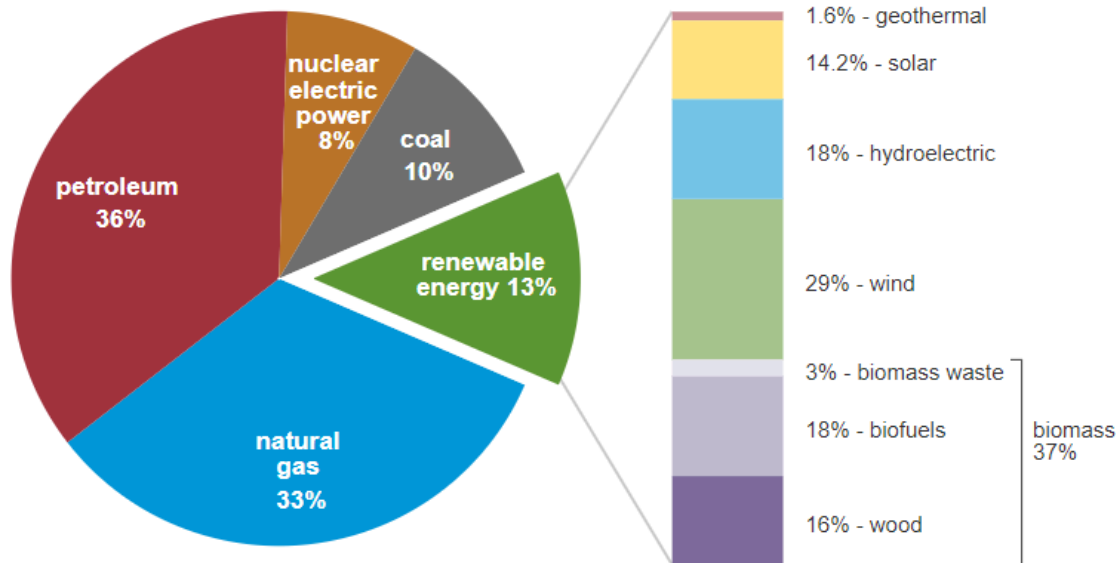


Fig. 2. The energy consumption portfolio of the U.S. in 2022 from “U.S. Energy Information Administration - EIA - Independent Statistics and Analysis.” *U.S. Energy Facts Explained - Consumption and Production - U.S. Energy Information Administration (EIA)*, EIA, Apr. 2023, www.eia.gov/energyexplained/us-energy-facts/.

Note: The table shows that there is a heavy reliance on biomass currently which makes up most of the renewable energy. However, solar, hydro, and wind are all getting close to the amount of biomass energy consumed, meaning that biomass is slowly losing relevance. In addition, U.S. energy needs are mostly covered by petroleum and natural gas meaning that a shift away from these fuel sources will take a long period.

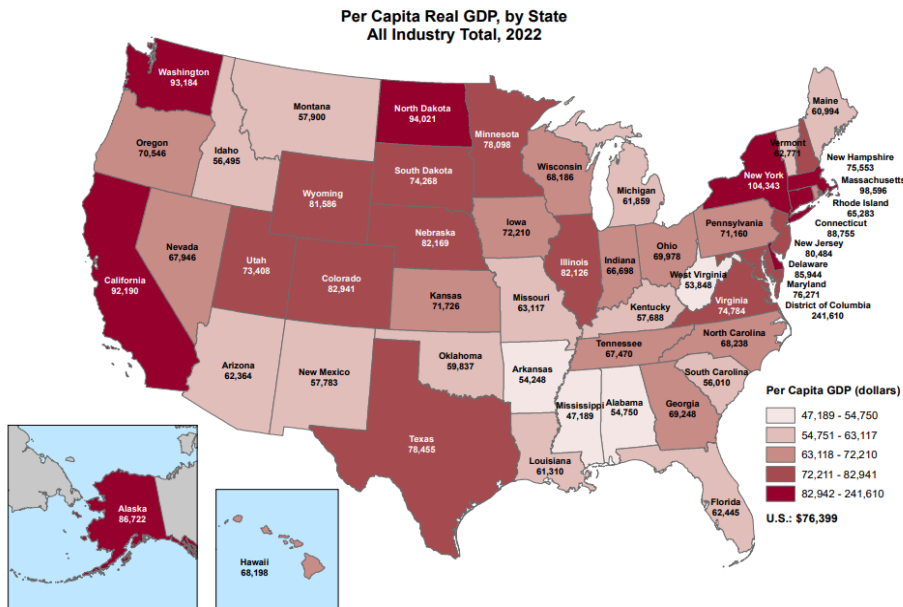


Fig. 3. A map of the U.S. displaying the per capita Real GDP by state from *Per Capita Real GDP, by State All Industry Total, 2022*, The University of Kansas, 2022, ipsr.ku.edu/ksdata/ksah/business/percapGDP.pdf.

Note: The map shows that Western states tend to have higher per capita real GDP compared to the Eastern States, except states like New York, and Massachusetts. In general, the North also tends to have a higher per capita GDP, as shown by the lighter colors at the bottom of the map compared to the top.

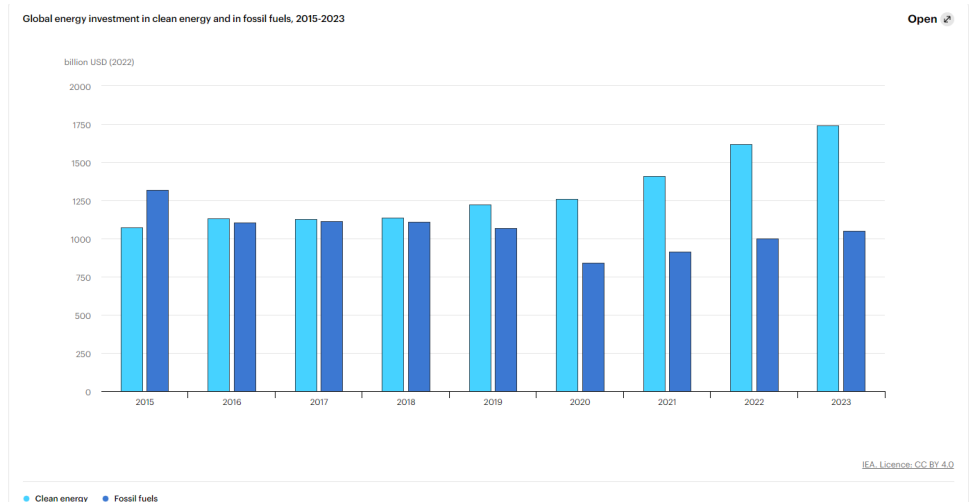


Fig. 4. Global investment into clean energy compared with fossil fuel investment over 8 years, with light blue being clean energy, and dark blue being fossil fuel energy from “Overview and Key Findings – World Energy Investment 2023 – Analysis.” *IEA*, IEA, 2023, www.iea.org/reports/world-energy-investment-2023/overview-and-key-findings.

Note: This table shows the growing trend of renewable energy investment around the globe, which also applies to the U.S. It also helps explain the development of cheaper solar, wind and other renewables. It also signals the growing obscurity of fossil fuels.

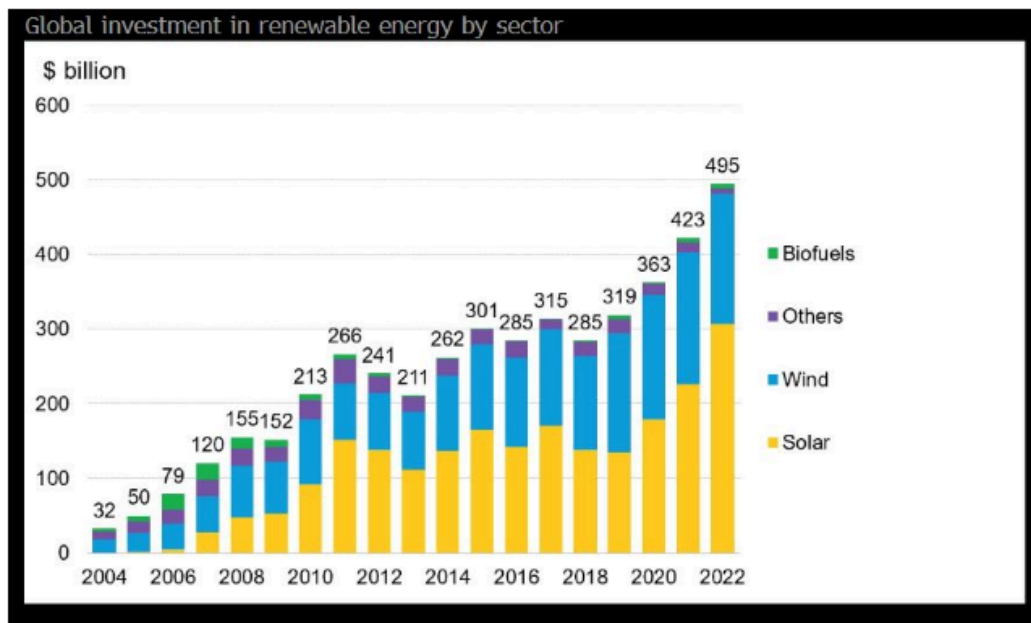


Fig. 5. Global investment into renewable energy by sector in billions of USD from “A Record \$495 Billion Invested in Renewable Energy in 2022.” *BloombergNEF*, Bloomberg, 2 Feb. 2023, about.bnef.com/blog/a-record-495-billion-invested-in-renewable-energy-in-2022/.

Note: In most of the table, wind energy has seen heavy investments while solar investments have increased over time, passing wind in total investments in 2016-2022. By contrast, other renewable energy sources like biofuels have not been as heavily invested in. This is significant as it shows that wind and solar are the most popular choices in terms of financing for many countries around the world.

Financing Solar

Throughout the U.S., there is existing renewable energy infrastructure in many places; however, the existing infrastructure cannot support full electrification. This also indicates that it does not provide enough power to support a switch to a fully renewable portfolio. In some states, less than 10% of their required energy is renewable (see Fig. 1). Other states with more existing renewable energy production can reach a 100% renewable portfolio more quickly, as shown by

the Western states where renewable energy makes up near 50% of their energy portfolio (see Fig. 1).

When taking this into account, it is important to note that some states are more well-suited to adopt renewable energy. For example, there is a strong positive correlation between total renewable infrastructure and GDP, meaning that states with a stronger GDP have a stronger financial base to invest in and better finance renewable energy (see Table. 1). Solar prices have also dropped significantly since the start of the 21st century: the average solar system cost \$10 per watt in the early 2000s; however, the value of solar solar photovoltaic has since decreased to \$0.1 per kWh (Nath, “The Economics of Solar Power”). This trend is likely to continue as solar technology improves.

Though solar has a significant influence on renewable energy now, other existing renewable energy sources are also making a difference in the shift toward more renewable energy sourcing. The majority of U.S. renewable energy comes from biomass, hydro, wind, and solar (see Fig. 2). Despite better solar technology, wind, biomass, and hydro still make up more renewable energy production than solar (see Fig. 2). However, solar has only recently become more efficient as mass production of solar panels has become more effective and existing technology has been improved. In the U.S., 46% of all new generating capacity in 2021 came from solar (Shinn and Lora, *Renewable Energy: The Clean Facts*). Solar is being developed at a rapid rate and investors have been quick to notice the changes that new solar technology has brought.

According to Table 1, 85.36% of the variation in the variable, Solar energy consumption, can be explained through the fixed effects model. This means that the Real GDP is heavily correlated with more solar energy consumption, meaning that the stronger a state’s economy is, the more consumption of solar energy there will be. Thus, there will be a larger market for financiers in this area. This is also supported by the data shown in Figures 1 and 3, as the states with a higher GDP tend to invest more in renewables which end up making up more of their energy portfolio. This isn’t without its exceptions, however, as North Dakota and New York have a high GDP yet a lower amount of their energy portfolio is made up of renewables (see Figs. 1 and 3).

Another factor that warrants consideration is that a highly educated populace does not have a significant effect on the amount of solar energy consumption in a given state (see Table 2). This doesn’t fully discount education’s impact on solar energy consumption; however, educated individuals tend to perceive climate change and other social and environmental issues as more significant (Keser, “Energy, Environment, and Education...”). This means that education may have an impact, but further research is needed to determine whether there is a true correlation between the two variables.

Solar also has many other factors that need to be considered while investing in solar, as temperature, sunlight, and weather can affect energy production from a solar panel (Ecoflow, “The Impact of Temperature...”). From outside data and this research, it is concludable that

financing solar is more effective in higher GDP states with good weather conditions, and is a technology that is still developing to be more cost-effective.

Financing Wind

Wind infrastructure is lacking when it comes to meeting all energy demands. However, it still makes up 29% of U.S. renewable energy production (see Fig. 2). The U.S. government has also invested \$12 billion in capital investment in wind (DOE, “U.S. Department of Energy Projects Strong ...”), making it the second most invested renewable energy by the government, with federal solar investment ranking first. With wind being one of the top substitutes for solar, it is one of the most important renewable energy sources to consider when financing.

Wind is currently one of the most cost-effective renewable energy sources with advances in scientific technology (DOE, “U.S. Department of Energy Projects Strong ...”). It is also able to be utilized in urban settings by building wind farms on the periphery of a city or by integrating a wind energy design into a building (Isugah et al., 614). This is similar to solar and makes it optimal for investment because it is usable in residential and urban settings.

Due to these factors, wind still carries a majority of the load in the renewable energy portfolio with 29% of energy consumption coming from wind energy (see Fig. 2). Global energy investment into wind has also skyrocketed in recent years because of new developing technology (see fig. 5). This includes the U.S. as they have invested the second most amount, falling behind China in investment (E&E, and Schonhardt, “China Invests “\$546 Billion...”).

When financing wind, the previous factors are important, but one important factor analyzed in this research was Real GDP and its correlation with wind energy consumption. The data showed, however, that Real GDP is not a significant variable, according to the 5% significance factor, when tied to wind energy consumption (see Table. 3). This is quite unexpected because, with a stronger economy, it would seem that there would be heavier consumption of wind energy due to heavier investment into wind energy. Yet, this is not the case according to the data shown. This could be explained by how the model accounts for 78.16% of the variation observed in the dependent variable; however, a stronger explanation would be that more investment is put into solar, thus meaning that solar is the renewable energy most affected by Real GDP (see fig. 2).

Fossil fuels also continue to make up a majority of the energy portfolio in the U.S. meaning that investment is also often put into other energy production sources. An estimated \$20 billion of taxpayer money continues to go into the fossil fuel industry (Senate.gov, “Sen. Whitehouse on...”). This means that, although wind energy is one of the more heavily invested in renewable energies, due to other sources of existing energy, other renewable energy sources, and new emerging technologies in other sectors, an increase in GDP is not significant to the increase in consumption of wind energy in the U.S.

Financing Hydropower

Hydropower has long been used by nations and is still currently one of the main renewable energy sources for the U.S. It makes up 18% of all renewable energy consumption, which is less than wind, but still significant (see Fig. 2). Even today, the Biden administration has continued to finance hydropower projects, with \$13 million financed towards development of hydropower (DOE, Biden-Harris Administration”).

Hydro is often a more consistent energy source than others because it relies on running water (DOE, “Hydropower Basics”). It is also longer lasting than other renewables because of the integrity of dams and it is also much more concentrated, making it easier to distribute (Paish, and Oliver, “Small hydro power...”). Yet, large-scale projects are often mainly funded by the government, for instance, the total construction cost for hydro was around \$2.5 billion in 2016 (HydroReview, “Hydro Has Highest Average ...”). This means that the amount of finance required to construct hydro is out of reach for many public investors due to the lack of financial assets required.

Our model accounts for 58.22% of the variation observed in the dependent variable, which in this case is hydroenergy consumption (see Table. 4). It is also significant that Real GDP is not a significant factor when it comes to hydroenergy consumption (see Table. 4). Instead, federal, and political subdivision come out to be significant, which is important because it shows how government finance in the U.S. is the main driving factor behind the industry. It is important to note that small-scale hydro can also be invested in by public investors, but the main production of hydro energy comes from larger-scale projects. This is represented in the data by the significance of the variables of federal and political subdivision on the dependent variable.

While hydro continues to be a part of the transition to a fully renewable energy portfolio, the fact remains that it is also restricted by location and size. Hydropower needs a consistent water supply and a large amount of land (MIT, “Why aren’t we looking...”). The investment into hydro is also often long-term, meaning that public investors have less reason to finance these projects due to the high risk for returns that may take many years to activate. Thus, hydro may be an important renewable energy source, but from a financial perspective, it is not one of the most effective renewable energies to finance.

Financing Biomass and Geothermal

Biomass and Geothermal are often less financed by investors as they are often only used in niche situations. Geothermal is heavily constricted by location and it is quite expensive to invest in (TWI, “What Are the Advantages and Disadvantages...”). On the other hand, biomass is also quite inefficient and can cause emissions if fuel is not burned effectively (Energylopedia, “Advantages and Disadvantages of Biomass”). Yet, despite these inefficiencies, they still need to be considered as they are still actively invested in and still make up a portion of the renewable energy portfolio in the U.S.

In the U.S. geothermal, has up to \$165 million in federal investments (DOE, “Doe to Invest up to \$165 Million...”). The U.S. is also the leading global producer of geothermal energy (EIA, “Use of Geothermal Energy...” and has invested \$1.029 billion in biomass (DOE,

“Department of Energy Recovery Act Investment...”). This shows that, although these energies are somewhat inefficient, they are still invested in.

With financing biomass, our model can account for 39.89% of the variation observed in the dependent variable. There are only two significant variables in the model, Real GDP and Federal (see table. 6). This means that, currently, biomass is most influenced by the government and is not as heavily financed by public investors. Biomass energy consumption also increases by 0.0741 million BTU for every \$1,000 rise in Real GDP, according to the link between biomass energy consumption and Real GDP (see table 6). However, the model only accounts for around 40% of the variation in the dependent variable, which in this case is biomass energy consumption meaning that the dependent variable is less correlated with the independent variables.

Biomass as energy overall has been in use for many years and still makes up 37% of renewable energy (see fig. 2). Yet, it is not entirely clean and is not financed through public means. It is often financed by government streams according to our data, as there are only two significant independent variables. However, biomass is also cheap and easy to use, and it has been used in the past as the main source of energy (Zhao et al., 1886). Biomass cannot be discounted as an important renewable energy to shift into, however, looking at it from a financial perspective, it is not the most efficient renewable energy source to invest in, in the modern world.

Financing geothermal energy is an entirely different matter. Our model explains 61.55% of the variation seen in the dependent variable, and there are only three significant variables within our data:

Real GDP, political subdivision, and federal. This shows that a rise of \$1,000 in Real GDP is associated with a corresponding decrease in geothermal energy consumption of -0.0101 million BTU (see Table. 5). It also means that the government is more involved in the financing of geothermal energy as well as indicated by the significance of the variables federal, and political subdivision (see table. 5). The fact that the more GDP increases, the less geothermal energy is consumed implies that financing geothermal in the U.S. is not extremely cost-effective.

This can be explained by the geographical limitations of geothermal as well as the limiting factor of construction. Geothermal is heavily limited by location as it relies on places with volcanoes, hot springs, and geysers (EIA, “Where Geothermal Energy Is Found...”). This means that this technology cannot be readily integrated into urban settings and is not as easily accessible to public financing. Despite the difficulty of financing geothermal, the government still invests money into the technology, albeit significantly less than solar and wind. It also still makes up 1.6% of renewable energy consumption in the U.S. Overall, geothermal has also been used around the world for 80 years and will still be part of the process of transitioning to all renewable energies (Fridleifsson, “Geothermal Energy for the Benefit...”).

Conclusion

After doing research and data gathering, we have drawn multiple conclusions from the research. One main finding would be that extensive education does not affect energy consumption, meaning that a populace with a bachelor's degree does not necessarily account for more renewable energy investment. Another major finding would be that wind and solar are the top contenders for public finance as they are the most easily accessible technologies. In addition, they are also able to be used in urban environments, making them more accessible to the general public. Furthermore, solar energy consumption is heavily correlated with Real GDP meaning that richer states often have more solar energy consumption, meaning that there should be more emphasis put on financing solar energy. Wind energy does not do the same, yet it is still the second most invested in renewable energy today in the modern world. Other renewable energies such as hydro, biomass, and geothermal on the other hand are more influenced by federal financing and are not as easily publicly financible. Although the U.S. government continues to invest in these technologies, it is still significantly less financed than other renewable energies such as wind and solar. With solar and wind technology developing at such a rapid rate, prices for wind and solar will inevitably decrease with time. Overall, the research done has brought forth some surprising and some expected outcomes, and it has shown that financing solar and wind is the most effective way to transition into renewable energies in the future in the U.S., and many countries around the world have already started in this shift.

Works Cited

- Advancing the Growth of the U.S. Wind Industry: Federal Incentives ...*, US Department of Energy, Apr. 2021,
www.energy.gov/sites/default/files/2021-04/us-wind-industry-federal-incentives-funding-partnership-opportunities-fact-sheet.pdf.
- “Advantages and Challenges of Wind Energy.” *Energy.Gov*, Office of Energy Efficiency & Renewable Energy, 2023,
www.energy.gov/eere/wind/advantages-and-challenges-wind-energy#:~:text=Wind%20power%20is%20cost%20effective,and%20technology%20of%20wind%20energy.
- “Advantages and Disadvantages of Biomass.” *Energypedia*, 29 Mar. 2016,
energypedia.info/wiki/Advantages_and_Disadvantages_of_Biomass.
- Bhattacharyya, Subhes C. “Energy Access Problem of the Poor in India: Is Rural Electrification a Remedy?” *Energy Policy*, Elsevier, 12 Oct. 2005,
www.sciencedirect.com/science/article/abs/pii/S0301421505002302#:~:text=This%20paper%20looks%20into%20the,energy%20mix%20of%20the%20poor.
- “Biden-Harris Administration Invests More than \$13 Million to Enhance Continued Deployment of Hydropower.” *Energy.Gov*, Department of Energy, 6 Sept. 2023,
www.energy.gov/articles/biden-harris-administration-invests-more-13-million-enhance-continued-deployment.
- “Clean Energy Investment Is Extending Its Lead over Fossil Fuels, Boosted by Energy Security Strengths - News.” *IEA*, IEA, 25 May 2023,
www.iea.org/news/clean-energy-investment-is-extending-its-lead-over-fossil-fuels-boosted-by-energy-security-strengths.
- The Department of Energy Hydrogen and Fuel Cells Program Plan*, U.S. Department of Energy,
www.energy.gov/eere/fuelcells/articles/department-energy-hydrogen-and-fuel-cells-program-plan. Accessed 16 Nov. 2023.
- “Department of Energy Recovery Act Investment in Biomass Technologies.” *Energy Efficiency & Renewable Energy*, U.S. Department of Energy,
www.energy.gov/eere/fuelcells/articles/department-energy-recovery-act-investment-biomass-technologies. Accessed 20 Oct. 2023.
- “Doe to Invest up to \$165 Million to Advance Domestic Geothermal Energy Deployment.” *Energy.Gov*, Department of Energy, 28 July 2022,
www.energy.gov/articles/doe-invest-165-million-advance-domestic-geothermal-energy-deployment#:~:text=DOE%20to%20Invest%20Up%20to%20%24165%20Million%20to%20Advance%20Domestic%20Geothermal%20Energy%20Deployment,-July%2028%2C%202022&text=WASHINGTON%2C%20D.C.%20%E2%80%94%20The%20U.S.%20Department,expand%20U.S.%20geothermal%20energy%20deployment.
- Electric Power Monthly - U.S. Energy Information Administration (EIA)*, EIA, 22 Nov. 2023,
www.eia.gov/electricity/monthly/.

- Ellerbeck, Stefan. "Government Spending on Clean Energy Is Ramping up Globally." *Pakistan & Gulf Economist*, 11 Apr. 2023, www.pakistangulfeconomist.com/2023/04/10/government-spending-on-clean-energy-is-ramping-up-globally-heres-why-it-matters/.
- Fridleifsson, I.B. "Geothermal Energy for the Benefit of the People." *Renewable and Sustainable Energy Reviews*, Elsevier, 7 May 2001, www.sciencedirect.com/science/article/abs/pii/S1364032101000028.
- Government Spending on Clean Energy Is Ramping up Globally. Here's Why It Matters*, World Economic Forum, 5 Apr. 2023, www.weforum.org/agenda/2023/04/clean-energy-government-investment-spending/.
- Guan, Yuru, et al. "Burden of the Global Energy Price Crisis on Households." *Nature News*, Nature Publishing Group, 16 Feb. 2023, www.nature.com/articles/s41560-023-01209-8.
- Guàrdia, Arnau Busquets, and Charlie Cooper. "The Delayed Impact of the EU's Wartime Sanctions on Russia." *POLITICO*, POLITICO, 3 Feb. 2023, www.politico.eu/article/numbers-delayed-impact-eu-european-union-war-sanctions-russia-vladimir-putin/.
- Hydro Review Content Directors. "Hydro Has Highest Average Construction Cost of Any Generating Technology in U.S." *Hydro Review*, 7 Aug. 2018, www.hydroreview.com/business-finance/business/hydro-has-highest-average-construction-cost-of-any-generating-technology-in-u-s/.
- "Hydropower Basics." *Energy.Gov*, U.S. Department of Energy, www.energy.gov/eere/water/hydropower-basics. Accessed 4 Nov. 2023.
- "The Impact of Temperature on Solar Panel Efficiency: How Heat Affects Your Solar Energy System." *EcoFlow CA Blog*, Ecoflow, 17 July 2023, blog.ecoflow.com/ca/effects-of-temperature-on-solar-panel-efficiency/#:~:text=The%20optimal%20temperature%20for%20solar%20panels%20is%20around%2025%C2%B0,%25%2C%20affecting%20overall%20energy%20production.
- Isugah, T.F., et al. "Renewable and Sustainable Energy Reviews - ResearchGate." *Renewable and Sustainable Energy Reviews*, Elsevier, 17 May 2014, www.researchgate.net/profile/Tf-Ishugah/publication/263048982_Advances_in_wind_energy_resource_exploitation_in_urban_environment_A_review/links/60b4a05945851557baaf225d/Advances-in-wind-energy-resource-exploitation-in-urban-environment-A-review.pdf.
- Keser, Ömer. *Energy, Environment, and Education Relationship, in Developing ...*, Taylor & Francis Online, 21 June 2010, www.tandfonline.com/doi/abs/10.1080/00908310390142181.
- Kirk, Karin. "Which State Is Winning at Renewable Energy Production? " *Yale Climate Connections*." *Yale Climate Connections*, 1 May 2023, yaleclimateconnections.org/2023/02/us-state-with-most-renewable-energy-production/.

- Liming, Huang. “Financing Rural Renewable Energy: A Comparison between China and India.” *Renewable and Sustainable Energy Reviews*, 12 Mar. 2008, [dpl6hyzg28thp.cloudfront.net/media/RenewableEnergyFinancing_IndiavsChina.pdf](https://doi.org/10.1016/j.rser.2008.03.001).
- Mazzucato, Mariana. “Financing Renewable Energy: Who Is Financing What and Why It Matters.” *Technological Forecasting and Social Change*, Elsevier, 8 June 2017, www.sciencedirect.com/science/article/pii/S0040162517306820.
- Nath, Trevir I. “The Economics of Solar Power.” *Investopedia*, Investopedia, 31 Jan. 2022, www.investopedia.com/articles/investing/061115/economics-solar-power.asp#:~:text=The%20cost%20of%20solar%20power,economic%20considerations%20surrounding%20solar%20energy.
- News, E&E, and Sara Schonhardt. “China Invests \$546 Billion in Clean Energy, Far Surpassing the U.S.” *Scientific American*, 30 Jan. 2023, www.scientificamerican.com/article/china-invests-546-billion-in-clean-energy-far-surpassing-the-u-s/#:~:text=The%20country%20spent%20%24546%20billion,billion%20in%20clean%20energy%20investments.
- “Overview and Key Findings – World Energy Investment 2023 – Analysis.” *IEA*, IEA, 2023, www.iea.org/reports/world-energy-investment-2023/overview-and-key-findings.
- Paish, Oliver. “Small Hydro Power: Technology and Current Status.” *Renewable and Sustainable Energy Reviews*, Elsevier, 11 Mar. 2002, www.sciencedirect.com/science/article/abs/pii/S1364032102000060.
- “The Paris Agreement.” *United Nations Climate Change*, United Nations, 2015, unfccc.int/process-and-meetings/the-paris-agreement.
- Per Capita Real GDP, by State All Industry Total, 2022*, The University of Kansas, 2022, ipsr.ku.edu/ksdata/ksah/business/percapGDP.pdf.
- “A Record \$495 Billion Invested in Renewable Energy in 2022.” *BloombergNEF*, Bloomberg, 2 Feb. 2023, about.bnef.com/blog/a-record-495-billion-invested-in-renewable-energy-in-2022/.
- Russia-Ukraine War Has Nearly Doubled Household Energy Costs Worldwide – New Study*, World Economic Forum, 20 Feb. 2023, www.weforum.org/agenda/2023/02/russia-ukraine-war-energy-costs/.
- “Sen. Whitehouse on Fossil Fuel Subsidies: ‘We Are Subsidizing the Danger’: U.S. Senate Committee on the Budget.” *Chairman Press | Chairman’s Newsroom | Chairman | U.S. Senate Committee On The Budget*, United States Senate Committee on the Budget, 3 May 2023, www.budget.senate.gov/charman/newsroom/press/sen-whitehouse-on-fossil-fuel-subsidies-we-are-subsidizing-the-danger-#:~:text=It’s%20not%20just%20the%20US,to%20the%20fossil%20fuel%20industry.
- Shinn, Lora. “Renewable Energy: The Clean Facts.” *Renewable Energy Definition - Sources, Clean Alternatives*, NRDC, 1 June 2022, www.nrdc.org/stories/renewable-energy-clean-facts#sec-types.

- “U.S. Department of Energy Projects Strong Growth in U.S. Wind Power Sector.” *Energy Gov*, Department of Energy, 24 Aug. 2023, www.energy.gov/articles/us-department-energy-projects-strong-growth-us-wind-power-sector#:~:text=According%20to%20the%20new%20reports,employing%20more%20than%20125%2C000%20Americans.
- “U.S. Energy Facts Explained - Consumption and Production - U.S. Energy Information Administration (EIA).” *U.S. Energy Information Administration - EIA - Independent Statistics and Analysis*, EIA, Apr. 2023, www.eia.gov/energyexplained/us-energy-facts/.
- “Use of Geothermal Energy - U.S. Energy Information Administration (EIA).” *U.S. Energy Information Administration - EIA - Independent Statistics and Analysis*, EIA, 20 Apr. 2023, www.eia.gov/energyexplained/geothermal/use-of-geothermal-energy.php.
- “What Are the Advantages and Disadvantages of Geothermal Energy?” *TWI*, TWI, 2023, www.twi-global.com/technical-knowledge/faqs/geothermal-energy/pros-and-cons.
- “Where Geothermal Energy Is Found - U.S. Energy Information Administration (EIA).” *U.S. Energy Information Administration - EIA - Independent Statistics and Analysis*, EIA, 15 Feb. 2023, www.eia.gov/energyexplained/geothermal/where-geothermal-energy-is-found.php.
- “Where Solar Is Found - U.S. Energy Information Administration (EIA).” *U.S. Energy Information Administration - EIA - Independent Statistics and Analysis*, EIA, 25 Apr. 2023, www.eia.gov/energyexplained/solar/where-solar-is-found.php.
- “Where Wind Power Is Harnessed - U.S. Energy Information Administration (EIA).” *U.S. Energy Information Administration - EIA - Independent Statistics and Analysis*, EIA, 20 Apr. 2023, www.eia.gov/energyexplained/wind/where-wind-power-is-harnessed.php.
- “Why Aren’t We Looking at More Hydropower?” *MIT Climate Portal*, 2 Mar. 2021, climate.mit.edu/ask-mit/why-arent-we-looking-more-hydropower.
- Zhao, Xiao, et al. “Biomass-Based Chemical Looping Technologies: The Good” *Energy & Environmental Science*, Royal Society of Chemistry, 19 May 2017, pubs.rsc.org/en/content/articlepdf/2017/ee/c6ee03718f.
- Özbay, Ferhat, and Ibrahim Duyar. “Exploring the Role of Education on Environmental Quality and Renewable Energy: Do Education Levels Really Matter?” *Current Research in Environmental Sustainability*, Elsevier, 13 Aug. 2022, www.sciencedirect.com/science/article/pii/S2666049022000639?ref=pdf_download&fr=RR-2&rr=82ddffa86ce63081.

A Tale of Two Cities: A Comparative Study to Tackle the Issue of Affordable Housing by Joya Trivedi and Oliver Ong

Abstract

As population growth reaches new highs in the US, so does the requirement for safe, sustainable housing. However, the supply of these affordable housing units is dramatically outpaced by the demand of individuals and families needing places to stay. The cities within the United States that struggle in particular are those with high population growth, high median incomes, and a high number of families in the area. This paper aims to delve into the difference between certain cities with their endeavors with affordable housing by identifying the key determinants of housing costs. The objective is to explore how city governments have reacted to these challenges, resulting in a comparative study between policies in two hallmark cities in the United States with distinct responses to the affordable housing shortage. Quantitative data measured by the quantity of affordable housing units has been collected as the result of policy action by these city governments. It has been determined that high regulation and lack of partnerships between developers and the public sector have been the most detrimental to the production of affordable housing units while developing relationships between the public and private sectors along with creating a clear policy framework were essential to supplying units for residents to live in. By utilizing the experiences of other cities, humanity may finally conquer the struggle towards affordable housing for all.

Introduction

Recent years have revealed a domestic migration pattern within the United States from both the east and west coasts to Florida, North and South Carolina, north Georgia, and Texas (Bureau). Texas consistently ranks as one of the fastest-growing states in the country, and the state had the largest migration inflow from 2000 to 2022 (Bureau). It is now the only state besides California with 30 million residents due to a constant flux of domestic immigration (Bureau). This shift in population hotspots has encouraged the increasing rate of urban development in Southern states along with the economic benefits that are associated; for example, Austin, Texas is considered a second Silicon Valley or “Silicon Hills” due to large volumes of high-end technology companies establishing their headquarters and major facilities there during the 1970s and 1980s.

While these changes have stimulated economic growth through job creation, diversification of the economy, and boosts to a technologically skilled workforce, this initially blessing population growth brought upon various unwanted consequences in Austin. The inflow of businesses and employees has driven up prices of goods and resources, which in turn have created labor shortages, and in particular, had a severely adverse effect on the amount of affordable housing (Torres et al.). The impact of such an influx has been observed within other locations such as our hometown of Plano, Texas, a mid-sized city with 284,579 residents, a variety of Fortune 500 companies, and more than 10,000 businesses (“About Plano, Texas”).

Because of a variety of factors to be explored later in this paper, around 18,000 families need affordable housing but are not able to receive effective amounts of aid.

The purpose of this paper is to find potential policy solutions to the affordable housing shortage in places such as Plano with high growth and high median incomes by comparing landmark cities with affordable housing, or a lack thereof. Finding policies was done through the comparison of factors such as median income, population growth, land value, and building regulation. Once we quantified these factors for Plano, we compared them to various cities in the United States until settling on San Francisco, California and Charlotte, North Carolina for their relative similarities to Plano and based on their metrics and housing market affordability or unaffordability. The importance of searching for policy examples stems from the previous focus on only the factors of housing affordability itself and the lack of progress in cities like Plano where demand continues to grow.

Determining Influences on Housing Costs

Three main elements determine the costs of delivering a housing unit: actual land value, construction costs, and profit needed to satisfy the builders of the unit (Glaeser and Gyourko 6). First of all, geography is essential because certain land is harder or easier to develop than others. Residential land along with land in and near urban centers is also generally more expensive than in rural areas due to increased demand and a reduced accessibility to construction on constricting lots (Zhang and Hou 20-23). This lowers the amount of affordable housing available where it is needed, increasing difficulty in finding homes for low-income households in areas they are needed the most.

Next, construction costs can be seen as an umbrella term for the sub-factors that influence the investments that developers would need to put in to create a unit. Land value can be seen as a contributor to construction costs as construction companies need to purchase lots they will build on. Furthermore, the costs of materials and labor in local markets must be considered along with the need to get projects approved based on local zoning laws. This leads to the importance of understanding the correlation between government regulation and housing prices.

The Wharton Residential Land Use Regulatory Index 2018 (WRLURI2018) is a cumulative index formed from surveys whose insights are put into twelve subindexes accounting for various potential governmental restrictions placed upon metroplexes (Gyourko et al. 7). The sub-indexes are then put into a weighted average to create WRULRI2018 values for forty-four metroplexes across the United States. Values greater than 0.637 mean an area is highly regulated, between -0.642 and 0.637 means moderately regulated, and anything less than -0.64 is lightly regulated (Gyourko et al. 58). The range for WRULRI2018 values is from -0.54 to 1.18, so there are no WRULRI2018 values below -0.64 (Gyourko et al. 61). Therefore, any value below -0.4 implies relatively lax regulations.

The final factor is the profit generated to appease builders. The constant profit margin of 17% is assumed (Glaeser and Gyourko 7), so we do not consider it in our comparative observations.

Understanding the importance of these factors determines how affordable a housing unit can be. Therefore policies should effectively address them by lowering costs in any sustainable way possible to ensure maximum output of affordable housing units.

Comparing Cities

To understand which policies were effective in increasing affordable housing units, we had to choose cities that were comparable to cities like Plano in the aforementioned influential factors. We chose two cities besides Plano to act as a negative and positive control. A negative control city would be one struggling with affordable housing while a positive control city would be able to be relatively affordable. To determine what was considered affordable, we chose the definition from First American Financial Corp., a real-estate services company. The definition states that an affordable housing market allows an average buyer to afford 50% or more of the houses for sale in that market (Picchi). We looked at a variety of cities throughout the United States based on an analysis carried out by First American Financial Corp. and looked for cities that were above and below the national median of 34%. This meant that an average buyer could purchase 34% of the houses of sale in the United States as a whole. Therefore a negative control city needed to be below this 34% while a positive control city needed to be above the 34%.

Next, we needed to find cities that were relatively similar to Plano looking at factors such as population growth, median household income, and construction costs. We calculated the growth rate of Plano over the past 10 years and found the median household income, the construction cost per acre, and the WRULRI2018 index value for the Dallas-Fort Worth metroplex. Once all of these numbers were found, we compared them to the cities in the First American Financial Corp. study to find cities similar enough to be considered negative and positive controls. We decided on San Francisco, California and Charlotte, North Carolina with affordability percentages of 2% and 46% respectively (Picchi).

Factor	Plano	San Francisco	Charlotte
¹ Population Growth	9.0%	0.9%	14.62%
Median Household Income	\$134,666	\$136,692	\$74,401
WRULRI2018 Index Value	0.17	1.18	-0.38
Land Value/Acre	\$354,130	\$3.2 million	\$325,000
Construction Costs/ft ²	\$110-\$140	\$440	\$145

Fig 1: Comparison of Affordable Housing Influences by City

¹Over the past ten years.

²Value for Dallas-Fort Worth-Arlington: Plano is located in the Dallas-Fort Worth metroplex.

Something to note is that some cities (Buffalo, New York, Pittsburgh, Pennsylvania, and Cleveland, Ohio) were above the 50% median for affordability (Picchi). However, these cities had declining populations, the opposite of the defining circumstance of Plano's rapid growth. Therefore, we chose Charlotte, NC due to its proximity to the 50% mark and its population growth rates.

Excess Regulation in San Francisco

Once the negative and positive control cities were chosen, we needed to look at different policies that worked or didn't, to determine what could be done in cities like Plano. Referring back to Table 1, San Francisco has a high median income, some population growth, is highly regulated, and has a high land value which contributes to high construction costs. Policies would not target the high median income to not stymie economic growth, and population growth is not regulated. This means that policy will target construction costs or find other ways to increase supply. As previously stated, San Francisco has very strict zoning laws and a high WRULRI2018 value. The correlation between this high regulation and housing affordability is direct as the majority of policies harming San Francisco residents restrict builders by placing unnecessary restrictions on construction projects.

The top policy put in place damaging affordability is requiring waterfront properties above a certain height to have voter approval (Bellisario et.al 21). While zoning laws such as these are used to preserve the cultural landscape of the city, extreme height limits would delay the construction of 3,690 units or prevent them from being built at all. Over twenty years, a total of 6,144 units would be negatively impacted. A second detrimental policy is the fees developers need to pay. First is a Transit Sustainability Fee which requires developers to pay a fee to fund San Francisco's transportation systems. This has forced developers to increase prices for housing units, causing an additional 2,160 families to be pushed out of their homes (Bellisario et.al 22). The second fee is the Child Care Fee which pushed rent for 837 units higher (Bellisario et.al 24). These fees are a result of the potential impacts of transit systems (Spencer 9) and the increasing need to provide child care for future residents of these developments (Yee 1) Finally San Francisco has required approval for projects that would demolish or convert the existing property, and developers have to find other space for anyone relocated by development. The increased time costs of these barriers to construction will prevent almost 2,000 units from being constructed. In total, around 14,831 housing units or more would be delayed or not built in twenty years.

Clear Framework, Dedicated Partnerships in Charlotte

In stark contrast to San Francisco, Charlotte is one of the least regulated cities from the WRLURI2018 surveys. This has prevented the aforementioned consequences, and this along with the definitive framework the city has implemented has allowed Charlotte to be very affordable relative to other domestic cities. Named Housing Charlotte, the framework was

developed in 2018 and is the guiding hand for Charlotte’s housing policies. Housing Charlotte consists of three key pillars: expanding the supply of affordable housing, preserving the affordability of existing units, and supporting family self-sufficiency (The City of Charlotte, *Housing Charlotte a Framework for Building and Expanding Access to Opportunity through Housing Investments* 26-28). Under these decisive pillars are various programs and methods utilized or to be utilized to expand affordable housing in all aspects. The chart below shows the three pillars and the most influential or impactful program and/or method for each.

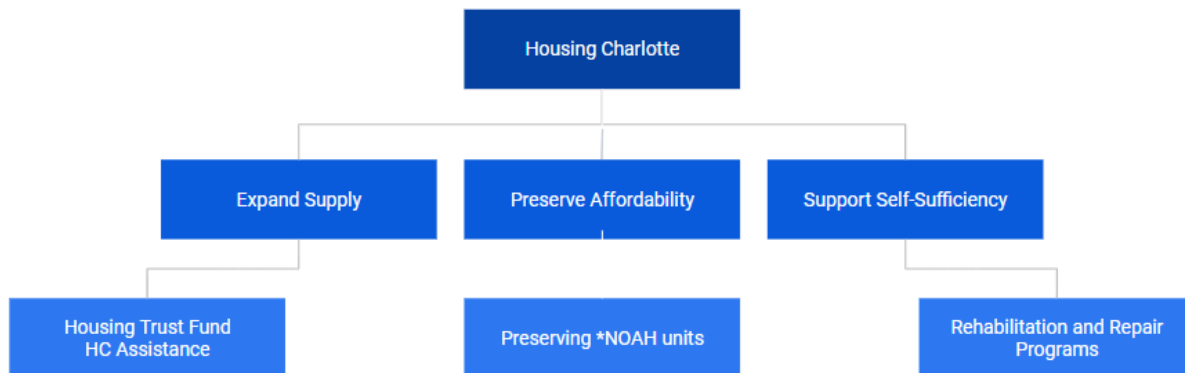


Fig 2: Housing Charlotte Framework

The first and most essential pillar is expanding the already existing supply of affordable housing units. This can be done in multitudes of ways, but the two most effective methods were creating a housing trust fund and creating mortgage aid programs. The purpose of the Housing Trust Fund is to provide gap financing for developers and offset some of the construction costs, allowing developers to expedite development and market units at a cheaper level for those who make below the AMI (area median income). This trust fund is funded primarily through the city government and voter-approved general obligation bonds. However, private sector contributions and partnerships with various faith-based communities, developers dedicated to affordable housing, investment firms, and more have proven to be essential as they have contributed around \$53 million to the fund. This has allowed \$1.6 billion to be utilized to construct affordable housing units (*Housing Trust Fund 20 Year Anniversary* 11) which resulted in 10,869 housing units added from 2002 to 2021 (*Housing Trust Fund 20 Year Anniversary* 4) and an additional 2,105 units in 2022 and 2023 (The City of Charlotte, “Workbook: Housing Dashboard”).

Adding to the Housing Trust Fund, there is the House Charlotte mortgage assistance program which gives opportunities for deferred loans and the option and forgivable loan options, either fully or in part, for qualified potential homebuyers (The City of Charlotte, “Affordable Housing”). Utilizing these programs, 6,365 units have been occupied since 2002 (The City of Charlotte, “Workbook: Housing Dashboard”).

First, home buyers must not have a total household income of more than 110% of the median income for Charlotte. Individuals or families with a total household income of 80% or

less of the AMI can qualify for a maximum of \$8,500 in assistance. Individuals or families with a total household income of 80% or less of the AMI and are public service employees, employed by public schools within the district, city employees, police officers, medical employees, and transit system employees, can qualify for a maximum of \$10,000 in assistance. Those who have a combined household income between 80% and 110% of the AMI can qualify for a maximum of \$5,000 in assistance in designated areas (Cornerstone Realty of the Carolinas) Assistance comes in deferred loans shown in the table below. Assistance comes in deferred loans shown in the table below.

At or Below 80% AMI	At or Below 80% AMI and Public Service Employee	Between 80% and 110% AMI
10-year forgivable loan *prorated by 20% per year starting Year 5	5-year forgivable loan *prorated by 33% per year starting Year 3	15-year forgivable loan *prorated by 20% per year starting Year 5.

Fig 3: Assistance for Various Income Levels

*Prorated meaning payments only due starting that year. For example, prorated starting Year 5 means mortgage payments are only due starting Year 5.

The next pillar is preserving the affordability of already existing units. Units can become less affordable over time due to expiring subsidies, tax rate increases, damage to units, and more. Therefore, it is essential to maintain the economic levels of rent low-income households pay. At-risk NOAH (Naturally Occurring Affordable Housing) units are monitored by a team formed in 2023 which will provide resources to repair units by partnering up with city government subsidies. The team also determines whether the individual or family living in units is qualified for tax relief and whether there is any potential abuse of deed restrictions. Thanks to this team, tax relief programs, and subsidies, 292 units have been preserved in a year (The City of Charlotte, “Workbook: Housing Dashboard”).

Finally, the last pillar is supporting family self-sustainability. The most effective method to help families is similar to preserving already affordable units: ensuring housing is maintained and safe for living and if not, rehoming these families. To do this, there are repair and rehabilitation programs designed specifically to do this. Repair programs are designed to mitigate maintenance costs for families and have been expanded to improve certain services such as electricity to lower utility expenses. Rehabilitation programs ensure that if a unit can’t be effectively repaired promptly, These programs have allowed 1,274 families to either stay in their homes or find new ones (The City of Charlotte, “Workbook: Housing Dashboard”).

In total, Charlotte's efforts have added 20,512 units since 2002. The pie chart attributes the contributions by various methods utilized by Charlotte, showing the successes of Charlotte's

efforts. These programs demonstrate the ability to address affordable housing effectively and sustainably.

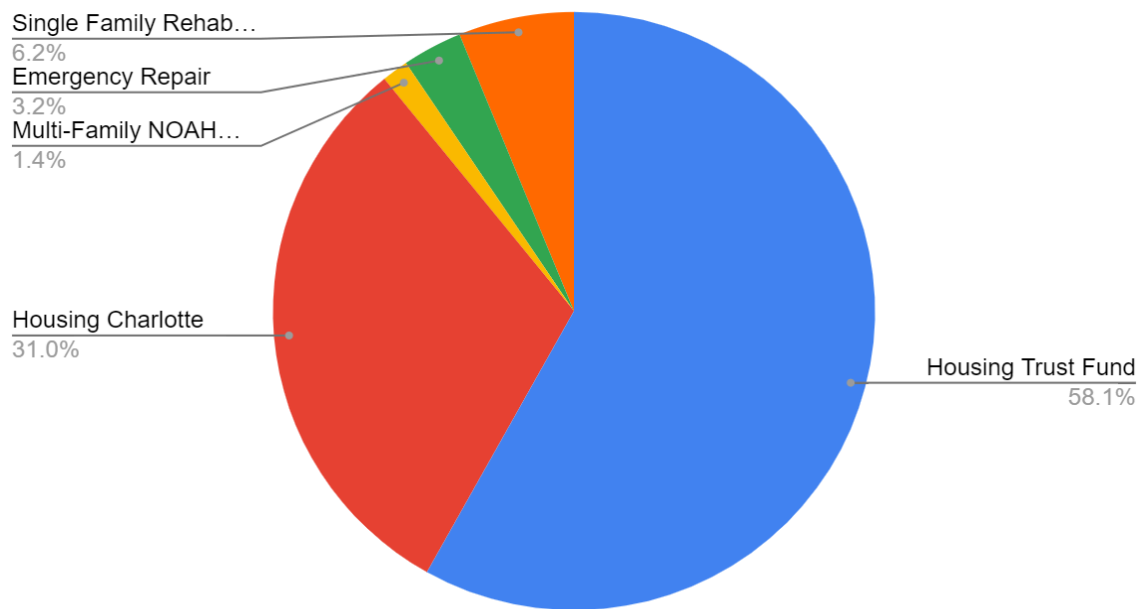


Fig 4: Contribution of Affordable Units by Program

Conclusion

As Plano and other cities continue to expand at an exponential rate, demand for affordable housing will continue to rise dramatically. Policymakers must maximize the presence of policies supportive of increasing and maintaining the housing supply. Thus, we advise that cities first create a framework such as the Housing Charlotte plan which allows Charlotte to address their residents' needs from multiple angles. The same three pillars are recommended because they ensure not only that units start low-cost but remain that way to provide long-lasting affordability for low-income individuals and families. Next, as shown by the donations and contributions by the Charlotte private sector, cities must work together with their developers, communities, and finance corporations to quicken the process of building new units. Not working with developers and going against their wishes led to transit sustainability and child care fees and the increase of building restrictions rather than a decrease. All of these are what severely injured San Francisco's housing situation.

Specific actions that would be beneficial include utilizing a housing trust fund and further helping developers with increasing the supply of housing units. This along with programs designed for every step of the housing process such as mortgage aid, preservation, and repair and rehabilitation programs would greatly benefit residents. Through these, cities can be one step closer to having adequate, low-cost housing for the populace.

Affordable housing is a complex issue that requires the consideration of multitudes of influential factors past the ones mentioned in this paper, and causes of the conditions of housing markets have other factors that affect them. For example, land value comes from geographical location, proximity to urban centers or farms, the owners of said land, and more. To fully grasp the housing shortage phenomenon, having a complete knowledge of the influences of affordable housing is essential. Additionally, the behavior of all buyers is not the same. For many, especially higher-income individuals and families, houses are considered assets. These entities would abhor the declining value of their investments and resist changes to their neighborhoods such as the addition of those in different economic backgrounds (Glaeser & Gyourko 20). Further study is required to decide on effective incentives to persuade current homeowners to be more lenient towards modifications in their area. Another thing to consider is the housing unit itself. While there are predeveloped standards and metrics designed to analyze the performance of housing units, there is little exploration of the effects of occupancy itself. A unit rated high in affordability and durability may prematurely break down once someone moves in (Marburger 6). Standards for housing units must be revised to include these considerations in all cities which can change the costs of production.

Overall, affordable housing is a challenge for any government to tackle because of the sheer amount of factors that come into consideration. On top of this, each city has its geography, regulations, land value, and more which requires specialized responses for every individual market. However, by learning from the example of others, positive and negative, the first steps towards solving this complex issue will be taken.

Acknowledgments

We want to thank Professor James Murdoch at the University of Texas at Dallas for his invaluable insight on this topic. We would also like to thank the Plano City Council for listening to our presentation on our study and the encouragement from all of the council members to pursue this topic.

Works Cited

- “About Plano, Texas.” *Plano Chamber of Commerce*,
www.planochamber.org/know-plano/about-plano/#:~:text=Plano%20is%20home%20to%20roughly. Accessed 8 Oct. 2023.
- Bellisario, Jeff, et al. *Solving the Housing Affordability Crisis*. Oct. 2016, pp. 21–22, 24,
www.bayareaconomy.org/files/pdf/BACEI_Housing_10_2016.pdf. Accessed 1 Oct. 2023.
- Bureau, US Census. “Texas Joins California as State with 30-Million-plus Population.”
Census.gov, 30 Mar. 2023,
www.census.gov/library/stories/2023/03/texas-population-passes-the-30-million-mark-in-2022.html#:~:text=Texas's%20population%20in%202022%20was,growing%2C%20largest%2Dgaining%20states.
- “Charlotte, North Carolina Population 2020 (Demographics, Maps, Graphs).”
Worldpopulationreview.com,
worldpopulationreview.com/us-cities/charlotte-nc-population. Accessed 22 Oct. 2023.
- Cornerstone Realty of the Carolinas. “House Charlotte Program.” *Charlotte NC Real Estate*,
nc-homeownership.com/buying-a-home/down-payment-assistance-programs/house-charlotte-program/. Accessed 25 Nov. 2023.
- “Cost to Build a House in Plano, Texas.” *Www.homeblue.com*,
www.homeblue.com/home-building/plano-tx-cost-to-build-a-house.htm. Accessed 15 Oct. 2023.
- Florida, Richard. “America’s Urban Land Is Worth a Staggering Amount.” *Bloomberg.com*, 2 Nov. 2017,
www.google.com/url?q=www.bloomberg.com/news/articles/2017-11-02/america-s-urban-land-is-worth-a-staggering-amount&sa=D&source=docs&ust=1702444621211893&usg=AOvVaw1ifpMYF5JBhdEcJWGcmZua. Accessed 19 Oct. 2023.
- Glaeser, Edward, and Joseph Gyourko. “The Economic Implications of Housing Supply.”
Journal of Economic Perspectives, vol. 32, no. 1, Feb. 2018, pp. 3–30,
<https://doi.org/10.1257/jep.32.1.3>. Accessed 28 Oct. 2023.
- Guzman, Gloria, and Melissa Kollar. “Income in the United States: 2022.” *Census.gov*, 12 Sept. 2023,
www.census.gov/library/publications/2023/demo/p60-279.html.
- Housing Trust Fund 20 Year Anniversary*. 2021,
www.charlottenc.gov/files/sharedassets/city/v/1/streets-and-neighborhoods/housing/documents/final-htf_20_year_anniversary-booklet-reformatted_1.pdf. Accessed 19 Nov. 2023.
- Marburger, Lindsey. *Metrics for High Performance Affordable Housing*. 2 Jan. 2010, p. 6,
programs.fas.org/energy/btech/policy/Metrics%20for%20High%20Performance%20Affordable%20Housing.pdf. Accessed 29 Oct. 2023.
- “Most Populated U.S. Cities: Median Household Income 2019.” *Statista*,
www.statista.com/statistics/205609/median-household-income-in-the-top-20-most-populated-cities-in-the-us/. Accessed 18 Oct. 2023.

- Picchi, Aimee. “The Typical Renter Can Only Afford to Buy a Home in Four Cities. Here’s Where They Are. - CBS News.” *Www.cbsnews.com*, 2 June 2023, www.cbsnews.com/news/real-estate-four-cities-where-renters-can-afford-homes/. Accessed 18 Oct. 2023.
- “Plano, Texas Population 2020 (Demographics, Maps, Graphs).” *Worldpopulationreview.com*, worldpopulationreview.com/us-cities/plano-tx-population. Accessed 15 Oct. 2023.
- “Plano, TX Land for Sale - 275 Properties.” *LandSearch*, www.landsearch.com/properties/plano-tx. Accessed 15 Oct. 2023.
- “San Francisco Metro Area Population 1950-2020.” *Www.macrotrends.net*, www.macrotrends.net/cities/23130/san-francisco/population. Accessed 18 Oct. 2024.
- Spencer, Robert. *San Francisco Transportation Sustainability Fee (TSF) Nexus Study Final Report*. May 2015, p. 9. Accessed 7 Nov. 2023.
- The City of Charlotte. “Affordable Housing.” *www.charlottenc.gov*, www.charlottenc.gov/Streets-and-Neighborhoods/Housing/Affordable-Housing. Accessed 21 Nov. 2023.
- The City of Charlotte. *Housing Charlotte a Framework for Building and Expanding Access to Opportunity through Housing Investments*. 2018, pp. 26–28, www.charlottenc.gov/files/sharedassets/city/v/1/streets-and-neighborhoods/housing/documents/housing-charlotte-framework.pdf. Accessed 10 Nov. 2023.
- The City of Charlotte. “Workbook: Housing Dashboard.” *Charlottenc.gov*, 2023, tableau.charlottenc.gov/t/Public/views/HousingDashboard/CityofCharlotteHousingServicesDashboardTable?%3Aembed=y&%3AisGuestRedirectFromVizportal=y. Accessed 25 Nov. 2023.
- Torres, Luis, et al. “High-Tech Hotspot Austin Works to Solve Labor, Housing Issues.” *Www.dallasfed.org*, 9 Mar. 2023, www.dallasfed.org/research/economics/2023/0309#:~:text=Today%2C%20thanks%20to%20university%2Dled. Accessed 14 Oct. 2023.
- Wright, Sarah. “How San Francisco Makes It Insanely Hard to Build Housing.” *The San Francisco Standard*, 13 Sept. 2022, sfstandard.com/2022/09/13/how-san-francisco-makes-it-insanely-hard-to-build-housing/. Accessed 18 Nov. 2023.
- Yee, Supervisor. *Child Care in Lieu Fee Increase and Application*. 2016, sfplanning.org/sites/default/files/documents/legis/code-summaries/150793.pdf. Accessed 15 Nov. 2023.
- Zhang, Ping, and Yilin Hou. “The Dynamics of Housing Price and Land Price in Urban versus Rural Contexts.” *Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association*, vol. 108, 2015, pp. 20–23, www.jstor.org/stable/90023175?seq=1. Accessed 16 Oct. 2023.

Biomass Conversion Methods for Green Energy Production in Rural Regions: A Review

By Joseph Tso, Salma Baig, Jack Schimler, and Jordan Penland

Abstract

Biofuels represent a sustainable substitute to fossil fuels, particularly in agricultural energy consumption, gaining importance as technology matures and scalability becomes achievable. This review investigates the research question: "Can biofuels be made more accessible through improvements in scalability, cost reductions, and minimizing labor or machinery requirements?" We analyzed the current systems for agricultural biomass conversion into bioenergy, comparing their benefits and limitations for agricultural producers. This analysis drew from a diverse range of databases, including Google Scholar, Scopus, Web of Science, and ResearchGate, using carefully selected keywords like "biofuel," "pretreatment method," "clean energy," and "lignocellulosic material." Out of 120 studies compiled, 20 were scrutinized for their relevance, highlighting key aspects of biofuel production processes. Findings suggest that biofuels can supplement or replace traditional agricultural energy sources. Investment in thermochemical biomass conversion and crop rotation methods integrating energy crops are viable options for sustainable energy transition, potentially enhancing soil health and farm productivity. While the accessibility of biofuels could be improved through measures targeting scalability enhancement, cost reduction, and minimizing labor or machinery, no significant economic or environmental impact difference was identified between different biomass-to-fuel conversion processes. This indicates that the choice between these technologies could be based on specific context or geographical characteristics, favoring a more localized approach to biofuel adoption. Future studies should explore these influencing variables to identify the most appropriate biofuel production system

Introduction

As the global population expands and more nations industrialize, the need for affordable, easily accessible energy intensifies. These energy demands of the modern world have historically been met by fossil fuels such as coal, oil and gas, but these resources are not only nonrenewable, but they are also the main source of greenhouse gas (GHG) emissions. These GHG emissions are projected to increase by 50% and become the fastest growing driver of climate change by 2050 (Di Lucia and Nilsson). The International Energy Agency suggests a key solution to reduce GHG emissions is renewable energy sources, which includes hydro-power, geothermal, nuclear, solar, wind, and in particular, biofuels.

Biofuels are made from biomass, an organic material derived from the byproducts of living organisms, which can be anything from soybeans to human wastewater (Vassilev et al.). Specifically, in the context of agricultural producing regions, biomass consists primarily of crop residues, waste materials, and dedicated energy crops such as corn and sugarcane. Biofuels work by storing the energy from the biomass in a more accessible form, like biogas or biocoal. However, unlike their fossil analogs, these biofuels can be produced from readily available

biomass in a relatively short time frame, meaning they are more sustainable. Furthermore, because of the similarities to existing fossil fuels, there is no need for the replacement of any infrastructure to deploy these biofuels,

Although there are benefits, biofuel production is not without challenges, as conversion of biomass, especially lignocellulosic materials such as wood, straw, and other agricultural residues, into usable fuel is a difficult and energy intensive procedure (Kumari and Singh). To break down the cell wall of the plant matter, refineries will use one of many pretreatment methods. The selection of an appropriate pretreatment method is guided by several factors: avoid reducing the size of the biomass particles, preserve the hemicellulose fractions, minimize the formation of degradation products, and employ a low-cost catalyst (Moodley). Pretreatment techniques are broadly classified into thermochemical and biochemical technologies, each with pros and cons, and their suitability is highly dependent on the specific biomass and the desired product (Mahmood et al.).

In the existing literature, several reviews have been conducted comparing these pretreatment methods, but none have considered the full environmental and ecological impact of these systems or the specific context of rural and agricultural regions (Alalwan et al.; Akia et al.; Mardiana et al.; Ho et al.). The objectives of this review are thus threefold: (1) to conduct an interdisciplinary review that will compare various thermochemical pretreatment methods with biochemical pretreatment methods, (2) to discuss the opportunities and challenges these systems create; and (3) to highlight the opportunities for efficient biomass conversions in rural agricultural production regions. The remainder of the paper will be structured as follows. Section 2 will outline the search methods used to obtain suitable articles for review, Section 3 will highlight gaps in the current literature and discuss future opportunities for biomass conversion, and Section 4 will discuss the research findings and provide a conclusive statement regarding the future of biofuels in rural and agricultural regions.

Methods

Biofuel production and pretreatment is a topic that is rapidly transforming, especially as new computational technologies change the industry. There is a considerable amount of research on the topic of biofuels, especially with more established pretreatment methods, but there is limited research on more recently developed methods. However, we expect the number of studies around these methods, especially in different languages, to increase considerably in the future as the popularity of these technologies grows across the globe. For the present study, reviewed articles were published in English, and used to gain a more widespread understanding of biofuel production, despite the limited scope of availability. This investigation was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines (Sarkis-Onofre et al.).

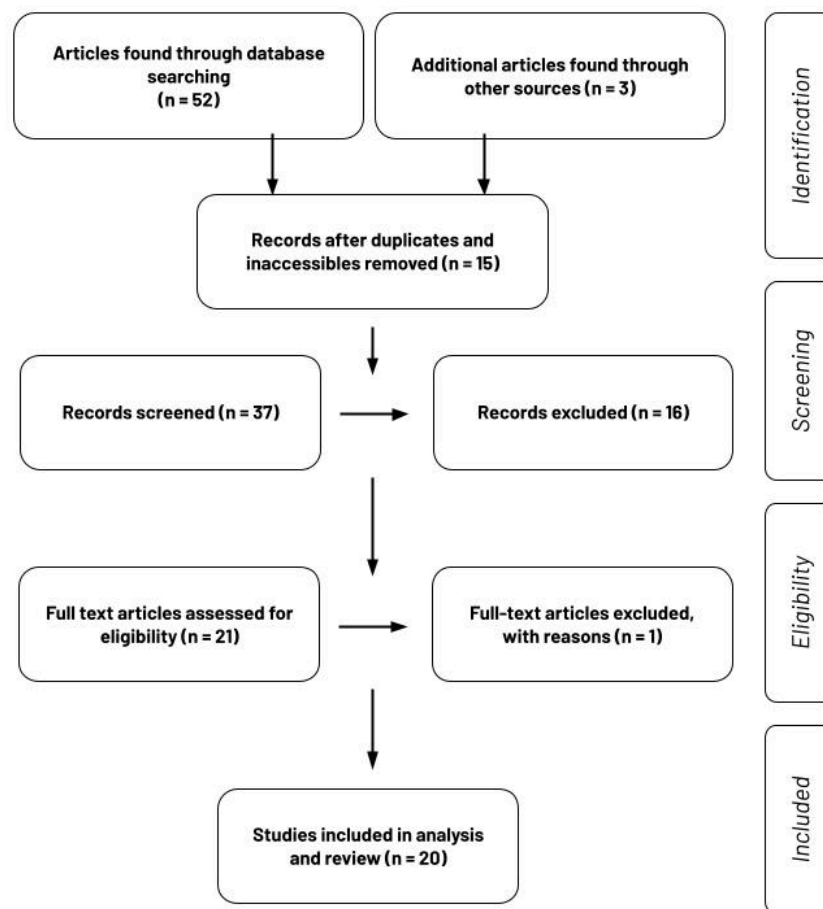


Figure 1: PRISMA Strategy (Self Created Image)

In this study, we conducted a comprehensive online search of the literature using various search engines, such as Google Scholar, Scopus, Web of Science, and ResearchGate. Figure 1 shows the PRISMA process (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), starting with the identification of the literature, the review exclusion processes, and the synthesis methods. Studies were collected based on their titles and abstracts in the identification phase to collect all relevant studies.

A total of 120 studies were compiled using a combination of keywords such as 'biofuel', 'pretreatment method', 'clean energy', and 'lignocellulose' in their titles and abstracts. Screening of topics and content excluded 85 of these articles based on lack of relevance. Finally, in the eligibility stage, 20 of these articles were selected and analyzed according to the relevance of their content to the main purpose of the study. The 20 articles chosen in this study were reviewed in different sections based on their topic and scope of analysis. Due to rapid developments in this field, more modern studies were chosen, although several older studies were also reviewed to provide historical context to better understand changes in the industry over the past decades. Of the 20 articles, 9 examined biochemical pretreatment methods and 10 examined thermochemical reactions. One study investigated a number of different pretreatment methods, including both

biochemical and thermochemical, but failed to assess the methods' overall impact and did not compare the methods.

Findings

Biochemical Conversions

Biochemical methods are pivotal techniques used to transform organic material into energy. These processes principally revolve around the utilization of microorganisms and enzymes to break down biomass into its constituent components, subsequently converted into energy-rich substances such as ethanol, biogas, or biohydrogen (Jenicek et al.). These biofuels serve as sustainable alternatives to fossil fuels, providing cleaner energy sources for various applications, including transportation, domestic heating, and industrial use.

Among biochemical methods, anaerobic digestion and fermentation are the most prominent and well-developed strategies. Anaerobic digestion employs bacteria in an oxygen-free environment to convert biomass into biogas, a methane-rich substance used as a renewable energy source (Mata-Alvarez et al.). On the other hand, fermentation, an age-old technique, utilizes yeasts or bacteria to convert sugars into ethanol, a biofuel used in various sectors such as transportation (Hill).

Fermentation

Fermentation is a complex biochemical process that involves the conversion of carbohydrates, such as sugars, into alcohol or acids by microorganisms under anaerobic conditions. The most common product of this process is ethanol, which is synthesized from natural proteins or sugars, such as acetyl-CoA, through a series of enzymatic reactions (Todaro and Vogel).

Process

The production of ethanol encompasses three major stages: fermentation, separation, and purification. During the fermentation stage, the microorganisms metabolize the sugars or proteins, producing ethanol and other byproducts, forming a broth. The separation stage involves the extraction of ethanol from the fermentation broth, usually through a process known as distillation. The final stage, purification, employs a method referred to as "distillation-rectification-dehydration," which ensures the ethanol produced is of the highest purity. This method also allows for the reuse of fermentation waste, thereby promoting a circular process in biofuel production (Ahorsu et al.).

There are several methods of fermentation, each with its unique operational parameters and advantages. Batch fermentation, the simplest of these methods, involves the addition of a hydrolysate, a liquid mixture containing nutrients, sugar, yeasts, and other ingredients, to a fixed medium within a reactor. The fermentation process proceeds without any interference until completion, after which the biofuel is harvested (Johnson). This process must remain undisturbed as it is happening due to the many reactions and processes inside the reactor.

Fed-batch fermentation, on the other hand, requires the intermittent addition of vital ingredients for conversion at different points of the process (Lee et al.). This method results in a surplus of bioenergy created, as well as a higher yield and reduced byproduct residue. Continuous fermentation employs a steady injection of fresh ingredients as the products are removed from the fermentation bioreactors. This method allows multiple batches of biofuel to be produced from a single reactor over an extended period (Li et al.). Dark fermentation, a more recent development, leverages the absence of light and waste biomass to produce biohydrogen, a carbon-free biofuel, using microorganisms (Sarangi and Nanda; Guo et al.).

Advantages/Disadvantages

The fermentation process offers several advantages. Primarily, it is a natural process caused by microbes, thus requiring minimal energy input. This factor allows it to be considered a sustainable and environmentally friendly method of biofuel production (Hölker and Lenz). Furthermore, fermentation can utilize food scraps and agricultural waste, thus reducing waste and promoting resource efficiency (Chelule et al.). Also, the process is time-efficient and can generate large quantities of biofuel, particularly when continuous and fed-batch processes are employed (Chelule et al.).

Despite these advantages, the fermentation process also presents some challenges. One of the primary limitations is that it operates optimally only at lower temperatures and reaction rates compared to thermochemical systems (Hoffman et al.). This constraint can limit the speed and efficiency of biofuel production (Chelule et al.). Another significant challenge is the risk of producing undesirable by-products, such as 5-hydroxymethylfurfural (HMF) and furfural, during the depolymerization of cellulose and hemicellulose (Ahorsu et al.). These compounds can inhibit microbial growth and metabolism, thus reducing the yield of ethanol (Ahorsu et al.).

Current Research

The field of biofuel production through fermentation has seen significant advances in recent years. In the paper by Chandrasekhar et al., the authors discuss the concept of electro-fermentation, an emerging bioprocess that regulates the metabolism of electrochemically active microorganisms with electrodes. They found that this process can increase product yields and reduce process failures. The authors also highlight the potential of electro-fermentation in the production of value-added chemicals and fuels, such as hydrogen, methane, biopolymers, and ethanol.

Henstra et al. explore the use of lignocellulosic waste for the production of biofuels through the fermentation process. They found that this waste has potential as a renewable resource for the production of biofuels. The authors also discuss the challenges and potential solutions associated with the use of lignocellulosic waste for the production of biofuels. Yaashikaa et al. discusses the use of lignocellulosic ethanol as a biofuel and the role of microbial and enzymatic processes in its production. They found that genetically modifying agricultural

biomass prior to the fermentation process can improve production by adding DNA. Similarly, the addition of external organisms, such as yeast, can increase the efficiency of the resulting biofuel.

Finally, Munasinghe and Khanal present a comprehensive review of the principles and technologies related to the production of biofuels from renewable resources. They discuss the role of biotechnological processes, including fermentation, in the production of these biofuels. They also highlight the challenges and potential solutions associated with the production of biofuels from renewable resources. In conclusion, current research in the field of fermentation for biofuel production is focused on improving the efficiency of the process, exploring new methods such as electro fermentation, and utilizing renewable resources for biofuel production.

Anaerobic Digestion

Anaerobic digestion refers to the biological degradation of organic materials, such as animal waste, wastewater biosolids, and food scraps, in an oxygen-free environment, resulting in biogas and digestate production (Laiq Ur Rehman et al.). Different types of anaerobic digesters treat different types of organic materials. The operation is carried out in secure containers known as reactors, which come in various designs to suit different site and feedstock conditions. These reactors house complex microbial communities that process the feedstock, produce biogas, and digestate.

Process

Hydrolysis is the first stage in anaerobic digestion, where specific bacterial species decompose organic polymers into simpler sugars (Menzel et al.). This is followed by acidogenesis, where these sugars and amino acids are converted into carbon dioxide, hydrogen, ammonia, and organic acids (Hill). Acetogenesis transforms these organic acids into acetic, carbon dioxide, and hydrogen, and finally, methanogenesis, which forms biogas from the intermediate products of the previous stages (Bajpai and Bajpai). The leftover material, including inedible material and dead bacteria, forms the digestate. (Laiq Ur Rehman et al.).

The types of digesters used in anaerobic digestion vary in their application. Farms use animal waste and food waste in their digesters, thus managing animal and food waste. (Güngör and Karthikeyan). Anaerobic digestion of animal and food waste reduces odor, helps manage nutrients, and generates income for farms and local communities. Stand-alone digesters cater to specific industries and have gained popularity due to the increased redirection of food waste from landfills (Guwy et al.). Digesters in wastewater and sewage treatment plants primarily treat wastewater solids. Digesters can be differentiated based on size, processing rate, operating temperature, and mixing strategies.

The anaerobic digestion process yields two beneficial outputs: biogas and digestate (Shi et al.). Biogas, consisting predominantly of methane and carbon dioxide, can be utilized for heat production, electricity generation, and cooling systems (Chen and Neibling). Digestate, the leftover material post-digestion, is a blend of solid and liquid matter that can be separated into solid and liquid segments. The solid fraction is helpful for animal bedding, soil amendment, and

as nutrient-rich organic compost and fertilizer. The liquid portion is also a valuable fertilizer and a possible feedstock for algae cultivation. Digestate products are not only cost-effective but also can be a source of revenue, enhancing the economic viability of anaerobic digestion (Abdo and Ackrill).

Advantages and Disadvantages

Anaerobic digestion offers several advantages: land conservation, soil improvement, water resource generation, and sustainable food production (Jenicek et al.). It also serves as a source of renewable energy, with feedstocks - material used in the methods - for digestion derived from biomass, a renewable resource (Jenicek et al.). However, the method also has its drawbacks. The initial and maintenance costs of the equipment are high, limiting its accessibility to smaller farms. (Mata-Alvarez et al.). The digestion process is time-consuming and demands regular maintenance to ensure efficient operation. Lastly, water used in the process requires subsequent treatment for reuse, a step that may not always be sufficiently carried out (Mata-Alvarez et al.).

New Research and Developments

New research is exploring the reusability of digestate in rural communities and farming operations, primarily as fertilizers (DeBruyn and Hilborn). Besides, digestate can be utilized for soil amendment, changing the soil's properties to enhance its productivity (Sun et al.). The material from digestate can also be repurposed to create building materials for farm use (Picuno). Furthermore, soft digestate materials are used to make farm animals bedding, contributing to their welfare. Finally, treated wastewater from the digestion process is recycled for irrigation (DeBruyn and Hilborn). Overall, the versatile use of biogas and digestate products from anaerobic digestion is paving the way for enhanced farm sustainability and efficiency. Such advancements provide a compelling case for the scalability of this method in the agricultural sector.

Thermochemical Conversions

One of the main ways in which biomass has been introduced into the energy sector has been the use of thermochemical conversion processes. Basically, the process is the heating of biomass in oxygen-deficient machinery to produce material ready to be turned into fuels (Ram and Mondal). These fuels are then used in transportation, heating, industrial and electrical applications in a more sustainable way than overusing nonrenewable resources.(Jeon et al.) Thermochemical methods are the most popular and mature method of converting biomass into energy. Thermochemical methods leverage more established processes, like combustion, gasification, and the lab-scale pyrolysis, thermochemical conversion is a system that offers feasible options for incorporating biofuels as a larger part of the energy market space (Ong et al.). Feedstocks can be natural scraps, commercial waste, or most commonly wood pellets. Wood

pellets are made up of common and effective lignocellulosic materials in many sustainable applications (R. Kumar et al.).

Gasification

Gasification, a prevalent method of thermochemical conversion, shares similarities with its predecessor, combustion, and finds application in electrical components such as fuel cells and generators (Molino et al.).

Process

Gasification follows a high-temperature pretreatment phase, which prepares the feedstock for conversion (P. Kumar et al.). This process occurs in a machine under steam influence, where several chemical reactions transpire concurrently, converting the feedstock into vapor. These reactions produce various compounds, including carbon monoxide and hydrogen, called syngas. (Wilhelm). In the gasification process, other byproducts are also generated, including slag or 'ash' resulting from feedstock heating. This waste has potential uses in infrastructure development, such as road construction (Abdollahi et al.). An additional byproduct of the process is carbon dioxide, formed when a steam and catalyst mixture interacts with syngas, a blend of CO and H₂ (Wilhelm). These syngas components and other products further undergo processing to generate biochemicals and energy, contributing to electricity production and other human necessities (Wilhelm). The process's complexity yields straightforward and consistent results analogous to other tested or applied conversion methods (Molino et al.). While gasification has potential for upscaling and wider fuel production, it still has associated challenges, including the production of potentially toxic outputs and system maintenance requirements.

Advantages and Disadvantages

Despite its potential, gasification has certain drawbacks. For instance, ash and soot, the slag products from feedstock combustion in an oxygen-deficient chamber, can lead to machine blockages, necessitating significant maintenance (Lu and Do). Tar build-up within the machinery may reduce product quality and efficiency, thereby raising operational costs and increasing labor time (Kushwah et al.). Consequently, these limitations decrease the competitiveness of gasification and hinder its adoption. Ongoing research aims to address these problems by enhancing machinery and developing innovative methods for overcoming these limitations (Kushwah et al.).

Recent Research

Research in the gasification sector has seen a great deal of progress in recent years, with a focus on improving efficiency, reducing environmental impact, and maximizing the usefulness of byproducts. In a study by Huber et al., the authors analyzed the steam gasification process of biomass in a dual circulating fluidized bed reactor. The researchers aimed to better understand the process kinetics, particularly the influence of temperature and steam to biomass ratio on

gasification performance. The results showed that temperature had a significant impact on carbon conversion efficiency and gas yield. Another recent study by Cloete et al. demonstrated an innovative method of integrating gasification with fuel cells. The authors analyzed a combined gasification and fuel cell system, which enhanced the overall process's efficiency by converting waste heat from the fuel cell into usable steam for the gasification process. They argued that such a system could effectively boost the efficiency of biomass energy utilization.

Pyrolysis

Process

Pyrolysis is a thermochemical conversion process that decomposes biomass into a liquid fuel, which holds significant potential for use in the transportation sector (Guda et al.). This transformation is facilitated by high temperatures and the absence of oxygen. The principle behind this process is rooted in the thermal decomposition of the biomass feedstock, resulting in a variety of end products, with bio-oil being the most prominent. In contrast to E10 gas, a fuel blend comprising 90% gasoline and 10% ethanol, the fuels derived from pyrolysis provide a more sustainable alternative as such fuels are composed entirely from biomass. This is due to the fact that the production of ethanol depends heavily on corn, a food crop, raising concerns about food security. Importantly, pyrolysis fuels offer a carbon-neutral scenario, as the carbon dioxide released during their combustion is equivalent to the carbon dioxide absorbed by the biomass during its growth phase.

Advantages and Disadvantages

Pyrolysis fuels are lauded for their potential role in mitigating carbon emissions, thus offering an environmentally friendly alternative to fossil fuels, particularly in the transportation industry. They are also recognized for their role in waste management, specifically in addressing the plastic waste problem. The pyrolysis process can convert waste plastics into valuable fuel by altering their chemical structure (Hamid et al.). However, the challenges associated with this process are considerable. On one hand, it requires vast tracts of land for large-scale production, leading to potential conflict with agricultural land use in rural areas. On the other hand, the process involves complex machinery that requires frequent maintenance and updates, contributing to increased operational expenses (Guda et al.). Moreover, the traditional pyrolysis process often yields bio-oil of relatively poor quality.

Innovative methods like catalytic pyrolysis have emerged to address these challenges. By utilizing catalysts, this process speeds up the pyrolysis reaction, leading to improved bio-oil quality while reducing the energy requirement and processing time (Miandad et al.).

Recent Research

Several studies have delved into various aspects of the pyrolysis process to maximize its efficiency, understand the resulting byproducts, and ascertain its environmental impact.

The study by Banks and A. Bridgwater is a comprehensive review of the process, exploring different biomass feedstocks and their impact on the output products. The paper emphasizes the importance of feedstock selection in determining the product's quality and the process's efficiency. The authors also discuss technological advancements in reactor design and operation, and the challenges faced in commercializing the process.

In another study by Resende, the authors compare different catalysts in the pyrolysis process. They examined the effect of various catalysts on the product's yield and composition, highlighting that zeolite-based catalysts are particularly effective in producing bio-oil with high calorific value and minimal oxygen content. The study underscores the importance of the catalytic upgrading process in improving the quality of pyrolysis products.

Mohan et al.'s research delves into the chemical properties of pyrolysis products, with a particular focus on bio-oil. His study offers a molecular characterization of the bio-oil, providing a deeper understanding of its composition and potential applications. This in-depth analysis is crucial for developing more efficient and sustainable uses for the resulting products of pyrolysis.

The environmental impact of pyrolysis is the main concern in the paper by Im-orb and Arpornwichanop. They conducted a life-cycle assessment of the pyrolysis process and found that it is an environmentally friendly method for managing biomass waste. The study emphasizes that the environmental impact can be significantly reduced by proper management of byproducts and efficient use of the energy produced in the process.

Analysis/Discussion

Comparative Analysis of Biochemical/Thermochemical Processes As Applied To Agriculture

To determine the best or most effective solution, one must understand the limitations and benefits of the different two systems. We know that biomass conversion is currently providing about 2-3% of the total energy used in the United States (King et al.). To highlight the differences in how biochemical and thermochemical processes can apply to different regions, this section analyzes how rural areas, in particular, can utilize energy derived from biomass. Agricultural areas present an important market for increasing the use of biomass in energy production.

Biofuel production techniques can revolutionize agricultural settings in multiple ways. Promising methods such as anaerobic digestion, algae biofuel production, and crop rotation can help utilize farm waste and crops more efficiently. Anaerobic digestion can turn agricultural waste into biogas, a form of biofuel, reducing waste and producing a sustainable source of energy. Similarly, cultivating energy crops like switchgrass, jatropha, or even algae in unused or less productive parts of farmland can provide an additional source of biofuel. Moreover, adopting crop rotation methods that incorporate energy crops could enhance soil health and increase farm productivity alongside biofuel production. The potential impact of these methods on energy sustainability in rural areas can be profound. These techniques can lead to self-sufficiency, reducing reliance on conventional energy sources, and reducing energy costs. They can also help

combat climate change by reducing greenhouse gas emissions and promoting a more circular economy, where waste is minimized and resources are used to their maximum potential.

The literature has analyzed the various factors for determining which system may be best for agricultural regions. Generally, thermochemical energy conversions are more mature than that of the biochemical processes. Out of the biochemical processes, fermentation is the main approach to converting biofuels. Biochemical processes have potential to scale up in agricultural regions. Importantly, anaerobic digestion is already in use on farms and utilized too. For example, digestate is incorporated into agricultural operations and incorporating biochemical processes is a healthy and economically sustainable way to run dairy farms (Chen and Neibling). Although some methods of thermochemical conversion may have higher efficiency and wider applications, the fuels work best in transportation and advanced electrical applications (Bridgwater et al.). While researchers continue to improve thermochemical machines to reduce the limitations of thermochemical products, the land that these machines require takes too much from fertile growing land in agricultural regions. A review of literature identifies biochemical processes as the most feasible to scale up in agricultural regions, because they 1) provide both energy and material usable in agricultural processes, 2) are already in use on farms, and 3) have reached technological maturity.

Biofuels offer several economic and environmental advantages over traditional fossil fuels like coal and natural gas. From an economic perspective, biofuels, while having higher upfront costs, can provide a stable, local source of energy that insulates farmers from fluctuations in energy prices over the long-term. Environmentally, biofuels generally emit fewer greenhouse gasses and other pollutants compared to fossil fuels. However, not all biofuels are created equal - the sustainability of biofuels can vary greatly depending on the feedstock and production method used. Biofuels could serve as a viable supplement or even replacement for fossil fuels in agricultural energy usage, especially as technologies mature and economies of scale are achieved.

Limitations and Future Research Opportunities

Despite the promising prospects, biofuel production techniques have their limitations. For instance, the energy return on energy invested (EROEI) can be low for some biofuels, making them less efficient than traditional energy sources. Biofuel production could also compete with food production for land and water resources, leading to potential food security issues. Current challenges in the biofuel industry include technical issues related to production efficiency, economic viability, and the sustainability of feedstocks. Future research can focus on these areas, seeking to develop more efficient production methods, exploring novel feedstocks that do not compete with food production, such as agricultural waste or algae, and developing holistic sustainability assessments to ensure that biofuel production is truly beneficial for the environment. There are also opportunities for research into policy and market mechanisms that can support the growth and sustainability of the biofuel industry

Conclusion

In sum, the literature has identified many factors that differentiate thermochemical and biochemical conversions. Biofuels provide an important tool to combat pollution caused by energy production by mitigating harmful toxins in the air. Biofuels would support the United Nations' Sustainable Development Goals, especially in helping environmental sustainability of rural areas. Despite their potential, this review of the literature identified a need for further research. For example, newer methods most commonly found in thermochemical conversion have a lot of potential, but their use is limited by the expenses required to provide feedstock. Currently, thermochemical processes also produce low-quality output. Further development and improvement in machines are needed to boost the efficiency of the processes and the quality of end products. For biochemical conversion, products may be incorporated into farm management and improve the conversion's sustainability and cost-effectiveness.

Nevertheless, these processes still have significant limitations: 1) not enough invested capital into machinery and inputs, such as feedstocks; 2) scalability is still low and newer technologies are still in development; and 3) low quality products with a lack of overall reliability. To overcome these challenges, land and space are needed to provide research and adequate material to scale them up. In addition, both systems require improved efficiencies. This ensures that these conversion processes can meet consumer demands as compared against the use of nonrenewable, mainstream fuels. One potential avenue for improvement is the use of enzymes to speed up the process.

Overall, bioenergy has the potential to supply a sustainable energy solution for many regions. The literature has shown its potential, for example, in agricultural areas. Its current usage is small compared to other nonrenewable sources. With time and research invested in promising biochemical and thermochemical methods, these methods have the potential to scale up as a viable alternative to fossil fuels. Biofuels can play a larger role in the fight against climate change and pollution. For instance, biofuels created through pyrolysis may provide a quality alternative to fossil fuels in the transportation industry specifically. Carbon neutrality is the primary argument to scale bioenergy, but considerations should look beyond just carbon neutrality. For example, biomass conversion may provide a sustainable process to reduce waste in both agricultural and industrial processes. Biomass conversion's overall effects on industries, such as the agricultural industry, should be considered in addition to its net effect on carbon neutrality. This review of literature has shown the potential for these biomass conversion technologies to supply such benefits to different industries by highlighting the effect on agricultural processes.

Works Cited

- Abdo, H., and Ackrill, R. (2021). On-farm anaerobic digestion: A disaggregated analysis of the policy challenges for greater uptake. *Energy Policy*, 153, 112258.
- Abdollahi, M., Yu, J., Hwang, H., Liu, P., Ciora, R., Sahimi, M., and Tsotsis, T. (2010). Process intensification in hydrogen production from biomass-derived syngas. *Industrial and engineering chemistry research*, 49(21), 10986–10993.
- Ahorsu, R., Medina, F., and Constanti, M. (2018). Significance and challenges of biomass as a suitable feedstock for bioenergy and biochemical production: A review. *Energies*, 11(12), 3366.
- Ahorsu, R., Medina, F., and Constantí, M. (2018). Significance and Challenges of Biomass as a Suitable Feedstock for Bioenergy and Biochemical Production: A Review. *Energies*, 11(12).
- Akia, M., Yazdani, F., Motaee, E., Han, D., and Arandiyan, H. (2014). A review on conversion of biomass to biofuel by nanocatalysts. *Biofuel Research Journal*, 1(1), 16–25.
- Alalwan, H., Alminshid, A., and Aljaafari, H. (2019). Promising evolution of biofuel generations. Subject review. *Renewable Energy Focus*, 28, 127–139.
- Bajpai, P., and Bajpai, P. (2017). Basics of anaerobic digestion process. *Anaerobic technology in pulp and paper industry*, 7–12.
- Banks, S., and Bridgwater, A. (2016). Catalytic fast pyrolysis for improved liquid quality. *Handbook of biofuels production*, 391–429.
- Bridgwater, A., Meier, D., and Radlein, D. (1999). An overview of fast pyrolysis of biomass. *Organic geochemistry*, 30(12), 1479–1493.
- Bundhoo, Z. (2017). Coupling dark fermentation with biochemical or bioelectrochemical systems for enhanced bio-energy production: A review. *International Journal of Hydrogen Energy*, 42(43), 26667–26686.
- Chandrasekhar, K., Kumar, A., Kumar, G., Kim, D.H., Song, Y.C., and Kim, S.H. (2021). Electro-fermentation for biofuels and biochemicals production: Current status and future directions. *Bioresource technology*, 323, 124598.
- Chelule, P., Mokoena, M., and Gqaleni, N. (2010). Advantages of traditional lactic acid bacteria fermentation of food in Africa. *Current research, technology and education topics in applied microbiology and microbial biotechnology*, 2, 1160–1167.
- Chen, L., and Neibling, H. (2014). *Anaerobic digestion basics*. University of Idaho extension, 6.
- Cloete, S., Giuffrida, A., Romano, M., Chiesa, P., Pishahang, M., and Larring, Y. (2018). Integration of chemical looping oxygen production and chemical looping combustion in integrated gasification combined cycles. *Fuel*, 220, 725–743.
- DeBruyn, J., and Hilborn, D. (2004). *Anaerobic digestion basics*. Ministry of Agriculture and Food.
- Di Lucia, L., and Nilsson, L. (2007). Transport biofuels in the European Union: The state of play. *Transport Policy*, 14(6), 533–543.
- Frolov, S. (2021). Organic waste gasification: A selective review. *Fuels*, 2(4), 556–650.

- Gaudreau, K. (2009). Biofuel basics. *Alternatives Journal*, 35(2), 15.
- Guda, V., Steele, P., Penmetsa, V., and Li, Q. (2015). Fast pyrolysis of biomass: Recent advances in fast pyrolysis technology. *Recent advances in thermo-chemical conversion of biomass*, 177–211.
- Güngör, K., and Karthikeyan, K. (2008). Phosphorus forms and extractability in dairy manure: A case study for Wisconsin on-farm anaerobic digesters. *Bioresource Technology*, 99(2), 425–436.
- Guo, X., Trably, E., Latrille, E., Carrere, H., and Steyer, J.P. (2010). Hydrogen production from agricultural waste by dark fermentation: a review. *International journal of hydrogen energy*, 35(19), 10660–10673.
- Guwy, A., Hawkes, F., Hawkes, D., and Rozzi, A. (1997). Hydrogen production in a high rate fluidised bed anaerobic digester. *Water Research*, 31(6), 1291–1298.
- Hamid, K., Sabir, R., Hameed, K., Waheed, A., and Ansari, M. (2021). Economic analysis of fuel oil production from pyrolysis of waste plastic. *Austin Environ Sci*, 6, 1053.
- Henstra, A., Sipma, J., Rinzema, A., and Stams, A. (2007). Microbiology of synthesis gas fermentation for biofuel production. *Current Opinion in Biotechnology*, 18(3), 200–206.
- Hill, Jason. "Environmental costs and benefits of transportation biofuel production from food-and lignocellulose-based energy crops: a review." *Sustainable agriculture* (2009): 125-139.
- Ho, D., Ngo, H., and Guo, W. (2014). A mini review on renewable sources for biofuel. *Bioresource technology*, 169, 742–749.
- Hoffman, S., Alvarez, M., Alfassi, G., Rein, D., Garcia-Echauri, S., Cohen, Y., and Avalos, J. (2021). Cellulosic biofuel production using emulsified simultaneous saccharification and fermentation (eSSF) with conventional and thermotolerant yeasts. *Biotechnology for biofuels*, 14(1), 1–17.
- Hölker, U., and Lenz, J. (2005). Solid-state fermentation—are there any biotechnological advantages?. *Current opinion in microbiology*, 8(3), 301–306.
- Huber, M., Koidl, F., Kreutner, G., Giovannini, A., Kleinhapfl, M., Roschitz, C., Hofbauer, H., Gruber, F., Krueger, J., and others (2008). Multi staged gasification systems-a new approach.. In *Proceedings of the World Bioenergy Conference and Exhibition on Biomass for Energy, Jönköping, Sweden, 27-29 May 2008* (pp. 222–226).
- Im-orb, K., and Arpornwichanop, A. (2020). Process and sustainability analyses of the integrated biomass pyrolysis, gasification, and methanol synthesis process for methanol production. *Energy*, 193, 116788.
- Jenicek, P., Koubova, J., Bindzar, J., and Zabranska, J. (2010). Advantages of anaerobic digestion of sludge in microaerobic conditions. *Water Science and Technology*, 62(2), 427–434.
- Jeon, P., Moon, J.H., Olanrewaju, O., Lee, S., Ling, J., You, S., and Park, Y.K. (2023). Recent advances and future prospects of thermochemical biofuel conversion processes with machine learning. *Chemical Engineering Journal*, 144503.

- Johnson, A. (1987). The control of fed-batch fermentation processes—a survey. *Automatica*, 23(6), 691–705.
- King, E., Teitelbaum, P., Barany, R., Stecura, S., Kelley, K., Campbell, W., and Weller, W. (1961). *Energy Production and Consumption in the United States: An Analytical Study Based on 1954 Data*. US Department of the Interior, Bureau of Mines.
- Kumar, P., Barrett, D., Delwiche, M., and Stroeve, P. (2009). Methods for pretreatment of lignocellulosic biomass for efficient hydrolysis and biofuel production. *Industrial and engineering chemistry research*, 48(8), 3713–3729.
- Kumar, R., Singh, S., and Singh, O. (2008). Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. *Journal of industrial microbiology and biotechnology*, 35(5), 377–391.
- Kumari, D., and Singh, R. (2018). Pretreatment of lignocellulosic wastes for biofuel production: A critical review. *Renewable and Sustainable Energy Reviews*, 90, 877–891.
- Kushwah, A., Reina, T., and Short, M. (2022). Modelling approaches for biomass gasifiers: A comprehensive overview. *Science of the Total Environment*, 834, 155243.
- Laiq Ur Rehman, M., Iqbal, A., Chang, C.C., Li, W., and Ju, M. (2019). Anaerobic digestion. *Water Environment Research*, 91(10), 1253–1271.
- Lee, J., Lee, S., Park, S., and Middelberg, A. (1999). Control of fed-batch fermentations. *Biotechnology advances*, 17(1), 29–48.
- Li, T., Chen, X.b., Chen, J.c., Wu, Q., and Chen, G.Q. (2014). Open and continuous fermentation: products, conditions and bioprocess economy. *Biotechnology journal*, 9(12), 1503–1511.
- Lu, G., and Do, D. (1994). Comparison of structural models for high-ash char gasification. *Carbon*, 32(2), 247–263.
- Mahmood, H., Moniruzzaman, M., Iqbal, T., and Khan, M. (2019). Recent advances in the pretreatment of lignocellulosic biomass for biofuels and value-added products. *Current Opinion in Green and Sustainable Chemistry*, 20, 18–24.
- Mardiana, S., Azhari, N., Ilmi, T., and Kadja, G. (2022). Hierarchical zeolite for biomass conversion to biofuel: A review. *Fuel*, 309, 122119.
- Mata-Alvarez, J., Macé, S., and Llabres, P. (2000). Anaerobic digestion of organic solid wastes. An overview of research achievements and perspectives. *Bioresource technology*, 74(1), 3–16.
- Menzel, T., Neubauer, P., and Junne, S. (2020). Role of microbial hydrolysis in anaerobic digestion. *Energies*, 13(21), 5555.
- Miandad, R., Rehan, M., Barakat, M., Aburiazaiza, A., Khan, H., Ismail, I., Dhavamani, J., Gardy, J., Hassanpour, A., and Nizami, A.S. (2019). Catalytic pyrolysis of plastic waste: moving toward pyrolysis based biorefineries. *Frontiers in Energy Research*, 7, 27.
- Mohan, D., Pittman Jr, C., and Steele, P. (2006). Pyrolysis of wood/biomass for bio-oil: a critical review. *Energy and fuels*, 20(3), 848–889.

- Molino, A., Chianese, S., and Musmarra, D. (2016). Biomass gasification technology: The state of the art overview. *Journal of Energy Chemistry*, 25(1), 10–25.
- Moodley, P. (2021). Sustainable biofuels: opportunities and challenges. *Sustainable Biofuels*, 1–20.
- Munasinghe, P., and Khanal, S. (2010). Biomass-derived syngas fermentation into biofuels: opportunities and challenges. *Bioresource technology*, 101(13), 5013–5022.
- Ong, H., Chen, W.H., Farooq, A., Gan, Y., Lee, K., and Ashokkumar, V. (2019). Catalytic thermochemical conversion of biomass for biofuel production: A comprehensive review. *Renewable and Sustainable Energy Reviews*, 113, 109266.
- Picuno, P. (2016). Use of traditional material in farm buildings for a sustainable rural environment. *International Journal of Sustainable Built Environment*, 5(2), 451–460.
- Pointner, M., Kuttner, P., Obrlik, T., Jäger, A., Kahr, H., and others (2014). Composition of corncobs as a substrate for fermentation of biofuels. *Agronomy Research*, 12(2), 391–396.
- Pomázi, I. (2012). OECD Environmental Outlook to 2050. The Consequences of Inaction. *Hungarian Geographical Bulletin*, 61(4), 343–345.
- Resende, F. (2016). Recent advances on fast hydrolysis of biomass. *Catalysis Today*, 269, 148–155.
- Safarian, S., Unnp'orsson, R., and Richter, C. (2019). A review of biomass gasification modelling. *Renewable and Sustainable Energy Reviews*, 110, 378–391.
- Sarangi, P., and Nanda, S. (2020). Biohydrogen production through dark fermentation. *Chemical Engineering and Technology*, 43(4), 601–612.
- Sarangi, P., and Nanda, S. (2020). Biohydrogen Production Through Dark Fermentation. *Chemical Engineering and Technology*, 43(4), 601-612.
- Sarkis-Onofre, R., Catalá-L'opez, F., Aromataris, E., and Lockwood, C. (2021). How to properly use the PRISMA Statement. *Systematic Reviews*, 10(1), 1–3.
- Shi, L., Simplicio, W., Wu, G., Hu, Z., Hu, H., and Zhan, X. (2018). Nutrient recovery from digestate of anaerobic digestion of livestock manure: a review. *Current Pollution Reports*, 4, 74–83.
- Sugiyama, S., Suzuki, N., Kato, Y., Yoshikawa, K., Omino, A., Ishii, T., Yoshikawa, K., and Kiga, T. (2005). Gasification performance of coals using high temperature air. *Energy*, 30(2-4), 399–413.
- Sun, Y., Xiong, X., He, M., Xu, Z., Hou, D., Zhang, W., Ok, Y., Rinklebe, J., Wang, L., and Tsang, D. (2021). Roles of biochar-derived dissolved organic matter in soil amendment and environmental remediation: a critical review. *Chemical Engineering Journal*, 424, 130387.
- Todaro, C., and Vogel, H. (2014). *Fermentation and biochemical engineering handbook*. William Andrew.
- Vassilev, S., Baxter, D., Andersen, L., and Vassileva, C. (2010). An overview of the chemical composition of biomass. *Fuel*, 89(5), 913–933.

- Wilhelm, D. J., et al. "Syngas production for gas-to-liquids applications: technologies, issues and outlook." *Fuel processing technology* 71.1-3 (2001): 139-148.
- Yaashikaa, P., Devi, M., and Kumar, P. (2022). Algal biofuels: Technological perspective on cultivation, fuel extraction and engineering genetic pathway for enhancing productivity. *Fuel*, 320, 123814.
- Yang, Y., and Sha, M. (2019). A beginner's guide to bioprocess modes—batch, fed-batch, and continuous fermentation. Enfield, CT: Eppendorf Inc, 408, 1–16.
- Zhang, M., Wang, F., Su, R., Qi, W., and He, Z. (2010). Ethanol production from high dry matter corncob using fed-batch simultaneous saccharification and fermentation after combined pretreatment. *Bioresource Technology*, 101(13), 4959–4964.

High School Transition: Hardships and Solutions By Anya Bhakta

Summary: The transition from middle school to high school is a pivotal moment in a student's educational journey. This shift isn't just about changing physical locations; it marks a shift in mindset, expectations, and academic demands. Recognizing these complexities, educational institutions have introduced transition programs to ease this change.

The academic expectations in high school, including heavier workloads, self-directed learning, and critical thinking, can be daunting, especially for students who were not sufficiently challenged in middle school. Additionally, emotional and social adjustments, coupled with these academic demands, can overwhelm students, leading to a notable percentage dropping out during the ninth grade.

To support freshmen, effective transition programs have become crucial. Elements such as familiarizing students with the school layout, peer mentoring, and bridging knowledge gaps have been identified as key components for success. However, while the general idea of transition programs is accepted, the specific qualities of successful programs have only been more recently studied. Recent research highlights the importance of starting early transition programs before the standard summer before high school. and involve not just high school educators but also those from middle and elementary schools.

Due to the cost and resources of such programs being prohibitive socioeconomically, a comprehensive freshman workbook, covering topics like navigating high school, study skills, test preparation, goal setting, time management, research skills, and digital literacy, emerges as a cost-effective supplement to transition programs. This workbook can serve as a valuable resource for schools with limited resources, providing essential information and empowering students to successfully transition to high school.

Article: The transition from middle school to high school is a critical point in a student's academic journey. It's a period marked by a whirlwind of challenges and adjustments that can deeply impact a young learner's experience. The shift to high school signifies more than just a change of physical surroundings; it signifies a shift in mindset, expectations, and academic demands (Roybal et al 475). The metamorphosis from being an 8th-grade upperclassman in a smaller pond to becoming a 9th-grade lowerclassmen in a much larger one can be both exciting and daunting. Recognizing the complexity of these challenges, many educational institutions have sought to ease the transition for incoming freshmen through the implementation of summer transition programs. This paper aims to identify components of effective transition programs that successfully bridge the gap between middle and high school. From the crucial role of mentorship and buddy systems to the innovative "school within a school" approach, this paper seeks to identify the components that equip freshmen with the tools they need to thrive in their new academic environment.

The academic expectations of high school can appear like a towering mountain. The demand for heavier workloads, the emphasis on self-directed learning, and the need to develop critical thinking skills can create a sense of academic unease. Students who confront academic

struggles or who have not been adequately challenged during their middle school years are especially susceptible to faltering in their high school journey ([Neild 59](#)). With the increase in graduation requirements and college readiness becoming a design of high school as opposed to a possible outcome, rigor at the high school level has increased tremendously ([Cohen and Smerdon 178](#)). This can be especially daunting for students who were not adequately challenged during their middle school years. Unlike middle school, where the hand-holding approach may have been more prevalent, high school assumes a certain level of independence and self-motivation ([Dillon 18](#)). The realization that high school demands a higher level of dedication for comparable results can be disheartening. The more rigorous curriculum, with its emphasis on self-study and efficient scheduling, presents a substantial obstacle for new freshmen, highlighting the importance of academic readiness during this critical juncture ([Longobardi et al 2](#)).

The realization that high school necessitates a higher level of dedication for comparable results can be disheartening, leaving some students grappling with feelings of inadequacy ([Neild 59](#); [Cohen and Smerdon 177](#)). In addition to this accelerated coursework, high school newcomers also manage intricate social dynamics, emotional turmoil, and behavioral adjustments ([Dillon 17](#)). These factors create a complex web of experiences that can potentially overwhelm and leave students feeling unprepared for the journey ahead ([Longobardi et al 1](#)). In fact, according to the U.S. Department of Education, only about 60 percent of freshmen from 1,700 schools nationwide make it to graduation ([Dillon 17](#)).

Educational institutions have recognized the need for comprehensive transition programs to support incoming freshmen as they navigate the complexities of high school. Summer transition programs are designed not only to provide academic preparation, but also to offer a supportive environment that addresses the diverse needs of new high school entrants ([Neild 70](#)). They recognize that the transition to high school is not just about adjusting to academic rigor; it's also about fostering emotional well-being, social integration, and the development of essential life skills ([Longobardi et al 2](#)). Additionally, while most high schools have a program dedicated to helping new freshmen familiarize themselves with their school, the risk of achievement loss and dropping out of school continues to occur for most students during this pivotal moment ([Akos and Galassi 212](#); [Andrews and Bishop 12-13](#)).

While having an extensive transition program lowers failure and drop out rates, the specific qualities of successful transition programs have only more recently been the subject of study and critique, specifically the student perception of the successfulness of such programs ([Mizelle and Irvin 58](#); [Akos and Galassi 212-13](#)). Current research suggests that being familiar with the layout of one's high school, having a peer mentor that has successfully completed the transition "buddy" with an incoming student, and having tutoring specifically aimed at bridging gaps in knowledge freshman students may experience are the most helpful components of a successful transition program. Additionally, beginning these programs earlier than the summer before freshman year is strongly encouraged for maximum effectiveness ([Mizelle and Irvin 59-60](#); [Akos and Galassi 217](#); [McIntosh and White 40-42](#)).

This paper aims to identify the most pivotal components of successful high school transition programs and attempts to address ways that these positive elements can be implemented early to increase student success.

Learning the School

One of the most noticeable transitions is the change in classroom dynamics. The shift from cozy, familiar middle school classrooms to the expansive and bustling corridors of high school can be disorienting ([Roybal et al 475](#); Thompson and Roberts 41-42). The once-intimate setting where teachers knew every student's name can give way to anonymity, and navigating the hallways can evoke feelings of being lost in a sea of unfamiliar faces. Negotiating this transition requires not only adapting to a more impersonal environment, but also finding one's niche within it. Significant increase in class sizes and the reduced teacher-student interaction can lead to feelings of isolation and disconnection among freshmen (42-43). This emotional strain can adversely affect students' self-esteem and motivation to engage in classroom activities, which ultimately impacts their academic performance and overall adjustment to high school ([Longobardi et al 2](#)).

Furthermore, the very structure of high schools, with added counselors and administrators, can be perplexing for incoming students ([Shaunessy-Dedrick et al. 109-110](#)). With a need to manage schedules, select courses, and juggle extracurricular activities, navigating the complex hierarchy of high school and understanding the protocols for seeking assistance can prove to be an unexpected challenge. These factors, which are often underestimated, can contribute significantly to the sense of unpreparedness that many freshmen experience ([Longobardi et al 2](#)). Additionally, understanding the requirements to graduate are fundamental to successful high school transitions. Educators should outline the courses, grades, and credits needed to advance past the ninth grade ([Cohen and Smerdon 177](#)).

Addressing this lack of connection is crucial as student engagement is both directly and indirectly affected by educators. Teachers who take time to individually address students build a sense of connection between freshmen students and their school ([Roybal et al 478-9](#)). Additionally, the diminishing role of parents in their children's lives, coupled with the influence of peer pressure, can lead to heightened isolation and risky behavior ([Neild 58](#)). While parental involvement decreases once a student reaches high school, schools and teachers that specifically reach out to parents to maintain involvement see a more successful middle to high school transition ([Cohen and Smerdon 181](#)).

All strong high school transition programs should inform students on how best to navigate their high school. Teachers should be given an opportunity to introduce themselves to their students prior to starting work in the classroom. Additionally, administration officials, such as guidance counselors and principals, should also have an opportunity to meet with students and explain their role in the school and in the student's academic journey and the overall credits and courses needed to succeed. Last, this information should be shared with parents.

The Buddy System

The shift from a reliance on parents to an exploration of independence may leave students feeling alone in dealing with the evolving dynamics of relationships and personal hardships. High school, particularly the ninth grade, becomes a phase where students engage in a process of readjustment and redefinition of their identities, often resulting in challenges to their emotional well-being ([Longobardi et al 2](#)). These factors, alongside the potential loss of social networks from middle school, amplify the emotional strain experienced by freshmen, which can have a detrimental impact on their mental health.

To alleviate the sense of isolation that accompanies this transition, it is imperative that every comprehensive transition program integrates a buddy system ([Neild 70](#)). This system involves pairing incoming freshmen with upperclassmen mentors prior to the commencement of the school year. These upperclassmen serve as guides, aiding the newcomers in acclimating to the high school environment. For instance, the International School of Beijing employs this approach, wherein new students are matched with middle school buddies who familiarize them with the campus and its intricacies ([International School of Beijing](#)). This system has proven effective in mitigating feelings of isolation, fostering a sense of belonging, and nurturing a positive acceptance of self and others (Whitby School).

Beyond its social benefits, the buddy system can also play a pivotal role in enhancing academic success for freshmen. In an ideal scenario, upperclassmen mentors offer tutoring, homework assistance, and guidance on effective teacher interaction, thereby potentially leading to improved grades ([Neild 70](#); [Roybal et al 483](#)). Additionally, the buddy system promotes the establishment of social connections, allowing upperclassmen to facilitate networking among their assigned freshmen buddies based on shared interests ([Dillon 16-18](#); [Roybal et al 483](#)). These connections can evolve into study groups, helping students reinforce weaker areas and collectively address academic and social challenges posed by the transition ([Neild 70](#)).

However, the implementation of an effective buddy system is not without its challenges. Ensuring the quality and consistency of mentor-mentee relationships can be complex. Issues such as limited access to communication tools, including phones and email, and scheduling conflicts may hinder effective interaction between buddies. To address these concerns, the introduction of incentives, such as extra credit for students and service awards for upperclassmen, can help ensure the commitment and active engagement necessary for a successful buddy system ([Neild 70](#)).

A key factor to an effective buddy system is the introduction of an in-person summer program. This program, similar to an orientation, offers new students an early opportunity to familiarize themselves with the school environment and, critically, to meet their assigned buddies before the commencement of classes ([Cohen and Smerdon 180-81](#)). Drawing inspiration from the successful model at the International School of Beijing, these summer programs can facilitate meaningful interactions, thereby fostering a stronger bond between freshmen and their mentors ([International School of Beijing](#)).

These pre-school activities can encompass personalized breakout sessions and group engagements, all aimed at nurturing a sense of camaraderie and introducing new students to the concept of peer support. Research has indicated that peers can serve as a safety net for students navigating challenging environments, enhancing their academic and social success (Gul et al.). By creating an environment of safety and support, these programs contribute to an atmosphere conducive to academic achievement.

The effectiveness of the buddy system heavily relies on the compatibility of mentor-mentee pairs. To maximize this compatibility, educators and school administrators should consider the alignment of interests and backgrounds. One approach involves the distribution of surveys among both underclassmen and upperclassmen, allowing them to articulate their hobbies, academic interests, and aspirations. Collating and analyzing these survey responses can facilitate the generation of pairings that share common traits and experiences, thereby enhancing the potential for a meaningful mentorship (Neild 70).

Strong transition programs should incorporate a buddy system between upperclassmen and lowerclassmen with a means to ensure compatibility and a continued relationship.

A School Within a School

While the transition from middle to high school is often thought of as a one-time event, the process actually begins at some point during middle school and continues into students' freshman year (Cohen and Smerdon 179-180). Helping students understand the academic expectations of high school, while acquiring essential study skills and habits, can significantly contribute to their success in high school, college, and beyond (Bottoms and Timberlake 5).

Addressing the significant gap in academic readiness between middle and high school is paramount to a successful transition. A 2006 survey involving 11,000 ninth graders revealed that a substantial percentage of students lacked the necessary skills for college-preparatory high school courses (1). This disparity is particularly evident for non-AP students, with freshmen in regular courses expressing inadequacy in high school-level writing, math, and science (1). The misalignment between middle and high school standards results in high ninth-grade failure rates (2).

To address this challenge, the "school within a school" concept emerges as a promising strategy. Tailored for students who are academically behind in core subjects, this approach swiftly addresses learning gaps and equips students with the foundational knowledge necessary for success in high school (Cauley and Jovanovich 18). By providing an intensive curriculum that addresses their specific academic deficits, the "school within a school" model offers a targeted solution to the challenge of academic readiness (18).

Schools that have successfully implemented the "school within a school" model have witnessed notable improvements in promotion and pass rates, accompanied by a reduction in dropout rates (18). By catering to individual learning needs and facilitating catch-up opportunities, this model proves instrumental in enabling students to meet graduation

requirements and achieve the necessary grades. Students feel “less bored in classes, believed that teachers were fair, and felt that teachers treated them respectfully” ([Roybal et al 482](#)).

The "school within a school" concept is also conducive to integrating engaging reading, writing, and hands-on real-world projects in math and science. These enrichment activities, while challenging to incorporate into the regular curriculum, can be seamlessly integrated within the specialized environment of the "school within a school" ([Bottoms and Timberlake 13](#)). However, the limited timeframe of typical one-day to one-week programs for incoming freshmen is insufficient to comprehensively address the knowledge discrepancies that many students face ([Cauley and Jovanovich 18-19](#)).

Although this component of transition programs is the most resource intensive, having a “school within a school” provides at-risk students with the means to succeed in the rigorous high school environment.

Beginning Early

Because the transition from middle school to high school is a pivotal juncture in a student's academic journey, it is crucial to provide them with the tools and skills necessary to bridge the readiness gap effectively. The most successful transition programs recognize the need to start early, imploring elementary, middle, and high school educations to work together with parents and students to ensure students experience a seamless transition between schools ([Mizelle and Irvin 60](#)). Visits to eighth graders by ninth grade counselors are viewed as particularly helpful ([59](#)). Holistic programs that consider a student's education from elementary to high school help prevent gaps that are challenging to bridge once they form ([Murphy 11-12](#)).

The issue with beginning early (as is the issue with most supplemental programs) is cost ([12](#)). While successful completion of intensive high school curriculum is a fundamental part of socioeconomic advancement, income segregation between districts strongly dictates the resources available to most students ([Baker 370](#); [Owens 2](#)).

The Workbook

While many schools employ one of two of the above components, a successful transition program will incorporate numerous strategies ([Roybal et al 480](#)). However, while most parents, students, and educators agree that transition programs and resources impact overall student success, implementing such strategies early and comprehensively is costly, and the cost of interventions to close gaps ultimately dictate their implementation ([Murphy 12](#); [Turner 245-246](#)).

A comprehensive freshman workbook designed to empower students with the aforementioned essential information would therefore be a strong accompaniment to any high school transition program. It could also add accessibility for schools that lack the resources to provide a comprehension transition program.

The ideal workbook would offer activities in various key areas to help students successfully transition to high school:

Navigating High School: Introducing students to the different kinds of teachers and administrators they will encounter and giving tips on how to navigate the larger campus.

Study Skills: Developing effective study habits is essential for high school success. The workbook can provide strategies for time management, note-taking, and organization, helping students become more efficient learners.

Test Preparation: High school often involves standardized tests and exams. The workbook can include sections on test-taking strategies, practice questions, and tips for managing test anxiety.

Goal Setting: Setting academic and personal goals is crucial. The workbook can guide students in setting SMART (Specific, Measurable, Achievable, Relevant, Time-bound) goals and tracking their progress.

Time Management: High school can be demanding, with multiple assignments and extracurricular activities. The workbook can teach students how to prioritize tasks, create schedules, and balance their academic and personal lives.

Research Skills: High school often requires research projects. The workbook can introduce students to research methods, citing sources, and creating bibliographies.

Digital Literacy: In today's digital age, understanding technology is crucial. The workbook can cover topics like online research, digital ethics, and using technology for academic purposes.

By incorporating these elements into a comprehensive freshman workbook, students will not only enhance their academic readiness but also develop essential life skills that will serve them well throughout their high school journey and beyond. This workbook can act as a valuable resource for schools looking to support their students during this critical transition period, especially those with limited resources to provide extensive transition programs.

Conclusion

The challenges faced by incoming freshmen during their transition to high school necessitate comprehensive transition programs that encompass a multifaceted approach. The establishment of a well-structured buddy system, accompanied by in-person summer programs and a dedicated "school within a school" framework, holds the potential to smooth many of the problems students find during this time. These programs provide the necessary support for incoming students to navigate the complexities of high school, fostering not only academic success but also emotional well-being and overall achievement. A great accompaniment to these programs is a comprehensive workbook that will act as a supplement or, if resources are unavailable, the main curriculum for a transition program.

Works Cited

- Akos, Patrick, and John P. Galassi. "Middle and High School Transitions as Viewed by Students, Parents, and Teachers." *Professional School Counseling*, vol. 7, no. 4, SAGE Publishing, Apr. 2004, p. 212.
- Andrews, Colin, and Penny Bishop. "Middle Grades Transition Programs around the Globe." *Middle School Journal*, vol. 44, no. 1, Sept. 2012, pp. 8–14.
- Baker, Joe G. "Class, Ability, Mobility: Economic and Academic Paths from Middle School to Early Adulthood." *Journal of Education Finance*, vol. 34, no. 4, Jan. 2009, pp. 355–71.
- Bottoms, Gene, and Allison Timberlake. *Preparing Middle Grades Students for High School Success*. 2002.
- "Buddy Systems Welcome New Students to the ISB School." *Blog.isb.cn*, blog.isb.cn/what-is-the-buddy-system-and-how-does-it-help-your-child-transition-into-isb#:~:text=Counselor%20Kyle%20Ottaviano.-.
- Cauley, Kathleen M., and Donna Jovanovich. "Developing an Effective Transition Program for Students Entering Middle School or High School." *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, vol. 80, no. 1, Sept. 2006, pp. 15–25.
- Cohen, Jennifer S., and Becky A. Smerdon. "Tightening the Dropout Tourniquet: Easing the Transition from Middle to High School." *Preventing School Failure: Alternative Education for Children and Youth*, vol. 53, no. 3, Apr. 2009, pp. 177–84, <https://doi.org/10.3200/psfl.53.3.177-184>.
- Dillon, Naomi. "The Transition Years." *Education Digest: Essential Readings Condensed for Quick Review*, vol. 74, no. 1, 2008.
- Ellerbrock, Cheryl R., et al. "Fostering a Developmentally Responsive Middle-To-High School Transition: The Role of Transition Supports." *The Middle Grades Research Journal*, vol. 10, no. 1, Apr. 2015, pp. 83–101.
- Ellerbrock, Cheryl R., and Sarah M. Kiefer. "Supporting Young Adolescents' Middle-To-High-School Transition by Creating a Ninth Grade Community of Care: Implications for Middle Grades Educators." *Middle School Journal*, vol. 45, no. 3, Jan. 2014, pp. 3–10.
- Ellerbrock, Cheryl, and Sarah Kiefer. "Supporting Young Adolescents' Middle-to High-School Transition by Creating a Ninth Grade Community of Care: Implications for Middle Grades Educators." *Journal*, vol. 45, no. 3, 2014, pp. 3–10. Accessed 21 Nov. 2023.
- Iver, Martha Abele Mac, et al. "Engaging Families to Support Students' Transition to High School: Evidence from the Field." *The High School Journal*, vol. 99, no. 1, 2015, pp. 27–45.
- Longobardi, Claudio, et al. "Student-Teacher Relationships as a Protective Factor for School Adjustment during the Transition from Middle to High School." *Frontiers in Psychology*, vol. 7, Dec. 2016.
- McIntosh, Julie, and Simon White. "Building for Freshman Success: High Schools Working as Professional Learning Communities." *American Secondary Education*, vol. 34, no. 2,

- Apr. 2006, pp. 40–49.
- Mizelle, Nancy B., and Judith L. Irvin. “Transition from Middle School to High School. What Research Says.” *Middle School Journal*, vol. 31, no. 5, Taylor and Francis, Jan. 2000, pp. 57–61.
- Murphy, Joseph. “Closing the Achievement Gap.” *The Phi Delta Kappan*, vol. 91, no. 3, 2009, pp. 8–12.
- Neild, Ruth Curran. “Falling off Track during the Transition to High School: What We Know and What Can Be Done.” *The Future of Children*, vol. 19, no. 1, 2009, pp. 53–76.
- Owens, Ann. “Income Segregation between School Districts and Inequality in Students’ Achievement.” *Sociology of Education*, vol. 91, no. 1, Nov. 2017, pp. 1–27.
- Roybal, Victoria, et al. “Effective Ninth-Grade Transition Programs Can Promote Student Success.” *Education*, vol. 134, no. 4, June 2014, pp. 475–87.
- Shaunessy-Dedrick, Elizabeth, et al. “Students’ Perceptions of Factors That Contribute to Risk and Success in Accelerated High School Courses.” *The High School Journal*, vol. 98, no. 2, 2015, pp. 109–37.
- Turner, Sherri. “Preparing Inner-City Adolescents to Transition into High School.” *Professional School Counseling*, vol. 10, no. 3, Feb. 2007, pp. 245–52.

How did Intellectuals Influence Public Opinion and the Course of the Dreyfus Affair in France at the Turn of the 19th Century and in Subsequent French Politics and Society? By Nikhil Kothari

“Zola not only revealed a judicial error, he denounced the conspiracy of all the forces of violence and oppression united to kill social justice, the republican idea, and free thought in France. His courageous words awakened France...he was a moment in the conscience of humanity.”⁴⁹

Anatole France, *Eulogy for Emile Zola*, October 5, 1902

Introduction

What is it about the Dreyfus Affair that has preoccupied us for over a century? Perhaps it is because social justice, democracy, and freedom still elude much of our world. Alfred Dreyfus was an innocent man. He was unjustly accused and convicted, and his case transcended the legal realm to divide France and portend extreme anti-semitism in Europe. The Dreyfus Affair raised questions about France’s identity as a nation. Would France remain in the past or move forward to the future? Would it become a secular republic bestowing on the individual the French Revolution values of “liberty, equality, and fraternity,” or remain steeped in the tradition of army, monarchy, and the Catholic Church? These were challenging questions; while Enlightenment philosophes such as Montesquieu and Voltaire concerned themselves with ideology and its relationship with society, with the Dreyfus Affair, intellectuals stepped in to wield influence on affairs of the state by swaying public opinion. History, intellectual petitions, and manifestoes became inextricably linked. Intellectual Emile Zola and other “Dreyfusards” defended truth, justice, and the public interest by forcing the people of France to reckon with the injustice committed against Dreyfus and rampant anti-semitism. On the other side, Edouard Drumont, Maurice Barres, and other “anti-Dreyfusards” promoted anti-semitism and fueled the rhetoric against Dreyfus. The intellectuals influenced public opinion and the trajectory of the Dreyfus Affair (and French politics and society more broadly) by filling the void left by transforming political and societal institutions and a corrupt army and legal system. Although the influence of the intellectual was a short-lived moment and waned in the early 1900s as radicalization came to dominate French politics and society, in the face of prominent opposition, the intellectuals used the power of reason, practiced activism, and skillfully engaged the press to stimulate crucial public debate. In doing so, they were able to resolve the Dreyfus Affair and catalyze political and social progress.

Historiography

The people of France in the late 1800s grappled with instability and uncertainty. The country was shaken by its humiliating defeat to the German confederation in the Franco-Prussian War in 1870-1871, which resulted in France ceding land to Germany and the balance of power in

⁴⁹ Michael Burns, *France and the Dreyfus Affair: A Documentary History* (Boston & New York: Bedford / St. Martin’s, 1999), 168.

Europe shifting in Germany's favor. The Treaty of Frankfurt ended the war on May 10, 1871, but France viewed its conditions as punitive. Alsace and Lorraine were annexed to the newly united German empire. In addition, France was charged reparations of five billion francs and the German Army occupied parts of France until it was paid. As a result, France experienced a hunger for retribution, aversion to foreign powers (especially Germany), reverence for the army, and the rise of nationalism.⁵⁰ The French Revolution had catalyzed the introduction of democratic ideals, but France was torn between tradition and progressiveness. Protestants attacked Catholic institutions and there was tension about separation of church and state in areas such as education. Economic challenges also ensued and when the Union Generale bank, a Catholic institution, crashed in 1881, the Catholic press blamed Jewish conspiracies.⁵¹ Jews were made fully emancipated citizens in 1791 after the French Revolution but a facade of equality hid undercurrents of anti-semitism and French society was ripe for xenophobia. In the political realm, the Third Republic was proclaimed in 1870 but there was ongoing debate about republican versus monarchical rule. As a result, during "the birth of the Third Republic to the height of the Dreyfus Affair, the nation was at war with itself."⁵² A confluence of powerful forces put stress on French politics and society and elevated the army which was viewed as a bastion of security.

The Dreyfus Affair

It was into this unstable era that Alfred Dreyfus was born in 1859 as the son of a wealthy textile industrialist from a Jewish Alsatian family. After the Franco-Prussian War in 1870, the family left Alsace to keep their French nationality. Dreyfus eventually joined the French army where he became a captain in 1889, and in 1893, he joined the General Staff. The Dreyfus Affair began in 1894 when the *bordereau* (Appendix 1), a secret memorandum torn into six pieces, was discovered by Marie Bastian, an employee for the French government. She found it in the wastebasket of Maximilian von Schwartzkoppen, the German military attaché in Paris. The note promised to provide information regarding a French cannon, artillery formations, and a firing manual for field artillery, among other items. She gave it to the Statistical Section, the department of the French Army responsible for military intelligence and national security, who concluded that the *bordereau* was written by a French officer.⁵³ They further concluded that the officer who wrote it must have had experience in artillery and had a connection with the General Staff. Captain Alfred Dreyfus fit the criteria. In reality, "Dreyfus, the only Jewish officer trainee on the General Staff, was fated to be singled out."⁵⁴ The Statistical Section fabricated evidence against Dreyfus because he was Jewish even though the handwriting experts that evaluated the *bordereau* were divided about whether Dreyfus wrote the note.

⁵⁰ George Whyte, *The Dreyfus Affair: A Chronological History* (Hampshire & New York: Palgrave Macmillan, 2005), 8.

⁵¹ Whyte, *The Dreyfus Affair: A Chronological History*, 7.

⁵² Burns, *France and the Dreyfus Affair: A Documentary History*, 1.

⁵³ Joseph Epstein, "Why Captain Dreyfus?," *The Weekly Standard*, May 21, 2012, <https://www.washingtonexaminer.com/weekly-standard/why-captain-dreyfus>.

⁵⁴ Louis Begley, *Why the Dreyfus Affair Matters* (New Haven & London: Yale University Press, 2009), 6.

Aside from his family, no one believed Dreyfus was innocent given the preeminent position of the army and the burgeoning anti-semitism shaping popular sentiment through the press. The conservative anti-Dreyfusards battled against the reformist Dreyfusards that were composed of the rising middle class including the Jews. The newspaper *La Libre Parole* announced the headline “High Treason: Arrest of the Jewish Officer A. Dreyfus” on November 1, 1894. A few days later on November 10, 1894, the newspaper published a cartoon of Dreyfus as “Judas” (Appendix 2).⁵⁵ The anti-Dreyfusards were winning against Dreyfus in the court of public opinion. Edouard Drumont was the anti-semitic editor of *La Libre Parole*, which he founded in 1892, with the sub-heading “France for the French.” The journal played an important role in turning the Panama scandal into a national attack on Jewish finance, which Drumont blamed on having corrupted the government.⁵⁶ Previously, in 1886, Drumont published *La France Juive* (Jewish France) where he wrote, “as you advance through life and become experienced, Jews will block your path everywhere.”⁵⁷ Drumont’s anti-semitic views surfaced in *La Libre Parole* where he used rhetoric and sensationalism to rouse public sentiment against Jews and vilify Dreyfus in the press. On December 22, 1894, the court-martial board unanimously declared Captain Alfred Dreyfus guilty of treason.⁵⁸ In his diary entry that he wrote on Devil’s Island on April 14, 1895, Dreyfus said, “until now, I worshiped reason; I believed in the logic of things and events; I believed, finally, in human justice!”⁵⁹ Dreyfus was despondent about his verdict of guilt without evidence, trial or “logic,” reflecting the farce of the French legal system and the power of the press to influence an unjust verdict.

Even though the leaders of the case discovered in 1896 that Major Ferdinand Walsin Esterhazy was the real traitor, a cover-up ensued. Georges Picquart, who became the head of the Statistical Section, came to Dreyfus’s defense. He discovered a new letter from the traitor to von Schwartzkoppen—sent from Esterhazy. Though Picquart initially thought Esterhazy was a second traitor, he realized that the handwriting matched the original *bordereau*. He told his superior General Charles-Arthur Gonse who proclaimed, “what does it matter to you if that Jew stays on Devil’s Island?” Picquart responded, “what you’re saying is vile. I don’t know what I will do, but of one thing I am certain - I will not take this secret to the grave.”⁶⁰ Motivated by the failure of justice, Picquart was determined to advocate for the truth. Dreyfus’s brother, Mathieu, his wife Lucie, and others had worked tirelessly to prove his innocence to no avail.

Zola and the Intellectuals

Bernard Lazare, a Jewish-French intellectual, journalist, and fighter against anti-semitism supported Dreyfus by using reason to point out the discrepancies in the case, folly of public opinion, and rampant anti-semitism. Lazare published the book *A Judicial Error* and paved the

⁵⁵ Burns, *France and the Dreyfus Affair: A Documentary History*, 34.

⁵⁶ David Rimoch, “The Affair or the State: Intellectuals, the Press, and the Dreyfus Affair,” (Senior Honors Thesis, University of Pennsylvania, April 2008), 16, https://repository.upenn.edu/hist_honors/8/.

⁵⁷ J.S. McClelland, *The French Right*. (New York: Harper & Row, Publishers, 1970), 102.

⁵⁸ Burns, *France and the Dreyfus Affair: A Documentary History*, 43.

⁵⁹ Burns, *France and the Dreyfus Affair: A Documentary History*, 59.

⁶⁰ Whyte, *The Dreyfus Affair: A Chronological History*, 467.

way for Zola to take the fight for Dreyfus into the public arena. As Lazare said in November 1896, Dreyfus “was arrested, then accused, tried, convicted, without it being known on what grounds he was arrested, and what accusation had been leveled against him, how he was tried, and why he was convicted...public opinion, without hesitation, acknowledged his guilt and heaped anger and exertion upon him.”⁶¹ Lazare pointed out the ambiguity of the situation, the unjust ways in which Dreyfus was treated and the public’s acceptance of it without asking questions. Lazare also asserted that Dreyfus was prosecuted because of his race saying “he is a soldier, but he is a Jew, and it is as a Jew that he was prosecuted...because he was a Jew, the voice of justice and truth could not be heard in his favor.”⁶² Lazare used facts to counter public opinion so that the public might start asking questions and identify inconsistencies. He attempted to create controversy around the situation and cause people to be more skeptical about how everything was handled. He also built a case to convince Zola to take action.

While Lazare raised the issue of Dreyfus’s innocence, Zola proceeded to publicize the argument in the press. On December 1, 1897, Zola wrote an article in *Le Figaro* called “The Syndicate” where he said “ours is a syndicate to act on public opinion, to cure it of the frenzy into which the foul press has whipped it up, and restore it to its age-old dignity and generosity.”⁶³ Zola used the same term of “syndicate” that the anti-Dreyfusards used in a derogatory way to describe Jews. While the mainline nationalistic press continued to advocate the anti-Dreyfusard perspective, and thereby shaped the course of the Dreyfus Affair, Zola strategically used the emerging alternative press to generate controversy. The Dreyfus Affair reached a turning point when Zola published “J’Accuse” on January 13, 1898 (Appendix 3) in Georges Clemenceau’s paper *L’Aurore*, which boldly asserted that the conviction had been ordered by the Army General Staff. In 1898, the circulation of daily newspapers such as *La Libre Parole* was 100,000, *La Croix* was 170,000, *Le Petit Journal* over a million, *Le Petit Parisien* over 700,000, and *Le Journal* about 450,000 for a combined total of 2,420,000. The circulation was significantly more than *Le Figaro* with 40,000 and the Dreyfusard *L’Aurore* with 25,000 for a total of 65,000.⁶⁴ Zola had help from Clemenceau who had 300,000 copies of *L’Aurore* printed, ten times its usual circulation.⁶⁵ Zola was able to sway public opinion even with the overall smaller circulation, proving the power of intellectual reason combined with the skillful use of the press could create impact and lead to change in society.

Zola’s “J’Accuse” raised the legality of the conviction and brought the ideological conflict into the public arena for debate. He stated, “I accuse the War Office of having conducted an abominable campaign in the press (especially in *L’Eclair* and *L’Echo de Paris*) in order to cover up its misdeeds and lead public opinion astray.”⁶⁶ He named the officers who had executed the cover-up and accused the General Staff of a “miscarriage of justice.”⁶⁷ Zola claimed that “it is

⁶¹ Burns, *France and the Dreyfus Affair: A Documentary History*, 76.

⁶² Burns, *France and the Dreyfus Affair: A Documentary History*, 77.

⁶³ Zola, *The Dreyfus Affair; ‘J’accuse’ and Other Writings*, 19.

⁶⁴ Rimoch, “The Affair or the State: Intellectuals, the Press, and the Dreyfus Affair,” 93.

⁶⁵ Rimoch, “The Affair or the State: Intellectuals, the Press, and the Dreyfus Affair,” 90.

⁶⁶ Zola, *The Dreyfus Affair; ‘J’accuse’ and Other Writings*, 52.

⁶⁷ Zola, *The Dreyfus Affair; ‘J’accuse’ and Other Writings*, 48.

a crime to poison the minds of the humble, ordinary people, to whip reactionary and intolerant passions into a frenzy while sheltering behind the odious bastion of anti-Semitism. France, the great and liberal cradle of the rights of man, will die of anti-Semitism if it is not cured of it.”⁶⁸ He concluded that “the truth is on the march, and nothing shall stop it.”⁶⁹ Zola elevated truth and the rights of man at a time when many others refused to do so. He also put the army on the defense by accusing it of a “crime.” Additionally, Zola talked about the possible death of France due to discrimination as a way to seize attention. With the issue now in the public arena, the public became responsible for justice and could demand that Dreyfus be absolved.

The poignancy of “J’Accuse” stirred up other intellectuals turning the Dreyfus Affair into an activist movement. Two petitions were issued asking for a revision of the trials and included 2,000 signatures.⁷⁰ On January 14th, a day after the publication of “J’Accuse,” Clemenceau published the names under “Manifesto of the Intellectuals” in *L’Aurore*.⁷¹ Along with Zola, Clemenceau, and Lazare, the Dreyfusards included Anatole France, Henri Poincare, Claude Monet, Charles Peguy, and Marcel Proust among others. On February 1, 1898, two weeks after J’accuse was published, Barres published an article in *Le Journal* titled “The Protest of the Intellectuals” in response to the Dreyfusard petition in order to mock the intellectuals. Cartoons of the intellectuals were published in the press (Appendix 4). Barres wrote, “in rejecting the intellectuals, we should pity them rather than offer them ill-will...I congratulate their resourceful recruiting sergeants for saying: ‘Give me your name and I will give you the title of ‘intellectual’...’”⁷² By discrediting the intellectuals, Barres attempted to undermine their activism and cast doubt on their claims. Scholars such as Barres also espoused anti-semitism stating that “Dreyfus is capable of betrayal, I deduce from his race.”⁷³ Nationalism also appeared in public debate and Barres stated that “the camp which supports Dreyfus as a symbol, would put into power those men whose intention it is to *remake France in the image of their own prejudices*. And as for me, I want to preserve France.”⁷⁴ For Barres, France came before the individual and he accused the intellectuals of trying to “remake” France. Though he accused the intellectuals of “prejudices,” ironically his verdict of guilt was based on prejudice, highlighting the irrationalism that dominated the Dreyfus affair. This irrationalism was present among many scholars and spilled into the public debate, which the intellectuals had to confront in order to mobilize support and resolve the Dreyfus Affair.

By the end of 1898, Zola had ignited a fire by turning the Dreyfus case into a public debate and setting it up for eventual resolution. However, he was sued for libel, found guilty, and forced to flee to England or go to prison. Zola endured harassment, kidnapping, assassination attempts, and setbacks in the literary world—even the cause of his death in 1902 became a

⁶⁸ Zola, *The Dreyfus Affair; ‘J’accuse’ and Other Writings*, 51.

⁶⁹ Zola, *The Dreyfus Affair; ‘J’accuse’ and Other Writings*, 52.

⁷⁰ Whyte, *The Dreyfus Affair: A Chronological History*, 206.

⁷¹ Whyte, *The Dreyfus Affair: A Chronological History*, 154.

⁷² Tom Conner, *The Dreyfus Affair and the Rise of the French Public Intellectual* (Jefferson, North Carolina: McFarland & Company, Inc., 2014), 177.

⁷³ Rimoch, “The Affair or the State: Intellectuals, the Press, and the Dreyfus Affair,” 83.

⁷⁴ McClelland, *The French Right*, 167.

matter of controversy.⁷⁵ The pressure from Zola and the intellectuals and influence on public opinion resulted in Dreyfus being offered a second trial. Though he was found guilty again, Dreyfusards had garnered world attention and French politicians were worried that the Paris Exposition of 1900 would be boycotted. President Emile Loubet pardoned Dreyfus in 1899. The Dreyfusards were not content since a pardon was not an acquittal. Eventually, on July 12, 1906, the Supreme Court of Appeal annulled the Rennes trial and declared Dreyfus's innocence. Dreyfus was reinstated in the army and awarded the Legion of Honor.

With the Dreyfus Affair, the idea emerged that intellectuals could defend society by championing French Revolution ideals and driving public debate; however, polarization evolved and the nationalist faction dominated by the 1920s, leaving France ill-equipped to deal with fascism and communism. The pressure of the industrial revolution, problems with Germany, and political instability created an environment where nationalism could flourish, calling for a return to the traditions of monarchy, Church, and army.⁷⁶ Additionally, even Dreyfusard intellectuals like Charles Peguy adopted nationalism to fight against Germany.⁷⁷ Peguy proclaimed "down with democracy! Down with universal suffrage! Long live national energy! Long live the racial instinct! Long live the French drum! Long live the French cannons, as thin and skinny as French adolescents! Long live war, which has become the hope of all true French people!"⁷⁸ The progress that France had made in terms of democracy and "racial" issues needed to be cast aside for a person to be "true French." There was a myopic focus on fighting. France fell in 1940 and according to political theorist Hannah Arendt, "what made France fall was the fact that she had no more true Dreyfusards, no one who believed that democracy and freedom, equality and justice could any longer be defended or realized under the republic."⁷⁹ Anti-semitism and social inequality grew and the waning intellectual influence was not there to counteract it. In a return to the ideologies of Drumont, Barres, and other anti-semites, Vichy France deported at least 75,670 Jews from France to concentration and extermination camps and did not actively rebel against the Nazis' anti-semitic policies.⁸⁰

Naturally, the question of whether the intellectuals generated any enduring change arises. In resolving the Dreyfus case, Zola and the intellectuals achieved social justice and a triumph of individual rights. Additionally, the enactment of the 1905 French law on the Separation of the Churches and State solidified a move to secularism, and parliamentary democracy was strengthened. Still, today, France struggles to prohibit discrimination based on ethnicity or religion, underscoring why the Dreyfus Affair still holds relevance. Perhaps the legacy of the Dreyfus Affair is the hope that reason, activism, and justice can prevail. Entrenched opposition, prejudice, and a biased press can be undermined if there is a significant investment in public discourse, and with the action of courageous activists such as Zola and the intellectuals.

⁷⁵ Conner, *The Dreyfus Affair and the Rise of the French Public Intellectual*, 220.

⁷⁶ Conner, *The Dreyfus Affair and the Rise of the French Public Intellectual*, 195.

⁷⁷ Conner, *The Dreyfus Affair and the Rise of the French Public Intellectual*, 214.

⁷⁸ Conner, *The Dreyfus Affair and the Rise of the French Public Intellectual*, 212.

⁷⁹ Hannah Arendt, *The Origins of Totalitarianism* (Boston & New York: Mariner Books, 1968), 93.

⁸⁰ "The JUST Act Report: France," U.S. Department of State, accessed May 7, 2023, <https://www.state.gov/reports/just-act-report-to-congress/france/>.

Works Cited

- Burns, Michael. *France and the Dreyfus Affair: A Documentary History*. Boston & New York: Bedford / St. Martin's, 1999.
- McClelland, J.S. *The French Right*. ed. George Steiner. New York: Harper & Row, Publishers, 1970.
- Kleeblatt, Norman L. *The Dreyfus Affair: Art, Truth, and Justice*. Berkeley: University of California Press, 1987.
- Snyder, Louis L. *The Dreyfus Case, A Documentary History*. New Brunswick: Rutgers University Press, 1973.
- Whyte, George. *The Dreyfus Affair: A Chronological History*. Hampshire & New York: Palgrave Macmillan, 2005.
- Zola, Emile. *The Dreyfus Affair; 'J'accuse' and Other Writings*. ed. Alain Pages. Translated by Eleanor Levieux. New Haven & London: Yale University Press, 1996.
- Arendt, Hannah. *The Origins of Totalitarianism*. Boston & New York: Mariner Books, 1968.
- Begley, Louis. *Why the Dreyfus Affair Matters*. New Haven & London: Yale University Press, 2009.
- Conner, Tom. *The Dreyfus Affair and the Rise of the French Public Intellectual*. Jefferson, North Carolina: McFarland & Company, Inc., 2014.
- Epstein, Joseph. "Why Captain Dreyfus?" *The Weekly Standard*, May 21, 2012.
<https://www.washingtonexaminer.com/weekly-standard/why-captain-dreyfus>.
- Rimoch, David. "The Affair or the State: Intellectuals, the Press, and the Dreyfus Affair." Senior Thesis, University of Pennsylvania, April 2008.
https://repository.upenn.edu/hist_honors/8/
- U.S. Department of State. "The JUST Act Report: France." Accessed May 7, 2023.
<https://www.state.gov/reports/just-act-report-to-congress/france/>.