|CURIEUX|

ACADEMIC JOURNAL

April Issue

Part 1 Issue 49

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: Discovery and Translational Applications of PARP Inhibitors for Cancer Therapeutics By Sophia Stavrakos

Page 27: Leading from Afar: Examining Remote-First Startup Leadership Practices and Their Impact on Company Culture and Performance By Nikhill Prerepa

Page 41: Navigating the Last Gasp: Success Factors in Business Adaptation By Zeen Chen

Page 61: The Theory of Natural Selection and How it Affects Humans By Fiona Allendorf

Page 62: The Impact of Music on Plants: A Review By Quinne Fang

Page 71: Working Memory, Depression, and Social Media: A Literature Review By Shivm Manrao

Page 83: The Efficacies of Music Therapy on Alzheimer's Disease Patients: A Review By Sofia Alexandrakis

Page 101: Black Holes By Arsh Garg

Page 110: What Happens When You Die? Life Beyond The Final Goodbye By Katherine Sheng

Page 117:The Determinant of Foreign Direct Investment: An Empirical Evidence from OECD Countries By Parth Sharma

Page 129: Syntactic Shifts in Contemporary American English: A Corpus-Based Study By Arjun Raj

Page 140: UAPTA - Urban Air Pollution Treatment Apparatus By Júlia Silva Djahjah

Page 157: Analyzing the Failure Modes of Neuromodulation By Anirudh Nalluri

Page 171: The Psychology and Effects of Workplace Ageism By Abhimanyu Shridhar

Page 184: Neural Mechanisms of Trauma and Anxiety Disorders: Insights from DSM-5 Classifications for Advancing Treatment Strategies By Euhyun Brooke Jahng

Page 195: miR-223 Regulating Neutrophils Via the NF-kB Pathway in Chronic Obstructive Pulmonary Disease By Jaelyn Chow

Page 209:The Industrial Revolution: From Artisanal Uniqueness and Craftsmanship to Reproducibility and Affordability By Taryn Hwang

Page 213: The Double Duty Defender: Vitamin D's Multifaceted Role in Skin Cancer Progression and Prevention By Isabelle Yang

Page 233: How Does Autopilot Control A Plane and How Can We Further Improve This Technological Concept? By Ethan Tang

Page 246: Artistic Currents and Commercial Tides: South-East Asian Art Expansion Through Trans-Oceanic Trade Networks in the First Millennium CE By Riya Rao

Page 259: Long-Term Exposure to Trauma: ACES By Saxon Mercado

Page 264: The Epigenetic, Environmental, and Microbial Intersection of Crohn's Disease By Sammy Naser

Page 275: Global Income Inequality and Upward Mobility By Emre Elgormus

Page 289: Treatment for Veterans with PTSD By Josephine connell

Page 293: Exploring Sustainability in Cosmetics for Policy Making and Marketing Communication: A Case Study on Sephora's approach and practices By Ege Yazlar

Page 311: The Politicization of Ultras in Soccer By Dylan Shah

Page 315: Peer Influence on Spending Habits: The Role of Group Chats on Social Media Among Bangladeshi Teenagers By Zahia Zakaria

Page 319: Understanding the Neuroscience of Play - the Impact of Play on Brain Development By Anahita Gupta

Page 331: Neurodiverse Activity Recognition Using Wearable Technology and Machine Learning By Aashna Parsa

Page 342: Impact of Dietary Management in Estrogen Regulation on the Prevention of Relative Energy Deficiency in Sport (REDs) By Katherine Mullane

Page 349: Unraveling Nemaline Myopathy: Genetic Mutations, Pathophysiology, and Emerging Therapeutic Approaches By Ishaan Murth

Page 357: Music Therapy: An Alternative Treatment for Tourette's Syndrome By Ellie Chen

Page 365: Black Holes in Our Universe By Arjun Konijeti

Page 379: Analyzing the Benefits and Drawbacks of the Effects of TikTok Virality on the Music Industry By Clara Xi

Page 387: Overcoming the Barriers to Scalability: Recent Advancements and the Current State of Quantum Computing By Satyajeeth Suresh Kannan

Page 401: Developing a Posture Detection System Using a Neural Network By Matthew Koo and Josh Park

Page 409: Influence of Hypothetical Planet 9 on Dwarf Planet Pluto By Baron Li, Sophia Kressy

Page 416: An Overview of Congenital Mutations and Chromosomal Abnormalities, Effect on Cancer and Quality of Life By Nirukta Balaji

Page 426: Proposing a CAR T-cell Therapy of Philadelphia Chromosome-Positive B-Cell Acute Lymphoblastic Leukemia (Ph+ B-ALL) via CD22 Apoptotic Pathway By Nehal R. Revuri

Page 442: "Law, Language, and Legitimacy: The Importance of Speaking Legal" By Aytaj Gurbanova

Page 450: Preliminary Investigation On the Visual Status of Older Adults in Community Care Center By Sophie Wei Li

Page 459: PTSD and Memory: A Comparative Study in a Child and Adult By Juliet Fox

Page 470: Innovations in Carbon Capture Technologies to Mitigate Climate Change Impacts from CO2 Emissions By Yuna Kim

Page 483: Asteroid Mining Robotics: A Novel Approach Using Adaptive Drill Bits for Optimal Resource Extraction By Anant Tyagi

Page 492: Dark Matter and the WIMPs as Its Prime Candidates By Yuvraj Arora

Page 515: Tamil-English Code-Mixed POS Tagging By Mohnish Sivakumar

Page 525: How Can Policymakers Guide Online Retailers to Nudge Consumption Towards Sustainable Products? By Eda Baser

Page 536: Using AI Models to Help Banks Save Over 170 Billion Dollars By Prediction Loan Defaults By Prashanth Venkatesh

Pag 551: Empowering Tomorrow's Voters: Evaluating How Access to Pre-registration Affects Youth Voter Turnout By Anna Gagne

Page 557: How Black Holes Kill People By Sunny Shin

Page 566: Biomedical Imaging Use Metastatic Cancer Research in Underdeveloped Regions By Jessica Wang and Ti'Aair Riggins

Discovery and Translational Applications of PARP Inhibitors for Cancer Therapeutics
By Sophia Stavrakos

Abstract

Cancer is a leading cause of death worldwide and is driven by mutations in cellular DNA. Uncontrolled cell division of these mutated cells ultimately leads to the formation of tumors. PARP inhibitors disrupt cellular DNA repair mechanisms, causing the error-prone non-homologous end-joining DNA repair mechanism to be used rather than the more accurate homologous recombination. This DNA repair dysregulation in cells with mutated BRCA1/2 genes results in apoptosis, moderating tumor progression and prolonging patient survival. This ultimately answers how PARP inhibitors are used as an immunotherapy for cancer treatment. This is presented through synthesizing primary and secondary research papers that have displayed such information. The research has shown cellular DNA repair mechanisms, the role of the PARP proteins and inhibitors, and future advancements that may lead to better cancer treatments. The data ultimately shows positive results in tumor progression but also seeks ways to combine PARP inhibitors with other treatments such as radiotherapy, targeted drugs, and other immunotherapies to augment its effects.

Introduction

Cancer is a leading cause of death worldwide without a cure. It can evolve for many reasons, including damage to a cell's DNA. This damage leads to the uncontrolled replication of these mutated cells, which leads to the formation of tumors. Tumor formation can begin with any cell, but cancerous cells can spread throughout the body through metastasis. Cancer is a highly adaptable disease and difficult to cure.

Consequently, most cancer therapies aim to slow tumor progression rather than attempt a cure. Currently, there are many types of therapies for cancer, including immunotherapy, chemotherapy, radiation, and surgery [1]. Regardless of the treatment utilized, when cancer metastasizes, the survival rates for patients significantly decrease. The survival rate for metastasized breast cancer is 29%, ovarian cancer is 18%, and prostate cancer is 28% [2-4]. However, one promising way to suppress tumor growth is using PARP inhibitors.

This paper reviews the development of PARP inhibitors and their mechanism of action. Through DNA repair research, PARP1 was identified as a potential protein to inhibit, and scientists wondered if their inhibition in conjunction with BRCA1/2 deficiency identified in some cancer cells would create a lethal combination. While testing this theory, researchers have identified many DNA repair mechanisms, particularly one that uses the PARP1 protein. In the future exploration of the therapy, it is being tested to see if it can be combined with another therapy for non-BRCA1/2 cancers while still creating synthetic lethality. Olaparib, rucaparib, talazoparib, and niraparib are the only FDA-approved PARP inhibitors, and they have proven useful in treating cancer.

While PARP inhibitors have shown promise in treating BRCA1/2 mutation-based cancers, there are several limitations of the scope of this research. For instance, clinical trials have been performed on ovarian, breast, and prostate cancers, but the efficacy of PARP inhibitors on other types of cancers is unknown. In addition, resistance mechanisms, such as the restoration of homologous recombination and the reversal of BRCA1/2 mutations, are not fully understood. These limitations pose challenges for long-term therapeutic success via reversing these mechanisms. Also, the data in this paper is sourced from clinical trials with limited populations; therefore, discrepancies in the efficacy of PARP inhibitors and their side effects may be found when applied to broader populations. Finally, the interaction between PARP inhibitors and other cancer treatments, such as chemotherapy and radiotherapy, requires further exploration to optimize combination therapies.

To conduct this review, I utilized databases, predominantly PubMed and Google Scholar, to source primary and secondary research studies on PARP inhibitors' mechanisms and clinical applications. Keywords such as “PARP,” “DNA repair mechanisms,” “homologous recombination,” and “non-homologous end joining” were used to find relevant studies. The focus was on research published within the past two decades to ensure relevant information. Studies were cross-referenced to ensure consistency in findings. I used peer-reviewed papers and clinical trial data to guarantee the accuracy of the findings discussed.

DNA Repair Mechanisms

DNA breaks are common, and six primary pathways have been identified for DNA repair. Of the six, four are for single-stranded breaks, and two are for double-stranded breaks [5]. There are two cellular mechanisms for repairing double-stranded breaks: homologous recombination (HR) and non-homologous end joining (NHEJ) [6, 7]. NHEJ can result in insertions or deletions of nucleotides through ligating the broken DNA ends with little or no homology [6]. It is an error-prone mechanism that fixes the DNA without a reference to what the original DNA strand looked like. However, HR repairs double-stranded breaks by using an undamaged DNA strand as a template. To repair the break, this template is used on the sister chromatid or homologous chromosomes, allowing for a replica of the original DNA strand [7].

Homologous Recombination

Homologous recombination is utilized by a cell in the S and G2 phases of the cell cycle to repair double-stranded DNA breaks. To begin the repair process, the DNA goes through DNA end resection to create duplicates with three prime-ended single-stranded tails that can be hundreds of nucleotides long. Recombination proteins then mediate a process that results in one of the single-stranded tails interacting with the sister chromatid, which is used as a template for the original DNA sequence. The single strand invades the double strand at a region of homology, creating a region of heteroduplex that includes the single strand attached to its complementary strand. The double strand opens up and creates a displacement loop (d-loop).

Once the single strand is attached to its complementary strand, the tail acts as a primer for DNA synthesis, with the complementary strand being the template to elongate the original strand. With this, a small bubble that contains the template, part of the new DNA, and the displaced strand forms and stays alongside the template. This bubble continues this way until new DNA complementary to the other tail is created, and then their bubble dissociates.

The newly created single strand is then captured by the three prime end tails of the broken DNA. This is then used as a primer for new DNA replication for the complementary strand to be created. If any gaps remain, they are filled by replication and ligation. This DNA repair mechanism is accurate because it uses a template to create the new DNA strand, and it is preferred to repair double-stranded breaks.

The BRCA1 and BRCA2 genes play a role in HR because BRCA1 helps promote the end resection of the broken DNA strands to become three prime-ended single-stranded tails. It does this by irritating 53BP1, which is a resection inhibitor. This process prevents 53BP1 from working so that the ends can be resected. BRCA2 then promotes RAD51 recombinase so that a RAD51-ssDNA filament can be formed from the resection product. This interaction prevents the three prime-ended single-stranded ends from being repaired by the single-strand annealing pathway because it leads to deletions between homologous repeats, which leads to damaged DNA.

It is important to note the correlation of BRCA1/2 genes with HR pathways because, without these genes, the pathway will not be favored over NHEJ. This later leads us to the concept of synthetic lethality in cancer cells, which will only work with the combination of the PARP inhibitor and BRCA1/2 mutations [7].

Non-Homologous End Joining

Non-homologous end joining is an error-prone DNA repair mechanism favored with the use of a PARP inhibitor. If DNA damage occurs in phase G2 of the cell cycle, then homologous recombination is possible because a copy of the DNA has already been made, so it can be used as a template. However, if DNA damage occurs before this is possible (like in G1) or if HR is disabled because of BRCA1/2 mutations, then NHEJ will happen because it does not require a DNA template.

In the case of double-stranded breaks, the DNA strands would have lost nucleotides in a process known as resection. Ku70 + Ku80 and DNA-Pkcs are three proteins called together to repair the broken DNA. These then call the Artemis, an endonuclease, which is phosphorylated by the DNA-Pkcs and trims the overhanging nucleotides of the overhanging DNA strands. Subsequently, Ligase IV, XLG, XRCC4, and more proteins are called to fix the double-stranded DNA break. Upon completion of the fix, all the molecules dissociate from the DNA.

However, because this repair mechanism does not use a template to copy the original DNA exactly, it is error-prone. And, because Artemis trims the overhanging ends, insertions or deletions can occur because it does not take the broken nucleotides into account, it just gets rid of them [6].

Base Excision Repair

One common single-stranded repair mechanism is base excision repair (BER). Within the BER process, a damaged DNA nucleotide is identified and is removed by a DNA glycosylase. A PARP protein then attaches to the open site and calls forth the necessary proteins to repair the damage. Inhibition of PARP would stop the BER repair mechanism, and the single-stranded DNA break will eventually result in a double-stranded break [8].

Mechanisms of PARP Proteins And Inhibitors

There are numerous PARP proteins, with PARP1 being the most common of the enzyme family. PARP1 uses donor NAD+ molecules during BER to bring negatively charged PAR groups onto their target DNA-binding proteins. Because of these PAR chains, PARPs can control cellular processes, including DNA repair, transcriptional regulation, RNA interface, etc [9]. As part of the BER process, PARP1 detects single-strand breaks by binding to the DNA next to the damage. It then synthesizes PAR chains on target proteins (PARylation), which leads to the recruitment of additional factors that complete the DNA repair process. When the repair is complete, autoparylation (part of PARP1) releases itself from the repaired DNA [10, 11].

Because of this process, PARP inhibitors can be used to induce apoptosis. If the PARP protein is inhibited from repairing a single-stranded break in BER by attaching to the damaged DNA, the single-stranded break will turn into a double-stranded break. From there, the cell must utilize either HR or NHEJ to fix the double-stranded breaks. In HR-deficient cancer cells, the addition of a PARP inhibitor would only allow for the NHEJ repair mechanism, resulting in a cell that accumulates mutations due to the error-prone nature of NHEJ repair. The HR-deficient cells would preferentially go through apoptosis as those mutations accumulate, resulting in slowing down cancer progression. This would prolong the development of tumor cells by killing the cancer cells instead of allowing them to replicate [12].

Synthetic Lethality

This leads us to the concept of synthetic lethality, which is when two conditions that are not lethal by themselves together become lethal when combined. An example of synthetic lethality is the use of PARP inhibitors, which inhibit BER in BRCA1/2 deficient cells, which are HR deficient. Separately, each of these factors does not cause apoptosis, but the combination of the PARP inhibitor and the HR-deficient cells is lethal because of the increased mutations leading to apoptosis [13, 14]. Additionally, chemotherapy can be used to induce synthetic lethality in tumor cells without an HR deficiency. Certain chemotherapy can damage a cell's ability to repair DNA damage, so when paired with a PARP inhibitor, it can lead to apoptosis [15]. Therefore, chemotherapy can have a similar effect as an HR deficiency with the use of PARP inhibitors.

Discovery of PARP Inhibitors with BRCA1/2 Mutants

It had been established that BRCA1 and BRCA2 play a role in homologous recombination of double-stranded DNA breaks, but it is not known if the BRCA1/2 mutations would enhance the effectiveness of PARP inhibitors. In 2005, a team of scientists hypothesized that by inhibiting PARP1 in BRCA-mutated cells, HR would be impaired, leading to damaged DNA and, ultimately, cell death. To block the expression of PARP1, they injected siRNA into embryonic stem cells in mice with BRCA1 and BRCA2 mutations. The same thing was done with wild-type cells from the same mouse, and the results showed that more BRCA1/2 mutated cells died than those without. Based on these results, they continued their experiment with PARP inhibitors.

A variety of cell types, including human breast cancer cells, embryonic stem cells, and Chinese hamster ovary cells, with wild-type BRCA genes and different variations of BRCA1/2 deficiencies. The results showed that across all the cells and PARP inhibitors that were tested, the BRCA1/2 mutated cells had an increased sensitivity to the drugs. This led to the hypothesis that PARP1 could be involved in the repair of single-stranded DNA breaks during BER and that inhibiting PARP would cause these single-stranded breaks to accumulate. If the single-stranded breaks become a replication fork, then a double-stranded break could arise. HR would typically fix this, but with mutated BRCA1/2 genes, the replication fork would collapse, and chromatid breaks would occur. The team hypothesized that some other (unknown at the time) error-prone repair mechanism would cause chromatid breaks and other damage, which would ultimately lead to apoptosis.

This difference in the reaction from wild-type versus BRCA1/2 mutated cells lead researchers to see a potential for PARP inhibitors to help in the treatment of cancer. It proved that PARP inhibitors are not toxic to wild-type cells, so the therapy of using PARP inhibitors could be more specific than other options, such as chemotherapy, since it only impacts the tumor BRCA mutated tumor cells. In the experiment, mice given PARP inhibitors remained healthy. Therefore, the same was expected for humans, and more importantly, the experiment proved that using the PARP inhibitors on BRCA1/2 deficient cells would result in apoptosis of those cells [16].

Initial Validation of PARP Inhibitors with BRCA Mutations

It was originally hypothesized that cells perform apoptosis after being exposed to PARP inhibitors due to the accumulation of single-stranded breaks, which eventually turned to double-stranded breaks. Double-stranded breaks are repaired by HR in cells with the BRCA genes, but in BRCA-deficient cells, these breaks would accumulate, and the cell would go through apoptosis. However, while this theory does highlight that PARP1 and HR are important in maintaining genomic stability, it does not mention the critical role of NHEJ in the synthetic lethality in HR-deficient cells when exposed to PARP inhibitors.

According to the theory, disabling BER by a method different from that of a PARP inhibitor would result in apoptosis in BRCA-deficient cells. To test this, an experiment that knocked out XRCC1, an essential protein in BER, in ovarian cancer cells (PEO1 and PEO4) was performed. These cells had different amounts of BRCA2 expression but came from the same

patient. When PARP1 was disabled in these cells, the survival rate of the BRCA2-deficient cells decreased while the survival rate of the BRCA2-expressing cells increased, as expected. However, when the XRCC1 gene was disabled, the survival rate of both cell lines remained unchanged, leading to the theory that PARP1 maintains genomic stability in a different way.

These results led to a new hypothesis, one that predicted that the inhibition of HR and PARP1 would deregulate NHEJ, which could lead to the conclusion that PARP inhibitors lead to increased activity of DNA-PK, NHEJ, and genomic instability because this DNA repair mechanism is error-prone. Therefore, if NHEJ was deactivated via some other method, the effect of the PARP inhibitor should be lessened.

To test this theory, the same researchers compared PEO4 cells (BRCA positive) and PEO1 cells (BRCA2 deficient) treated with the PARP inhibitor ABT-888 for a variety of proteins. As a control, they tested for Hsp90, which was there for both cell lines. They then tested for the presence of pADPr to see how the presence of PARP was associated with the presence of other proteins. Then, they tested for DNA-PKcs proteins, which play a role in NHEJ, and it was shown that these proteins had increased activity in cells that contained the PARP inhibitor. They also tested for H2AX p-S139, which is utilized by NHEJ to repair a double-stranded DNA break. The tests showed an increased expression of H2AX p-S139 as well.

When PARP inhibitors were added to the cells with BRCA1/2 mutations, an increased expression of DNA-PKcs and H2AX p-S139 led the researchers to conclude that NHEJ was utilized by the cells to restore the double-stranded DNA breaks. Furthermore, since NHEJ is error-prone, the increased mutations in the DNA over time lead to apoptosis [17].

Clinical Trials of PARP Inhibitors

Olaparib, rucaparib, talazoparib, and niraparib are the only FDA-approved PARP inhibitors. Each of these drugs has been approved for demonstrating the ability to slow down the progression of different types of tumors. More drugs are being tested and awaiting approval, and they are hoping to have even better results with fewer side effects.

Clinical Trials in Ovarian Cancer

Olaparib (AZD 2281, KU-0059436) is an oral PARP inhibitor drug that induces synthetic lethality in tumor cells with BRCA1/2 mutations. A phase 2 trial of olaparib included 54 women aged 18 or older with BRCA1/2 deficient ovarian cancer, where each woman had previously received a median of three chemotherapy regimens. The study was split up into two cohorts, each with 27 women, with the first cohort receiving 400 mg twice a day and the second cohort receiving 100 mg twice a day. In the first cohort, the ORR was 11 of 27 patients (41%), with a 95% CI of 25-59%. The most common side effects were fatigue (56%), nausea (56%), vomiting (22%), and anemia (15%). In the second cohort, the ORR was 6 of 27 patients (22%), with a 95% Cl of 11-41%. The most common side effects were fatigue (30%) and nausea (41%) [18]. This study shows us that olaparib was effective in the treatment of BRCA1/2 deficient ovarian

cancer. However, it also demonstrated that the higher the dose, the more effective the drug, but the higher the probability of side effects.

In the SOLO1/GOG 3004 trial, olaparib was used as a maintenance therapy, meaning the treatment of their cancer was successful, and olaparib was used as maintenance to keep the cancer away for patients with BRCA1/2 mutated advanced ovarian cancer. This trial provides the survival rates for the patients after a seven-year follow-up appointment. Patients receive chemotherapy and then take olaparib (260 people) or a placebo (131 people). The patients with olaparib received a median treatment of 24.6 months, while those with the placebo received a median treatment of 13.9 months, and the median follow-up times were 88.9 and 87.4 months. The hazard ratio for this initial trial was 0.55. After seven years, for the patients who had first received chemotherapy, 67.0% of the patients who also received olaparib were alive, while only 46.5% who received the placebo were alive. In both groups, the side effects of myelodysplastic syndrome or acute myeloid leukemia were uncommon [19]. This trial showed promise in PARP inhibitor's long-term benefits and that they have increased efficacy when paired with chemotherapy.

The ATHENA trial tests the efficacy of rucaparib as a maintenance treatment for patients with stage III-IV BRCA1/2 mutated, HR deficiency, or high-risk clinical characteristics of ovarian cancer. These patients were responding to chemotherapy and going through surgical extraction of tumors. Of the patients, 427 were given 600 mg twice a day of the drug, while 111 were given the placebo. Of those who received rucaparib, the median time without the progression of tumors was 28.7 months (CI of 95%), while only 11.3 months (CI of 95%) for those who received the placebo and were HR deficient. However, for those without HR deficiency and with olaparib, it was only 12.1 months and 9.1 months with the placebo. For the side effects, 28.7% of those on rucaparib developed anemia, while 0% with the placebo. And 14.6% of patients on rucaparib developed neutropenia, while 0.9% with the placebo [20]. This trial further proves the efficacy of PARP inhibitors when used on BRCA1/2 mutated, HR deficiency, or high-risk clinical characteristics of ovarian cancer. There is also promising data for the use of PARP inhibitors on breast cancer.

Clinical Trials in Breast Cancer

In addition to ovarian cancer, PARP inhibitors have also been proven to treat breast cancer. In a phase 2 trial for metastatic triple-negative breast cancer, 451 patients were given atezolizumab (the PARP inhibitor) and nab-paclitaxel, and another 451 patients were given a placebo and nab-paclitaxel. Both groups of patients continued this course of treatment until the cancer stopped progression or until too many toxic side effects happened. For those who received atezolizumab and nab-paclitaxel, the median time without progression was 7.2 months, while without atezolizumab, it was only 5.5 months (CI of 95%). For patients with PD-L1–positive tumors, the median time without progression for those who received atezolizumab was 7.5 months, and for those with the placebo, it was 5.0 months. Additionally, the median overall survival rate with atezolizumab was 21.3 months, and only 17.6 months without it (CI of 95%).

And, for patients with PD-L1–positive tumors, the median overall survival rate for those who received atezolizumab was 25.0 months, and for those with the placebo, it was 15.5 months. No new side effects were discovered compared to other PARP inhibitors, but 15.9% of patients with atezolizumab and nab-paclitaxel experienced them, and only 8.2% with the placebo and nab-paclitaxel had them [21]. This study shows how using PARP inhibitors in combination with another immunotherapy drug has been proven beneficial for the treatment of breast cancer.

Olaparib has been tried in HER2-negative metastatic breast cancer with germline BRCA1/2 mutations. OlympiAD is a phase 2 trial that sought to discover the efficacy of olaparib versus a chemotherapy treatment for patients with hormone receptor-positive or triple-negative breast cancer with BRCA mutations. The patients were 18 and older and had all previously received at least two chemotherapy lines as treatment. The ratio was 2:1 of patients receiving 300 mg twice a day of olaparib to chemotherapy (of the physician's choice) in 21-day cycles of either capecitabine, vinorelbine, or eribulin. Of the 302 patients, the median age was 44; 50% had triple negative breast cancer, 71% had prior chemo treatments, and 28% had prior platinum treatments. The results after 77% data maturity showed that the progression-free survival (PFS) of patients who took olaparib was significantly longer than patients who took TPC, 7.0 versus 4.2 months. The response rate was 59.9% in olaparib and 28.8% in TPC. Grade 3 or higher adverse effects were less common in olaparib, only 36.6% versus 50.5% who took TPC. Because of these adverse effects, 4.9% who took olaparib and 7.7% who took TPC had to stop the trial [22]. This trial proves that PARP inhibitors were more successful in the treatment of breast cancer compared to chemotherapy.

An open-label phase 3 trial tested talazoparib (the PARP inhibitor) compared to a physician’s choice of single-agent therapy, including capecitabine, eribulin, gemcitabine, or vinorelbine all in 21-day intervals. The 431 patients in the trial all had advanced breast cancer and BRCA1/2 mutations, and 287 were given talazoparib while 144 were given one of the other therapies. In the group with talazoparib, the median progression-free survival was higher at 8.6 months compared to 5.6 months (Cl of 95%). The talazoparib group had a higher objective response rate, which was 62.6%, compared to the 27.2% of the standard therapies group. Additionally, 32% of patients with talazoparib received nonhematologic grade 3 adverse effects, while 38% of patients who received standard therapies did. However, grade 3-4 adverse effects occurred in 55% of the talazoparib patients and 38% of the standard therapy patients, with the most side effects being anemia [23]. These trials all showed that PARP inhibitors can also be used for the treatment of breast cancer, and they have also been used for prostate cancer.

Clinical Trials in Prostate Cancer

In addition to ovarian cancer, PARP inhibitors have also been proven to treat prostate cancer. Olaparib was the first PARP inhibitor to be used in clinical trials with prostate cancer in a phase two trial that was conducted with patients who had metastatic, castration-resistant prostate cancer (mCRPC). There were 50 patients in the trial, and they were each given 40 mg of the drug twice a day. All of the patients had prior treatments, including abiraterone, enzalutamide, and/or

cabazitaxel. The result showed that 16 of the 49 patients had a response (33%), with a 95% CI of 20-48, and 12 of the parties received the treatment for over six months. Within the group, 16 of the 49 patients who had been tested (33%) had homozygous deletions, deleterious mutations, or both in their DNA repair genes, including mutations in BRCA1/2, ATM, Fanconi's anemia genes, or CHEK2. Of the patients with these mutations, 88% had a response to the PARP inhibitor, including all of the patients (7) with BRCA1/2 mutations. The most common side effects were anemia (20%) and fatigue (12%) [24]. These results show that olaparib is most effective when used in cancers that have a mutation in DNA repair genes.

Rucaparib has been tested in mCRPC during an ongoing phase 2 study called TRITON2. In this study, patients were selected to have mCRPC and a mutation in BRCA1, BRCA2, ATM, CDK12, or other genes that affect DNA repair mechanisms. In addition, patients must have undergone 1-2 lines of androgen receptor–directed therapy and 1 line of taxane-based chemotherapy for their tumors. Starting on February 15, 2019, 126 patients receiving 600 mg daily of rucaparib had at least 16 weeks of follow-ups. Of these patients, the mutations were 62 with BRCA2, 7 with BRCA1, 41 with ATM, 14 with CDK12, and 12 with other DDR mutations. The ORR for BRCA1/2 mutations was 56.5% with somatic alterations and 40.0% with germline alterations. 3 ATM patients had at least a 30% reduction in target lesion diameter, and 3 ATM and 2 CDK12 patients had at least a 50% decrease in PSA level. However, for ATM and CDK12 patients, only 1 ATM and 1 CDK12 patient had a prostate-specific antigen (PSA) response, meaning the cancer had evolved. For PSA progression, in BRCA patients, it took 5.7-7.5 months; in ATM patients, it took 2.8-4.6 months; and for CDK12 patients, it took 2.8-4.6 months [25]. We can see that rucaparib was most effective with BRCA1/2 mutated tumors and had barely any effect on the other DNA repair genes tumors.

GALAHAD is a continuous phase 2 trial of niraparib for patients with mCRPC and a DNA repair defect (DRD) who have already been through taxane and androgen receptor-targeted therapy. Through a plasma or tissue-based test, patients were found to have biallelic DRDs in BRCA1/2 (BRCA), ATM, FANCA, PALB2, CHEK2, BRIP1, or HDAC2. On May 23, 2019, 165 patients were given 300 mg daily of niraparib, with 81 having DRD (46 for BRCA1/2), 51/81 having a measurable disease at baseline (29 for BRCA1/2), and 47% of the tumors had metastasized. The median follow-up for BRCA1/2 was 7.3 months, and non-BRCA was 6.4 months. With BRCA1/2 mutationstrialsORR was 41%, and the CRR was 63%. The duration of the ORR was 3.5-9.2 months, and 7/12 responses were ongoing for patients without BRCA1/2 mutations. For BRCA1/2, the median PFS was 9.2 months, and the OS was 1,2.6 months. In non-BRCA cancers, the ORR was 2/22, with both patients having the FANCA gene, the duration of each being 3.8 and 6.5 months, and the CRR was 17%. The side effects included anemia (29%), thrombocytopenia (15%), and neutropenia (7%) [26]. These clinical trials of ovarian, breast, and prostate cancer have demonstrated the efficacy of PARP inhibitors.

Immunogenic Effects of PARP Inhibitors

In addition to affecting DNA repair mechanisms, it has been shown that PARP inhibitors can also affect the immune system positively towards the treatment of cancer. Understanding these effects can help lead to better and more specialized treatments.

DNA Damage Response and T-Cell Recruitment

PARP inhibition results in DNA damage to the cell; it usually leads to increased levels of cytosolic DNA, which is DNA that is found in the cytoplasm of a cell [27]. cGAs pathway detects the cytosolic DNA and produces secondary messenger cyclic GMP, which then activates STING. TBK1 and IFI16 are then recruited, which activates nuclear translocation of IRF3 and NF-kB and their transcription factors, as well as other genes that create an immune response. Examples are type 1 interferons and T-cell-recruiting chemokines, which lead to increased expression of tumor-infiltrating T-cells. However, it is unknown what the phenotypes of the t-cells are [28-33].

T-Cell Differentiation

The nuclear factor of activated T-cells (NFAT) regulates CD4 T-cell differentiation through differential gene expression [34]. PARP1 modulates NFAT activity by binding and PARylating it, which regulates NFAT's nuclear import, export, and DNA binding ability [35, 36]. Because of PARP1's role in t-cell differentiation, inhibition of PARP reduces the expression of NFAT-reliant cytokines. IL-2 and IL-4 are impacted, which leads to effects in the later t-cell differentiation process, such as a bias towards a specific phenotype [37, 38-40]. It was commonly found that a Th1 phenotype was preferred against a Th2 phenotype with PARP inhibition, but there are conflicting models [35-37]. Olaparib treatment for airway inflammation resulted in an increased expression of Th1, but the use of treatment for inflammatory arthritis showed reduced expression of TH1 [39, 41].

Decrease in T-Cell Exhaustion

Exhausted T-cells in cancer have high levels of immune checkpoint receptors like PD-1, Tim-3, and Lag-3 [42]. However, the treatment of PARP inhibitors results in the downregulation of immune checkpoint receptors like PD-1, Tim-3, and Lag-3 on CD8 T-cells, suggesting a reduced propensity for T-cell inhibition and exhaustion [43].

Enhanced Antigen Presentation

Tumoral DNA can also activate the cGAS/STING pathway in dendritic cells, improving their antigen-presenting ability. This was also replicated with Olaparib on BRCA1 deficient triple-negative breast cancer cells, which showed increased antigen presentation and recruitment of CD8 t-cells [43].

Sensitivity to NK Cells

PARP inhibitors increase tumor cell sensitivity to natural killer (NK) cell-mediated apoptosis via TRAIL (TNF-related apoptosis-inducing ligand). In tumor cells, HMGB1 leaks into the cytoplasm and promotes cell survival pathways. One of these pathways is to stimulate autophagy, which aggressively reduces degrading damage to organelles or proteins that would otherwise trigger apoptosis. By PARPi suppressing the HMGB1 pathways, it reverses the resistance to apoptosis [44].

Limitations to PARP Inhibitors

Even though PARP inhibitors have shown promise in the treatment of cancer, it has been shown that there are limitations to the drugs, especially when it comes to cells building up resistance to them.

Resistance to PARP Inhibitors

Although PARP inhibitors have shown success in prolonging the advancement of tumors, resistance to PARP has come up as an issue. However, rather than just one resistance mechanism, tumors may have acquired several resistance mechanisms.

Drug Efflux

In order for PARP inhibitors to kill tumor cells, they must enter the cell. With this, an increase in drug efflux may develop, which would lead to resistance because the drug is getting removed from the cell. MDR1 is a multidrug resistance protein that is encoded by ATP binding cassette subfamily B member 1 (ABCB1), which has been shown to be involved in the efflux of many chemotherapy drugs. It has been found that up-regulation of ABCB1 encoding p-glycoprotein efflux pumps as a result of using olaparib for a long period of time leads to drug resistance. However, by inhibiting ABCB1, it can help reverse the resistance [45].

Reversal of BRCA1/2 Mutations

PARP resistance can also come from the genomic reversal of BRCA1/2 mutations. Through somatic reversion mutation, those who have BRCA1/2 mutations may recover their function. This would lead to the resistance of PARP inhibitors, as they only work on cells with those mutations. Throughout many studies, 23,375 patients with a variety of cancer types were analyzed in regard to the likelihood of pathogenic/likely pathogenic germline HR mutated genes [46-48]. It was found that some of the mutations had been reversed after chemotherapy and/or PARP inhibitor resistance, including 3.8% of BRCA1, 3.5% of BRCA1, and 2.0% of PAB2 [48]. In ovarian and breast cancer, the occurrence of mutations that restore BRCA1/2 function is as high as 40% and 50%, respectively [49-51]. It has been shown that mutated DNA repair pathways that include BRCA2 often lead to reversion of the genes because of the accumulation of deletions months after the DNA sequences that underwent the DNA repair mechanisms [52].

Restoration of Homologous Recombination

Resistance to PARP inhibitors can be a result of HR function being restored. It was found that loss of TP53BP1 or the amplification of MRE11A restored HR function through increasing DNA terminal excision [51]. This is done through genomic changes because 53BP1 is a gene that is thought to control which double-stranded break mechanism gets used in cells [53]. Additionally, C20orf196, FAM35A, CTC-534A2.2, and REV7 are genes that together make up the protein Shielfin, which affects 53BP1/RIF1/MAD2L2, which, by limiting DSBs resection, could promote DSBs terminal connection [54, 55]. Studies have shown that CST-Pola-primase fill-in synthesis is a process in which the role of 53BP1 and Shielfin depend. As a result, it contributes to determining how effective PARP inhibitors are in BRCA1 mutated cells because HR is hindered by the antagonization of BRCA2/Rad51 loading in these cells [56-58]. It has also been found that DYNLL1 reduces DNA terminal excision activity in BRCA1 deficient cells, which would cause them to go through terminal excision, leading to HR. So, the reduction of DYNLL1 can lead to the resistance of PARP inhibitors [59].

Ethical Considerations

Although PARP inhibitors show promising results for the treatment of cancer, there are ethical considerations to keep in mind when deciding which drug to use and which patients should receive the treatment.

Off Target Effects

Through clinical trials and use of PARP inhibitors, many off target effects have been revealed. A network meta-analysis of PARP inhibitors including talazoparib, olaparib, rucaparib, conventional therapies, placebo, and PARP inhibitors + Al. Throughout all of these therapies, anemia, neutropenia, and thrombocytopenia were discovered to be blood related off target effects. Additionally, fatigue, anorexia, nausea, vomiting, and constipation were other common side effects shown throughout the therapies [60].

Safety Profile

A network meta-analysis of PARP inhibitors, including talazoparib, olaparib, rucaparib, conventional therapies, placebo, and PARP inhibitors + Al, reviewed the safety profile of these therapies. The safety profiles of these drugs were analyzed based on the amount of grade 3 or higher general and specific off-target effects. Based on their results, the ranking of lowest to the highest probability of getting these grade 3-5 effects goes: olaparib (57%), conventional therapy (50%), talazoparib (45%), rucaparib (75%), niraparib (77%), and PARP inhibitors + angiogenesis inhibitors (94%). These rankings are important for clinicians and researchers to take into account when determining which PARP inhibitors to use on patients [60].

Discovery of Predictive Biomarkers

However, discovering predictive biomarkers can help ease these ethical considerations of PARP inhibitors by showing patients if PARP inhibitors will be effective on their tumor cells.

For example, studies have found that RAD51 could potentially be a predictive biomarker [61-64]. Through immunohistochemistry, the RAD51 foci can be detected, and the detection of this gene has been associated with whether BRCA1/2 is present or not. Based on this, you could predict if a patient's cancer will have a response to PARP (or other DNA damage treatment) [65, 66]. A study showed that the downregulation of RAD51 sensitized cancer stem cells to PARP inhibitors, which prevented tumor growth in triple-negative breast cancer cells [67]. RAD51 was later tested against gene sequencing to see which was better at predicting resistance to PARP inhibitors and was proven more useful [68, 69].

Improvements and Future Advancements

In order for PARP inhibitors to be a more effective drug, there needs to be a way to reverse the resistance that has occurred in some patients. Luckily, promising ways of doing this have been discovered.

Combination of PARP Inhibitors with other Targeted Drugs

Studies have shown promising results when using PARP inhibitors with other targeted drugs because they have been shown to reduce the resistance to PARP inhibitors and/or improve the use of synthetic lethality. One example is the use of PARPi alongside mitogen-activated protein kinase (MAPK) kinase (MEK) inhibitors. Studies have shown that in RAS-mutated tumors, RAS-MEK-ERK signaling is upregulated. It was also found that PARP inhibitor-resistant cells have active RAS pathways and that the efficacy of PARPi and MEK inhibitors does not rely on BRCA1/2 or TP53 mutations [70]. In addition, MEK inhibitors can help reverse the resistance to PARP inhibits, which shows how the combination of the two therapies had positive results [71].

PARP, in Combination with Radiotherapy

Because of the way radiation causes DNA damage to cells, radiation has been proven to be a useful therapy in cancer treatment because of the way that it leads to apoptosis through SSBs and DDBs [72]. It has been shown that inhibiting PARP1 can increase the cytotoxicity of ionizing radiation because of how it inhibits NF-kB activation [73]. Moreover, using PARP inhibitors in combination with ionizing radiation has enhanced its effects in vitro in breast cancer, ovarian cancer, cholangiocarcinoma, and soft tissue sarcoma [74-78]. There have also been investigations into maximizing the efficacy of olaparib with another therapy in vitro. It was found that taking a low dosage of olaparib with a combination makes for more effective combination therapy. This is best shown in HR-deficient cells [79]. It was also shown that in tumor cells with an HR deficiency, p53 was a gene that made for an effective biomarker. This information could be useful for finding what patients would benefit more from combination therapy in comparison to using one or the other [80].

PARP in Combination with Immunotherapy

PARP inhibitors have also shown promise in combination with immuno-oncology (IO) therapy, which helps eliminate cancer cells rather than targeting the tumor directly by activating the immune system [81]. PARP inhibitors could improve the response to IO therapy by turning "cold" into "hot" tumors. PARP inhibitors also increase the expression of PDL1 and increase the DNA damage to the DNA repair pathways, leading to more neoantigens being generated [82]. In a clinical trial, Pembrolizumab was combined with niraparib and has shown its efficacy in advanced TNBC, ovarian cancer, and mCRPC. Pembrolizumab is an immunotherapy that inhibits PD-1, therefore blocking it from interacting with its ligands PD-L1/PD-L2, and niraparib is a PARP inhibitor. This combination received FDA approval for metastatic cancers, which sparked other treatment combinations to be tested [83-85].

Methods

To find the necessary research papers and data for this review, I looked on databases including PubMed and Google Scholar. When looking for sources on these databases, I used keywords including "PARP", "DNA Repair Mechanism", "Non-Homologous End Joining", and "Homologous Recombination". I also made sure that such sources were recent and gave information consistent with what I had learned from other sources.

The inclusion criteria for the selected papers were based on their relevance to the therapeutic use of PARP inhibitors in cancers with BRCA1/2 mutations, the mechanisms of DNA repair (homologous recombination and non-homologous end joining), and their role in clinical outcomes. I used peer-reviewed studies, clinical trials, and reviews from reputable journals. Clinical trial data was found from studies that focused on the efficacy of FDA-approved PARP inhibitors, such as olaparib, rucaparib, talazoparib, and niraparib, across different types of cancers, including ovarian, breast, and prostate cancers.

In addition to primary research articles, this review incorporated secondary sources to provide context on the mechanisms of PARP inhibitors and applications in cancer therapeutics, as well as DNA repiar mechanisms. All selected studies were verified to ensure that they were consistent in data. By synthesizing this information, this review presents a thorough analysis of the discovery, mechanism, limitations, and future directions for PARP inhibitors in cancer treatment.

Conclusion/Discussion

This review covered the many roles of PARP inhibitors in cancer therapy, particularly for BRCA1/2 mutated tumors. Through PARP inhibitors impairing single-stranded DNA break mechanisms in combination with an HR deficiency, synthetic lethality occurs, leading to cancer cells performing apoptosis and prolonging tumor growth. Clinical trials with drugs like olaparib, rucaparib, and niraparib have demonstrated promising results in ovarian, breast, and prostate cancers, especially in patients with BRCA1/2 mutations. These results show the efficacy of PARP inhibitors as a targeted therapy, which can better the treatment of cancer. Future research

should focus on overcoming resistance to PARP and PARP inhibitors in combination with other therapies to enhance their efficacy and allow for those without BRCA1/2 mutations to use the drug.

However, there are still limitations to the therapy. There have been challenges relating to drug resistance, including the restoration of homologous recombination pathways and BRCA1/2 reversion mutations. These mechanisms limit the long-term effectiveness of PARP inhibitors, so further research is necessary into overcoming resistance. Additionally, while current studies have shown efficacy in cancers with BRCA1/2 mutations, their impact on broader cancer types without this mutation is less well understood, limiting their application in diverse patient populations.

The review also highlights that PARP inhibitors have increased efficacy when combined with other therapies, including radiotherapy or immunotherapy. These combinations can enhance the efficacy of the inhibitors and also reduce the likelihood of resistance. So, future research should focus on overcoming resistance to PARP inhibitors and the use of PARP inhibitors in combination with other therapies to enhance their efficacy and allow for cancer types without BRCA1/2 mutations to use the drug.

In conclusion, while PARP inhibitors have shown positive effects for certain cancers, continued investigation into resistance mechanisms and potential combination therapies is necessary to expand their use and improve patient care. The development of biomarkers to predict responsiveness to PARP inhibitors will be crucial for personalizing treatment and enhancing therapeutic success.

Works Cited

1. National Institute of Cancer. Types of Cancer Treatment. August 23, 2023. https://www.cancer.gov/about-cancer/treatment/types
2. American Cancer Society. Survival Rates for Ovarian Cancer. March 1, 2023. https://www.cancer.org/cancer/types/ovarian-cancer/detection-diagnosis-staging/survival-rates.html
3. American Cancer Society. Survival Rates for Breast Cancer. January 27, 2024. https://www.cancer.org/cancer/types/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-survival-rates.html
4. John Hopkins Medicine. Prostate Cancer Prognosis. https://www.hopkinsmedicine.org/health/conditions-and-diseases/prostate-cancer/prostate-cancer-prognosis
5. Plummer R. Perspective on the pipeline of drugs being developed with modulation of DNA damage as a target. Clin Cancer Res 2010; 16: 4527–4531.
6. Benjamin M. Stinson and Joseph J. Loparo. Repair of DNA Double-Strand Breaks by the Non-homologous End Joining Pathway. Annu Rev Biochem. 2021 Jun 20; 90: 137–164. Published online 2021 Feb 8. doi: 10.1146/annurev-biochem-080320-110356
7. Andrei Kuzminov. Homologous Recombination—Experimental Systems, Analysis and Significance. EcoSal Plus. 2011 Dec; 4(2): 10.1128/ecosalplus.7.2.6. doi: 10.1128/ecosalplus.7.2.6.
8. Hans E. Krokan and Magnar Bjørås. Base Excision Repair. Cold Spring Harb Perspect Biol. 2013 Apr; 5(4): a012583. doi: 10.1101/cshperspect.a012583.
9. Hakme A, Wong HK, Dantzer F, et al.The expanding field of poly(ADP-ribosyl)ation reactions. ‘protein modifications: beyond the usual suspects’ review series. EMBO Rep 2008; 9: 1094–1100.
10. Langelier MF, Eisemann T, Riccio AA, et al. PARP family enzymes: regulation and catalysis of the poly(ADP-ribose) posttranslational modification. Curr Opin Struct Biol 2018; 53: 187–198.
11. Murai J, Huang SY, Renaud A, et al. Stereospecific PARP trapping by BMN 673 and comparison with olaparib and rucaparib. Mol Cancer Ther 2014; 13: 433–443.
12. Mateo J, Lord CJ, Serra V, et al. A decade of clinical development of PARP inhibitors in perspective. Ann Oncol 2019; 30: 1437–1447.
13. Bryant HE, Schultz N, Thomas HD, et al. Specific killing of BRCA2鄄deficient tumors with inhibitors of poly (ADP鄄ribose) polymerase [J]. Nature, 2005,434(7035):913-917.
14. Bryant HE, Petermann E, Schultz N, et al. PARP is activated at stalled forks to mediate Mre11鄄dependent replication restart and recombination [J]. EMBO J, 2009,28(17):2601-2615.
15. Tentori L, Graziani G. Chemopotentiation by PARP inhibitors in cancer therapy [J]. Pharmacol Res, 2005,52(1):25-33.

16. Hannah Farmer, Nuala McCabe, Christopher J. Lord, et al.Targeting the DNA repair defectin BRCA mutant cells as a therapeutic strategy. May 2005. Nature 434(7035): 917-21. DOI:10.1038/nature03445
17. Anand G. Patela, Jann N. Sarkariab, and Scott H. Kaufmanna. Nonhomologous end joining drives poly(ADP-ribose) polymerase (PARP) inhibitor lethality in homologous recombination-deficient cells. Edited by Philip C. Hanawalt, Stanford University, Stanford, CA, and approved January 20, 2011 (received for review September 13, 2010).
18. Andrew Tutt, Mark Robson, Judy E Garber, et al. Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and advanced breast cancer: a proof-of-concept trial. Lancet. 2010 Jul 24;376(9737):235-44. doi: 10.1016/S0140-6736(10)60892-6. Epub 2010 Jul 6.
19. Paul DiSilvestro, MD, Susana Banerjee, MD, PhD, Nicoletta Colombo, MD, PhD, et al. Overall Survival With Maintenance Olaparib at a 7-Year Follow-Up in Patients With Newly Diagnosed Advanced Ovarian Cancer and a BRCA Mutation: The SOLO1/GOG 3004 Trial. J Clin Oncol. 2023 Jan 20; 41(3): 609–617. Published online 2022 Sep 9. doi: 10.1200/JCO.22.01549.
20. Bradley J. Monk, MD, Christine Parkinson, MD, Myong Cheol Lim, MD, PhD, et al. A Randomized, Phase III Trial to Evaluate Rucaparib Monotherapy as Maintenance Treatment in Patients With Newly Diagnosed Ovarian Cancer (ATHENA–MONO/GOG-3020/ENGOT-ov45). Journal of Clinical Oncology. Volume 40, Number 34.
21. Peter Schmid, M.D., Ph.D., Sylvia Adams, M.D., Hope S. Rugo, M.D., et al. Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer. Published October 20, 2018. N Engl J Med 2018;379:2108-2121. DOI: 10.1056/NEJMoa1809615. VOL. 379 NO. 22
22. Mark E. Robson, Seock-Ah Im, Elżbieta Senkus, et al. OlympiAD: Phase III trial of olaparib monotherapy versus chemotherapy for patients (pts) with HER2-negative metastatic breast cancer (mBC) and a germline BRCA mutation (gBRCAm). Journal of Clinical Oncology. Volume 35, Number 18_suppl.
23. Jennifer K. Litton, M.D., Hope S. Rugo, M.D., Johannes Ettl, M.D., et al. Talazoparib in Patients with Advanced Breast Cancer and a Germline BRCA Mutation. Published August 15, 2018. N Engl J Med. 2018; 379:753-763. DOI: 10.1056/NEJMoa1802905. VOL. 379 NO. 8
24. Joaquin Mateo, Suzanne Carreira, Shahneen Sandhu, et al. DNA-Repair Defects and Olaparib in Metastatic Prostate Cancer. N Engl J Med. 2015 Oct 29; 373(18): 1697-708. doi: 10.1056/NEJMoa1506859.
25. W. Abida, D. Campbell, A. Patnaik, et al. Preliminary results from the TRITON2 study of rucaparib in patients (pts) with DNA damage repair (DDR)-deficient metastatic castration-resistant prostate cancer (mCRPC): Updated analyses. Volume 30, Supplement 5, V327-V328, October 2019

26. M.R. Smith, S.K. Sandhu, W.K. Kelly, et al. Pre-specified interim analysis of GALAHAD: A phase II study of niraparib in patients (pts) with metastatic castration-resistant prostate cancer (mCRPC) and biallelic DNA-repair gene defects (DRD). Volume 30, Supplement 5, V884-V885, October 2019
27. Shen J, Zhao W, Ju Z, et al. PARPi Triggers the STING-dependent immune response and enhances the therapeutic efficacy of immune checkpoint blockade independent of BRCAness. Cancer Res 2019; 79: 311–319.
28. Pantelidou C, Sonzogni O, De Oliveria Taveira M, et al. PARP inhibitor efficacy depends on CD8+ T-cell recruitment via intratumoral STING pathway activation in BRCA-deficient models of triple-negative breast cancer. Cancer Discov 2019; 9: 722–737.
29. Shen J, Zhao W, Ju Z, et al. PARPi Triggers the STING-dependent immune response and enhances the therapeutic efficacy of immune checkpoint blockade independent of BRCAness. Cancer Res 2019; 79: 311–319.
30. Chabanon RM, Lord CJ and Postel-Vinay S. PARP inhibition enhances tumor cell-intrinsic immunity in ERCC1-deficient non-small cell lung cancer. J Clin Invest 2019; 129: 1211–1228.
31. Sen T, Rodriguez BL, Chen L, et al. Targeting DNA damage response promotes antitumor immunity through STING-mediated T-cell activation in small cell lung cancer. Cancer Discov 2019; 9: 646–661.
32. Huang J, Wang L, Cong Z, et al. The PARP1 inhibitor BMN 673 exhibits immunoregulatory effects in a Brca1–/– murine model of ovarian cancer. Biochem Biophys Res Commun 2015; 463: 551–556.
33. Wang Z, Sun K, Xiao Y, et al. Niraparib activates interferon signaling and potentiates anti-PD-1 antibody efficacy in tumor models. Sci Rep 2019; 9: 1853.
34. Macian F. NFAT proteins: key regulators of T-cell development and function. Nat Rev Immunol 2005; 5: 472–484.
35. Olabisi OA, Soto-Nieves N, Nieves E, et al. Regulation of transcription factor NFAT by ADP-ribosylation. Mol Cell Biol 2008; 28: 2860–2871.
36. Valdor R, Schreiber V, Saenz L, et al. Regulation of NFAT by poly(ADP-ribose) polymerase activity in T cells. Mol Immunol 2008; 45: 1863–1871.
37. Olabisi OA, Soto-Nieves N, Nieves E, et al. Regulation of transcription factor NFAT by ADP-ribosylation. Mol Cell Biol 2008; 28: 2860–2871.
38. Saenz L, Lozano JJ, Valdor R, et al. Transcriptional regulation by Poly(ADP-ribose) polymerase-1 during T cell activation. BMC Genomics 2008; 9: 171.
39. Ghonim MA, Pyakurel K, Ibba SV, et al. PARP inhibition by olaparib or gene knockout blocks asthma-like manifestation in mice by modulating CD4+ T cell function. J Transl Med 2015; 13: 225.
40. Sambucci M, Laudisi F, Novelli F, et al. Effects of PARP-1 deficiency on Th1 and Th2 cell differentiation. Sci World J 2013; 2013: 375024.

41. Gonzalez-Rey E, Martínez-Romero R, O'Valle F, et al. Therapeutic effect of a poly(ADPribose) polymerase-1 inhibitor on experimental arthritis by downregulating inflammation and Th1 response. PLoS One 2007; 2: e1071.
42. Jiang, Y., Li, Y. & Zhu, B. T-cell exhaustion in the tumor microenvironment. Cell Death Dis 6, e1792 (2015). https://doi.org/10.1038/cddis.2015.162
43. Ding L, Kim HJ, Wang Q, et al. PARP inhibition elicits STING-dependent antitumor immunity in brca1-deficient ovarian cancer. Cell Rep 2018; 25: 2972–2980.e5.
44. Yang M, Liu L, Xie M, et al. Poly-ADPribosylation of HMGB1 regulates TNFSF10/ TRAIL resistance through autophagy. Autophagy 2015; 11: 214–224.
45. Rottenberg S, Jaspers JE, Kersbergen A, et al. High sensitivity of BRCA1-deficient mammary tumors to the PARP inhibitor AZD2281 alone and in combination with platinum drugs. Proc Natl Acad Sci U S A. 2008;105(44):17079e17084.
46. Heeke AL, Xiu J, Elliott A, et al. Actionable co-alterations in breast tumors with pathogenic mutations in the homologous recombination DNA damage repair pathway. Breast Cancer Res Treat. 2020;184(2):265e275.
47. Konstantinopoulos PA, Ceccaldi R, Shapiro GI, et al. Homologous recombination deficiency: exploiting the fundamental vulnerability of ovarian cancer. Cancer Discov. 2015;5(11): 1137e1154.
48. Zong H, Zhang J, Xu Z, et al. Comprehensive analysis of somatic reversion mutations in homologous recombination repair (HRR) genes in A large cohort of Chinese pan-cancer patients. J Cancer. 2022;13(4):1119e1129
49. Ang JE, Gourley C, Powell CB, et al. Efficacy of chemotherapy in BRCA1/2 mutation carrier ovarian cancer in the setting of PARP inhibitor resistance: a multi-institutional study. Clin Cancer Res. 2013;19(19):5485e5493.
50. Weigelt B, Comino-Me´ndez I, de Bruijn I, et al. Diverse BRCA1 and BRCA2 reversion mutations in circulating cell-free DNA of therapy-resistant breast or ovarian cancer. Clin Cancer Res. 2017;23(21):6708e6720.
51. Waks AG, Cohen O, Kochupurakkal B, et al. Reversion and non-reversion mechanisms of resistance to PARP inhibitor or platinum chemotherapy in BRCA1/2-mutant metastatic breast cancer. Ann Oncol. 2020;31(5):590e598.
52. Tobalina L, Armenia J, Irving E, et al. A meta-analysis of reversion mutations in BRCA genes identifies signatures of DNA end-joining repair mechanisms driving therapy resistance. Ann Oncol. 2021;32(1):103e112.
53. Mirman Z, de Lange T. 53BP1:a DSB escort. Genes Dev. 2020; 34(1e2):7e23.
54. Setiaputra D, Durocher D. Shieldin - the protector of DNA ends. EMBO Rep. 2019;20(5):e47560.
55. Panier S, Boulton SJ. Double-strand break repair: 53BP1 comes into focus. Nat Rev Mol Cell Biol. 2014;15(1):7e18.

56. Mirman Z, Sasi NK, King A, et al. 53BP1-shieldin-dependent DSB processing in BRCA1-deficient cells requires CST-Polaprimase fill-in synthesis. Nat Cell Biol. 2022;24(1):51e61.
57. Noordermeer SM, Adam S, Setiaputra D, et al. The shieldin complex mediates 53BP1-dependent DNA repair. Nature. 2018;560(7716):117e121.
58. Dev H, Chiang TWW, Lescale C, et al. Shieldin complex promotes DNA end-joining and counters homologous recombination in BRCA1-null cells. Nat Cell Biol. 2018;20(8):954e965.
59. He YJ, Meghani K, Caron MC, et al. DYNLL1 binds to MRE11 to limit DNA end resection in BRCA1-deficient cells. Nature. 2018;563(7732):522e526.
60. file:///Users/sophiastavrakos/Downloads/atm-09-15-1229.pdf
61. Carreira S, Porta N, Arce-Gallego S, et al. Biomarkers associating with PARP inhibitor benefit in prostate cancer in the TOPARP-B trial. Cancer Discov. 2021;11(11):2812e2827.
62. Gachechiladze M, Skarda J, Soltermann A, et al. RAD51 as a potential surrogate marker for DNA repair capacity in solid malignancies. Int J Cancer. 2017;141(7):1286e1294.
63. Castroviejo-Bermejo M, Cruz C, Llop-Guevara A, et al. A RAD51 assay feasible in routine tumor samples calls PARP inhibitor response beyond BRCA mutation. EMBO Mol Med. 2018;10(12):e9172.
64. Pellegrino B, Herencia-Ropero A, Llop-Guevara A, et al. Preclinical in vivo validation of the RAD51 test for identification of homologous recombination-deficient tumors and patient stratification. Cancer Res. 2022;82(8):1646e1657.
65. Cruz C, Castroviejo-Bermejo M, Gutie´rrez-Enrı´quez S, et al. RAD51 foci as a functional biomarker of homologous recombination repair and PARP inhibitor resistance in germline BRCA-mutated breast cancer. Ann Oncol. 2018;29(5): 1203e1210.
66. Tobalina L, Armenia J, Irving E, et al. A meta-analysis of reversion mutations in BRCA genes identifies signatures of DNA end-joining repair mechanisms driving therapy resistance. Ann Oncol. 2021;32(1):103e112.
67. Liu Y, Burness ML, Martin-Trevino R, et al. RAD51 mediates resistance of cancer stem cells to PARP inhibition in triplenegative breast cancer. Clin Cancer Res. 2017;23(2):514e522
68. Castroviejo-Bermejo M, Cruz C, Llop-Guevara A, et al. A RAD51 assay feasible in routine tumor samples calls PARP inhibitor response beyond BRCA mutation. EMBO Mol Med. 2018;10(12):e9172.
69. Pellegrino B, Herencia-Ropero A, Llop-Guevara A, et al. Preclinical in vivo validation of the RAD51 test for identification of homologous recombination-deficient tumors and patient stratification. Cancer Res. 2022;82(8):1646e1657.
70. Sun C, Fang Y, Yin J, et al. Rational combination therapy with PARP and MEK inhibitors capitalizes on therapeutic liabilities in RAS mutant cancers. Sci Transl Med. 2017;9(392):eaal5148.

71. Yang B, Li X, Fu Y, et al. MEK inhibition remodels the immune landscape of mutant KRAS tumors to overcome resistance to PARP and immune checkpoint inhibitors. Cancer Res. 2021; 81(10):2714e2729.
72. Sharma RA, Plummer R, Stock JK, et al. Clinical development of new drug-radiotherapy combinations. Nat Rev Clin Oncol. 2016;13(10):627e642.
73. Veuger SJ, Hunter JE, Durkacz BW. Ionizing radiation-induced NF-kappaB activation requires PARP-1 function to confer radioresistance. Oncogene. 2009;28(6):832e842.
74. Boudra MT, Bolin C, Chiker S, et al. PARP-2 depletion results in lower radiation cell survival but cell line-specific differences in poly(ADP-ribose) levels. Cell Mol Life Sci. 2015;72(8): 1585e1597.
75. Mao Y, Huang X, Shuang Z, et al. PARP inhibitor olaparib sensitizes cholangiocarcinoma cells to radiation. Cancer Med. 2018;7(4):1285e1296.
76. Bi Y, Verginadis II, Dey S, et al. Radiosensitization by the PARP inhibitor olaparib in BRCA1-proficient and deficient highgrade serous ovarian carcinomas. Gynecol Oncol. 2018;150(3): 534e544.
77. Mangoni M, Sottili M, Salvatore G, et al. Enhancement of soft tissue sarcoma cell radiosensitivity by poly(ADP-ribose) polymerase-1 inhibitors. Radiat Res. 2018;190(5):464e472.
78. Abbotts R, Topper MJ, Biondi C, et al. DNA methyltransferase inhibitors induce a BRCAness phenotype that sensitizes NSCLC to PARP inhibitor and ionizing radiation. Proc Natl Acad Sci U S A. 2019;116(45):22609e22618.
79. Verhagen CVM, de Haan R, Hageman F, et al. Extent of radiosensitization by the PARP inhibitor olaparib depends on its dose, the radiation dose and the integrity of the homologous recombination pathway of tumor cells. Radiother Oncol. 2015;116(3):358e365.
80. Sizemore ST, Mohammad R, Sizemore GM, et al. Synthetic lethality of PARP inhibition and ionizing radiation is p53- dependent. Mol Cancer Res. 2018;16(7):1092e1102.
81. Riley RS, June CH, Langer R, et al. Delivery technologies for cancer immunotherapy. Nat Rev Drug Discov. 2019;18(3): 175e196.
82. Wang Z, Sun K, Xiao Y, et al. Niraparib activates interferon signaling and potentiates anti-PD-1 antibody efficacy in tumor models. Sci Rep. 2019;9:1853.
83. Vinayak S, Tolaney SM, Schwartzberg L, et al. Open-label clinical trial of niraparib combined with pembrolizumab for treatment of advanced or metastatic triple-negative breast cancer. JAMA Oncol. 2019;5(8):1132e1140.
84. Konstantinopoulos PA, Waggoner S, Vidal GA, et al. Single-arm phases 1 and 2 trial of niraparib in combination with pembrolizumab in patients with recurrent platinum-resistant ovarian carcinoma. JAMA Oncol. 2019;5(8):1141e1149.
85. Sandhu SK, Hussain M, Mateo J, et al. PROfound: phase III study of olaparib versus enzalutamide or abiraterone for metastatic castration-resistant prostate cancer (mCRPC) with homologous recombination repair (HRR) gene alterations. Ann Oncol. 2019;30:ix188eix189.

Leading from Afar: Examining Remote-First Startup Leadership Practices and Their Impact on Company Culture and Performance By Nikhill Prerepa

Abstract

This research explores the evolving leadership practices of remote-first startups and their significant impact on organizational culture and performance in virtual work environments. In response to the global shift towards remote work, this study argues that remote-first startups are redefining traditional leadership models to effectively manage distributed teams. Employing a comprehensive mixed-methods approach, quantitative surveys and qualitative interviews were conducted to investigate the prevalence and effectiveness of transformational and servant leadership styles in these settings. The findings highlight the crucial role of clear and frequent communication, facilitated by digital tools, in fostering collaboration and maintaining alignment across diverse geographical locations. This study contributes empirical insights into the adaptation of leadership strategies for virtual work environments, offering actionable implications for startup leaders aiming to optimize team dynamics and organizational outcomes amidst the rise of remote-first organizational structures**.**

Keywords Remote-first startups, Technological, Transformational leadership, Culture

Introduction

In the fast-evolving landscape of startups, a notable trend has emerged: the rise of remote-first organizations. This paradigm shift challenges traditional notions of leadership, company culture, and performance metrics. Imagine a scenario where physical office spaces are replaced by virtual workspaces spread across continents, where face-to-face meetings are substituted by video calls, and where collaboration thrives amidst geographical diversity. This transformation raises crucial questions about the leadership practices adopted by remote-first startups and their profound implications on organizational culture and overall performance.

This study investigates the leadership practices employed by remote-first startups and examines how these practices influence company culture and overall performance. Specifically, it seeks to uncover the strategies and methodologies remote-first startups utilize to lead their teams effectively in virtual environments, and to analyze the resultant. The research question is *what leadership practices are being used by remote-first startups, and how do these practices affect the company culture and overall performance*?

To explore these dynamics comprehensively, this research employs a mixed-methods approach. Quantitative data is gathered through surveys distributed among employees of remote-first startups, focusing on leadership styles, communication strategies, and perceptions of company culture and performance. Qualitative insights are derived from in-depth interviews with executives and team members, providing nuanced perspectives on the effectiveness of remote leadership practices.

Preliminary findings highlight the prevalence of transformational and servant leadership styles in remote-first startups, emphasizing empowerment, trust-building, and fostering a collaborative ethos among distributed teams. Effective communication emerges as a cornerstone, facilitated by digital tools and structured virtual meetings that ensure clarity and alignment across geographies. Moreover, a strong emphasis on results-oriented performance metrics and flexible work arrangements underscores the adaptability and resilience of remote-first organizational cultures.

This study contributes to the existing literature by offering empirical insights into the unique leadership dynamics of remote-first startups. It advances our understanding of how innovative leadership practices tailored for virtual environments can shape company culture positively and drive overall organizational performance. By bridging theoretical frameworks with practical implications, this research aims to inform strategic decision-making in startup ecosystems navigating the shift towards remote work models.

This paper proceeds as follows: Section 3 provides a comprehensive review of relevant literature, exploring current theories and empirical research on remote leadership and its impact on organizational outcomes. Section 2 outlines the methodology employed in this study, detailing the survey design, participant selection criteria, and data collection procedures. Section 4 presents the analysis of survey findings and interview insights, offering a detailed examination of leadership practices and their effects on company culture and performance. Section 5 delves into discussions and implications drawn from the study, highlighting key findings and their implications for remote-first startups. Section 6 acknowledges the limitations of this research and suggests avenues for future research directions. Finally, Section 7 synthesizes the conclusions derived from the study, summarizing key insights and practical implications for leaders and policymakers in the remote-first startup ecosystem.

Research Methodology

The research problem addressed in this study revolves around the evolving leadership practices within remote-first startups and their impact on organizational culture and performance in virtual work environments. As organizations increasingly transition to remote-first models, understanding effective leadership strategies becomes crucial for sustaining productivity and fostering a positive work culture amidst geographical dispersion. This study employs a mixed-methods research design. This choice is justified by the need to comprehensively explore both the quantitative metrics of leadership effectiveness and the qualitative nuances of leadership practices within remote-first startups. The research aims to provide a holistic understanding of how different leadership styles influence organizational dynamics and performance outcomes. The research is primarily explanatory in nature. It seeks to explain the relationships between leadership practices in remote-first startups, organizational culture, and overall performance. By integrating quantitative surveys with qualitative interviews, the study aims to uncover underlying mechanisms and contextual factors that shape leadership effectiveness in virtual work

environments. The target population includes employees and executives of remote-first startups globally.

The sample size was determined using a stratified sampling method to ensure representation across various organizational roles and geographic locations. Inclusion criteria encompassed individuals currently employed in remote-first startups, while exclusion criteria excluded those not directly involved in leadership roles or lacking sufficient tenure in their current roles. Primary Data consists of Quantitative data collected through structured surveys distributed electronically to employees of remote-first startups. Qualitative data was gathered through semi-structured interviews with founders, executives, and team members, focusing on their experiences and perceptions of leadership practices. Secondary Data consists of sources including academic literature, industry reports, and organizational documents such as internal policies and leadership communications from selected remote-first startups. The survey instrument included validated scales to measure leadership styles and organizational culture dimensions. Interview guides were developed based on the research objectives to ensure consistency and depth in qualitative data collection.

Data Collection Procedures: Surveys were administered electronically using platforms that ensured anonymity and confidentiality. Interviews were conducted virtually or in-person, recorded with consent, and transcribed verbatim for analysis. Data collection emphasized flexibility to accommodate diverse time zones and scheduling constraints of remote participants. Qualitative Analysis underwent thematic analysis to identify recurring patterns and themes related to leadership practices and their impact on organizational culture and performance. Quantitative Analysis was analyzed using statistical software (SPSS) to compute descriptive statistics and inferential tests, correlating leadership styles with organizational outcomes. The Mixed-Methods Analysis integrates qualitative themes with quantitative findings enabling a comprehensive understanding of how leadership practices influence organizational dynamics in remote-first startups. Triangulation of data sources, member checking, and peer debriefing were employed to enhance the validity and reliability of findings. Quantitative scales were validated through reliability tests, ensuring consistency in measurement.

Participants were provided with detailed information about the study's purpose, procedures, and their rights, obtaining informed consent prior to data collection. Measures were taken to anonymize data and ensure participant confidentiality throughout the research process, adhering to data protection regulations. Ethical approval was obtained from the relevant institutional review board or research ethics committee, ensuring compliance with ethical standards and guidelines for research involving human subjects. In conclusion, this study employs a rigorous mixed-methods approach to investigate leadership practices in remote-first startups. The chosen methodology is well-suited for addressing the research questions by combining quantitative metrics with qualitative insights. By exploring how leadership styles influence organizational culture and performance in virtual environments, this research aims to contribute empirical evidence and practical implications for enhancing leadership effectiveness in remote-first settings.

Literature Review

Previous studies on leadership in remote-first startups highlight key themes such as transparent communication, flexibility, technology use, empathy, and company culture. These studies provide insights into various leadership practices and their impact on organizational outcomes. Studies on leadership in remote-first startups underscore the importance of transparent communication, flexibility, and technology. Effective leaders prioritize open communication channels and flexible work arrangements to enhance team cohesion and productivity (Chee & Zainal, 2022). Leveraging technology facilitates real-time collaboration and project management, contributing to streamlined workflows and higher productivity levels (Ensley et al., 2006). Additionally, emotional intelligence and empathy are crucial for remote leaders, fostering higher employee satisfaction and engagement (Yao, 2019).

A strong company culture positively impacts employee retention and performance in remote settings (Hendratmi et al., 2024). Startups with well-defined cultural values and a supportive environment experience lower turnover rates and increased productivity. Furthermore, cultural values drive innovation and agility, enabling startups to adapt quickly to market changes and drive growth (Brüggeman, 2014).

Clear communication and technology are fundamental to effective remote leadership across studies. Emotional intelligence consistently predicts leadership success in remote settings. The studies reviewed offer diverse insights into leadership practices in startup contexts. Chee and Zainal (2022) focus on evolving themes like authenticity and ethics in Malaysia, whereas Ensley et al. (2006) explore various leadership styles in the USA, emphasizing the effectiveness of shared leadership models (Chee & Zainal, 2022; Ensley et. al., 2006). Yau (2019) addresses visionary leadership and management challenges in uncertain environments, contrasting with Hendratmi et al. (2024) who examine empowering leadership's impact on innovation in Indonesia (Yau, 2019; Hendratmi et. al., 2024). These studies highlight how leadership practices vary across cultures and methodologies, influencing company culture and performance in startup settings. This thematic literature review integrates findings across leadership practices, technology use, company culture, and performance outcomes in remote-first startups. By synthesizing these insights, the review provides a comprehensive understanding of effective strategies for enhancing leadership and organizational dynamics in virtual work environments. In the context of the increasing prevalence of remote-first startups, this review offers actionable insights for leaders to refine their leadership approaches, foster a positive company culture, and optimize team performance in remote settings. Understanding these dynamics is crucial for navigating the challenges and opportunities associated with remote work, ultimately driving sustainable growth and success for startups in today's digital age. Refer to table 1 of the selected literature review.

Table 1: Relevant Literature Review

Study	Theme	Methods	Country / Market	Factors/ Dimensions/ Variables	Findings
Chee and Zainal, 2022	Leadership Literature: emerging thematic evolution of the leadership literature	Qualitative	Malaysia	- Narrowed view point (ie. Limited perspectives) - Authenticity - Service - Ethics - Inclusivity	1. Various leadership theories seem to be overwhelmed with different thematic thoughts 2. The increased emergence of value-based leadership theories such as authentic, service, servant, ethical, inclusive, and later the paradoxical yin-yang leadership style 3. No one style fits all
Ensley et al., 2006	Shift from vertical to shared leadership: ideas of leadership that's centralized like CEOs are being challenged	Quantitative	USA	Types of leadership: -Directive Leadership -Transactional Leadership -Transformational Leadership -Servant Leadership	1. Effectiveness of Shared Leadership: Shared leadership improves organizational performance, especially in startups and dynamic environments. 2. Continued Role of Vertical Leadership: Traditional vertical leadership styles remain important alongside shared leadership approaches. 3. Optimal Integration: Combining both shared and vertical leadership enhances overall team and organizational outcomes.
Yao, 2019	Navigating leadership dynamics in startup environments	Qualitative	Very broad market	-Visionary leadership - Compassion -Balancing -leadership ideals -management styles -High uncertainty and chaotic nature -Leadership-As-Practice	1. Co-founders emphasize visionary leadership and compassion for their team members. 2. Balancing leadership ideals with practical management styles is a significant challenge. 3. The startup environment is characterized by high uncertainty and chaos, influencing leadership dynamics.
Hendratmi et al., 2024	Exploring Shared Leadership and Venture Performance	Qualitative	Indonesia	- Empowering Leadership -Member Integrity, -Ethic -Innovativeness and Reflexivity as mediating factors -Business Venture Performance Levels	1. Empowering leadership, member integrity, and an Islamic perspective on shared leadership positively influence innovativeness and reflexivity. 2. Innovativeness and reflexivity mediate the relationship between these leadership factors and business venture performance levels. 3. Shared leadership skills enhance organizational innovativeness and reflexivity, thereby improving business performance.

Qiu, 2018	Measuring Entrepreneurs hi p Performance and Policy Priorities in Vietnam	Quantitative	Vietnam	-Entrepreneurial Attitudes -Entrepreneurial Abilities, -Entrepreneurial Aspiration - risk acceptance -opportunity perception -internationalization technology absorption -Comparative analysis	1. Vietnam shows significant challenges in several pillars of the GEDI, particularly in risk acceptance, opportunity perception, internationalization, and technology absorption. 2.Comparative analysis with Thailand and Indonesia reveals specific areas where Vietnam lags behind or performs better. 3.The study identifies nine bottlenecks within Vietnam's GEDI profile that require urgent policy attention for improvement. 4. The Penalty for Bottleneck (PFB) methodology is proposed as a strategic approach to prioritize and optimize policy interventions aimed at enhancing Vietnam's entrepreneurship performance.
Brüggema n , 2014	Role of entrepreneuri al leadership in the trajectory of startups, emphasizing how leadership styles evolve and impact organizational outcomes as firms grow	Qualitative	USA	- Different leadership styles -leadership shapes -Strategies fostering innovation	1. Transformational leadership (inspiring and motivating employees towards a shared vision) is more prevalent and effective in startups where innovation and agility are crucial. 2. Startups tend to have a more flexible and risk-tolerant culture fostered by entrepreneurial leaders who encourage experimentation and learning from failures.
Men, et al., 2018	Focusing on how executive leadership shapes corporate culture and uses strategies like symmetrical, transparent, and visionary communicatio n	Qualitative	China	-Culture Types -Innovation/adventur e -Inclusiveness/partici pation -Supportiveness -Aggressiveness/com petitiveness	1. Five prevalent types identified: innovation/adventure, openness, inclusiveness/participation, supportiveness, aggressiveness/competitiveness. 2. Startup leaders shape culture through their traits, styles, and behaviors. 3. Effective strategies include symmetrical, transparent, authentic, and visionary approaches,

Pangestu, et al., 2020	Entrepreneurial leadership and competitive advantage between business innovation and startup performance.	Quantitative		- Business models -Operational efficiencies - Entrepreneurial Leadership -Leadership styles -Behavior	1. Business Model Innovation and Startup Performance: The study confirms a significant positive relationship, indicating that startups that innovate their business models tend to achieve better performance outcomes. 2. Entrepreneurial Leadership and Business Model Innovation: There is a significant association between entrepreneurial leadership and business model innovation, highlighting how leadership qualities influence the ability to innovate within startups.
Akpaneko ng, 2023	Impact of leadership behaviors and styles on startup performance, emphasizing the relevance of entrepreneurial leadership in startup growth and development.	Qualitative	USA	- Leadership behaviors -Qualities - Personal attributes	1. Impact of Leadership Behavior on Startup Culture: Significant influence was observed, with leadership behaviors shaping the values and norms within startup organizations. 2. Emerging Leadership Behaviors in Response to Startup Challenges: Identified changes in leadership styles, including the emergence of humble leadership, adapting to the dynamic startup environment.
Jenkins, 2020	Leaders at biotechnology organizations establishing a positive culture and achieving high organizational performance during a phase of rapid growth	Qualitative	USA	- Organizational Culture - Leadership Practices -Strategies	1. Leaders emphasized building multi-leveled trust and treating employees as "people" or "family," fostering a sense of equality and belonging. 2. Themes emerged around maximizing employee potential and the critical role of communication in linking culture to organizational performance. 3. key factors contributing to a positive culture, - trust-building, fostering innovation through risk-taking, and responsiveness to employee feedback.
Prabawati, et al., 2022	Impact of perceived organizational support, organizational culture, and servant leadership on contextual performance	Qualitative	Indonesia	-Contextual Performance -Behavior -Effectiveness - Organizational Support:	1. Contribution of Perceived Organizational Support, Organizational Culture, and Servant Leadership- a significant positive relationship between perceived organizational support, organizational culture, and servant leadership with contextual performance. 2. Supportive organizational environments, strong cultural values, and servant leadership behaviors contribute positively to employees' contextual performance in startup settings.

Men, et al., 2021	CEOs' use of motivating language influences employee relationships within startup organizations	Quantitative	China	-Motivating Language -Communication -Psychological Need Satisfaction -Organizational Identification	1. Direct Influence: CEO use of motivating language directly enhances employee-startup relationships. 2. Indirect Influence: Motivating language also indirectly influences relationships by increasing psychological need satisfaction and organizational identification among employees. 3. Theoretical Contributions: The study advances theoretical understanding of executive leadership's role in public relations and communication within Chinese startup contexts, leveraging motivating language theory to explain relational dynamics.
Joshi and Achuthan, 2018	Navigating leadership challenges in Indian high-tech startups amidst evolving government policies and financing landscapes.	Qualitative	India	-Types of Entrepreneurs - Leadership styles and strategic orientations. - Exploration vs. Exploitation -Leadership Traps	1. Leadership Lessons-Draws insights from transformational leadership, complexity theory, entrepreneurial leadership, and contingency leadership, highlighting their relevance and application in the Indian high-tech startup context. 2. Characteristics of Successful Entrepreneurs-Explores the traits and skills required for success in high-tech startups, emphasizing innovation, human resources management, and adaptability. 3.Case Study of Infosys-Analyzes how Infosys evolved from a startup to an industry leader, focusing on founder roles as mentors and investors, and examines recent challenges between founders and current management.
Udoyiu, et al., 2024	Innovation culture and leadership culture on entrepreneurial competitiveness in selected areas of Nigeria	Quantitative	Nigeria	-Innovation Culture -Attitudes -Creativity -Leadership -Competitiveness	1. Impact of Innovation Culture-Significant positive effect on entrepreneurial competitiveness, showing the importance of fostering a culture that encourages innovation and adaptive responses to market demands. 2. Impact of Leadership Culture-Significant positive effect on entrepreneurial sustainability, indicating that effective leadership styles and practices contribute to the long-term survival and growth of SMEs. 3. Emphasizes the need for SME owners to embrace innovative trends and cultivate a culture that supports continuous improvement + suggests collaboration with government initiatives to enhance human capacity building for better entrepreneurial leadership and performance.

Interviews

The interviews were conducted from the last weeks of July through early August and involved a diverse group of participants, ranging from founders and executives to team members from various remote-first startups. These conversations were essential for gaining firsthand insights into how leadership operates in virtual work environments and how it influences company culture and performance.

Each interview was structured around a thoughtfully prepared set of questions aimed at exploring different facets of remote leadership. It delved into topics such as the core leadership principles guiding their organization, strategies employed for managing teams spread across different locations, evolution of leadership practices in response to remote work dynamics, insights gained from managing remote teams, and the overall effects of these practices on organizational dynamics.

Ensuring ethical integrity throughout the interview process was of utmost importance. The interview questionnaire was meticulously crafted to uphold confidentiality, obtain informed consent from participants, and ensure anonymity where needed. Questions were neutrally framed to elicit genuine and candid responses, avoiding any influence or bias that could sway the insights shared by interviewees.

Questions	**Verbatims**	**Dimensions**
Q1: How does your organization's leadership philosophy shape the management of remote teams and what specific approaches do you employ?	*"As I often say, 'Leadership isn't about managing from the top down; it's about supporting from the bottom up."* (P1, Male, 24 yrs) *"We make it a priority to create opportunities for team members to interact and collaborate, whether through virtual team-building activities or regular informal catch-ups."* (P2, Female, 38 yrs) *"Our leadership philosophy emphasizes trust, flexibility, and clear communication, shaping our remote management by setting clear goals, using digital tools, and fostering regular check-ins to keep the team connected and engaged." (P4, Female, 42 yrs)*	-Communication -Team Cohesion -Flexibility
Q2: From your perspective, what are the primary priorities and values that influence leadership decisions in your remote-first startup?	*"The primary priorities and values that guide our leadership decisions are trust, flexibility, and innovation…Trust is fundamental"* (P3 , Male, 46 yrs) *"We value empowering our team with autonomy, fostering a culture of open communication, and remaining flexible to adapt to evolving needs and challenges, ensuring that every team member feels supported" (P1, Male, 24 yrs)*	-Trust -Flexible -Communication

Q3: Can you discuss how leadership strategies have evolved to address the unique challenges of remote work over time?	*"Leadership strategies have evolved to embrace flexibility and empathy in remote work environments, as Sheryl Sandberg aptly puts it: 'Leadership is about making others better as a result of your presence and making sure that impact lasts in your absence." (P1, Male, 24 yrs)* *"Embracing technology to bridge distances and prioritizing meaningful connections have become central to our approach" (P4, Female, 42 yrs)*	-Flexibility -Technology -Goal-intended
Q4: What valuable lessons have your leaders learned from managing remote teams?	"*One of the most valuable lessons we've learned from managing remote teams is the importance of maintaining clear and open communication. In a remote setting, it's easy for messages to get lost or misinterpreted"* (P4, Female, 42 yrs) *" A key lesson I learned the hard way was the necessity of investing in the right technology and tools. Reliable communication platforms and project management tools are indispensable for keeping remote teams connected and productive" (P5, Male, 37 yrs)*	- Clear Communication - Technology
Q5: What strategies do you use to ensure transparent/effective communication within your company?	*"For real-time interactions, we rely on instant messaging platforms and video conferencing. For project management and updates, we use collaborative tools that allow everyone to track progress and share documents seamlessly. These tools help ensure that communication is not only effective but also accessible to everyone, regardless of their location."* (P4, Female, 42 years) *"Our strategy includes bi-weekly sync-ups where we dive into both metrics and team feedback, supplemented by a 'communication charter' that outlines our norms for clarity and responsiveness" (P6, Female, 32 yrs)*	-Transparency -Feedback -Communication
Q6: Can you share a memorable instance where effective communication improved team cohesion and created a successful project outcome?	*"We implemented a series of collaborative workshops and feedback loops, using interactive digital platforms to facilitate brainstorming and discussions" (P7, Male, 29 yrs)* *"By introducing a shared project dashboard and hosting regular cross-departmental sync meetings, we created a transparent environment where updates and feedback flowed seamlessly." (P8, Male, 51 yrs)* *"By maintaining transparent and frequent communication, we quickly identified and resolved issues, aligned our efforts, and supported one another. This approach not only ensured a smooth transition but also strengthened inter-departmental relationships, culminating in a successful CRM implementation that improved our overall efficiency." (P9, Female, 33 yrs)*	-Feedback -Transparent -Communication -Efficient -Frequency

Q7: How do the company's culture, values, and support for remote work contribute to	*" The values we uphold, such as open communication and mutual respect, create a supportive environment where everyone feels connected, regardless of location." (P10, Male, 30 yrs)*	-Communication -Respect -Flexibility -Continuous
your personal job satisfaction	*"The company's commitment to work-life balance and flexibility is a significant factor in our success with remote work" (P11, Female, 22 yrs)* *"The company's emphasis on continuous learning and development plays a key role in supporting remote work." (P7, Male, 29 yrs)*	

Discussion

This study provides a multidimensional exploration of leadership practices in remote-first startups, shedding light on their impact on company culture and performance. The analysis reveals that leadership in these organizations transcends traditional management paradigms, emphasizing adaptability, trust, and transparent communication.

1. Technological Adaptation and Leadership Effectiveness

A central theme in the findings is the integration of technology to mitigate the challenges of remote work. Tools facilitating real-time collaboration, asynchronous communication, and progress tracking have proven indispensable. Leaders leveraging these technologies not only enhance team efficiency but also foster inclusivity, ensuring all team members, regardless of location, feel equally engaged. This technological fluency distinguishes successful remote-first startups and underscores the necessity for continuous investment in digital infrastructure.

2. Cultural Cohesion Across Distances

The study highlights the unique challenge of nurturing a cohesive organizational culture within geographically dispersed teams. Effective leaders employ deliberate strategies to sustain shared values, such as routine team-building activities, virtual town halls, and recognition programs. By prioritizing cultural alignment, leaders ensure that remote teams operate under a unified vision, which enhances productivity and employee satisfaction.

3. Leadership Styles: Transformational and Servant Leadership

The prevalence of transformational and servant leadership styles underscores the shift toward more empathetic and inclusive approaches. Transformational leaders inspire and align their teams with a shared vision, while servant leaders prioritize individual growth and well-being.

These leadership styles cultivate trust, flexibility, and resilience, enabling organizations to navigate the complexities of remote work successfully.

4. Implications for Leadership Development

The findings suggest that leadership in remote-first startups requires a blend of technical skills, emotional intelligence, and strategic foresight. Training programs focusing on these competencies can prepare emerging leaders to manage distributed teams effectively. Additionally, creating mentorship opportunities and knowledge-sharing platforms can enhance leadership practices across the startup ecosystem.

Implications

This study explores how leadership practices influence company culture and performance within remote-first startups. Current research often overlooks the unique dynamics of these startups, focusing instead on traditional office environments or general remote work. Understanding leadership in these contexts is crucial as remote-first setups become increasingly prevalent. The research provides specific insights into effective leadership tailored for these startups, advancing knowledge in organizational behavior and entrepreneurship. The following are implications for different stakeholders.

1. **For Scholars and Researchers:** This study opens up a new area of exploration by focusing on leadership in remote-first startups, which is often overlooked compared to traditional office environments. It gives academics fresh perspectives on how leadership shapes company culture and performance in these modern, remote settings. This could spark new research and discussions, helping to build a more complete understanding of leadership dynamics in today's evolving work landscape.
2. **For Leaders and Managers in Remote-First Startups:** For the leadership team managing a remote-first startup, this study offers practical advice for making your leadership more effective. It emphasizes the importance of clear communication, using technology to keep teams connected, and building a culture of openness and understanding. By adopting these strategies, you can improve how your team works together, boost morale, and drive better results, all while navigating the unique challenges of managing a remote team.
3. **For Organizational Behavior Experts:** For those studying organizational behavior, this research adds a valuable layer to the existing theories by focusing specifically on remote-first startups. It provides new insights into how leadership practices affect remote teams, helping to refine and expand theories about organizational culture and behavior. This could be instrumental in developing more relevant and effective strategies for managing today's remote and hybrid work environments.
4. **For Policymakers and Business Consultants:** Policymakers and consultants can use the insights from this study to improve how remote work is structured and supported. The research highlights key strategies like fostering transparent communication and using technology to keep teams cohesive. This can help in crafting better policies and offering more effective consulting advice, ultimately making remote work more productive and satisfying for everyone involved.

Limitations

There are a few things to keep in mind about this study's limitations. First, we worked with a relatively small group of people, which means the results might not cover all the different experiences you might find in remote-first startups. This makes it harder to generalize the findings to a wider audience or to other types of remote-first organizations. Additionally, the research was focused on a particular geographic area, so the insights might not fully apply to startups in other regions with different cultures and working styles.

Another point to consider is that the data we gathered relied on self-reports. While this method gives us a lot of useful information, it also means there's a chance that responses could be biased. People might answer in a way that they think is expected or desirable, rather than sharing their true feelings and experiences. This can impact how accurate and detailed our findings are.

For future studies, it would be helpful to include a larger and more diverse group of participants from various locations and backgrounds. This would help make the findings more relevant and applicable to a broader range of remote-first startups. Also, using a mix of data sources—like combining self-reports with observations or input from other sources—could provide a more complete and accurate picture of leadership practices in remote-first startups around the world.

Conclusion

This research underscores the transformative impact of innovative leadership practices on remote-first startups, offering insights into how they navigate the challenges of virtual environments. Leaders in these organizations prioritize adaptability, trust, and transparent communication, leveraging technology and fostering cultural alignment to drive performance and employee satisfaction. The study also highlights the significance of transformational and servant leadership styles in creating an inclusive, empowered workforce.

By emphasizing a people-first approach, these startups are not only reshaping traditional notions of leadership but are also setting benchmarks for organizational success in the digital age. Future research should expand on these findings by exploring cross-cultural dynamics and longitudinal impacts of leadership practices in remote settings. As remote-first startups continue to redefine the workplace, their leadership strategies hold valuable lessons for organizations across industries, emphasizing that adaptability, empathy, and innovation are paramount in the modern work environment.

Works Cited

Mahajan, K., Pal, A., & Desai, A. (2023). Revolutionizing Fan Engagement: Adopting Trends and Technologies in The Vibrant Indian Sports Landscape. *International Journal of Management Thinking*, *1*(2), 116-135.

Chee, C. Y., & Zainal, S. R. M. (2022). An Integrative Review of Past Prominent Leadership Literature: Future Research Directions in Malaysia. *Global Business & Management Research*, *14*.

Ensley, M. D., Hmieleski, K. M., & Pearce, C. L. (2006). The importance of vertical and shared leadership within new venture top management teams: Implications for the performance of startups. *The leadership quarterly*, *17*(3), 217-231.

Yau, L. (2019). *Effective leadership practices of early-stage startup co-founders* (Doctoral dissertation, Northeastern University).

Hendratmi, A., Ratnasari, R. T., & Fitriyah, N. L. (2024). The Role of Shared Leadership in Fostering Innovation and Performance in Indonesian Startups. *Journal of System and Management Sciences*, *14*(11), 375-394.

Qiu, T. (2018). INFLUENCE OF CULTURAL PRACTICES ON SOCIAL SUPPORT OF FEMALE ENTREPRENEURS'STARTUPS. *Journal of Developmental Entrepreneurship*, *23*(01), 1850007.

Brüggemann, H. (2014). *Entrepreneurial leadership styles: a comparative study between Startups and mature firms* (Bachelor's thesis, University of Twente).

Men, L. R., Chen, Z. F., & Ji, Y. G. (2018). Walking the talk: An exploratory examination of executive leadership communication at startups in China. *Journal of Public Relations Research*, *30*(1-2), 35-56.

Phangestu, J., Kountur, R., & Prameswari, D. A. (2020). The moderating effect of entrepreneurial leadership and competitive advantage on the relationship bbetween business model innovation and startup performance. *Journal of Business and Retail Management Research*, *14*(3).

Akpanekong, E. O. (2023). Impact of Leadership for Startup Companies.

Jenkins, D. G. (2020). An Intrinsic Case Study of How Leaders at a Startup Biotechnology Organization Established a Positive Culture and High Organizational Performance. Northeastern University.

Prabawati, C., & Wafa, Z. (2022). The Role of Perceived Organizational Support, Organizational Culture, and Servant Leadership in The Contextual Performance of The Startup Employees. *Journal of Educational, Health & Community Psychology (JEHCP)*, *11*(2).

Men, L. R., Chen, Z. F., & Ji, Y. G. (2021). Cultivating relationships with startup employees: the role of entrepreneurs' leadership communication. *Management Communication Quarterly*, *35*(4), 518-545.

Joshi, D., & Achuthan, S. (2018). Leadership in Indian high-tech start-ups: Lessons for future. *The Future of Leadership: Addressing Complex Global Issues*, 39-91.

Udoyiu, E. U., Emerole, G. A., Okebaram, S. M., & Ikoro, I. E. Impact of Innovation and Leadership Culture on Entrepreneurial Competitiveness and Sustainability: A Study of Selected SMES In South-South Nigeria. *World*, *2579*, 0544.

Navigating the Last Gasp: Success Factors in Business Adaptation By Zeen Chen

Abstract

This paper explores the concept of the "last gasp" in businesses, diving into the question of why some companies succeed in their final efforts to adapt to significant market changes while others fail. The study aims to identify the factors contributing to the success or failure of these adaptation efforts by focusing on four case studies: IBM, Nokia, AMC, and Blockbuster. The methodology involves a review of academic literature, including Adner and Kapoor's (2011) analysis of technology S-curves and innovation ecosystems, Perez's (1998) exploration of technological revolutions and societal adaptation, and Henderson's (1993) examination of radical versus incremental innovations. IBM's ongoing success amid technological shifts underscores the critical role of strategic foresight and adaptable leadership. In contrast, Nokia's inability to innovate and keep pace with market demands led to its downfall, highlighting the risks of inertia. AMC's successful pivot to enhance the movie-watching experience in response to streaming services demonstrates effective resource utilization and adaptability. On the contrary, Blockbuster's failure to ultimately even recognize the streaming revolution exemplifies the dangers of slow adaptation and untimely leadership. The research concludes that the success of a company's last-gasp efforts is influenced by the leader's decision-making, resource management, and the ability to overcome inertia. These findings provide valuable insights for businesses navigating technological disruptions and striving for long-term relevance.

Introduction

The term *last gasp* is often used to describe taking the final actions before something dies out completely. However, in the context of businesses, there remains a difference between the normal adaptation with shifts in business focus and a last-gasp effort. In the period of technological change, when a new technology is replacing the old technology, the last gasp is defined as"the extension of the old technology …by individual incumbent firms to prolong the life of their old technology investments." (Adner, Kapoor, 2011: 626) A last-gasp situation usually allows much less time to change, not only so, but the goal that deviates it from a normal adaptation is that it focuses on surviving instead of long-term growth or innovation, moreover, they occur in different phases of a company's history: an immediate, possibly transformative, change that proceeds the downfall of a company is a last-gasp effort, but a regular announcement of a shift in business focus in situations such as trying to win a competition with another company or adapting to more customer needs are not considered last gasps.

When new innovations outperform older technologies and yet companies want to hold on to their traditional technology, the last gasp is usually considered "the extension of the old technology… by individual incumbent firms to prolong the life of their old technology investments" (Adner & Kapoor, 2011). In their study, Adner and Kapoor (2011) examined an I-line in the semiconductor lithography industry, while other companies used the extended G-line generation by increasing the size of the lenses, only switching focus to the I-line when extending

the G-line was not possible anymore. Extending the G-line is an example of the results of the company's last gasp. Gas and electric cars reflect another example of the last gasp in a theoretical situation, as electric cars have superior performance and are more eco-friendly than gas cars. They also need the infrastructures to uphold their place, such as more easily accessible charging stations and short waiting times for charging, a developed system of automatic driving, and more.

Furthermore, if gas car companies make significant technological breakthroughs for better fuel economy that exceeds the eco-friendliness of electric cars during a time with an inevitable trend of replacements of gas-powered cars with electric ones, that will also be considered a last gasp. This concept of the last gasp is crucial as it signifies an important decision or action that could change the outcome, especially when facing unavoidable failure or decline. It represents a final opportunity to save an existing industry. Moreover, even big companies sometimes make radical changes to avoid being replaced by smaller companies who made the change. For example, the radical development of generative artificial intelligence, which began around 2022, has made an enormous impact not only on technological industries but also on many other sectors.

Nevertheless, questions remain regarding the term *last gasp*. For example, what underlying factors are at the core of the success rate of a company's last gasp? Moreover, what determines whether a company engages in the last gasp? In today's ever-evolving technology era, these questions are crucial for determining which companies will succeed or fail. Other researchers have also talked about the factors of last-gasp efforts of different companies, which will also be taken into account in this paper in the Literature Review section. Meanwhile, this research focused on answering the question: Why are some firms successful in their effort for the last gasp while others are not? An understanding of the factors and their contributive outcomes will emerge by examining four companies: IBM, Nokia, AMC, and Blockbuster. Through examining four different companies, this research aimed to analyze some key factors that affect a company's success overall and in a last-gasp situation, focusing on factors such as resources, adaptability, and leadership decisions.

Literature Review (needs much more evidence and paper references)

In their study examining *Innovation Ecosystems and the Pace of Substitution: Re-examining Technology S-curves,* Adner and Kapoor (2011) discussed new technologies and their adaptability to the current market ecosystem. They further expounded on how traditional firms in the old technology industry responded to this change, whether allocated resources to extending old technology, in other words, the last gasp, or seizing opportunities for new technologies. With constant changes within the industry, Adner and Kapoor noticed companies considering factors to successfully transition into a new technological revolution or the development of traditional technologies. However, a gap in literature remains involving the success rate of a traditional company's last gasp. Why are some firms successful in co-existing with the new technology? At the same time, some are getting left out of the stage. In the current

research, these factors are explored, as well as their adaptations of both the new and the traditional firms in the market ecosystem.

Perez (1998) examined technological revolutions, paradigm shifts, and social-institutional changes, “the long wave phenomenon” in the context of the coupling and decoupling of “techno-economic origin and socio-institutional solution.” These changes explain how society adapts to technological innovation in four parts: (a) the explanation of the technology and why it evolves, (b) how technology affects the entire social system, especially the productive system, (c) how the two spheres, “techno-economical origin” and “socio-institutional solution” match together, and (d) the nature of new technologies, and how it can help the society to better adapt to it. Understanding the process of adapting new technologies and how it’s integrated into society remains essential to how technologies evolve. Perez’s study included society’s adaptation to new technologies that have been developing for decades. However, the evolution of technology is currently growing at an exponential rate, so the technological adaptation that the business made in the 1990s needs a bit more modifications to better suit today’s society. What’s more, new technologies today can also affect the industry that might not be so related to them. That’s why it’s best to study the more recent developments in the techno industry to better understand how society is adapting to evolving technologies today. In the current study, the investigation involves the most recent evolution of technology and how society is adapting to it, along with the different results of companies that made different decisions regarding the oncoming new technologies.

In a publication examining radical technology innovations, Henderson (1993) articulated the “competitive implications of radical and incremental innovations” overlooking the effect of heterogeneous investment strategies when facing innovation. Henderson used two data types from the photolithographic alignment equipment industry., which included investment strategies in big firms. Henderson demonstrated consistency between the firm’s success and its research productivity. Because neoclassical theory and organizational theory were too generalized and encompassed firms, Henderson merged the theories, suggesting that smaller companies focus on disrupting the industry by making radical innovations. On the other hand, larger companies, who possess historical market power and organizational inertia, favored incremental changes.

Methods

This research aims to provide a comparative case study approach to explore the concept of the "last gasp" and its four distinct relationships with different companies and their efforts to survive and adapt. These relationships could potentially illustrate different outcomes for different companies. The four companies—IBM, Nokia, AMC, and Blockbuster—were selected at specifically to showcase the different relationships that companies can have with the last gasp to exemplify and provide a comprehensive analysis of the "last gasp" concept. IBM’s history, marked by adaptability and resilience, shows how a company can reinvent itself according to the ever-changing market conditions. On the contrary, Nokia’s downfall resulted from its failure to swiftly adapt to the new smartphone revolution. The successful last gasp of AMC through the

competition with the emerging streaming services and the pandemic contrasts with the unsuccessful last gasp of Blockbuster against the uprising of streaming services.

Case Selection

The companies were chosen to represent four distinct scenarios that companies may face in relation to the "last gasp":

- **IBM**: A successful company that never actually went through a last-gasp situation, its capability to perceive the market and always make the right decision puts it on top of the technology industry. IBM serves as an example of a company that did not undergo a last-gasp situation and yet has been successful throughout its history, which is one of the four scenarios that a company may face with the last gasp.
- **Nokia**: Although now involved in 5G technology, Nokia did not undergo a last-gasp effort when it lost the No.1 title and even a place to stand in the mobile phone industry as it didn't make any significant changes over such a short period of time of its downfall. Its selection highlights a company that, while can still be operational, did not regain its former stature, and ultimately failed to succeed in recovering from a major breakdown. IBM and Nokia were used as a pair for comparison here to illustrate the different outcomes for companies that do not experience a last-gasp moment.
- **AMC**: This company represents a successful last-gasp effort, as it adapted to the rise of streaming services by evolving its business model to prevent further loss from the ongoing technology. Its selection shows how a company can survive and thrive by implementing a well-timed survival strategy.
- **Blockbuster**: The company failed in its last-gasp attempt to adapt to the streaming era as it did try to expand its business beyond renting physical DVD copies by providing services online, ultimately going out of business. Comparing AMC with Blockbuster illustrates the two different results with the implementation of a last-gasp effort.

The companies were selected based on their alignment with the scenarios described. This allowed for a diverse representation of industries and outcomes, ensuring that each relationship captured a different aspect of corporate survival.

Data Analysis

The analysis was conducted through a comparative framework that focused on identifying patterns and differences in the strategic responses of each company. IBM and Nokia were compared to assess how companies that did not undergo last-gasp efforts fared, with a particular focus on long-term resilience versus loss of market dominance.

AMC and Blockbuster were analyzed as counterparts, providing insight into how timely adaptation (AMC) contrasted with delayed responses (Blockbuster) impacted their fates. Each case was evaluated based on the effectiveness of their strategies, considering both quantitative

outcomes (e.g., revenue, market share) and qualitative factors (e.g., management decisions, external pressures).

Results

IBM

The sale of its PC division to Lenovo in 2005 marked a pivotal shift from a product-oriented to a service-oriented business model. The company also invested significantly in cloud computing and generative AI. The company's strategic focus on emerging technologies and the quick decision-making by the leaders enable IBM to maintain its leadership position in the industry and not get eliminated by the evolution of technology.

Nokia

The company's decline from the world's biggest cellphone manufacturer began with its inability to compete efficiently in the smartphone market. Its slow adaptation to the industry, enhanced by internal management issues, quickly dragged Nokia down almost to bankruptcy.

AMC

The company's innovation in theater experiences, including the creation of IMAX and dine-in theaters, changed its product from the movie itself to the experience of watching a movie, which highlighted its adaptability to the emerging streaming market.

Blockbuster

The rise of streaming services challenged Blockbuster's dominance in the video rental market. However, because Blockbuster did not recognize this potential opponent, it struggled to adapt its traditional rental model to the digital age, which led to its bankruptcy in 2012.

Discussion

IBM

IBM is an international technology and consulting company. It develops software and provides technology consultation solutions. Since the 1950s, the company has led the technology revolution. IBM is known for innovating and creating cutting-edge technology products, including mainframe computers, cloud computing, and generative AI research.

Founded by merging three companies, IBM was created by the combination of the Computing Scale Company, the Tabulating Machine Company, and the Time Recording Company (Reed, 2020). Its original name was Computing-Tabulating Company (CTR), specializing in producing equipment inside a standard office such as machines for accounting and calculating machines, time recorders for business, and punch card systems (Reed, 2020)., Therefore, in the early stages IBM "focused on standard office products rather than being the invention-driven company" (Reed, 2020). Under Thomas Watson Jr.'s leadership, IBM's ability

to pivot during the technology era was reinforced when the company transitioned from the typewriter era to the computer era, setting the foundation for future market dominance at that time.

However, competitors like Microsoft and Apple competitors, during the PC domination period, posed a significant challenge to IBM's traditional business models. Except for IBM, most of the well-known brands in the PC industry began to slowly gain market recognition from 1990 to 1992 (see Figure 1). IBM's net income decreased considerably (see Figure 2). Recognizing the shifting paradigm, and because IBM previously had earned all the resources and fame, including the intelligent employees, the company focused on a different approach, providing technological services and solutions instead of hardware. In 1992, IBM officially formed "IBM Consulting Group, as a new management consulting organization with service lines in Business Transformation and IT Strategy Consulting." (IBM Consulting). Not out of their surprise, Apple came out with its Macintosh Color Classic in 1993.

Moreover, IBM sold the computer division to Lenovo on May 3, 2005 (Acquisition of the IBM PC Business to Lenovo). This transition from a product-oriented approach to a service-oriented approach reflected the company's agility and willingness to evolve to fit in with the ever-changing times. Its innovations have been further exemplified by its commitment to cloud-computing and generative AI language processing in recent times. By embracing emerging technologies and investing in research and development, IBM positioned itself as a pioneer and leader in the digital age, offering cutting-edge solutions to businesses worldwide. The company's rapid development in the industry is supported by the quick adaptation to the changing era and by the strategic collaboration and partnership it has developed over the years with companies like Microsoft and Samsung, giving them more resources at hand for development and innovation.

Figure 1: Most of the PC brands began to gain some market recognition starting from the 1980s and 1990s, posing considerable threats to IBM

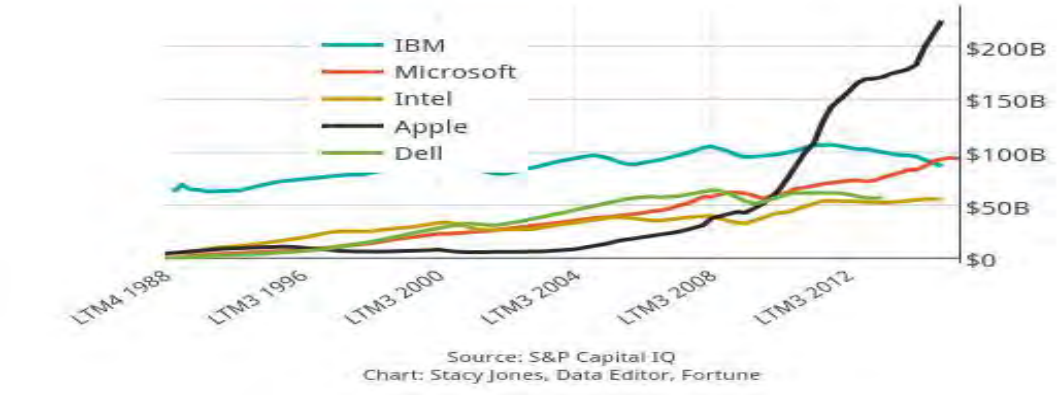
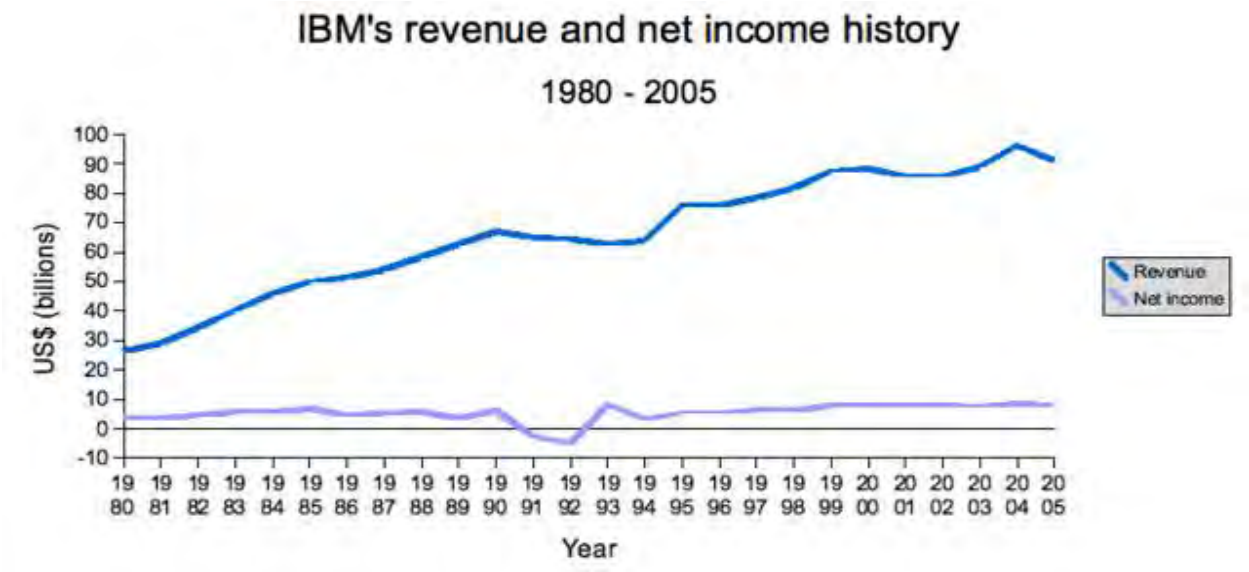


Figure 2: During the period when other PC industry competitors gained market recognition, IBM's revenue and net income took a considerable hit



What IBM did was not the last gasp. Instead, the company went on another path of directing itself into another business approach when it sold the computer division to Lenovo. IBM changed its product completely, so it is no longer competing in the hardware industry and was more associated with service industries. Nevertheless, the last gasp requires a company to surpass a new technology with an old one in the same industry, so the actions taken by IBM are not a last gasp.

Several factors contributed to IBM's success. Lou Gerstner, the company's leader at that time, was credited for shifting IBM's focus to a product-centered. Before his appointment, Gerstner worked outside the company and had no affiliation with IBM's leadership team, which allowed him to bring a fresh perspective free from the constraints of the company's traditional culture. Secondly, IBM's transformation into a service company required an awareness of customers' needs. With its extensive history of inventing products and systems in the tech industry, IBM possessed invaluable insights into the market and recognized what was missing to satisfy customers. Finally, IBM's talented employees played a major role in its success in becoming a service-oriented company. Successful consulting demands the skill to deal with complex problems-and satisfy customers. Both were crucial in promoting and building the company's reputation in the new industry. In summary, IBM's shift to a service company was driven by a leader with a fresh perspective, the company's deep understanding of customer needs, and the expertise of its employees.

Changing the business focus of a large company like IBM requires significant effort, but this shift was crucial for the company to thrive during the technological revolution. IBM leveraged its existing resources effectively and avoided the risk of being outpaced by competitors, even as new technologies emerged. If IBM were to make a last gasp when it is

competing in the smartphone and PC industry with other companies such as Apple or Google, what it might do is put most of its resources into researching a better computer that has higher graphic resolution, higher process speeds, better internet connection, and more. The company is one of the perfect examples of avoiding being compelled to do the last gasp and giving itself a bigger chance of survival by adapting to something it is more familiar with.

In conclusion, IBM's journey from a hardware company to the global technology and consulting giant it is today was built on its adaptability to ever-changing times, commitment to innovation, and strategic responses to market dynamics. It will continue to lead and influence the future of technology and businesses.

IBM Timeline: The time frame within the red box is when IBM made critical changes to survive and stay competitive



NOKIA

Nokia Corporation is a Finnish international telecommunications electronic company that reached its peak in the early 2000s. It created several breakthroughs in the telecommunication category and became the largest telephone manufacturer in the world before smartphones appeared. With its origin dating back to 1865 as a paper mill, the company has left a deep mark on the road of global connectivity.

Nokia started as a pulp mill in 1865 in a small town in Finland founded by Fredrik Idestam. Its journey toward one of the biggest telecommunication companies started with the production of rubber, forestry, cable, and electronics. It went through several transitions throughout its history. In the 1980s and 1990s, the company experienced several major developments in the telecommunication area. Nokia also sold the rubber division of the company to turn itself into a telecom-based company. It reached the pinnacle of telecom history in 1998 when it became the world's number one phone manufacturer, holding 49.4% of the entire mobile

phone market share (Verma, 2023). The downfall of Nokia occurred shortly after the company's golden days, with the emergence of smartphones and internal issues that tore the company up, as a result, the telecom division was later sold to Microsoft but was bought back and revived by HMD in 2016 for continuing the development in 5G technologies.

The company started to change into the electronic generation in the 1900s when the leadership of Kari Kairamo took over. Nokia made breakthroughs in the 1980s when it developed the world's first international cellular data and made the first mobile phone called Mobira Senator. In the 1990s and 2000s, Nokia became the world's biggest manufacturer of mobile phones. Nokia's annual revenue went from less than 10,000M euros in 1997 to over 50,000M euros in 2007, as shown in Figure 3. With its durability and innovation in the telecommunication category, Nokia captured the hearts of consumers worldwide. However, not long after its peak era, it stepped down from the historical stage for several reasons. First, Nokia's rise in the mobile phone industry is the result of its ability to innovate, for example, 3G technology and camera phones were all first introduced by Nokia, but as they focused more on manufacture instead of innovation to fulfill the demand for their phone, along with its decreasing popularity shortly after its peak, which is the reason why the price of Nokia's phone dropped as shown in Figure 4, and allowed other companies such as Apple and Google to enter the competition with their newly introduced smartphones, in which Apple and Android dominated the market not long after the start of the downward trend for Nokia as shown in Figure 5. Second, the internal struggle with the top management caused the company to make poor decisions, in 5 years, Nokia's CEO changed twice and did not give time for employees to adjust to the new vision proposed by their new leader, there's also a problem with the decision-making speed of the top management, it took them too long to make up a decision, and because of that, "Way before the release of the iPhone, Nokia's research department had already come up with the idea. But because of the corporate culture that was prevailing, it never saw the light of the day"(Verma 2023). Last but not least, the ultimate wrong decision to focus on traditional phones instead of coming out sooner with new phones containing new OS systems, allowing other companies in the smartphone industry like Samsung to become more established. A positive example of innovation is Apple, as shown in Figure 6, Apple's revenue growth soared not long after the company invented the iMac, iPod, and iPhone, while Nokia's cellphone market share was being pushed out by the smartphone industry.

Figure 3: Nokia reached its peak during the 2007-2008 period

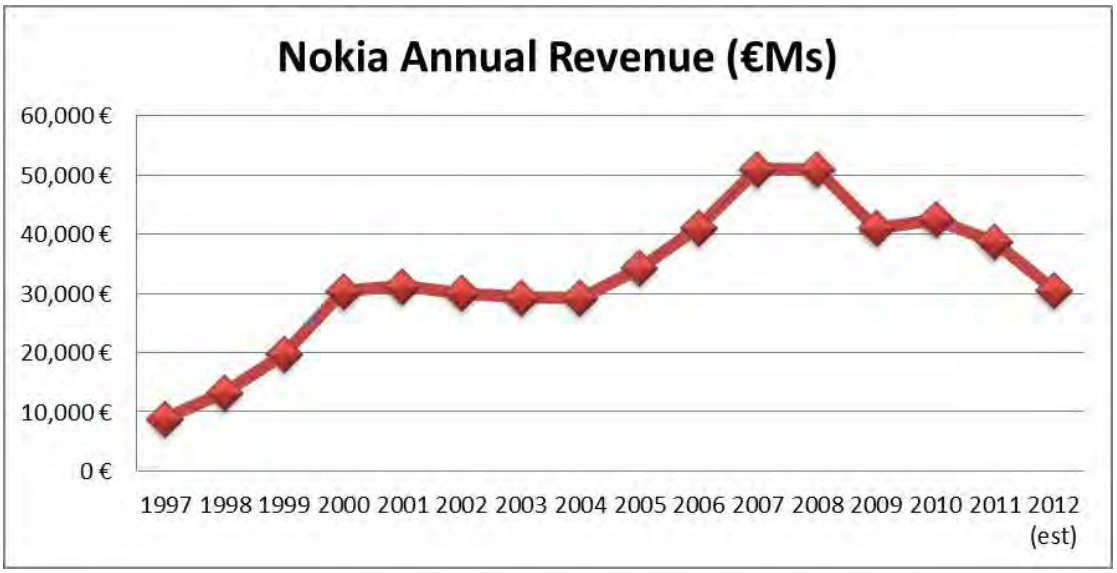
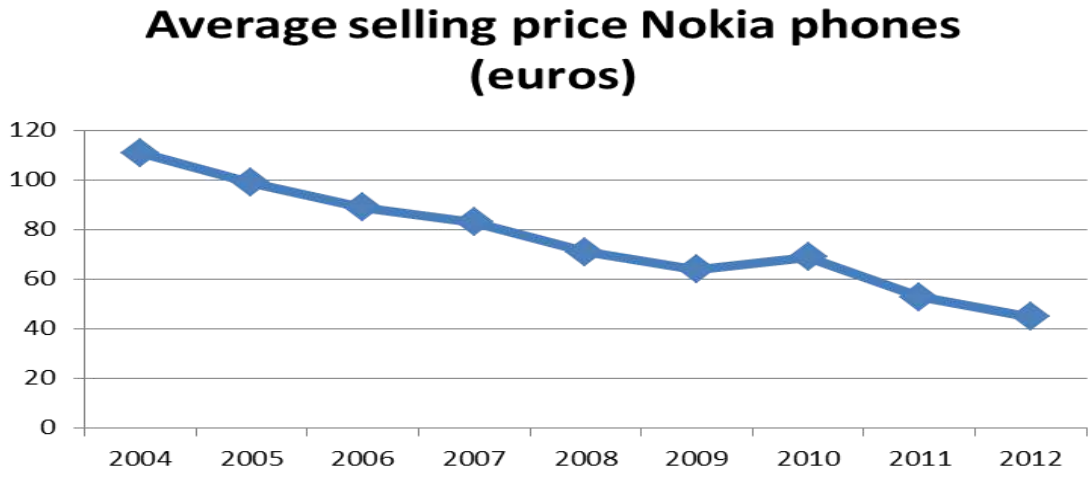


Figure 4: Before 2007 and 2008, the decrease in the prices of Nokia phones was due to the mass production, but after 2008, it was due to the emergence of iPhone



In response to shifting market dynamics and its unpreparedness, Nokia underwent a series of strategic transformations to secure its position in the digital age. The company forged partnerships with Microsoft, adopting the Windows Phone platform to regain market share from iOS and Android devices after losing the chance to work with Android OS to stay in the competition. Additionally, Nokia expanded its portfolio into network infrastructure technologies. Despite these efforts, Nokia ultimately failed to keep pace with the rapidly evolving era of smartphones. It sold its mobile phone division to Microsoft in 2014, marking the end of an era for the Finnish giant. However, in recent years, Nokia Network, a network division of the brand Nokia, has undergone a resurgence, focusing on network infrastructure, 5G technology, and

digital health solutions, HMD has also released an Android OS-based smartphone under the brand name Nokia (Verma 2023). Nokia has again positioned itself as a leader in the telecommunications industry, committed to not repeating the same mistake twice.

Figure 5: Symbian, Nokia's operating system, was surpassed by IOS and Android not long after its peak period

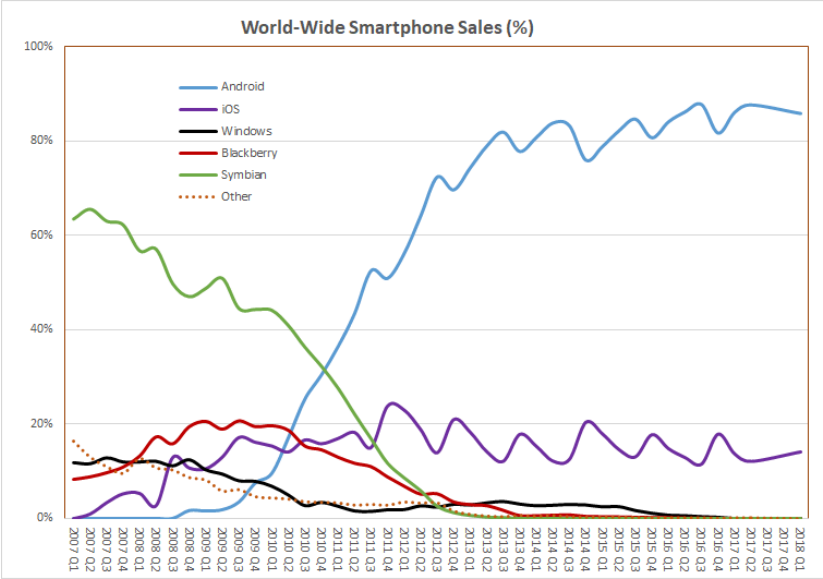
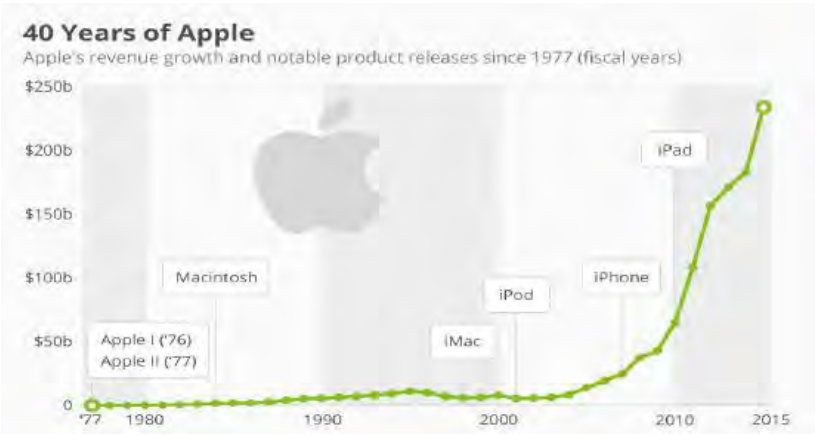


Figure 6: the growth of the Apple Comp corresponded with the downhill of Nokia



Nokia Timeline: The time frame within the red box indicates the downfall of the Nokia



AMC

AMC is an American movie theater founded in 1920 by Maurice Edward and Barney Dubinsky. Historically known for its movie theaters and cable television network, AMC is an example of a successful provider of theatrical experiences. It became the world's first megaplex theater in 1995, but despite this, it faced a rough journey to try and adapt to the changing society. AMC enjoyed significant success as a provider of cinema experiences and movies. However, with the emerging technology of video streaming and the changing consumer wants, formidable challenges were presented to test AMC's adaptiveness towards an ever-changing society.

The rise of streaming services such as Netflix and Amazon Prime changed the traditional way consumers view businesses like AMC, and the company has to overcome several setbacks to succeed in the new society. First off is a decline in theater attendance, as the streaming service becomes more convenient and less expensive, consumers opt in more for the convenience of streaming movies and enjoying the cheap cost of streaming from the comfort of their homes than the inflexibility of the movie schedules of movie theaters, including AMC. Illustrated in Figure 7, it's not so obvious but there's a little dip between 2004 and 2006, when Netflix became popular in the US, which is a sign that the streaming service did impact AMC. Moreover, due to the COVID-19 pandemic, theaters were forced to close down due to safety measures, accelerating the shift to digital viewing. The figure shows a gigantic revenue drop during the year 2020. Second, the shifting of consumer preferences, with the new generation slowly dominating the market, makes them more accustomed to on-demand access to a wide variety of movies and TV shows instead of the traditional inflexible viewing options provided by movie theaters. Third, there is competition in the prices of streaming and movie theaters. Nowadays, streaming services offer a very competitive price for accessing a vast library of content, giving AMC the pressure to readjust its price range and business model to remain competitive. As we can see from Figure 8, even the most expensive streaming service costs only around $20 a

month, but an average American person “spends over 21 hours per week (on streaming services) – equivalent to the commitment of a part-time job” (Durrani, 2024), which is well worth the price, while movie theater tickets for one movie were fluctuating between $9 and $12 since 1995 as shown in Figure 9, with that comparison, we can understand why streaming services is such a big competitor for movie theaters.

Lastly, technological advancements, with the widespread adoption of smart TVs and higher-speed internet, make it much easier and faster for people to stream at home using online streaming services. To counter all of this, AMC made a brilliant decision to shift its main product from the movie to its associated experience. “(Partnering) with IMAX to bring IMAX large-screen digital projection” in 2007, “launched dine-in theaters” in 2008, and “partnered with Sony and RealD to implement digital and 3D technology” in 2009 (AMC history) are all efforts made by the company aimed to enhance the theatrical experience and substantially differentiate the experience that consumers receive from AMC than at-home viewing options, appealing to the audience with its memorable entertainment experiences. The reason this change made by AMC could count as a last-gasp effort is that it fits with the criteria, as it is a big change made at a time when the company is suffering loss due to the rise of technology.

Despite all the adjustments AMC had made to attract more customers, the rise of the COVID-19 pandemic had been a critical hit to the company’s business, forcing them to close down most theaters temporarily due to the severe spread of the virus. While the company implemented safety measures in theaters and explored alternative revenue income sources, such as online video releases like streaming services, the pandemic inevitably weakened the whole company. Still, luckily, AMC at that time had enough cash on hand to operate through the pandemic, even when they closed down all of their stores for a certain period (Gren, 2022).

The company’s willingness to embrace digital innovation and enhance its movie experience is the main adaptability keeping it prospering. Looking ahead, AMC will remain a prominent player in the movie theater industry while leveraging its strength and staying attuned to fit the needs and preferences of its consumers.

Figure 7: Two major revenue drop results from streaming services and COVID-19



Figure 8: the cost-effectiveness of streaming services is so much more than movie theaters since going to the movie once costs around $10, but having streaming services gives one access to thousands of movies for only $20 per month.

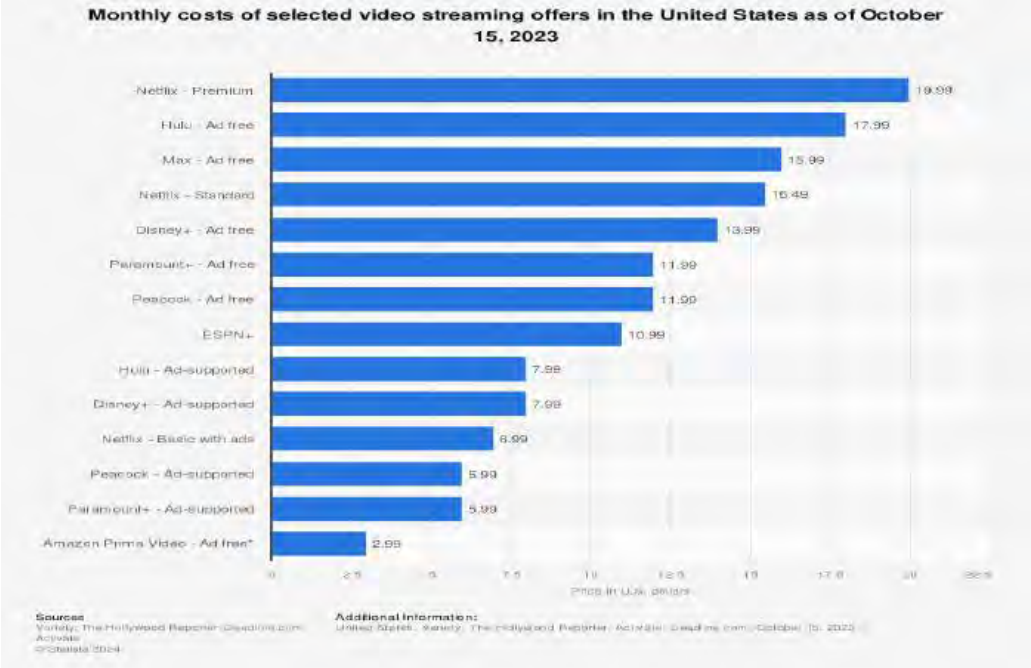
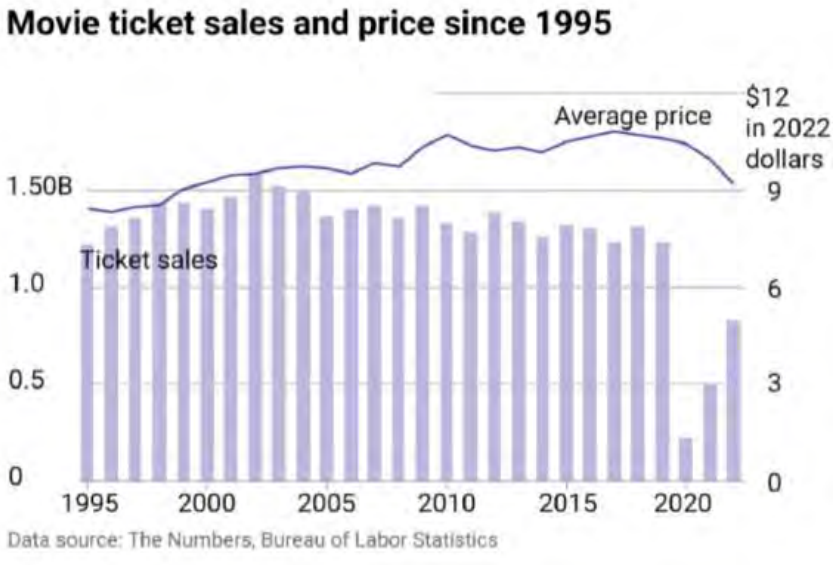
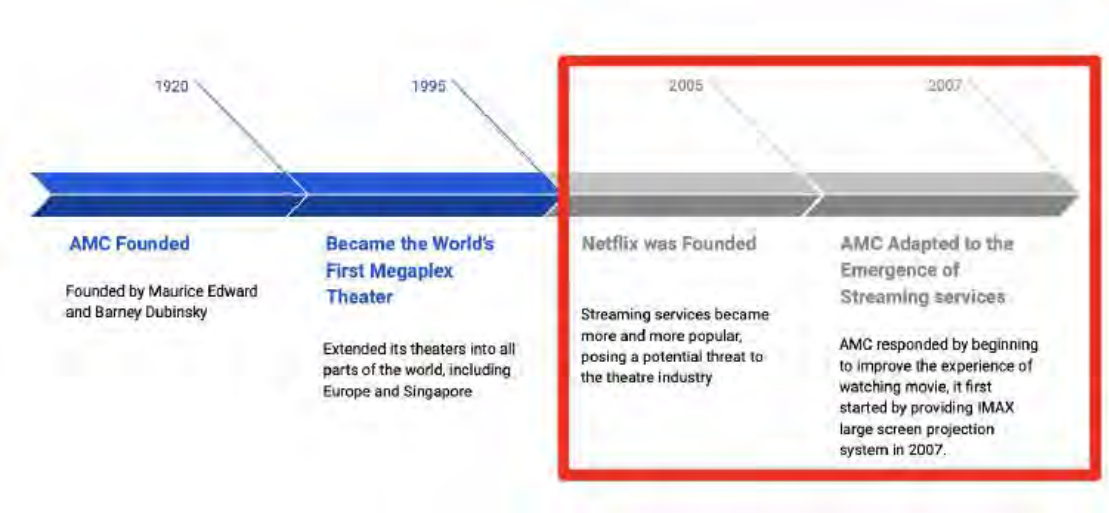


Figure 9: As mentioned in Figure 8, the cost of one movie ticket fluctuates around $9 in the US.



AMC Timeline: The time frame within the red box is when AMC initiates a successful last-gasp effort



BLOCKBUSTER

Blockbuster is a multimedia brand and former video rental store chain. Founded in 1985, the company grew into a global phenomenon with thousands of stores worldwide during the 1990s and early 2000s. However, the company is too reliant on its traditional business model of renting physical copies of DVDs and has become increasingly outdated. Technological advancement, shifting consumer preferences, and a failed attempt to enter the new streaming industry all contributed to its eventual downfall in the 2010s.

Blockbuster's downfall was not just a result of one factor but a convergence of various challenges stemming from technological advancements, changes in consumer preferences, and a slow response to market dynamics. Blockbuster was synonymous with video rental at its peak, having thousands of stores worldwide. However, as the digital era emerged, the company struggled to adapt its traditional business model to the rapidly evolving landscape of media consumption The rise of on-demand video streaming platforms, propelled by advancements in technology and widespread access to high-speed internet, fundamentally transformed how consumers accessed and consumed entertainment content. Consumer preferences of streaming movies and TV shows at home without visiting the rental store were becoming increasingly popular and easier to access than ever. Not only for its convenience, Blockbuster stores failed to compete with video streaming in terms of availability range. While Blockbuster had a wide variety of movies and TV shows, a single Blockbuster store could not hold that much content. As a result, even though Blockbuster had the resources, it was not easy for individuals to get a hold of the movies they wanted, which was no match for the watch-as-you-go content offered by the streaming platforms.

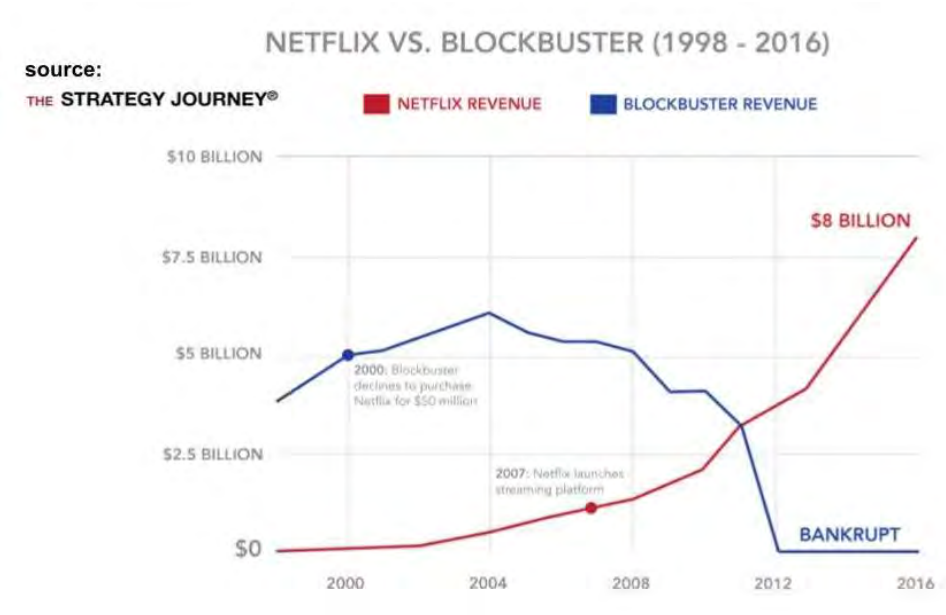
Moreover, the pricing offered by streaming platforms directly challenged Blockbuster's business model. While Blockbuster was still in business, it charged “ $2.99 for new releases and $1.99 for older films. All movies will cost 99 cents for additional days" (Fritz, 2011). Still, people nowadays spend “over 21 hours per week” (on streaming services), which, if we did the

math correctly, would cost a lot more than streaming services (Durrani, 2024). Also, coupled with the convenience of access to the internet, streaming services' cost-effective model further eroded Blockbuster's competitive edge.

Technological advancement is also one aspect that is driving Blockbuster out of business. With the widespread use of faster and more reliable online digital content, the traditional way of accessing a movie or TV show by DVD is becoming increasingly inconvenient. BlockBuster faced pressure to adapt to online business models and invested in digital library content. Still, because of the company's slow response to the changing market dynamic, it reduced its capacity to remain competitive. When video streaming first came out, the CEO of Blockbuster made a very arrogant decision to not consider the industry as its competitor and refused to purchase Netflix and get involved in the video streaming industry until it was too late to do so. In 2004, Blockbuster launched Blockbuster On Demand and Blockbuster Total Access along with getting rid of its late fees (Unglesbee, 2019) in response to the steadily growing revenue and fame of Netflix, as shown in Figure 10, but failed to get a foothold in the new video streaming category as it realized its wrong decision too late. All this effort went in vain as Blockbuster went bankrupt in 2012 because of the growing video streaming industry.

Ultimately, Blockbuster's downfall serves as a caution to the importance of adaptability and innovation in the face of technological disruption. The company's failure to anticipate and respond effectively to the shifting dynamics of media consumption led to its eventual downfall, highlighting the need for businesses to continuously evolve to meet the ever-changing needs and preferences of consumers in the modern age.

Figure 10: since when Netflix launched its streaming services in 2007, its revenue has been growing steadily, while Blockbuster was going downhill steadily



Blockbuster Timeline: The time frame within the red box showed the late realization and the unsuccessful last-gasp of Blockbuster, leading to its bankruptcy



Conclusion

The current research emphasized the determining factors of the success rate of a company's last gasp. The contrasting history of the four different companies, IBM, Nokia, AMC, and Blockbuster highlighted the importance of decision-making in the success of the last gasp. IBM's history of smoothly navigating through turbulent eras without the need for making a last gasp tells us the importance of leaders with forethought. By making the right decisions most of the time, IBM is consistently staying in the top tiers of the market. In contrast, Nokia's downfall from the No. 1 cellphone producer serves as a negative example of the failure of innovation once it dominated the market.

AMC and Blockbuster are examples of companies that examined the importance of leveraging resources in the highly competitive market. While both AMC and Blockbuster were influenced by the uprising of the new video streaming industry, they had very different outcomes regarding adapting to the new market. AMC switched focus very quickly and successfully captured the preferences of the consumers through strategic changes in the products they sell. Instead of focusing on the movie, AMC aimed to give the audience a better movie-watching experience. On the other hand, Blockbuster's reluctance to embrace the fact that movie streaming is their competitor ultimately sealed the company's fate of elimination. In summary, the success rate of the last gasp of a company is influenced by its ability to utilize its resources effectively, the foresight of the leadership, and the effectiveness of breaking free from industrial inertia and embracing innovation.

Acknowledgments

I would like to express my deepest gratitude to Dr. Andrew Foley, who taught me how to write a research paper and guided me through the research and writing of this paper with his invaluable insights and expertise. I am also incredibly thankful to Dr. Timothy Belton, my writing coach, for his advice about my paper and constructive feedback, which significantly improved the quality of my work.

Special thanks to Ms. Harishree Khunt for her assistance in the publishing process, ensuring that my paper reached a wider audience. I am also grateful to Ms. Yamini Yoshi, my program manager, who helped me connect with the wonderful mentors and coaches mentioned above.

This research and writing journey would not have been possible without the support of the Lumiere program, which provided me with the resources to engage in this academic opportunity. Thank you all for your encouragement and support throughout this process.

Works Cited

R. Adner, R. Kapoor. Innovation ecosystems and the pace of substitution: Re‐examining technology s‐curves. Strategic Management Journal. 37, 625–648 (2015). https://doi.org/10.1002/smj.2363

C. Perez. Technological revolutions, paradigm shifts, and socio-institutional change. In Globalization, Economic Development and Inequality: An Alternative Perspective (E. Reinert, Ed.), doi:https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://carlotaperez.org/wp-content/downloads/publications/organizational-change/TRs_TEP_shifts_and_SIF_ch.pdf&ved=2ahUKEwjc37bznY-HAxUtdDABHZ8vDWQQFnoECBcQAQ&usg=AOvVaw2aHVPxLMeeIoepCKM02xHv (1998).

R. Henderson. Underinvestment and incompetence as responses to radical innovation: Evidence from the photolithographic alignment equipment industry. RAND Journal of Economics. 24, 248–270 (1993).

TheStreet Staff. History of IBM. https://www.thestreet.com/personal-finance/history-of-ibm (2024).

The evolution of IBM: A century of resilience and innovation. The Business Anecdote. https://www.thebusinessanecdote.com/post/the-evolution-of-ibm-a-century-of-resilience-and-innovation (2023).

Microsoft inches past IBM in revenue. Fortune. https://fortune.com/2015/08/05/microsoft-inches-past-ibm-in-revenue (2015).

IBM consulting. Wikipedia. https://en.wikipedia.org/wiki/IBM_Consulting#:~:text=In%201992%2C%20the%20company%20formed,manager%20of%20IBM%20Consulting%20Group (2023).

Acquisition of the IBM PC business by Lenovo. Wikipedia. https://en.wikipedia.org/wiki/Acquisition_of_the_IBM_PC_business_by_Lenovo#:~:text=The%20acquisition%20of%20IBM%20Personal,completed%20on%20May%203%2C%202005 (2023).

Macintosh color classic. Wikipedia. https://en.wikipedia.org/wiki/Macintosh_Color_Classic (2023).

File: IBM's revenue and net income. Wikipedia. https://en.wikipedia.org/wiki/File:IBM%27s_revenue_and_net_income.png (2023).

Broad Institute. Gerstner Family Foundation. https://gerstner.org/program-areas/broadinstitute.org (2023).

Gerstner: Changing the culture at IBM. Harvard Business School Working Knowledge. https://hbswk.hbs.edu/archive/gerstner-changing-culture-at-ibm-lou-gerstner-discusses-changing-the-culture-at-ibm (2023).

Lou Gerstner. Wikipedia. https://en.wikipedia.org/wiki/Lou_Gerstner?scrlybrkr=bb81019e (2023).

IBM history. History Tools. https://www.historytools.org/companies/ibm-history (2023).

IBM consulting. Wikipedia. https://en.wikipedia.org/wiki/IBM_Consulting#:~:text=In%201992%2C%20the%20company%20formed,manager%20of%20IBM%20Consulting%20Group (2023).
The history of Nokia. Feedough. https://www.feedough.com/the-history-of-nokia (2023).
Nokia timeline. History Draft. https://historydraft.com/story/nokia/timeline/349 (2023).
Why did Nokia fail? Feedough. https://www.feedough.com/why-did-nokia-fail (2023).
Sprint versus Nokia: A tale of two telecom turnaround stories. Seeking Alpha. https://seekingalpha.com/article/685991-sprint-versus-nokia-a-tale-of-2-telecom-turnaround-storie (2023).
Disruptive innovation. Disruptive Innovation. https://disruptiveinnovation.se/?p=131 (2023).
File: Worldwide smartphone sales share. Wikipedia. https://en.wikipedia.org/wiki/File:World_Wide_Smartphone_Sales_Share.png (2023).
Apple's revenue since 1977. Statista. https://www.statista.com/chart/4574/apples-revenue-since-1977 (2023).
History of AMC Entertainment. The Street. https://www.thestreet.com/investing/history-of-amc-entertainment (2023).
AMC history. AMC Theatres. https://www.amctheatres.com/corporate/amc-history (2023).
How AMC Entertainment survived the pandemic. Industry Leaders Magazine. https://www.industryleadersmagazine.com/how-amc-entertainment-survived-the-pandemic (2023).
North American box office gross revenue since 1980. Statista. https://www.statista.com/statistics/187069/north-american-box-office-gross-revenue-since-1980 (2023).
SVOD monthly subscription costs the US. Statista. https://www.statista.com/statistics/1110896/svod-monthly-subscription-cost-us (2023).
26. Streaming stats. Forbes. https://www.forbes.com/home-improvement/internet/streaming-stats (2023).
How movie ticket prices and sales have fluctuated over time. Giggster. https://giggster.com/blog/how-movie-ticket-prices-and-sales-have-fluctuated-over-time (2023).
Blockbuster: Its rise and fall. Startup Talky. https://startuptalky.com/blockbuster-its-rise-and-fall (2023).
A timeline of Blockbuster's ride from megahit to flop. Retail Dive. https://www.retaildive.com/news/a-timeline-of-blockbusters-ride-from-megahit-to-flop/564305 (2023).
Blockbuster, Netflix, Coinstar. Forbes. https://www.forbes.com/2010/05/18/blockbuster-netflix-coinstar-markets-bankruptcy-coinstar_slide.html?sh=2e810de75464 (2010).
Comparing revenue for Blockbuster and Netflix 1998-2016. Slow Reveal Graphs. https://slowrevealgraphs.com/2022/10/24/comparing-revenue-for-blockbuster-and-netflix-1998-2016 (2022).
Blockbuster bankruptcy. Los Angeles Times. https://www.latimes.com/business/la-xpm-2011-may-28-la-fi-0528-ct-blockbuster-20110528-story.html (2011).
Streaming stats. Forbes. https://www.forbes.com/home-improvement/internet/streaming-stats (2023).

The Theory of Natural Selection and How it Affects Humans
By Fiona Allendorf

The theory of natural selection applies to all species, but some humans don't think it applies to them. Has the growth of civilization affected natural selection in humans? Or are we being selected against now?

The theory of natural selection ties in with the theory of evolution. In short, natural selection is survival of the fittest. Organisms with beneficial traits are more likely to survive and pass on their DNA. This is the opposite for species with harmful traits. Through this process, species evolve towards their environment.

The purpose of life is simply to reproduce. This may not align with what humans deem to be the purpose of their life, but nonetheless from a biological standpoint it is true. Animals' goal is to survive long enough to pass on their DNA. The more offspring you produce, the more likely you are for one of your own to survive. It is commonly believed that organisms are just trying to survive, but for what purpose? No matter what species you are, death is inevitable. For species that don't have concepts such as careers, what is driving them to live? It is the instinct to reproduce.

When people think about beneficial traits in animals, we think of strength, speed, and other similar adaptations. This is because these traits are what help an organism survive in the wild. But what about in society? What traits are beneficial to surviving in society or producing offspring? People often think about factors such as beauty when choosing people to marry or have children with. With this in mind, being beautiful could be a beneficial trait. So, are less attractive people being slowly selected against? It is a possibility — one that may never be confirmed or denied based on the rate of evolution.

There is also another question: are humans cheating natural selection? By using resources like medicine, we are stopping natural death. Physical or mental characteristics that would have died out in the wild flourish in society. By saving those who would die in the wild, are we creating a weaker population? Or, are we simply redefining what a strong population is?

Humans have changed the world in many ways. We have caused global warming; we have built infrastructure. These changes to our environment around us change the way animals around us evolve as well. For example, birds who can build their houses on telephone poles or insects that can get into houses are more likely to have access to shelter and food. There is also the factor of human emotions. Let's say you are looking on an endangered species list and choosing which organizations to donate your money to. You see one cute animal and one less attractive animal. Which will you give your money to? Most people would choose the cute one. Over time, animals might evolve so only the ones deemed "cutest" survive. This is all because humans have a great impact on our earth. We are an invasive species.

Do humans cheat natural selection? I don't believe we do. Instead, I think we change the parameters of what makes a trait beneficial — for we have become a crucial part of the environment to which species live and evolve around.

The Impact of Music on Plants: A Review By Quinne Fang

Abstract

Music is observed to have a positive effect on humans. Plants, like us, are also living objects that breathe and grow. Scientists are in disagreement as to whether plants are capable of perceiving and reacting to music because they lack nervous systems. This raises an intriguing question: could music potentially serve as an alternative to fertilizers, or even pesticides? I conducted a literature review to assess current knowledge regarding the impact of music on plants. Results yielded prominent studies that investigated the effect of different types of sound waves (i.e., music) on marigold growth, alfalfa seed germination, and lettuce seed germination. Ultimately, it can be concluded that the impact of music on plants differs from plant to plant. In most cases, low frequency, soft, and/or harmonious sound promotes plant growth whereas disruptive or non rhythmic sound hinders plant growth. These findings suggest that music treatments might open a new eco-friendly pathway for farmers to plant and grow crops. Future research should further explore the specific mechanisms by which plants respond to sound in order to determine the most effective sound frequencies for promoting plant health.

Keywords Plant, music, sound, sound waves, agriculture, pesticide

Introduction

Plants are one of the most critical and abundant groups of organisms on earth. Plants comprise 80% of all biomass on Earth, weighing in around 450 gigatons of carbon (Gt C) (Bar-On, 2017). Ultimately this provides the foundational support for ecosystems. They provide the basis for healthy habitats for wildlife and humans, releasing oxygen, absorbing carbon dioxide, and regulating the water cycle. Additionally, plants support food webs from the bottom up (Weathers et. al., 2013), serving as primary producers that sustain herbivores, which in turn support carnivores and omnivores. Furthermore, plants also comprise a significant proportion of the biodiversity on the planet (Vellend et. al., 2017).

However, data supporting plant abundance indicates a decline of 69% over the past 50 years (Whiting, 2022). According to the 2020 State of the World's Plants and Fungi Report, about 40% of the world's plants are at risk of extinction (Igini, 2022). Plant biodiversity is also in jeopardy. Such declines are attributed to challenges such as pest infestation, extreme weather, deforestation, and climate change (Ellis et. al., 2021). In particular, climate change is a growing threat to plant survival and diversity. Climate change has been attributed to limiting plants' accessibility to nutrients and thereby stunting their growth (Thuiller et. al., 2005). Therefore, drought-resistant crops have been designed to combat challenges posed by climate change on crop yields. In addition to climate change, pests are simultaneously threatening plants thereby necessitating the use of pesticides and fertilizers. Such chemicals are widely used in agriculture to bolster plant growth and/or to control weeds, insect infestation, and diseases. Although the use of such chemicals can control pests efficiently, they can impose significant negative impacts on

the environment and human health, for example via soil degradation and water contamination (Marsh, 2024). More environmentally friendly alternatives are being sought to protect and grow plants. With this, scientists have recently begun experimenting with a possible replacement for fertilizers, pesticides, and other chemicals – perhaps unexpectedly, music.

Early research has demonstrated that music may have a positive effect on humans. Music can enhance physical and mental wellbeing: improving heart rate, motor skills, brain stimulations and immune system enhancement (Revecchini, 2021). Stemming from these preliminary studies, researchers are now trying to extend such applications to plants. Conceptually, the extension of music from humans to plants has arisen because both are organisms that breathe and grow. Research has led scientists to believe plants have the ability to perceive and react to music. Experiments involved the audible playing of different types of music and to test plants' ability to respond. Ultimately, researchers concluded that plants can detect vibrations that act as a stimulus to them. But how can plants perceive and respond to sound waves if they lack the ability to "hear"? If so, does it mean music can be a potential replacement for harmful chemicals and pesticides?

To enhance our understanding of a new and burgeoning field, I conducted a literature review to investigate the current body of knowledge regarding how plants perceive and respond to sound waves—in particular, music— in their environment. My objective is to identify major findings in this field of study, including whether specific plant species and/or musical genres have been a primary focus for research. The scope of the study focuses on the effects of different types of music on growth rate of flower (marigold) and crop (lettuce). The literature review is limited to open-access research articles, or articles otherwise accessible elsewhere, and therefore may omit additional relevant studies. First, I provide a summary of the ecological and anatomical explanations for how plants perceive and respond to sound waves. I then summarize prominent research findings regarding the relationship between music and plants. Lastly, I provide suggestions for how this field of study can be further extended to address various challenges faced by plant growth and production for both wild and domesticated plant species.

Methods

I found peer-reviewed journal articles on the impact of music on plant growth by searching Google Scholar (www.google.com/scholar) using the keywords: ('music' OR 'sound') AND ('plants' OR 'grass' OR 'trees' OR 'flowers' OR 'crops'). Articles were deemed relevant if they involved experiments, review-style discussion, and/or data regarding how music affects the growth of a specific plant species. Articles focused solely on plants or sounds, individually, were excluded because such information ultimately does not explain the correlation between those two topics. For each paper I recorded the plant species and the music type investigated. I also recorded the geographic location of the study, specifically whether the study occurred *in situ* or *ex situ*, and any available environment variables (e.g., temperature). Finally, I recorded the difference in rates of plant growth between controlled and experimental groups, which were analyzed and compared in detail in my broader discussion. The data were recorded in any

manner provided (e.g., plant height, seedling growth), although consistency in both music delivery and plant growth metrics were not always uniform. Ultimately, the gathered information was compared to determine whether there are any generalizable findings and/or if there is still disagreement amongst this field of study.

Results

Chemicals and Plant Health

Plants made their first appearance on Earth approximately 470 million years ago during the Ordovician period. They have been living on earth longer than any other organisms except single-celled organisms. Today, there are about 391,000 species of known plants, with about 2,000 new plant species being discovered every year (Dasgupta, 2016). Plants are perhaps most ecologically noteworthy because of their contributions to oxygen production: the Amazon and other rainforests generate around 28% of oxygen produced on land whereas about 50-80% of oxygen are produced by marine algae and phytoplanktons, microscopic organisms found in aquatic environments (EarthSky. How much oxygen comes from the ocean?, 2020). Nevertheless, plants are important contributors to other key ecosystem processes. For example, they play a crucial role in carbon sequestration, a process by which carbon dioxide is removed from the atmosphere. Today, carbon sequestration is leveraged by environmental scientists, industrial professionals, and communities and governments alike to slow down or reverse the effects of climate change. Forests account for about 44% of carbon sequestration annually, capturing about 2.7 billion metric tons of carbon dioxide (Bai & Cotrufo, 2022). Beyond this, humans and other animals consume plants' nutrients to survive; They are also the source of numerous medication compounds.

Despite plants' contributions to the Earth, their population, size, and biodiversity are gradually declining due to human behaviors such as habitat destruction, pollution, climate change, and over-exploitation of plant resources (Igini, 2022). Air pollutants can affect plant health by directly settling onto leaf surfaces. This can induce malignant spots and cause leaves to premature, ultimately making plants vulnerable and ill. Scientists predict, due to the high demands of production, Earth could lose up to 23% of natural habitat in the next 80 years (Nairn, 2020). Rapid loss of habitats could accelerate the extinction rate of already vulnerable species. Deforestation leads to increases in carbon dioxide levels, increases the likelihood of plants being infected with disease and indirectly affects their biomass production and the density of crop stands.

Domesticated plants, and other plant species grown and harvested for food, are also critical for human livelihoods and wellbeing. As a crucial component in agricultural production, fertilizers and pesticides help to minimize crop losses and enhance the yield and quality of food, ultimately making it more affordable. Today, around 85% of agriculture producers regularly use pesticides to increase their fruit and vegetable production (Khanal & Singh, 2016). The term 'pesticide' encompasses a variety of forms: insecticides, fungicides, herbicides, rodenticides,

molluscicides, and nematicides. Insecticides are specifically designed to prevent insects from destroying crops and plants. Fungicides are used to kill malignant fungi or their spores. Others may specifically target weeds, rodents, and other harmful pests.

Although fertilizers and pesticides are targeted for use on land, they can contaminate nearby water sources such as rivers, ponds, or oceans through runoff. Scientists find that 64% of global agricultural land (roughly 24.5 million km^2) is at risk of pesticide pollution from multiple active ingredients, and 31% of this land is at high risk (Tang et. al., 2021). Once the water becomes contaminated, it can kill fishes and damage marine ecosystems. These chemicals also affect humans: some can disrupt nervous systems, others can irritate skins or eyes; research has shown pesticide and herbicide exposures increase cancer risks (Alvanja et. al., 2013) and pregnancy harm.

Despite these chemicals' short term effectiveness, the harm they bring to our environment is irreversible. They ubiquitously impact biotic and abiotic aspects of the environment. Ironically, their use to optimize plant growth can ultimately harm the environment plants should thrive in. Nevertheless, the growing amount of environmental harm today requires innovative and sustainable solutions – for example, "growing" plants with music.

Music's Impact on Plants

Music by definition is an art of sound in time that expresses ideas and emotions in significant forms through the elements of rhythm, melody, harmony, and color. Although it is an uncommon research topic, its effects on human mental health, behavior, and verbal memory have been analyzed thoroughly. Studies have demonstrated that Baroque music, with a tempo of 60 beats per minute (bpm), significantly improves memorization and learning by activating both brain hemispheres. Additionally, classical music, particularly compositions by Mozart, has been shown to enhance cognitive performance (Ciborowska et. al., 2021). The documentary film "Alive Inside" highlights how doctors utilize music to treat dementia and Alzheimer's disease with the objective of improving the patient's quality of life (Rossato-Bennett, 2014). However, exposure to music is not a panacea for all human mental and cognitive deficits. In Sally Anne Gross's book "Can Music Make You Sick?", Gross describes modern day musicians' experiences of creating music. She explains how the industry's frequently unstable working conditions create an ideal environment that fosters mental health struggles in music creators (Gross, 2024).

Music has also reached its way to impact animals: Researchers have studied the effect of music on animal physiology and behavior. It is also being discussed as a potential solution for improving animal welfare. In one study, researchers sought to determine if music could be utilized to alleviate stress in animals. These researchers used cattle as their focal animal. Amongst the cattle that were exposed to classical or lullaby music, more than 10 cows showed relaxed behaviors and 1 fell asleep in the lullaby group. Alternatively, only 1 cow showed relaxed behaviors amongst the cattle exposed to heavy metal music. The benefits from such relaxation may not solely be mental; Cows in the classical and lullaby group had greater milk yield (Ciborowska et. al., 2021).

Plants, despite their quiet nature, actively perceive and respond to their environments. They sense factors like temperature, light (including day length), touch, and gravity, and react accordingly. Plants display phototropism, or reaction to light. They sense light through special molecules called photoreceptors, which are made up of a protein linked to a light-absorbing pigment called a chromophore. When the chromophore absorbs light, it causes a change in the shape of the protein, thereby altering its activity that leads to the plant's growth movement. Temperature, another environmental stress, triggers the plant's signaling pathways and alters protein shape, affecting membrane fluidity, metabolism, and cytoskeleton rearrangements (Ruelland & Zachowski, 2010). Plants can also sense and react to the direction of gravity, allowing them to align themselves appropriately, which is a process called Gravitropism.

A number of experiments have been conducted and data have been recorded on the impact of different music types on plant species. Chowdhury and Gupta (2015) investigated the effect of light Indian music, meditation music, and noise on the growth and health of marigolds (Chowdhury & Gupta, 2015). The experiment was conducted *ex situ* and divided into three subgroups that differ in music type: Light Indian music, Meditation music, and noise. In each subgroup, two plants were planted in two plots with the same basic conditions required for plant growth (e.g., soil, water, nutrients). One plot was treated with soundwaves, and the other acted as a control group for comparison. In the Light Indian music group, the treated marigold plant performed better overall than the one untreated. Marigold plants treated with Light Indian music displayed a higher number of buds, flowers, and leaves: In week 1, two plants were similar in size. At the end of week 4, the treated marigold showed significant differences in flower size, bud count, and leaf size. Results from the meditation music group were similar to those in the Light Indian music group. The treated plant always grows faster and better than the untreated. After exposure to noise, the treated plant, at the end of week 4, started to wither and bend away from the direction the noise was coming from. The plant was greener on the side with less noise exposure. In summary, calm music which low frequency sound waves facilitate positive effects on marigold plant growth. Contrastingly, noise, a non-rhythmic and unharmonious audio frequencies, has a negative effect on growth of marigold plants.

Adding on to Chowdhury and Gupta (2015), Lai and Wu (2020) conducted experiments regarding effects of music on the germination and seedling growth of alfalfa and lettuce plants (Lai & Wu, 2020). In total, nine music types were used: Control (no music), Gregorian Chant, Baroque, Classical, Jazz, Rock, Nature Sound, New-age, and Waltz. Each plant was placed in a box along with a cell phone playing music. Researchers recorded the decibel (db) value of each music treatment. After 7 days, lettuce seeds treated with Nature sound were found to have the highest germination percentage (100%), while classical music seeds had the lowest germination percentage (93.3%). This observation contradicts my hypothesis on classical music having positive effects on plant growth. In alfalfa seeds, classical, nature sound, and waltz had the highest rate of germination (100%), and rock had the lowest rate of germination (91.7%). This finding is consistent with previous findings claiming that rock music, due to its intense vibrations, hinders plant growth. However, rock music did not make a huge negative impact on

germination of lettuce seeds as much as classical music did; This implies hardcore rhythms like rock music may not always disrupt plant development, and classical music, so-called "harmonious music", may not always improve plant health. Thus, we can conclude the rhythmic qualities of music alone do not fully account for its impact on seed germination. There is no direct correlation between db value and germination percentage.

Discussion

The results of this literature review demonstrate that plants respond differently to various frequencies of sound waves. The research identified in this literature review differs by plant type, music type, environment, and experimental approach. Overall, plants tend to avoid disruptive sound frequencies: In Chowdhury and Gupta (2015), marigold plants treated with unrhythmic sounds (i.e., noises) started to wither after 4 weeks — a finding made even more astounding given the fact that marigolds lack any structural organs that would allow them to process sound waves (Chowdhury & Gupta, 2015). Alternatively, plants have been observed to favor soft and relaxing sound waves (e.g., meditation and light Indian music) by exhibiting significant growth in leaf size. From Lai and Wu (2020), it can be concluded that the impact of music on plant growth is influenced by both plant type and musical genre (Lai & Wu, 2020). I had hypothesized that all music considered "harmonic" (e.g., classical music) would unanimously benefit plant health, but the literature demonstrates that the impact of classical music varies: classic musical was found to have a slight negative effect lettuce seed germination, but a positive effect on alfalfa seed germination (Chowdhury & Gupta, 2015). However, a handful of variables could explain this unexpected outcome: environment, temperature, soil nutrients, or music range. The studies above do not specify these variables, therefore they may be uncontrolled in these experiments. Repeating these studies, with clear and stringent variable control, will benefit our understanding of how common these phenomena are.

Notably, the two studies described above both rely upon visual observations of seeds and plant growth to determine the effect of different types of music. Broadly, both studies agree that music *can* bolster plant growth as determined by observations of plant height, leaf/bud count, and/or germination percentage. Furthermore, Chowdhury and Gupta (2015) hypothesize that music enhances plant growth by promoting faster nutrient absorption from the soil and improving metabolite production, ultimately leading to a higher growth rate (Chowdhury & Gupta, 2015). However, the two studies disagree on the *factors* that affect plant growth and seed germination. One on hand, Chowdhury and Gupta (2015) conclude that all sound that has soft, rhythmic, and audible frequencies (i.e., music) can accelerate seed germination, plant growth, and plant development (Chowdhury & Gupta, 2015). Nevertheless, specific research is lacking on nutrient absorption and metabolite production in the presence of music. On the other hand, Lai and Wu (2020) propose that the rhythmic elements of music alone do not fully account for its impact on seed germination (Lai & Wu, 2020). They suggest that future research should measure the various aspects of music—such as rhythm, melody, harmony, repetition, variation, and instruments—to determine how each element, either individually or in combination, might

influence seed germination or plant growth. Thus, because these two studies only provided visual observations, more quantitative data should be measured to ensure the reliability of the result.

My findings provide preliminary support that music can be a more environmentally-friendly replacement for fertilizers, particularly as a means to bolster plant growth without harming ecosystem health. These results contribute new and unique strategies to the agricultural field by providing evidence and data that reinforce the availability of alternative agricultural practices. Such data-supported insights could persuade agricultural factories to reduce chemical use on crops and adopt natural alternatives in order to mitigate harmful offshoots of environmental practices, such as chemical runoff into waterways.

Impactful future research will likely need to address three key points: First, research should focus on the long-term effects of these new methods. By repeatedly measuring plant growth over multiple plant generations and seasonal variations, we may be able to conclude the statistical significance of any benefits offered by music and ultimately lay groundwork for any large-scale policy development and implementation to benefit both agriculture and the environment. Second, continued research will need to determine the degree of specificity to which music impacts plant growth. Future research could aim to determine whether different crop varieties and/or species uniquely respond to different music genres; from this, farmers could strategize optimal growing conditions specific, or generalized, to their crops. Additionally, if music has the potential to bolster agricultural plant growth, then there may be a chance that it also bolsters the growth of non-target plants (e.g., weeds). This could introduce additional challenges to farmers, thereby placing the overall utility of this approach in question. Third, whereas musical research has focused on determining whether it has beneficial effects on plant growth, there is still a large gap in knowledge regarding whether music can also ward off pests. Pesticides and fungicides have proven extremely efficient and sufficient in protecting plant physical form and health. Thus, even if music offers a strong alternative for plant growth, there are lingering challenges to agriculture that are still most efficiently solved through chemical means.

As we continue to face environmental challenges, it is clear that sustainable agriculture is a significant part in contributing to a healthier future. This literature review uniquely shows that true progress comes not just from higher crop yields, but from preserving the ecosystems that support life. Clear solutions are available, but it is up to us to decide whether we will take action and make a change: whether to continue pursuing short-term gains or choose a sustainable future where thriving farms coexist with healthy ecosystems.

Acknowledgements

I would like to acknowledge the assistance of Lumiere Education and Lumiere mentor Mary Burak for their guidance, support, and feedback throughout the research process.

Works Cited

Alavanja, M. C. R., Ross, M. K., & Bonner, M. R. (2013). Increased cancer burden among pesticide applicators and others due to pesticide exposure. CA a Cancer Journal for Clinicians, 63(2), 120–142. https://doi.org/10.3322/caac.21170

Anindita Roy Chowdhury and Anshu Gupta (2015). Effect of Music on Plants – An Overview, International Journal of Integrative Sciences, Innovation and Technology (IJIIT), 4(6), 30 – 34.

Bai, Y., & Cotrufo, M. F. (2022). Grassland soil carbon sequestration: Current understanding, challenges, and solutions. Science, 377(6606), 603–608. https://doi.org/10.1126/science.abo2380

Bar-On, Y. M. (2017, July 3). The biomass distribution on Earth. PNAS. https://www.pnas.org/doi/full/10.1073/pnas.1711842115

Ciborowska, P., Michalczuk, M., & Bień, D. (2021). The effect of music on livestock: cattle, poultry and pigs. Animals, 11(12), 3572. https://doi.org/10.3390/ani11123572

Dasgupta, S. (2016, May 12). How many plant species are there in the world? Scientists now have an answer. Mongabay Environmental News. https://news.mongabay.com/2016/05/many-plants-world-scientists-may-now-answer/

EarthSky. (2020, June 24). How much oxygen comes from the ocean? | Earth | EarthSky. EarthSky | Updates on Your Cosmos and World. https://earthsky.org/earth/how-much-oxygen-comes-from-ocean/#:~:text=Scientists%20estimate%20that%2050-80%25%20of%20the%20oxygen%20production,can%20photosynthesize%2C%20that%20is%2C%20convert%20sunlight%20into%20energy.

Ellis, E. C., Antill, E. C., & Kreft, H. (2012b). All is not loss: Plant biodiversity in the Anthropocene. PLoS ONE, 7(1), e30535. https://doi.org/10.1371/journal.pone.0030535

Gross, S. (n.d.) Can music make you sick? | Groundbreaking research on Mental health in the music industry. Sally Anne Gross. https://www.sallyannegross.com/can-music-make-you-sick

Igini, M. (2022, November 8). All you need to know about endangered plants species. Earth.Org. https://earth.org/endangered-species-of-plants/#:~:text=Deforestation%2C%20mass-agriculture%2C%20logging%2C%20and%20livestock%20farming%20are%20among,medical%20and%20food%20species%20at%20risk%20of%20extinction.

Khanal, G., & Singh, A. (2016). Patterns of Pesticide Use and Associated Factors among the Commercial Farmers of Chitwan, Nepal. Environmental Health Insights, 10s1, EHI.S40973. https://doi.org/10.4137/ehi.s40973

Lai, Y.-N., & Wu, H.-C. (2020). Effects of different types of music on the germination and seedling growth of alfalfa and lettuce plants. AGRIVITA Journal of Agricultural Science, 42(2), 197–204. https://doi.org/10.17503/agrivita.

Marsh, J. (2024, May 23). Pros and cons of pesticides in agriculture. Environment Co. https://environment.co/pesticides-pros-and-cons/

Nairn, C. (2020, November 12). 3% of Earth's natural habitats could be gone by 2100, study

finds. World Economic Forum. https://www.weforum.org/agenda/2020/11/earth-natural-habitats-destroyed-biodiversity-loss/

Rebecchini, L. (2021). Music, mental health, and immunity. Brain Behavior & Immunity - Health, 18, 100374. https://doi.org/10.1016/j.bbih.2021.100374

Rossato-Bennett, M. (Director). (2014). Alive Inside [Film]. Projector Media.

Ruelland, E., & Zachowski, A. (2010). How plants sense temperature. Environmental and Experimental Botany, 69(3), 225–232. https://doi.org/10.1016/j.envexpbot.2010.05.011

Tang, F. H. M., Lenzen, M., McBratney, A., & Maggi, F. (2021). Risk of pesticide pollution at the global scale. Nature Geoscience, 14(4), 206–210. https://doi.org/10.1038/s41561-021-00712-5

Thuiller, W., Lavorel, S., Araújo, M. B., Sykes, M. T., & Prentice, I. C. (2005). Climate change threats to plant diversity in Europe. Proceedings of the National Academy of Sciences, 102(23), 8245–8250. https://doi.org/10.1073/pnas.0409902102

Vellend, M., Baeten, L., Becker-Scarpitta, A., Boucher-Lalonde, V., McCune, J. L. Messier, J., Myers-Smith, I. H., & Sax, D. F. (2017). Plant biodiversity changed across scales during the anthropocene. Annual Review of Plant Biology, 68(1), 563–586. https://doi.org/10.1146/annurev-arplant-042916-040949

Weathers, K. C., Ewing, H. A., Jones, C. G., & Strayer, D. L. (2013). Controls on Ecosystem Structure and Function. In Elsevier eBooks (pp. 215–230). https://doi.org/10.1016/b978-0-08-091680-4.00011-1

Whiting, K. (2022, October 18). 6 charts that show the state of biodiversity and nature loss - and how we can go nature positive. World Economic Forum. https://www.weforum.org/agenda/2022/10/nature-loss-biodiversity-wwf/

Working Memory, Depression, and Social Media: A Literature Review
By Shivm Manrao

Abstract

Social media, depression, and working memory have all been shown to be related to dopamine. In fact, some of these aspects have been shown to be directly related to each other, using dopamine as a connector. Despite this, there has been little research into the relationship between all three. This literature review hopes to dive deeper into a possible three-way relationship and uncover the neural mechanisms behind it. This article revolves around Almarzouki, et al, a 2022 paper that found that healthy university students suffered no working memory impairment after using social media, however, those with depression surprisingly did show impairment. To understand why this is, we took a closer look at the neural mechanisms behind working memory, depression, and the reward system. Working memory is linked with the prefrontal cortex, which is a part of the mesolimbic and mesocortical dopamine systems. Depression can affect the mesolimbic system by altering connectivity between brain regions and decreasing the population activity of dopaminergic neurons. Depression can therefore have an effect on both the reward system and working memory. There are many plausible theories explaining Almarzouki, et al's results that stem from this framework, and although there is evidence to base these theories off of, they are all still hypothetical, and further research is needed to verify any theory.

Keywords Depression, working memory, dopamine, social media, prefrontal cortex, mesolimbic system

Introduction

There are a myriad of possible mental disorders, and each one can have a unique but destructive impact on the lives of many. One of these is depression. Depression is a mood disorder that can be caused by both environmental and genetic factors. Depression itself can come in many different forms, each with unique characteristics. Major depressive disorder (MDD) is the most common and the most severe (1). Symptoms include anhedonia, or the inability to experience pleasure, suicidal thoughts, and excessive weight loss or gain (1). Persistent depressive disorder (PDD), is long-term depression that lasts two years or longer. Bipolar disorder is characterized by extreme episodes of overwhelming joy followed by overwhelming depression (1). There are many more types of depression, however for the purposes of this review, we will focus solely on MDD, as it is the most prevalent and the simplest to study.

One understudied symptom of depression is the effect it has on intelligence, specifically working memory. Working memory was first used by Miller et al (1960) (2) as a term to refer to the memory used to carry out tasks. The concept of working memory eventually became popular after Baddeley and Hitch (1974) (3) defined it as a brain system that stores and manipulates

information, allowing for higher-level tasks such as problem-solving and decision-making. Although classified as a type of short-term memory (4), working memory is more concerned with manipulation of information, while short-term memory is focused just on storing that information (5). Eventually, the concept would evolve into what is known as Baddeley's model of working memory, which is widely used by psychologists today.

Excessive social media use is one of many factors that can predict depression. Social media use has skyrocketed over the past decade. Nearly 81% of Americans said they have used YouTube, and 69% say they have used FaceBook. (6) For the purposes of this study, we will be defining social media as any website, app, or platform that can stimulate a dopamine release either in the form of positive social interaction or through short-form entertainment. This includes platforms such as Instagram and Facebook that simulate social interactions and platforms such as TikTok or YouTube that provide repetitive, short-form entertainment.

While depression, social media, and working memory may seem to have few connections with each other, they are in fact all related to dopamine. Dopamine is a neurotransmitter that is most famously known for controlling the reward center. However, it also controls mood, sleep, learning, and working memory (7). It's produced in the ventral tegmental area (VTA), midbrain, and arcuate nucleus in the brain (7), and there are four major dopaminergic pathways in the brain. The nigrostriatal pathway plays a role in movement, and dysfunction of this pathway can lead to Parkinson's disease. Dopaminergic fibers originate from the substantia nigra and lead to the caudate nucleus and the putamen in the basal ganglia. The mesolimbic pathway plays a role in emotion and the reward system. It's activated by pleasurable situations and drives people to seek out those pleasurable situations. Dopaminergic fibers originate in the VTA and spread out to the amygdala, pyriform cortex, lateral septal nuclei, and the nucleus accumbens. Pleasurable situations can also stimulate dopamine release in the prefrontal cortex. The mesocortical pathway controls cognitive and emotional behavior. Dopamine fibers start from the VTA and travel to the frontal cortex and septohippocampal regions. The right balance of dopamine in the prefrontal cortex, mediated by the mesocortical pathway, is critical for working memory. The fourth pathway, the tuberoinfundibular pathway, plays a role in lactation and inhibition of prolactin production, but isn't relevant for our purposes (9, 7).

Dopamine and working memory are closely tied together, as both D1 and D2 dopamine receptors play a role in working memory (7). Working memory is also closely associated with the prefrontal cortex. This goes hand-in-hand with the mesocortical pathway, which controls working memory by mediating levels of dopamine in the pre-frontal cortex (7). This mediation is critical, as too much or too little dopamine in the prefrontal cortex can impair working memory (7). Furthermore, the dopamine system is shown to be directly engaged during the delay epoch of working memory tasks (8). In other words, dopamine is critical for holding information in the mind in between receiving that information and being asked to recall it.

Social media and depression are both related to dopamine as well. Social media provides the same dopamine spikes as any other pleasurable stimuli. When social media users receive a positive stimulus (a like, comment, or direct message), dopamine receptors within the striatum

and the VTA become activated (9). These dopamine spikes are almost addictive—they make us constantly come back for more. Over time as we keep coming back and experiencing these dopamine spikes, our baseline dopamine levels decrease. The brain works to restore homeostasis, and if there are constant spikes in dopamine levels, our resting dopamine levels must decrease to compensate (9).

One of the hallmarks of depression is decreased dopamine. Patients with depression often have increased levels of monoamine oxidase A (MAO-A), an enzyme that breaks down neurotransmitters including serotonin and dopamine, resulting in lower levels of dopamine in the brain. This effect can go the opposite way as well, explaining the correlation between excessive social media use and depression. Lowered dopamine levels can lead to depression, and as our baseline dopamine levels decrease from excessive social media usage, we start to experience less pleasure from things that would have previously been pleasurable. Anhedonia develops, and eventually it leads to MDD.

Since social media, depression, and working memory are all related to dopamine in some way, the question arises whether these three factors exert an influence on each other when combined together. Two of these factors may have an influence on each other, but what is the relationship between all three? How does dopamine availability cause clinically depressed patients to exhibit different working memory capabilities compared to normal people in response to repetitive dopamine exposure via social media? This is the question this review article hopes to answer.

While there has already been a lot of research into how two of these factors exert an influence on each other, there is very limited literature on all three. Examining how depression and social media interact to affect working memory could potentially lead to new insights into the mental disorder and how it affects not just mood, but cognition and intelligence as well.

Furthermore, this review paper hopes to shed more light on a new side effect of mood disorders. By understanding that people with depression do not just have their mood out of order, but also their basic intelligence, we may be able to better empathize with depressed patients and treat their mood disorder.

This paper will examine the possible relationship between how social media can change the working memory capabilities of depressed patients. Considering all three factors are related to dopamine, it is possible that they are all related. This paper hopes to uncover if there is a relationship or not, and if there is, speculate on the possible neural mechanisms behind the interaction. This speculation may open up future avenues of research into this topic.

This paper will not be without its limitations, however. As stated earlier, this study will be focusing on major depressive disorder, and “depression” will often refer to only MDD. Social media will be confined to platforms that drive dopamine release. Therefore, platforms such as LinkedIn, although considered by some as social media, will be excluded from this study.

Since this is a review paper, the availability of recent literature and research will be a large limitation. Research may be outdated, or it might simply not exist yet. Furthermore, depression is a complex disorder and there are many possible causes, both environmental and

genetic. Therefore, while social media might play a factor, it's impossible to say for sure that any relationship found is due to one factor or another. Similarly, the brain is an extremely complex organ and there are many other factors that affect working memory, not just depression or social media use. Finally, since this paper will largely be theoretical, there is a possibility that there is no connection at all. Depression and working memory could use completely different dopaminergic pathways, and it would be difficult to say for sure without *in vivo* tests.

Results

Almarzouki, et al (10) tested whether social media use had any effects on working memory. Participants were split into two groups, one who spent some time on social media, and one who spent some time painting. The participants were then asked to perform a spatial working memory (SWM) task. The groups then swapped, with those that spent time on social media previously now painting and vice versa, and another SWM task was conducted afterward. When considering the sample as a whole, there was no significant difference in errors made in the SWM task regardless of whether participants painted or used social media prior to the task. However, when considering at least moderately depressed participants only, hours spent on social media significantly predicted errors in the SWM test. However, Social Media Disorder Scale (SMDS) scores did not predict errors in the whole sample nor the depressed sample, but it did predict depression and anxiety prevalence. Finally, the study mentioned that participants who were depressed made more errors after exposure to social media, while those who weren't depressed didn't.

In order to theorize on the neural mechanisms behind these results, we first need to understand working memory in greater detail. Baddeley's model of working memory, which is the most widely used model today, consists of one master component and 3 slave components. The master component, or the central executive, is thought to coordinate the other slave components and allocate memory resources to each slave component. (3) The slave components consist of the phonological loop, the visuospatial sketchpad, and the episodic buffer. The phonological loop allows for speech and language comprehension, while the visuo-spatial sketchpad processes visual and spatial information. At first, Baddeley proposed his model with only two slave components (11) but eventually added a third (12). The episodic buffer has limited storage capacity and is thought to integrate information from long-term memory and perception and store that information as integrated episodes. (3)

Due to the higher-level cognitive nature of working memory, it comes as no surprise that working memory is tied closely with the prefrontal cortex. In fact, the central executive component is thought to be a direct function of the prefrontal cortex. (3) The prefrontal cortex plays a role in executive functions such as attention shifting and updating or monitoring information, and damage to the prefrontal cortex often results in executive dysfunction (3). Most notably, delay-period activity in prefrontal neurons has been found to correlate with storage processes of spatial working memory. Monkeys were put through the delay-response task where they were required to remember spatial information, such as the location of an object, during a

time when the object is not present, called the delay period. Delay period activity in prefrontal neurons was present only when the monkey recalled the object's position correctly. In incorrect trials, delay period activity was either cut off or not present at all. (3)

In addition, a dual-task paradigm on monkeys further supports the role of the prefrontal cortex as the central executive. The dual-task paradigm forces subjects to perform two tasks at the same time. Subjects show worse performance on both tasks in this paradigm than if they were performing each separately. This is theorized to be because of the brain's limited capacity of task-processing resources. When performing two tasks at a time, the brain is forced to share these resources between the tasks, resulting in impaired performance on each. The central executive in Baddeley's working memory model is what allocates resources between tasks. The dual-task paradigm used in Watanabe and Funahashi (2014, 2015) (13) involves two tasks that activate overlapping neural populations in the prefrontal cortex. The higher the overlap of these regions, the greater the dual-task interference effect and the worse the monkeys performed on each task.

Now that we have a better understanding of working memory, we can look at ways it can be disrupted in depression. Numerous studies have shown that working memory is indeed impaired in MDD, however, the specific components that are affected are still unclear (15). Some studies have proposed a dysfunction in the central executive, causing a faulty allocation of resources (15), while other studies have found that both the central executive and the slave components of working memory are impaired (16). Other studies have taken a different route and have shown how information updating is impaired in MDD. One study found that depressed participants showed greater interference from irrelevant negative words than control groups in both normal and sad moods (17). Levens, et al. supports this bias for negative material in MDD, as they showed that depressed participants integrated sad content quicker than controls, and were quicker to remove happy content from working memory while slower to remove sad content. They also found that depressed participants were slow to link emotional content in general (18).

Depression affects not only working memory, but also various connections between neural regions in the mesolimbic dopamine system, as shown in Figure 1. These regions include the prefrontal cortex (PFC), the ventral tegmental area (VTA), which synthesizes dopamine and distributes dopamine, the nucleus accumbens (NAc), which regulates motivation and projects to executive and emotional regions within the brain (19), and the anterior cingulate cortex (ACC), which functions in emotional awareness and expression, decision making, and attention allocation (20). Those with MDD most notably display decreased connectivity between the PFC and VTA, the PFC and NAc, the VTA and NAc, and the ACC and NAc. On the other hand, patients with MDD show increased connectivity between the ACC and the PFC, and the ACC and the VTA (19).

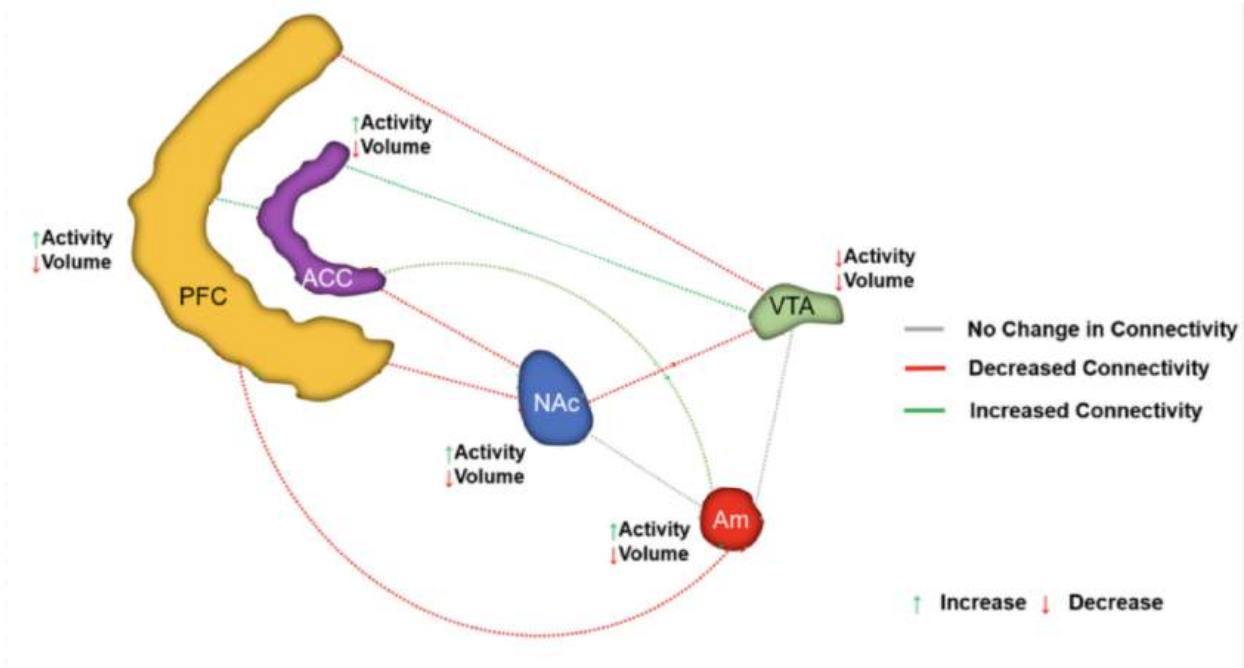
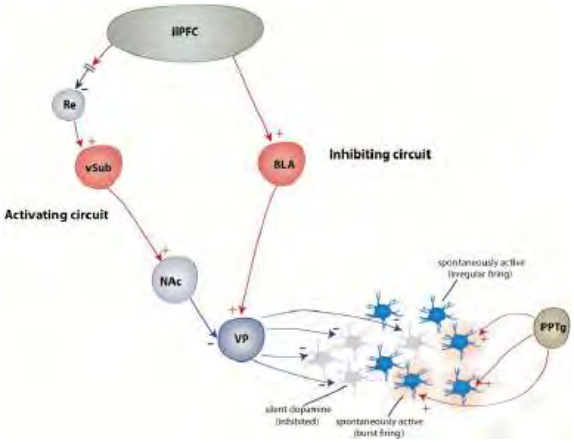


Figure 1: changes in connectivity and activity in depression across different brain structures, adapted from (19)

In order to understand the specific neurobiology of how depression affects the reward system, we must first look at how the reward system works in healthy adults (Figure 2A). Mesolimbic dopaminergic neurons can either be spontaneously active or inactive, and the ratio of active to inactive neurons is called population activity. On average at any given time, about one-half of mesolimbic dopaminergic neurons are spontaneously active. The ventral pallidum (VP) keeps the other half inactive through constant GABAergic signals. The ventral subiculum of the hippocampus (vSub) can increase the number of active dopamine neurons by sending signals to the NAc, which inhibits the VP. The infralimbic subregion of the medial prefrontal cortex (ilPFC) can inhibit the nucleus reuniens of the thalamus (Re). Since the Re activates the vSub, this ilPFC inhibition eventually decreases population activity. The basolateral amygdala (BLA) is also under ilPFC control. The ilPFC activates the BLA, which then activates the VP, decreasing population activity (21)



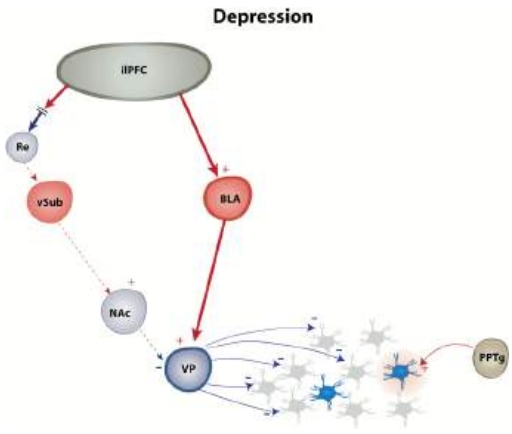


Figure 2A Figure 2B

Figure 2: An overview of the regulation of the dopamine system in healthy and depressed adults. (21) Figure 2a shows the mesolimbic reward system in healthy adults, Figure 2b shows the mesolimbic system in depressed adults

Spontaneously active neurons can either fire in slow, irregular patterns or in bursts of activity. The pedunculopontine tegmental nucleus (PPTg) is the main driver of burst firing, but can only drive burst firing in active neurons. Since the vSub is also involved in processing context, it can control population activity and therefore burst firing based on the environment. In a calm state, the vSub maintains low population activity through the NAc-VP pathway. Any signals from the PPTg will not trigger burst firing, as most neurons are inactive. However, in an arousing situation, the vSub can increase population activity. Whenever a stimulus for dopamine release is detected, the PPTg will be able to trigger burst firing, as most neurons are in the active state due to the vSub activation (21).

In depression, the dopamine system is downregulated (Figure 2B). In rats, the chronic mild stress (CMS) model is meant to mimic MDD in humans. CMS in rats caused a decrease in dopamine population activity in the VTA, similar to humans. However, inactivation of the ilPFC or the BLA-VP pathway reverses this change, indicating that either an increase in activity of the ilPFC or hyperactivation of the BLA-VP pathway as a whole leads to the decrease in population activity that was observed (21).

Discussion

Although there was no relationship between social media use and performance on working memory tasks for healthy participants, there was one for depressed participants. Depressed participants made more errors after being exposed to social media, and hours spent on social media per week predicted errors made as well. In addition, working memory was found to be closely tied to the prefrontal cortex. In fact, the prefrontal cortex is thought to perform the functions of the central executive and allocate memory resources to the other components of

working memory. In depression, working memory is impaired, however the exact mechanisms are still unclear. Whether it is through a faulty allocation of memory resources by the central executive, or a bias towards negative material, depressed patients tend to perform worse on working memory tasks than healthy individuals. Depression also affects the reward system in the brain. The connectivity between different brain structures in the mesolimbic dopamine pathway is altered in depression. Furthermore, overactivity of the ilPFC and the BLA-VP pathway leads to lower population activity in dopaminergic neurons, contributing to depressive symptoms.

It is unknown why social media may affect the working memory of only depressed participants, however we can still theorize on the mechanisms behind these results. The study offered some of their own theories behind the results. Perhaps in the SWM test, solving the problems correctly and moving on implied a reward, or had some positive aspect to it. It's known that people with MDD have impairments in regard to updating positive information in working memory, so that could explain why depressed participants performed worse. However, that doesn't account for the decrease in performance after exposure to social media. Perhaps the participants had difficulty refocusing their attention and updating their working memory with the SWM test after using social media, since the SWM test seemed to have positive aspects. However, if this were the case, they should theoretically also have difficulty after painting as well. Another proposed theory was that social media had a "draining" effect on participants, and depressed participants were more vulnerable to this drain than healthy participants. This theory is more plausible, especially since some studies have shown impairment of central executive function in depression. Impairment of the memory-allocating function of the central executive could have meant that it was difficult to switch to a SWM test right after using social media. The flaw with this theory is, again, the depressed participants should have had an equally difficult time when switching from painting to the test. This brings into play the nature of social media. Social media provides a larger release of dopamine than painting. This dopamine release could have some effect on working memory that results in impairment. This effect could only be present in depressed adults, and not healthy ones, accounting for the difference in results.

The reward system has been shown to be faulty in MDD. Overactivity of the ilPFC causes overactivation of the BLA-VP pathway, eventually reducing population activity of dopaminergic neurons. When someone with MDD uses social media, the PPTg attempts to induce burst firing in active neurons, however since the population activity is low, most neurons are inactive. Since the mesolimbic pathway has dopaminergic fibers originating in the VTA and traveling to the PFC, and since fewer dopaminergic neurons are burst firing in the VTA, dopamine levels in the PFC could be altered. The mesocortical pathway usually mediates levels of dopamine in the PFC, as even slightly too much or too little dopamine can impair working memory. Thus, this alteration of levels of dopamine in the PFC could impair working memory, leading to the results shown. Future testing would be necessary to track how dopamine travels through the brain and test this theory.

This article's aim was to establish a connection between social media, depression, and working memory, and unravel the neural circuitry behind said relationship. While there was a

study that showed the effect of depression and social media on working memory, one study is not nearly enough to come to a conclusion on a definite relationship. Assuming there is a connection, however, my hypothesis was that dopamine availability would be the main connector between depression, social media, and working memory. While this is in part true, and dopamine availability in the PFC is a plausible explanation for the results found, there are also other possible mechanisms, such as a working memory biased against positive stimuli.

Confirming this relationship and the "priming effect" depression has on social media's effect on working memory is one future avenue of study. Furthermore, additional research is needed to test the theories proposed in this article. There have been articles examining how we can measure dopamine neurotransmission in real-time. (22) Some of these techniques could be adapted to track dopamine in MDD throughout the prefrontal cortex, allowing us to observe changes during dopamine-inducing activities and working memory tasks.

This paper is not without its limitations. A large portion of this paper was theoretical, and even though these theories were based on contemporary research and literature, they are still just theories. They could be completely spot on, or they could not be further from the truth, and until additional research is conducted, we cannot say anything for certain. Furthermore, the studies cited in this paper have their own limitations. For one, Almarzouki et al. studied only university students, meaning their results may be limited to young adults who repeatedly use working memory in an academic context. Other common limitations, such as the availability of recent research, should also be taken into account.

Social media is used by millions, yet the side effects of disordered social media use have not been highlighted until recently. Constant dopamine release can lead to mood disorders, which can be tricky to treat, especially with the addictive nature of social media and dopamine. Depression, one of the most potent side effects of social media, is vastly misunderstood by the public. Many who have never seen the disorder up close don't understand that those with depression are mentally unable to experience a healthy mood state. Even now, new side effects of depression are being discovered, and our view of the disorder is constantly changing. As our understanding of the effects and symptoms of depression improves, so will our ability to treat it.

Methods

The two main databases used to search for literature were Google Scholar and PubMed. Keywords such as "dopamine," "depression," and "working memory" were used to filter articles based on relevance. Furthermore, articles connecting two aspects of the investigation were also searched for to determine relationships or correlations among aspects. For example, search terms such as "social media and depression" or "depression and dopamine" were used to determine how social media, depression, and dopamine are all related.

All studies relating to the research question were considered, however studies with higher relevance were more likely to be included than studies focusing on other topics. Furthermore, studies published recently (within the past 10 years) were far more likely to be included, as older

studies may not have updated information. Studies that focused on depressive disorders other than MDD were not included due to the differing nature of each depressive disorder.

All the gathered information was first written down in bulleted notes, arranged by review article. From there, the most important and relevant information was chosen to be highlighted in the results section. Discussion sections from the gathered articles were analyzed, and certain information from these sections was included in my discussion section. In addition, my own theories were synthesized from the results and included as well.

Articles that were published in a journal were mainly used for this review. Almost all of those articles have been peer-reviewed, and the ones that haven't were reviewed by editors with significant education and experience in the field. In addition, any limitations that were either stated by the paper or thought of when reviewing the paper were taken into account.

Works Cited

D. F. Bruce, A. McPhillips. Types of depression and how to recognize them. *WebMD*. https://www.webmd.com/depression/depression-types. (2024)

G. A. Miller, G. Eugene, K. H. Pribram. Plans and the structure of behavior. *Systems Research for Behavioral Science*. 369-382 (1968).

A. Baddeley, G. Hitch. Working memory. *Current biology*. **20**, R136-R140 (2010).

S. Funahashi. Working memory in the prefrontal cortex. *Brain sciences.* **7**, 49 (2017).

N. Cowan. What are the differences between long-term, short-term, and working memory?. *Progress in brain research.* **169**, 323-338 (2008).

B. Auxier. Social media use in 2021. *Pew Research Center*. www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/. (2021).

G. Ayano. Dopamine: receptors, functions, synthesis, pathways, locations, and mental disorders: review of literatures. *Journal of Mental Disorders and Treatment*. **2**, (2016).

S. M. Landau, R. Lal, J. P. O'Neil, S. Baker, W. J. Jagust. Striatal dopamine and working memory. *Cerebral Cortex*. **19**, 445-454, (2009).

V. Fernandez. Social media, dopamine, and stress: converging pathways. *Darthmouth Undergraduate Journal of Science*. https://sites.dartmouth.edu/dujs/2022/08/20/social-media-dopamine-and-stress-converging-pathways/. (2022).

A. F. Almarzouki, R. A. Alghamdi, R. Nassar, R. R. Aljohani, A. Nasser, M. Bawadood, R. H. Almalki. Social media usage, working memory, and depression: an experimental investigation among university students. *Behavioral sciences*. **12**, 16, (2022).

A. Baddeley. Working memory. *Science.* **255**, 556-559 (1992).

A. Baddeley. The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*. **4**, 417-423, (2000).

K. Watanabe, S. Funahashi. Neural mechanisms of dual-task interference and cognitive capacity limitation in the prefrontal cortex. *Nature Neuroscience*. **17**, 601-611, (2014).

K. Watanabe, S. Funahashi. A dual-task paradigm for behavioral and neurobiological studies in nonhuman primates. *Journal of neuroscience methods*. **246**, 1-12, (2015).

E. J. Rose, K. P. Ebmeier. Pattern of impaired working memory during major depression. *Journal of Affective Disorders.* **90**, 149-161, (2006).

G. Christopher, J. MacDonald. The impact of clinical depression on working memory. *Cognitive neuropsychiatry*. **10**, 379-399 (2005).

J. Joormann, I. H. Gotlib. Updating the contents of working memory in depression: interference from irrelevant negative material. *Journal of Abnormal Psychology*. **177**, 182-192, (2008).

S. M. Levens, I. H. Gotlib. Updating positive and negative stimuli in working memory in depression. *Journal of Experimental Psychology: General*. **139**, 654-664, (2010).

R. A. Serafini, K. D. Pryce, V. Zachariou. The mesolimbic dopamine system in chronic pain and associated affective comorbidities. *Biological Psychiatry*. **87**, 64-73, (2020).

MasterClass. How the anterior cingulate cortex affects cognitive function. *MasterClass*. https://www.masterclass.com/articles/anterior-cingulate-cortex-explained#41Zlt71XlCIHwnMg5mrHyG. (2021).

P. Belujon, A. A. Grace. Dopamine system dysregulation in major depressive disorders. *The international journal of neuropsychopharmacology*. **20**, 1036-1046, (2017).

R. D. Badgaiyan. Detection of dopamien in “real time”. *Frontiers in Neuroscience*. **7**, https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2013.00125/full, 2013).

The Efficacies of Music Therapy on Alzheimer's Disease Patients: A Review
By Sofia Alexandrakis

Abstract

The aging population is expected to increase substantially in the near future, bringing with it increased concerns for the number of individuals diagnosed with dementia, specifically Alzheimer's Disease. Alzheimer's Disease is classified as the progressive deterioration of the brain, resulting in memory loss, confusion, and several behavioral symptoms. While various pharmacological medications exist to slow the progression of the disease, there is no current cure to reverse its effects. Thus, researchers have investigated inexpensive non-pharmacological treatments lacking undesirable side-effects that could serve as potential alternatives to current medications. Music therapy has recently emerged as a potential memory and behavioural improving strategy to Alzheimer's patients. Music therapy is a highly dynamic intervention that focuses on active engagement, communication, and learning. This review will explore the different ways music therapy has been implemented with the intention of improving the quality of life and memory of Alzheimer's patients.

Introduction

Alzheimer's Disease is the most common form of dementia among older adults (National Institute of Aging, 2023), affecting more that 55 million people worldwide (World Health Organization, 2024) and more than 6 million Americans (National Institute of Aging, 2023). Alzheimer's Disease is defined by the progressive degeneration of the brain that affects memory, thinking, behavior, and one's ability to carry out simple daily activities. While extensive research is ongoing, the development of Alzheimer's Disease is generally caused by abnormal chemical changes that affect the tau proteins of the brain. In a healthy brain, tau binds to the microtubules that support neurons, guiding the transfer of nutrients to the cell body to the axon and dendrites. However, in an Alzheimer's brain, tau separates from the microtubules and attach to each other, causing tau tangles. Tau continues to spread when levels of beta-amyloid are abnormally high and thus clump together into plaques. Tau tangles and amyloid plaques block the neuron's transport system, which disrupts neuron communication. This eventually leads to brain atrophy, defined by widespread brain cell death causing brain shrinkage (National Institute of Aging 2024). There are three main stages of Alzheimer's: mild, moderate, and severe. The mild stage involves "wandering and getting lost, trouble handling money and paying bills, taking longer to complete normal daily tasks, and personality and behavior changes (National Institute of Aging, 2023)." The moderate stage involves brain damage in areas responsible for "language, reasoning, conscious thought, and sensory processing," and memory loss and confusion continues to worsen (National Institute of Aging, 2023). The severe stage involves the presence of tangles and plaques throughout the brain and significant brain atrophy. Severe Alzheimer's patients lack communication skills and independence (National Institute of Aging, 2023). There are no current cures to reverse the progression of Alzheimer's disease, but antidepressants, antipsychotics, and

anti-anxiety drugs are helpful to stabilize symptoms, and provide physical and emotional comfort to patients (National Institute of Aging, 2023). Examples of drug medications include cholinesterase inhibitors, which prevent the breakdown of acetylcholine, responsible for memory and thinking, and Lecanemab and donanemab, which help reduce amyloid plaques in the brain (National Institute of Aging, 2023). However, all pharmacological interventions have side effects, including "nausea, vomiting, diarrhea, allergic reaction, a loss of appetite, headaches, confusion, dizziness, and falls" (National Institute of Aging, 2023). Thus, researchers have recently investigated non-pharmacological interventions to assist with the progression of Alzheimer's disease.

Music therapy has posed as a promising alternative to pharmacological intervention methods against the progression of Alzheimer's disease. Numerous studies have been conducted regarding the benefits of music therapy on Alzheimer's patients or other forms of dementia. Music therapy generally entails the exposure of audible music to cognitively impaired patients in hopes of improving social interactiveness, recall, agitation, anxiety, depression, etc. Therapy sessions are consistently given to patients over a set number of weeks to track progress. However, each study has differed regarding the implementation and types of music therapy tested on participants. Music therapy is often implemented passively in which the participant is exposed to music sitting down in a relaxed, calm position. Music therapy is also implemented actively, in which the participants listen to high-energy music and actively move along with the music, whether they dance, clap, or move their arms. Numerous studies also investigated whether there was a difference regarding a difference between individualized and non-individualized music. Through individualized music therapy, researchers attempted to choose specific music genres or songs that had deep personal connections with the participants. Through personalized music, researchers hoped to enhance and generate increased positive effects of the therapy session. Lastly, numerous studies conducted studied the difference between implementing music therapy among participants grouped in the same room or individually on their own. These studies mainly focused on how music therapy affected the social communications between patients and the researchers, and how this may have had an effect on depression and anxiety.

General Music Therapy

Dementia

Due to the significant cognitive deficits of dementia and Alzheimer's patients, there has been a dire need to subside and improve their rapidly degrading symptoms. Recently, researchers have emphasized how music therapy could potentially serve as an alternative to current pharmacological treatments. Researchers have investigated the effects of incorporating music intervention to the usual regimen of individuals with dementia. This section focuses on general music therapy, including both active and receptive listening, aimed at improving patients' mood,

motivation, and socialization, and to 'stimulate sensory, motor, and cognitive aspects' (Raglio et al. 2015).

Jenny van der Steen et al. studied the effects music therapy has on the emotional wellbeing of individuals with overall dementia. This topic of study was researched because these researchers noticed the limited effects of pharmacological medications and the lack of research regarding nonpharmacological treatments (van de Steen et al., 2017). This study consisted of 16 individual studies for a total of 620 participants. Five studies had individualized music intervention, while the other studies consisted of group music intervention. All studies consisted of both active and receptive music therapy. The studies revealed that there was little-to-no evidence that music therapy had an effect on emotional wellbeing, quality of life, behavioral problems, cognition, anxiety, and social behavior. There was some evidence regarding a reduction in depression; however, there was no decrease in agitation or aggression. There was also a lack of evidence regarding the long-term and sustainability of music therapy's effects (van de Steen et al., 2017). The data collected demonstrates the need for future studies to involve larger sample sizes, and an examination of the effects' duration compared to the sessions' duration and the number of sessions. Perhaps the limited efficacy of music therapy in this study was due to investigating dementia in general. Therefore, looking at patients with a specific form of dementia, like Alzheimer's disease, would generate results that imply a greater efficacy of music therapy.

Alzheimer's Disease

Therefore, researchers have investigated how general music therapy has affected the well-being of Alzheimer's disease patients specifically. Alicia Clair et al. specifically researched the effects that music therapy has on severely regressed Alzheimer's patients. This topic was studied to better understand the difference music therapy may have on patients with further developed Alzheimer's, as most other studies conducted research regarding patients with mild to moderate Alzheimer's (Clair et al., 1990). To conduct this study, three male subjects with severe Alzheimer's participated in weekly, 30-minute music therapy sessions for 15 months. The data was collected during the last 11 weeks of the study, which revealed that despite the subjects' deterioration during the intervention, they were able to function with others in the group and interact. There was an increase in the subjects' musical participation. The nursing staff suggested that the music therapy sessions were one of the only opportunities the patients were able to successfully interact with others due to their low-functioning ability (Clair et al., 1990). Overall, the study demonstrated the feasibility of music therapy among Alzheimer's patients, but additional research is required to confirm the beneficiary results of the therapy. Hence, there have been numerous studies that have focused on the symptomatic relief of music therapy on Alzheimer's patients.

Due to limited pre-existing research, Stephane Guetin et al. specifically studied music therapy's effect on anxiety and depression among affected patients (Guetin et al., 2009). The study was conducted using a control group of 15 patients, who received no music intervention,

and a treatment group of 15 patients. The treatment group received weekly sessions of receptive music therapy, in which the music style was chosen by the patient, while the control group participated in reading sessions. Treatments lasted a total of 16 weeks. Anxiety, measured using the Hamilton Scale, and depression, measured using the Geriatric Depression Scale, were both evaluated at weeks 1, 4, 8, 16, 24 after treatment began. Guetin et al. found significant improvements in anxiety and depression from week 4 to week 16. They also found that the music therapy effects prolonged for up to 8 weeks after the treatment ended, at 24 weeks (Guetin et al., 2009). Their findings reveal that music therapy is an approachable method in reducing anxiety and depression among patients of mild to moderate Alzheimer's Disease.

Despite improvements to well-being, other groups have studied how music therapy has specifically affected Alzheimer's patients' memory. Prickett et al. specifically studied how music affected recall of material both sung and spoken for Alzheimer's patients. This topic was studied to better understand how music therapy can be used to slow memory loss or improve memory of Alzheimer's patients (Prickett et al., 1991). To conduct this study, 10 Alzheimer's patients participated in three 20-minute sessions. These sessions were videotaped with a close up of the patient's face. These videos were evaluated through listening and watching to count the number of words the patients correctly recalled. The researchers found that patients recalled words to songs much better than spoken words, and that long-familiar songs were recalled with much greater accuracy than a newly presented song. Additionally, most patients attempted to sing, hum, or keep time while the therapist sang (Prickett et al., 1991). Overall, the data demonstrated that long-familiar songs could be used to stimulate participants into participation. With consistent practice, the patients are highly capable of learning new songs. The data generally poses music therapy as a potential therapy to improve memory of Alzheimer's patients.

Similarly, Nicholas Simmons-Stern et al. studied music's effect specifically on memory due to the limited pre-existing research regarding this topic (Simmons-Stern et al., 2010). To conduct this study, 13 Alzheimer's patients and 14 healthy older adults participated in a single 30-minute session. The stimuli consisted of four-line excerpts of 40 children's songs, in which 20 were spoken lyrics while the other 20 were sung lyrics. The study revealed that the Alzheimer's patients demonstrated better recognition accuracy for sung lyrics rather than spoken. However, healthy older adults showed no significant difference between sung and spoken lyrics (Simmons-Stern et al., 2010). The data demonstrates that areas in the brain subserving music may still be spared in individuals with Alzheimer's Disease. It shows that music allows for better attention and memory among Alzheimer's patients, and thus music therapy is a potentially effective therapy that should be implemented.

Likewise, Muireann Irish et al. conducted a study regarding music's role as a memory enhancer specifically for mild Alzheimer's patients in comparison to healthy elderly people. (Irish et al., 2006). To conduct this study, 10 Alzheimer's patients and 10 healthy participants were seen twice. One session involved exposure to music, specifically Vivaldi's 'Spring' movement from 'The Four Seasons,' while the other session was in silence. The study demonstrated notable improvement in the Alzheimer's patients' recall during the

Autobiographical Memory Interview. There was also a significant reduction in anxiety, which is a possible explanatory factor for increased memory recall (Irish et al., 2006). Overall, these studies suggest music therapy as a promising non-pharmacological therapy for Alzheimer's patients' memory recall. Unlike drug therapies, non-pharmacological therapies are less expensive and result in limited negative side effects. Thus, music therapy should be more widely implemented to slow down and prevent cognitive decline among Alzheimer's patients.

Either

Researchers have investigated the effects of general music therapy on either dementia in general or specifcally Alzheimer's disease in the same study.

Specifically, Owen Clute investigated the effects that music therapy explicitly had on social interaction and mood (Clute, 2017). Clute researched this topic because he noticed the lack of existing research regarding this topic of study. Thus, his research consisted of four female participants ages 84, 90, 92, and 96. The patients participated in two music therapy sessions a week for six weeks, for a total of eleven sessions. The music genres varied, along with variation in live and recorded music offered. The patient's dance movements were restricted, limiting the music intervention to strictly receptive therapy. Throughout the study, Clue found that the first session was not very successful, but the participants started to enjoy the later sessions. He also found that social interaction of the patients improved, as they continually asked the researchers about the song titles or singers' names. The patients seemed increasingly happy as the sessions progressed. The patients also laughed and moved along with the music, indicating an improvement in mood (Clute, 2017). The data collected confirms that music therapy improves social interaction and mood among patients suffering from dementia. However, there was no improvement in physical and cognitive abilities (Clute, 2017). Thus, the data suggests that the study must be conducted over longer periods of time with more participants, as the results do not apply on a large scale.

Additionally, Amelia Gulliver et al. similarly studied the general effects of a specialized Music Engagement Program (MEP) on quality of life, emotional well-being and depression of Alzheimer's disease and dementia patients. MEP is an attractive therapy approach due to its accessibility and feasibility, and its highly interactiveness and intent to bring people together (Gulliver et al., 2021). This study was conducted due to a lack of cost-effective, evidence-based programs that could be broadly applied and sustained in large communities of dementia-affected individuals. Technology is an expensive pursuit, so MEP requires no musical instruments and equipment, and the staff is trained at a low cost (Gulliver et al., 2021). To conduct this study, 16 participants with dementia participated in 8 weekly group singing MEP sessions, lasting about 45-60 minutes each. The sessions were led by highly experienced music facilitators, who found that the music would elicit strong emotions and memories of the past, causing the participants to sometimes cry. The participants experienced a reduction in depression, stability in emotional wellbeing, decrease in verbal aggression, and increase in social activity. The participants seemed to enjoy themselves, laughing, clapping, and smiling throughout the sessions. Family members

and staff were highly complementary of the program and did not report any negative consequences (Gulliver et al., 2021). Overall, the study demonstrates the positive effectiveness of MEP on patients with dementia or Alzheimer's Disease, and thus should be incorporated into daily life. However, implementing the program posed to be a challenge, as the residents would often wander off during assembly (Gulliver et al., 2021). In terms of future MEP implementation, it would be difficult to continue the sessions without a facilitator due to the lack of equipment that MEP requires (Gulliver et al., 2021).

Altogether, the studies discussed demonstrate the effects that general music therapy has on dementia. Although individuals with general dementia do not demonstrate positive results to the extent that Alzheimer's patients do, music therapy poses as an inexpensive therapy that can sustain or limit the degrading symptoms of cognitively impaired individuals.

Music Therapy Types

Researchers have explored various methods of implementing music therapy. One method is known as group music therapy, which involves therapy of multiple individuals with dementia at once. It is effective because it allows the patients to socialize and interact with each other regarding the music (Takahasi et al., 2006). Another type of implementation method is individualized music, which involves the patient listening to music tailored to their interests or music with a sense of personal emotional attachment to it (Leggieri et al., 2019). Individualized music therapy is a hopeful intervention that can trigger memories and improve autobiographical memory. Dual-task therapy has also proved to be a possible implementation method, which involves singing and playing instruments while walking and responding to the therapist's cues. Dual task therapy poses as a beneficial intervention by "strengthening the attention network" and by allowing for greater brain activity (Chen et al., 2018). This section involves studies that have compared the efficacy of multiple implementation methods of music therapy in hopes of finding the best type of therapy for cognitively impaired individuals.

Group

Hsin Chu et al. studied how group music therapy impacts on depression and its delay of the deterioration of cognitive functions in elders with dementia (Chu et al., 2014). To conduct this study, 104 participants with dementia were divided into two groups: an experimental group and a control group. The experimental group received twelve 30 minute sessions of music therapy over the course of six weeks. The control group, meanwhile, received usual care. The data was collected before the study, after the 6th and 12th therapy sessions, and one month after the study. The study revealed immediate improvements in depression and slight improvements to cognitive function and short-term recall. Cortisol levels, however, did not significantly decrease. The study demonstrated its greatest impacts on patients with mild to moderate dementia as opposed to severe dementia (Chu et al., 2014). Overall, the data implies that group music therapy is a noninvasive, inexpensive therapy that is beneficial to individuals with mild to moderate dementia.

However, the data found from Chu et al.'s study regards the short-term effects of group music therapy on Alzheimer's disease patients. Short-term benefits alone are insufficient for nursing homes seeking to implement music intervention as a long-term therapy and as a possible alternative to current pharmacological treatments. Thus, Takiko Takahashi et al. studied the long-term effects of group music therapy for those with moderate and severe dementia (Takahashi et al., 2006). To conduct this study, 24 elderly patients with dementia were put into a treatment group, while 19 patients were put into a control group. The treatment group received active group music therapy, which involved singing and participation in concerts, and used school songs from the early 1900s. After two years, the researchers found no statistically significant difference between the pre and post music group's salivary cortisol or intelligence. Additionally, they found that the music group's systolic blood pressure was much lower in comparison with the control group. The music therapy group also better maintained physical and mental health than the control group. Aside from psychological effects, the participants enjoyed singing and playing the instruments, which helps prevent cardiac and cerebral diseases (Takahashi et al., 2006). The study demonstrates that group music therapy is an effective intervention for patients with dementia long-term, not just short-term, and should be implemented in nursing homes as a possible addition or alternative to pharmacological treatments for cognitively impaired individuals.

Dual-Task

Other groups have explored the effectiveness of musical dual-task training in patients with mild to moderate dementia on attention control, dual-task performance, fall efficacy, agitation, and balance (Chen et al., 2018). The researchers investigated whether dual-task therapy could serve as an effective music therapy approach. Another motivating factor was that cohesive protocols for music-based dual-task training did not exist (Chen et al., 2018). In 2018, Yu-Ling Chen et. al divided 28 individuals into two research groups. One group received the music dual-task training while the other was a control group receiving no musical therapy. Their findings revealed that the music dual-task training group showed significantly improved performance in attention control and fall efficacy, and decreased agitation, whereas the control group expressed no improvement (Chen et al., 2018). The findings demonstrated that music dual-task training can provide an effective approach to music-based therapy for cognitively impaired individuals.

Individualized

Individualized music therapy has also posed as a promising therapy that can trigger personal memories associated with familiar music. Kristinr Gustavsen Mads et al. studied how individualized music therapy affected the well-being and social behavior of individuals with dementia. This group of researchers studied this topic because the results of previous studies may have undermined the effectiveness of music therapy due to the degenerative nature of dementia. Thus, their goal was to implement small, single-case designs that allow for increased validity

(Mads et al. 2022). To conduct this study, patients with dementia participated in 10 weeks of music therapy and the effects were calculated for each session using meta-analysis. The researchers found a 48% increase in wellbeing, a 32% increase in social behavior towards significant others, and a 93% increase in nonverbal social interactions during the therapy sessions. They also found that the sessions increased in expressions of enjoyment and happiness, allowed for a preservation of the patient's identity, and increased joint interaction and turn taking, implying mutual engagement. However, there was no increase in long-term well-being, questioning the effectiveness of music therapy when implemented over extended periods (Mads et al. 2022). Overall, the study demonstrates that single-case designs are effective for generating valid results regarding individualized music therapy. However, more research must be done regarding the longevity of individualized music therapy's effects.

Comparisons

Numerous studies have conducted different types of music therapy implementation to compare their efficacies. Maria Gomez-Gallego et al. studied the differences between the effects of active group music therapy and receptive group music therapy on Alzheimer's disease patients. The researchers recognized that active and receptive music tasks engage different areas of the brain, which can result in varying effects and improvements from either therapy (Gomez-Gallego et al., 2021). To conduct this study, 90 Alzheimer's patients living in nursing homes were randomly assigned active, receptive, or usual care. Assessments regarding cognition, living activities, behavior, and motor function revealed that active music therapy patients had improved the most, while receptive music therapy patients were mostly stabilized. Regarding active music therapy participants, 85.7% showed cognition improvements, 92.9% showed behavior improvements, and 46.4% showed function improvements. Regarding receptive music therapy participants, 11.8% showed cognition improvements, 42.9% showed behavior improvements, and 14.3% showed function improvements. Regarding patients receiving their usual care, 6.3% showed cognition improvements, 12.2% showed behavior improvements, and 17.4% showed function improvements (Gomez-Gallego et al., 2021). The data reveals that music therapy, especially active music therapy, is a useful Alzheimer's treatment and should be incorporated into usual therapies. However, this study does not assess the longevity of active music therapy's effects, an important factor to be considered before its implementation.

Alfredo Raglio et al. similarly compared active music therapy and receptive music therapy efficacies on behavioral and psychological symptoms on dementia (Raglio et al., 2015). To conduct this study, 120 participants with dementia were randomized into three groups: active music therapy, receptive music therapy, and usual care. The data collected revealed little differences in behavior. Additionally, the NPI global score fell 28% in the active music therapy group, 12% in the listening music therapy group, and 21% in the standard care group. The active music therapy group also demonstrated a positive trend in communication and relationships (Raglio et al., 2015). The study demonstrated that active and listening music therapy, despite Maria Gomez-Gallego et al.'s findings, do not have significant effects on behavior and

psychological dementia. Again, the limited results of Raglio et al.'s findings may be due to studying dementia in general. Thus, further studies must be conducted regarding the active and receptive music therapy effects on Alzheimer's patients specifically to confirm the data results.

Taking a different approach to implementation methods, Ana Maseda et al. studied the efficacy differences between multisensory simulation (MMSE) and individualized music therapy on mood, behavior, and biomedical parameters. They studied this topic to better understand the effects of non-pharmacological treatments on Alzheimer's patients (Maseda et al., 2018). To conduct this study, 21 patients over the age of 65 were randomized to either the MMSE or individualized music therapy groups. Both groups completed 30-minute sessions twice a week for twelve weeks. The outcomes of the study were recorded before, during, and after the therapy intervention. The researchers found that both groups showed immediate effects on mood and behavior. They noticed that the Alzheimer's patients were happier, more talkative, related to people more, more attentive, less bored, and more relaxed. However, the individualized music therapy group seemed more relaxed and happier, while the MMSE group performed a better visual follow-up of the stimuli. Additionally, both groups also experienced a decrease in heart rate and an increase in oxygen saturation (Maseda et al., 2018). Overall, the data demonstrated that non-pharmacological therapies, like multisensory stimulation and individualized music, pose as effective intervention strategies; however, the longevity of the results remains inconclusive.

Similarly, Eva M. Arroyo-Anllo et al. studied the differences between individualized and non-individualized music therapy on aspects of self-consciousness, including personal identity, anosognosia, affective state, body representation, prospective memory, introspection, and moral judgements (Arroyo-Anllo et al, 2013). This topic was studied because pharmacological interventions are limited in their effectiveness in treating Alzheimer's patients (Arroyo-Anllo et al, 2013). To conduct this study, 20 Alzheimer's patients were divided into two groups: individualized and non-individualized music therapy. The patients were given therapy sessions three times a week for twelve weeks, totalling 36 sessions. The researchers found that the individualized music therapy group experienced a stabilization or improvement in all aspects of self-consciousness. On the other hand, the non-individualized music therapy group experienced a decrease in all self-consciousness aspects, except for body representation and affective state (Arroyo-Anllo et al, 2013). Overall, Maseda et al. and Arroyo-Anllo et al.'s findings stress the importance of using familiar music during music therapy interventions. The data imply that individualized music therapy should be considered as a self-consciousness enhancer for Alzheimer's disease patients.

In conclusion, the studies collectively imply that active and individualized music therapies most effectively improve Alzheimer's patients social behavior, well-being, mood, and cognition. Additional research needs to be conducted regarding group music therapy and its corresponding efficacies.

MUSIC THERAPIES	Style of music training	Effect on memory	Short vs Long term effects	AD/Dementia
Group	Using music in a group setting to increase social interactions	Slight improvement	Both	Both
Dual-Task	Simultaneous music training and physical movement	N/A	Short-term	Dementia
Individualized	Tailoring music to patients' life experiences	Improvement	Short-term	Both
Active	Music training involving clapping, dancing, arm movements, etc.	Significant improvement	Short-term	Both
Receptive	Calm and relaxed music training	Slight improvement	Short-term	Both

Figure 1. Comparing music therapy implementation methods and their efficacies

Communication

Music therapy implementation has facilitated increased social behaviors and connectedness among Alzheimer's patients. These social behaviors include lyric reading, singing, and communicating with therapists, resulting in improvements to language and speaking abilities among Alzheimer's patients. This section focuses on the social impacts of music therapy on cognitively impaired individuals

Melissa Brotons et al. studied how music therapy specifically affected the language functioning of Alzheimer's patients. This topic was studied because it is unclear whether verbal improvements reflect specific improved language skills or if simple interactions with therapists are responsible for the improvements (Brotons et al. 2000). To conduct this study, the Alzheimer's patients were assessed for cognitive functioning using the Mini-Mental State Examination (MMSE) and for language ability using the Western Aphasia Battery (WAB). The participants were divided in groups of 2-3 people and participated in 20–30-minute sessions twice a week for a total of 8 sessions. Four of the sessions were music therapy based and the other four were conversational sessions. The participants were all retested after every 2-week interval. The researchers found significant improvements on speech content and fluency performance, while the Aphasia Quotient for music and conversions did not statistically differ (Brotons et al. 2000). The data demonstrates that additional research is required for the use of music therapy and conversational sessions as effective non-invasive intervention methods for Alzheimer's patients' communication skills. Additionally, similar to the active music therapy, perhaps this study's limited results is due to studying dementia in general. Thus, other studies specifically focusing on Alzheimer's patients have found more promising impacts to social behavior.

Thus, other studies found that music therapy had significant communication effects on Alzheimer's patients specifically. Jihui Lyu et al. studied the effects that music therapy has on cognition, psychiatric symptoms, and activities of AD patients. This topic was studied because pharmacological therapies have lacked evidence of clinical efficacy on Alzheimer's patients in the last couple decades, and new therapies must be investigated (Lyu et al., 2018). To conduct this study, 298 Alzhiemer's patients were divided into three groups: singing, lyric reading, and a control receiving no treatment. All groups received intervention for three months, and were tested on cognitive function, neuropsychological symptoms, and activities after three months and six months. The study revealed that music therapy is actually more effective for improving verbal fluency and enhancing memory rather than lyric reading. The researchers also found that the therapy alleviated psychiatric symptoms and caregiver distress, with no significant effect on the daily activities of the patients (Lyu et al., 2018). The study demonstrates that music therapy is a potential therapy for improving the communication and social wellbeing of Alzheimer's patients in addition to improving their psychological wellbeing.

On the other hand, Masayuke Satoh et al. investigated how intentionally involving communication to music therapy through singing has improved the cognitive function of Alzheimer's patients. This topic was studied because there is a lack of scientific evidence regarding how singing and verbal involvement in music therapy has affected Alzheimer's patients, and thus more studies must be conducted to confirm results (Satoh et al., 2015). To conduct this study, ten Alzheimer's patients participated in a singing training music therapy group, while ten other Alzheimer's patients were put into a control group receiving no treatment. The patients tested had a mean age of 78.1 years old. The music therapy group trained using karaoke and a unique voice training method once a week for 6 months. The participants were assessed using neuropsychological batteries and functional magnetic resonance imaging (fMRI) once before, twice during, and once after the intervention. The study revealed that the music therapy group's time for completion of the Japanese Raven's Colored Progressive Matrices was significantly reduced. The researchers also found that there was a decrease in the neuropsychiatric inventory score, a prolongation of sleep time, and increased activity in the right angular gyrus and left lingual gyrus (Satoh et al., 2015). The study demonstrates that singing training music therapy is effective for improving the neural cognitive processing of Alzheimer's patients. However, more studies must be conducted regarding singing training to confirm the results.

In conclusion, music therapy poses as a promising intervention to the social life of Alzheimer's patients. Although not very effective for individuals with dementia, patients specifically with Alzheimer's dementia experienced significant improvements to communication and social well-being. Additionally, incorporating social aspects to music therapy, like singing and lyric reading, has proven to be beneficial to the cognitive function of Alzheimer's patients.

Agitation

As research has explored music therapy's effects on Alzheimer's patients, improvements in agitation consistently emerge as a prominent theme in music intervention studies.Common signs of agitation include pacing, hand wringing, inability to sit or lie still, rapid speech, increased psychomotor activity, repeated expression of distress, and crying (Brotons et al., 1996). This section focuses on how music therapy has played a role in reducing these behaviors.

Hanne Mette O Ridder et al. studied how individualized music therapy affects the agitation of patients specifically with moderate/severe dementia in nursing homes. This topic was studied because previous studies have found music therapy to be effective, but this group of researchers deemed these interventions are procedurally insufficient (Ridder et al., 2013). To conduct this study, 42 patients with dementia were randomized into either a treatment group or a control group. The treatment group received six weeks of individualized music therapy, while the control group received six weeks of standard care. The researchers found that agitation disruptiveness increased in the control group and decreased in the treatment group. The researchers also found a significant increase in the prescription of psychotropic medication for the control group during the six weeks (Ridder et al., 2013). The study demonstrated that music therapy is vital for reducing agitation among individuals and dementia. More importantly, the study demonstrates that music therapy can pose as a potential alternative to pharmacological medications for dementia.

Melissa Brotons et al. similarly studied the effects music therapy has on agitation for Alzheimer’s patients. Their goal of the study was to also determine if there was a difference in the effects of music therapy between the patients with or without music background, and if there was difference in the agitation scores of music therapists and caregivers before and after the study (Brotons et al., 1996). This group studied this topic because there is no current cure for Alzheimer’s Disease, as drugs have limited effects in managing behavior and can have worsening side effects. Thus, the researchers sought to alleviate symptoms, decrease stress, and improve the quality of life of Alzheimer’s patients (Brotons et al., 1996). To conduct this study, 20 Alzheimer’s patients were randomized into groups of three or four people. The groups participated in 30 music therapy sessions twice a week, for a total of five sessions. The caregivers rated the subjects’ agitation behaviors during the music therapy sessions and for an additional twenty minutes following the session. The researchers found that there was no statistically significant difference between the patients with music background and those without. The study also revealed that the patients were significantly less agitated during and after the music therapy sessions, as the mean agitation scores decreased from about 10 before the session to about 7 after the session. Specific agitation behaviors were also assessed, including pacing, which decreased from 16% to 11%, hand wringing, which decreased from 8% to 7%, and crying, which decreased from 9% to 1%. However, rapid speech increased from 13% to 15%, psychomotor activity increased from 16% to 19% and distress increased from 20% to 27%. The researchers also found that there was no statistically significant difference regarding the agitation scores of the music therapists and caregivers (Brotons et al., 1996). The results of this study

similarly compare to Ridder et al.'s study, concluding music therapy positively affects agitation of Alzheimer's.

Overall, both studies demonstrate that music therapy is an effective behavioral intervention for managing agitation among Alzheimer's patients. However, various unanswered questions remain, including whether the sessions were effective due to their structured format or whether music therapy was just causing fatigue that hence decreased agitation (Brotons et al. 1996).

Pathology/Mechanisms

While studies exploring the physical effects of music therapy on Alzheimer's patients are important for future long-term implementation, it is also crucial to explore the pathology of music therapy. Studies have investigated how music therapy has specifically affected the brain functions and its systems.

AM Kumar et al. studied how music therapy affects the concentration of melatonin, norepinephrine, epinephrine, serotonin, and prolactin in the blood of Alzheimer's patients. They studied this topic because music therapy is known to release certain neurohormones from the specifics of their neurohormonal systems that have not been fully investigated (Kumar et al. 1999). To conduct this study, 20 male Alzheimer's patients participated in 30–40-minute morning sessions of music therapy 5 times a week for four weeks. Blood samples were taken after the four weeks of study, and then again, an additional two weeks later. The researchers found that melatonin concentration significantly increased after 4 weeks, and even more after six weeks. They also found that norepinephrine and epinephrine levels increased significantly after 4 weeks but returned to normal levels after 6 weeks. However, prolactin and platelet serotonin levels remained unchanged (Kumar et al. 1999). Overall, the data confirms that the release of melatonin during music therapy sessions is an explanatory factor to the calming effects observed in other studies investigating music intervention.

Currently, there is exceptionally limited research regarding the pathology of music therapy, thus demanding the conduct of additional studies to validate Kumar et al.'s results. Additional research must also work to find any other beneficiary effects of music therapy on other brain mechanisms.

Other Implementation Methods

After analyzing the variety of implementation methods and effects of music therapy, there remains studies that differ from traditional music therapy, but still involve music to the patients' treatment regimen. Specifically, studies have investigated if implementing music therapy in addition to pharmacological treatments is any more beneficial to the patients' wellbeing. Other therapies investigated have involved other art forms, like drawing and painting, or have involved other technologies, like videos.

Music therapy has emerged as a potential alternative to pharmacological treatment for many years, but researchers have also explored the effects of combining pharmacological and

non-pharmacological treatments in an Alzheimer's patient's treatment regimen. Anna Rita Giovagnoli et al. sought to clarify if providing Alzhiemer's patients with both active music therapy and the drug memantine could improve language compared to just using drugs alone. This topic was studied due to the lack of existing research regarding the joint effects of music therapy and medications (Giovagnoli et al., 2018). To conduct this study, 45 moderate Alzheimer's patients already receiving stable doses of acetylcholinesterase inhibitors participated in active music therapy and 20 mg per day of memantine for 24 weeks. Language, overall cognition, psycho-behavior, social, and functional aspects were assessed at weeks 12 and 24. The researchers found improvement in the depression and appetite scores of patients receiving both music therapy and memantine rather than just memantine. There were also fewer patients in the music therapy and memantine group that indicated a worsening in their NPI scores. Additionally, daily activity, social relationships, and cognitive performance remained relatively stable. However, the music therapy and memantine group had no increased language improvements than using memantine alone (Giovagnoli et al., 2018). Overall, the study reveals that incorporating music therapy along with medication can work to enhance the benefits of pharmacological interventions and further improve the daily life of Alzheimer's patients.

Aside from pharmacological interventions, therapies involving other art forms aside from music have been considered as a means for inhibiting the progression of Alzheimer's symptoms. Elodie Pongan et al. studied the effects of choral singing compared to painting sessions on chronic pain, mood, quality of life, and cognition in Alzheimer's patients. This topic was studied because the efficacy differences regarding music and painting remain debatable, and must be investigated further (Pongan et al., 2017). To conduct this study, 31 mild Alzheimer's patients were grouped in a 12-week singing group, while another 28 mild Alzheimer's patients participated in 12 weeks of painting sessions. Chronic pain, anxiety, depression, and quality of life were assessed before, after, and one month after the sessions ended. The researchers found that both the singing and painting group experienced significant reductions in pain, depression, and anxiety, and improvements in quality of life, digit span, and inhibitory processes. Additionally, verbal memory performance remained stable in the singing group and decreased in the painting group (Pongan et al., 2017). Overall, the study implies that both painting and singing are useful interventions for improving the quality of life for mild Alzheimer's patients. However, singing posed as a slightly more beneficial therapy and should be considered for long-term implementation among cognitively impaired individuals.

Similarly, other researchers have studied how involving other technologies to traditional music therapy has provided any additional improvements to Alzheimer's patients' wellbeing. Ivan Rubbi et al. specifically studied how video music therapy affects the quality of life of Alzheimer's patients. They studied this topic because pharmacological treatments have limited effects on treating Alzheimer's, and management of the disease has shown previous growth through studies involving non-pharmacological treatments (Rubbi et al., 2016). To conduct this study, 32 patients were treated with two cycles of 6 video-music therapy sessions. The sessions consisted of folk music and videos to help stimulate recall. Mini Mental State Examinations

(MMSE) were administered before the test to divide the participants into test groups, and participants were assessed after the sessions using the quality-of-life scale (QOL-AD). The researchers found that the patients with MMSE stages 1, 2, and 3 (mild to moderate Alzheimer's) reported an improvement in quality of life, including physician health, energy, mood, memory, mental well-being, and daily activities. The therapy also reduced anxiety and loneliness and improved their communication and social life. However, participants with MMSE stage 4 (severe Alzheimer's) did not report any change after participating in the video music therapy sessions (Rubbi et al., 2016). The study demonstrates that video music therapy is a useful intervention among patients with mild to moderate Alzheimer's. However, other non-pharmacological or pharmacological interventions must be looked into to address the symptoms of severe Alzheimer's patients.

In conclusion, different implementation methods apart from traditional music therapies, like video and painting, or the incorporation or pharmacological treatments, have proven to be beneficial to Alzheimer's patients' well-being. Additional studies should explore and investigate other implementation methods that differ from general receptive music therapy that could be administered to nursing homes to help slow down and prevent future cognitive decline.

Conclusion

Alzheimer's disease is a growing concern among a large percentage of the world's population. Thus, developing effective strategies for treatment and therapeutics represent an important field requiring increased attention. Developing consistent methods to address the behavioral side effects of dementia would alleviate the challenges many patients face in their daily lives. Most studies regarding music therapy result in positive cognitive, behavioural, and pathological effects, demonstrating the need to prescribe music therapy among cognitively impaired individuals. Increased opportunity for Alzheimer's patients to participate in music therapy would ultimately lead to regressed memory loss and improved quality of life. Despite music therapy's undoubtedly positive effects among Alzheimer's patients specifically, the intervention's benefits are inconclusive regarding patients with dementia in general. Thus, future studies must investigate other treatments and therapies for dementia that can be implemented beneficially long-term.

Works Cited

Brotons, M., and P. K. Pickett-Cooper. "The Effects of Music Therapy Intervention on Agitation Behaviors of Alzheimer's Disease Patients." Journal of Music Therapy, vol. 33, no. 1, 1 Mar. 1996, pp. 2–18, https://doi.org/10.1093/jmt/33.1.2.

Brotons, Melissa, and Susan Koger. "The Impact of Music Therapy on Language Functioning in Dementia." Journal of Music Therapy, vol. 37, no. 3, 1 Sept. 2000, pp. 183–195, https://doi.org/10.1093/jmt/37.3.183.

Chen, Yu-Ling, and Yu-Cheng Pei. "Musical Dual-Task Training in Patients with Mild-To-Moderate Dementia: A Randomized Controlled Trial." Neuropsychiatric Disease and Treatment, vol. Volume 14, no. 354, May 2018, pp. 1381–1393, https://doi.org/10.2147/ndt.s159174.

Chu, Hsin, et al. "The Impact of Group Music Therapy on Depression and Cognition in Elderly Persons with Dementia: A Randomized Controlled Study." Biological Research for Nursing, vol. 16, no. 2, 2014, pp. 209–17, www.ncbi.nlm.nih.gov/pubmed/23639952, https://doi.org/10.1177/1099800413485410. Accessed 2 Sept. 2024.

Clair, Alicia Ann, and Barry Bernstein. "A Preliminary Study of Music Therapy Programming for Severely Regressed Persons with Alzheimer's-Type Dementia." Journal of Applied Gerontology, vol. 9, no. 3, Sept. 1990, pp. 299–311, https://doi.org/10.1177/073346489000900305.

Clute, Owen. "An Investigation into the Short-Term Effects of Music Therapy for Patients with Probable Alzheimer's Disease or Another Form of Dementia." The Young Researcher, vol. 1, no. 1, 30 June 2017, pp. 129–138.

Gómez-Gallego, María, et al. "Comparative Efficacy of Active Group Music Intervention versus Group Music Listening in Alzheimer's Disease." International Journal of Environmental Research and Public Health, vol. 18, no. 15, 30 July 2021, p. 8067, https://doi.org/10.3390/ijerph18158067.

Groene, R. W. "Effectiveness of Music Therapy 1:1 Intervention with Individuals Having Senile Dementia of the Alzheimer's Type." Journal of Music Therapy, vol. 30, no. 3, 1 Sept. 1993, pp. 138–157, https://doi.org/10.1093/jmt/30.3.138. Accessed 4 Oct. 2024.

Gulliver, Amelia, et al. "The Music Engagement Program for People with Alzheimer's Disease and Dementia: Pilot Feasibility Trial Outcomes." Evaluation and Program Planning, vol. 87, 1 Aug. 2021, p. 101930, www.sciencedirect.com/science/article/pii/S0149718921000252?via%3Dihub, https://doi.org/10.1016/j.evalprogplan.2021.101930. Accessed 6 Oct. 2024.

Irish, Muireann, et al. "Investigating the Enhancing Effect of Music on Autobiographical Memory in Mild Alzheimer's Disease." Dementia and Geriatric Cognitive Disorders, vol. 22, no. 1, 23 May 2006, pp. 108–120, www.karger.com/Article/Abstract/93487, https://doi.org/10.1159/000093487. Accessed 13 Oct. 2024.

Kumar, A. M., et al. “Music Therapy Increases Serum Melatonin Levels in Patients with Alzheimer’s Disease.” Alternative Therapies in Health and Medicine, vol. 5, no. 6, 1 Nov. 1999, pp. 49–57, pubmed.ncbi.nlm.nih.gov/10550905/. Accessed 8 Sept. 2024.

Lyu, Jihui, et al. “The Effects of Music Therapy on Cognition, Psychiatric Symptoms, and Activities of Daily Living in Patients with Alzheimer’s Disease.” Journal of Alzheimer’s Disease, vol. 64, no. 4, 24 July 2018, pp. 1347–1358, content.iospress.com/articles/journal-of-alzheimers-disease/jad180183, https://doi.org/10.3233/jad-180183. Accessed 2 Sept. 2024.

Madsø, Kristine Gustavsen, et al. “Observing Music Therapy in Dementia: Repeated Single-Case Studies Assessing Well-Being and Sociable Interaction.” Clinical Gerontologist, vol. 45, no. 4, 29 Sept. 2021, pp. 1–15, https://doi.org/10.1080/07317115.2021.1978121.

Maseda, Ana, et al. “Multisensory Stimulation and Individualized Music Sessions on Older Adults with Severe Dementia: Effects on Mood, Behavior, and Biomedical Parameters.” Journal of Alzheimer’s Disease, vol. 63, no. 4, 30 May 2018, pp. 1415–1425, content.iospress.com/articles/journal-of-alzheimers-disease/jad180109, https://doi.org/10.3233/jad-180109. Accessed 13 Oct. 2024.

National Institute of Aging. “What Happens to the Brain in Alzheimer’s Disease?” National Institute on Aging, 19 Jan. 2024, www.nia.nih.gov/health/alzheimers-causes-and-risk-factors/what-happens-brain-alzheimers-disease. Accessed 30 Nov. 2024.

National Institute on Aging. “Alzheimer’s Disease Fact Sheet.” National Institute on Aging, National Institutes of Health, 2023, www.nia.nih.gov/health/alzheimers-and-dementia/alzheimers-disease-fact-sheet. Accessed 30 Nov. 2024.

National Institute on Aging. “How Is Alzheimer’s Disease Treated?” National Institute on Aging, 12 Sept. 2023, www.nia.nih.gov/health/alzheimers-treatment/how-alzheimers-disease-treated. Accessed 30 Nov. 2024.

Pongan, Elodie, et al. “Can Musical or Painting Interventions Improve Chronic Pain, Mood, Quality of Life, and Cognition in Patients with Mild Alzheimer’s Disease? Evidence from a Randomized Controlled Trial.” Journal of Alzheimer’s Disease : JAD, vol. 60, no. 2, 2017, pp. 663–677, www.ncbi.nlm.nih.gov/pubmed/28922159, https://doi.org/10.3233/JAD-170410. Accessed 29 Sept. 2024.

Prickett, C. A., and R. S. Moore. “The Use of Music to Aid Memory of Alzheimer’s Patients.” Journal of Music Therapy, vol. 28, no. 2, 1 June 1991, pp. 101–110, academic.oup.com/jmt/article-abstract/28/2/101/906612, https://doi.org/10.1093/jmt/28.2.101. Accessed 7 Oct. 2024.

Raglio, Alfredo, et al. “Effect of Active Music Therapy and Individualized Listening to Music on Dementia: A Multicenter Randomized Controlled Trial.” Journal of the American Geriatrics Society, vol. 63, no. 8, Aug. 2015, pp. 1534–1539, https://doi.org/10.1111/jgs.13558.

Raglio, Alfredo, and Osmano Oasi. “Music and Health: What Interventions for What Results?” Frontiers in Psychology, vol. 6, 2 Mar. 2015, https://doi.org/10.3389/fpsyg.2015.00230. Accessed 28 Nov. 2024.

Ridder, Hanne Mette O., et al. “Individual Music Therapy for Agitation in Dementia: An Exploratory Randomized Controlled Trial.” Aging & Mental Health, vol. 17, no. 6, 27 Apr. 2013, pp. 667–678, www.ncbi.nlm.nih.gov/pmc/articles/PMC4685573/, https://doi.org/10.1080/13607863.2013.790926. Accessed 8 Sept. 2024.

Rubbi, Ivan. “Efficacy of Video-Music Therapy on Quality of Life Improvement in a Group of Patients with Alzheimer’s Disease: A Pre-Post Study.” Acta Bio-Medica : Atenei Parmensis, vol. 87, no. 4-S, 2016, pubmed.ncbi.nlm.nih.gov/27874842/. Accessed 29 Sept. 2024.

S, Guétin, et al. “Effect of Music Therapy on Anxiety and Depression in Patients with Alzheimer’s Type Dementia: Randomised, Controlled Study.” Dementia and Geriatric Cognitive Disorders, 2009, pubmed.ncbi.nlm.nih.gov/19628939/. Accessed 29 Sept. 2024.

Satoh, Masayuki, et al. “Music Therapy Using Singing Training Improves Psychomotor Speed in Patients with Alzheimer’s Disease: A Neuropsychological and FMRI Study.” Dementia and Geriatric Cognitive Disorders Extra, vol. 5, no. 3, 4 Sept. 2015, pp. 296–308, www.ncbi.nlm.nih.gov/pmc/articles/PMC4608626/, https://doi.org/10.1159/000436960. Accessed 11 Oct. 2024.

Simmons-Stern, Nicholas R., et al. “Music as a Memory Enhancer in Patients with Alzheimer’s Disease.” Neuropsychologia, vol. 48, no. 10, 7 May 2010, pp. 3164–3167, www.ncbi.nlm.nih.gov/pmc/articles/PMC2914108/, https://doi.org/10.1016/j.neuropsychologia.2010.04.033. Accessed 13 Oct. 2024.

Takahashi, Takiko, and Hiroko Matsushita. “Long-Term Effects of Music Therapy on Elderly with Moderate/Severe Dementia.” Journal of Music Therapy, vol. 43, no. 4, 1 Dec. 2006, pp. 317–333, https://doi.org/10.1093/jmt/43.4.317.

van der Steen, Jenny T, et al. “Music-Based Therapeutic Interventions for People with Dementia.” Cochrane Database of Systematic Reviews, vol. 5, no. 5, 2 May 2017, www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD003477.pub3/abstract, https://doi.org/10.1002/14651858.cd003477.pub3. Accessed 12 Oct. 2024.

World Health Organization. “Dementia.” World Health Organization, 15 Mar. 2023, www.who.int/news-room/fact-sheets/detail/dementia. Accessed 30 Nov. 2024.

Black Holes By Arsh Garg

Abstract

This project explores how objects behave around black holes and how bodies, both big and small, interact with black holes. Objects, primarily consisting of gasses and other matter, are attracted and move towards the black hole due to its immense gravitational pull. These objects start revolving around the black hole and form an excretion disk where these objects get heated to extreme temperatures. They eventually fall into the black hole (past the event horizon) and are unable to escape. Black holes force the object to revolve around it at high speed and temperature due to its gravitational force, forming an accretion disc. They eventually get absorbed into the black hole and cannot escape. Black holes also interact with other black holes. They attract each other and eventually merge into each other which results in them combining to form a bigger black hole and releasing large amounts of gravitational waves.

Introduction

As quoted by Albert Einstein, " Black holes are where God divided by zero." Black holes are weirder than anything else in the universe. A black hole is thought to contain a singularity, which is a point of infinite density. This point breaks down so many concepts and theories. Does the singularity really exist? This paper studies the interaction of black holes with everything and anything in its surroundings, ranging from small objects to celestial bodies the size of the sun to other supermassive black holes.

What are black holes, and how do they form?

Black holes are regions of space where gravity is so strong that nothing, not even light, can escape from them. Black holes consist of a point of infinite density known as a singularity. It is surrounded by an event horizon—the boundary beyond which nothing can escape. Black holes can vary in size, with stellar black holes forming from individual stars, and supermassive black holes existing at the centers of galaxies, containing millions to billions of times the mass of the Sun ("The Editors of Encyclopedia Britannica", 2024). -

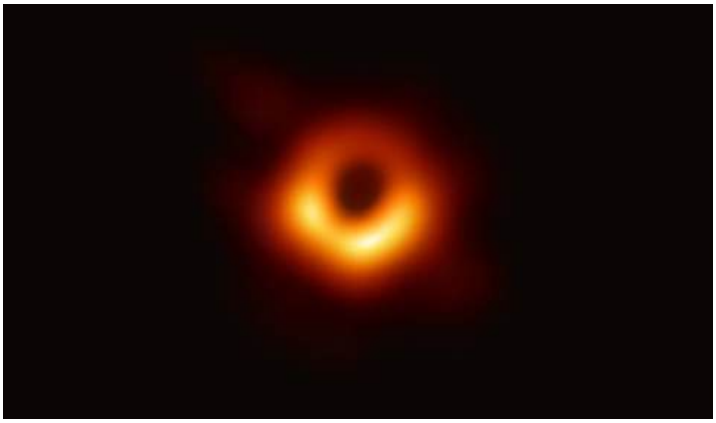


Figure 1: This is the first ever image of a supermassive black hole at the centre of galaxy M87 captured using the event horizon telescope. The black hole is outlined by emission from hot gas swirling around it under the influence of strong gravity near its event horizon.
Image taken from (NASA Science, 2019).

Black holes are formed for massive stars that run out of fuel to continue fusion reactions. Nuclear fusion is a process in which two light atomic nuclei combine to form a heavier nucleus, releasing a significant amount of energy in the process. This reaction occurs under extreme temperatures and pressures, typically found in the cores of stars, where hydrogen nuclei (protons) fuse to create helium and other heavier elements. During fusion, the mass of the resulting nucleus is slightly less than the sum of the original nuclei, and this mass difference is converted into energy according to Einstein's equation $E = mc^2$. Fusion is the source of a star's energy, providing the heat and light that power the universe, and this radiation is also responsible for balancing out the gravitational forces acting on the star.

When temperatures reach around 10 million K, hydrogen atoms fuse into helium in the core, releasing energy. This reaction is necessary for a star's survival as the outward pressure from nuclear fusion balances the inward gravitational pull, allowing the star to maintain a stable size. Stars spend approximately 90% of their life in this phase. The duration varies with mass: smaller stars (like the Sun) can last billions of years, while massive stars may last only millions. In high mass stars, due to extreme pressure and temperature, the fusion process continues beyond helium. It continues to form carbon, oxygen, neon, magnesium, silicon, and elements of higher atomic numbers until iron. The iron fusion reaction absorbs energy leading to core collapse and a supernova explosion (“Stellar Evolution”, 2024).

After a supernova explosion all that remains is the super dense core of the star and hot gas of the nebula. If what remains of the core of the star after the supernova explosion has a mass less than about three times the Sun's mass, then it forms into a neutron star. The force of gravity is so huge that the protons and electrons combine to form neutrons. As a result, neutron stars have a mass greater than that of the Sun but are only about 20 kilometers in diameter, making them extraordinarily dense. If what remains of the core of the star after the supernova explosion has a mass greater than about three times the Sun's mass, then it collapses into a black hole. (“How can a star become a black hole?”, n.d.)



Figure 2: This is the image of a supernova explosion. This explosion usually results in the formation of neutron stars or black holes (Ezzy Pearson, 2024)

What do black holes consist of?

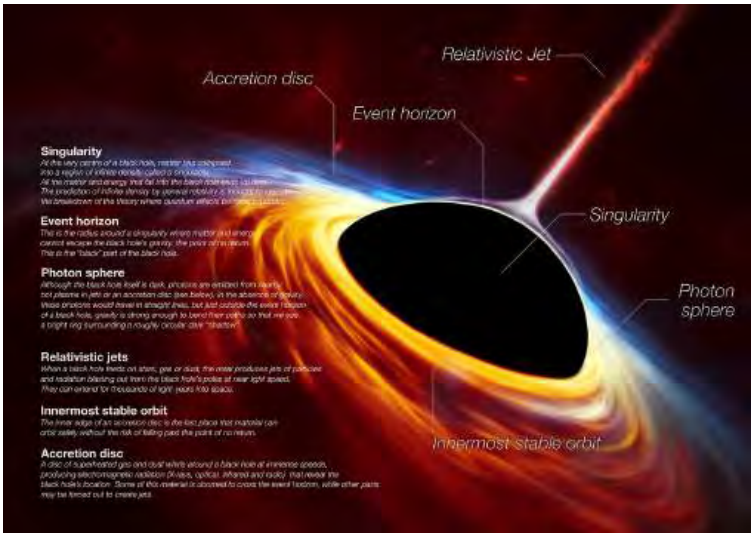


Figure 3: This is the various elements of a black hole and shows the complex structure of black holes ("Anatomy of a Black Hole", 2019)

· **Event Horizon**

The event horizon acts like the surface of a black hole and marks the boundary where the escape velocity surpasses the speed of light, the maximum speed at which anything can travel in the universe. Nothing can escape once it enters the event horizon, not even light. Since no light can leave, black holes themselves neither emit nor reflect it, making them invisible to outside observers. However, black holes can still be detected by observing the light from surrounding matter that has not fallen past the event horizon yet.

· **Accretion Disk**

A black hole is visible due to the light emitted by its accretion disk, a structure that consists of all the matter the black hole pulls in and consumes, a process known as accretion. Black holes can grow by absorbing gas from nearby stars, or by merging with other black holes. Stellar-mass black holes are usually paired with stars and draw gas from the star into the black hole, while supermassive black holes capture gas from stars that wander too close. As this gas falls in, it spins around the black hole at immense speeds and heats up to intense temperatures. As matter moves from the outer regions of the disk to its inner edge, it eventually falls into the event horizon. Black holes that have already consumed all nearby matter and are isolated lack an accretion disk and are much harder to detect and study.

If we could observe the accretion disk closely, we would notice how unusual its shape is when viewed from different angles. This distortion occurs because the black hole's intense gravitational field bends the space-time fabric, and forces light to follow a curved path. This phenomenon is called gravitational lensing. Light from the top of the disk behind the black hole appears as a hump above it, while light from the bottom of the far side of the disk forms another hump below. The size and shape of these humps changes depending on the viewing angle, and they become invisible when the disk is viewed directly face-on.

· **Corona**

This region around a black hole is considered as one of the most extreme physical environments in the universe. Powerful magnetic fields from the inner accretion disk stretch outward, and form a hot, turbulent cloud known as the corona. The temperatures inside this corona reach billions of degrees. Particles in the corona move at speeds close to that of light, and the corona emits X-rays with far higher energies than those coming from the accretion disk. However, astronomers are still working to understand the exact size, shape, and other properties of this corona.

· **Particle jets**

An unusual phenomenon occurs near the inner edge of the accretion disk In black holes of all sizes. Two jets that shoot out in opposite directions sometimes unexpectedly capture some of the material falling into the black hole. Particles inside these jets are propelled at nearly the speed of light, although astronomers still do not fully understand the mechanics behind them. Supermassive black holes, like those at the centres of most large galaxies, have massive particle jets that can extend hundreds of thousands of light-years. When one of these jets points in our direction, we tend to see it more clearly due to a phenomenon known as Doppler beaming, which makes the jet coming toward us much brighter and significantly dims the one moving away.

· **Singularity**

According to predictions of General Relativity, a point of infinite density is present at the centre of a black hole; this point is known as a singularity. Falling into this point is the ultimate

fate of anything that crosses the event horizon. The singularity could either be a real physical structure or merely a mathematical concept, but astronomers are uncertain which it is. The prediction of a singularity may indicate the limits of General Relativity, implying that quantum effects—currently not accounted for in the theory—might play a crucial role in providing a more complete understanding of gravity.(“Anatomy of a Black Hole”,n.d.)

How do objects behave around black holes?

Objects encounter intense gravitational forces that dramatically alter their behavior near a black hole. These objects accelerate and heat up as they spiral toward the event horizon, drawn in by the black hole's tremendous gravity. Around the black hole, matter frequently forms an accretion disk, where it orbits at extraordinarily high speeds and releases a lot of radiation, including X-rays. The closer an object approaches to the event horizon, the stronger the gravitational pull grows, stretching and compressing matter in a process known as spaghettification.

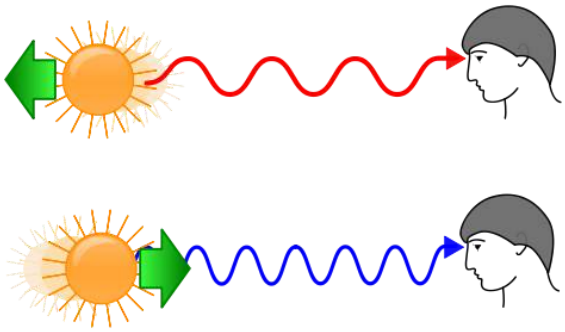


Figure 4: the figure above shows us how objects become redshifted. It happens due to the light waves spreading as the light source moves away from the observer rapidly. For example, if a lightbulb were to move very rapidly through space, the light would appear blue as it approaches you and then become red after it passes. (Christopher Crockett, 2021)

The enormous gravity of the black hole causes time to behave differently as an object approaches the event horizon. A phenomenon known as gravitational time dilation occurs as an object gets closer to the event horizon, giving the impression to a distant observer that time is slowing down for it. Reaching the event horizon, the object seems to freeze and starts to become redder and eventually disappears. This happens because the light coming from it becomes too redshifted to see. Once an object crosses the event horizon, It is impossible for it to escape. From this point, the object becomes invisible to an outside observer as no information or light can come back out of the black hole. This unique behaviour makes black holes some of the most mysterious and fascinating objects in the universe. (Chelsea Gohd, 2023)

How do black holes behave around other black holes?

How can supermassive black holes get close enough to spiral together and coalesce? In calculations, when the converging holes reach the so-called final parsec — a distance of about

one parsec, or 3.26 light-years — their progress stalls. They should essentially orbit each other indefinitely after that point.

“It was thought in-spiral times could be as high as … the age of the universe,” said Stephen Taylor, an astrophysicist at Vanderbilt University. “People were concerned that you might not get any merging black holes.”

Evidence has accrued that they do merge. Last year, observations of the subtle movements of pulsating stars known as pulsar timing arrays revealed a background hum of gravitational waves in the universe — ripples in the fabric of space-time. These gravitational waves most likely come from tightly orbiting supermassive black holes within a parsec of each other that are close to merging. So how do they merge?

Recently, scientists proposed that certain types of dark matter might help solve this problem by draining angular momentum from the black holes, nudging them into a collision course. One idea involves *self-interacting dark matter* (SIDM), a complex form of dark matter hypothesized by physicist Gonzalo Alonso-Álvarez and his team at the University of Toronto.(Alonso-Álvarez et al., 2024) Unlike cold dark matter, which would passively respond only to gravity, SIDM particles can interact with each other, behaving like a viscous fluid. This "stickiness" could create drag on the black holes, slowing them and drawing them together to trigger a merger. A separate team in September proposed that *fuzzy dark matter*, composed of ultralight particles that form wave-like structures, could achieve similar results.(koo et al., 2024**)** As black holes move through these waves, the dark matter could drain their orbital energy, bringing them closer.

Alternative explanations suggest that regular stars or gas disks might contribute to black hole mergers. For instance, stars in nearby orbits could pass close enough to the black holes to drain angular momentum, gradually driving them closer. Another proposal is that gas disks surrounding each black hole might attract material from a wider disk, creating a gradual loss of energy that could lead to a merger. Additionally, the presence of a *third black hole* in the system could provide a "three-body interaction" strong enough to destabilize the black hole pair and speed up their collision.

Astrophysicists plan to test these theories using data from pulsar timing arrays and the future Laser Interferometer Space Antenna (LISA), a gravitational wave observatory expected to launch in 2035. If self-interacting dark matter is indeed at work, gravitational wave signals should show distinctive patterns at close distances between black holes, providing clues about the nature of dark matter and black hole dynamics.

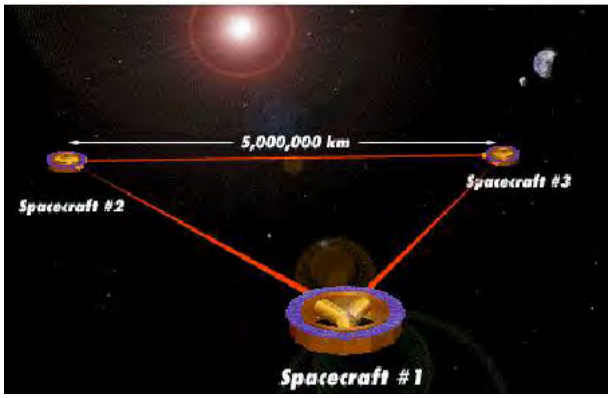


Figure 5: The idea is to have three spacecraft flying in an equilateral triangle formation. The length of the triangle side will be five million kilometers.GWs will stretch or contract the space between the spacecraft. These changes in distance will be detected using laser interferometry, just like the ground-based IFOs but the interferometer arm length is a million times longer and so a million times more sensitive. (LISA--Laser Interferometer Space Antenna,n.d.)

This ongoing research could not only solve the final-parsec problem but also reveal new properties of dark matter, shedding light on 85% of the universe's mass that remains invisible yet profoundly influential. (O'Callaghan, 2024)

Black hole effects on the space time fabric

Black holes represent one of the most extreme environments in the universe, characterized by gravitational fields so intense that even light cannot escape them, and by densities that push the boundaries of our understanding. Researching the effects of black holes on the space-time fabric may help us understand some of its properties. To accomplish this, we need to study the space-time fabric. The concept of space-time fabric was developed from Albert Einstein's theories of relativity and expanded through Hermann Minkowski's four-dimensional model, which interweaves space and time as a unified, interdependent framework. According to Einstein's General Theory of Relativity, gravity emerges as the curvature of space-time, and black holes represent regions of such extreme gravitational curvature that they fundamentally distort surrounding space and time.

Examining the mathematics governing black holes, focusing on the solutions to Einstein's field equations that describe them, gives us two metrics for rotating and non-rotating black holes. The Schwarzschild metric is used to characterize non-rotating black holes, providing insight into their gravitational influence and space-time distortion. The Kerr metric, which describes rotating black holes, adds another layer of complexity by accounting for rotational energy. The Kerr metric introduces the phenomenon of frame-dragging, wherein space-time itself appears to be "dragged" along with the black hole's rotation. This phenomenon has been observed around rotating black holes and represents a unique feature of how space and time interact with them. Alongside frame-dragging, gravitational time dilation occurs near black holes. This is the effect where time slows down significantly for an observer near the event horizon compared to an

observer farther away. This dilation has been observed in high-gravity environments and represents a tangible demonstration of how gravity warps time as well as space.

General relativity successfully describes the behavior of massive objects and the gravitational fields they generate but struggles at the scale of black holes, where quantum effects cannot be ignored. Alternatively, quantum mechanics explains the behavior of particles at microscopic scales but lacks a framework for incorporating gravity in extreme scenarios. This discrepancy becomes especially evident when examining singularities. Within the singularity, density and gravitational forces theoretically become infinite, creating a point where space and time cease to have conventional meaning. At this point, the laws of physics as currently understood break down. The singularity presents a paradox: while general relativity predicts its existence, quantum mechanics denies it, as such a state defies the principles of information and energy conservation central to quantum theory.

One of the pivotal issues around is the so-called "information paradox," brought to prominence by Stephen Hawking's discovery of Hawking Radiation. This radiation, a form of energy emitted by black holes, implies that black holes emit a faint glow due to quantum effects near the event horizon. Over vast timescales, Hawking Radiation could theoretically cause black holes to lose mass and eventually evaporate completely. This raises an unsettling question: if a black hole containing vast amounts of information ultimately evaporates, what happens to the information it holds? In classical terms, information would be irrevocably lost, contradicting quantum mechanics, which maintains that information is conserved even if matter is transformed. This paradox remains one of the most profound unresolved issues in theoretical physics. Efforts to resolve it have driven much of the research into developing a quantum theory of gravity that could unify general relativity and quantum mechanics. (Aliza & Sophia, 2023)

Conclusion

In conclusion, we have discussed what black holes are and how they form. We have reviewed in detail all the components that make up a black hole. We have also seen the major effects black holes have on tiny objects, other black holes, and the spacetime continuum itself. This helps us to understand one of the most confusing elements of our universe just a little better. Understanding black holes may be the key that helps us to understand the laws that govern our universe. There are multiple areas where research would go a long way in helping our understanding of the universe. One of the areas is dark matter. We still do not know much about dark matter, but what we do know has helped us solve many issues in our theories (like the “Final-Parsec” problem). Another area that requires research is the singularity of black holes, and understanding its properties would help us in maybe understanding the true laws of the universe.

Works Cited

“First Image of a Black Hole - NASA Science.” *NASA*, NASA, 2019, science.nasa.gov/resource/first-image-of-a-black-hole/

"black hole" Britannica, The Editors of Encyclopaedia. . *Encyclopedia Britannica*, 26 Dec. 2024, https://www.britannica.com/science/black-hole.

"Stellar evolution." Wikipedia contributors. *Wikipedia, The Free Encyclopedia*. Wikipedia, The Free Encyclopedia, 27 Oct. 2024. https://en.wikipedia.org/wiki/Stellar_evolution.

“Hubblesite: Black Holes: Gravity’s Relentless Pull Interactive: About This Site: Encyclopedia.” *STScI Home*, www.stsci.edu/~marel/black_holes/encyc_mod3_q12.html.

Pearson, Ezzy. “What Is a Supernova? A Complete Guide to Exploding Stars and How Astronomers Observe Them.” *BBC Sky at Night Magazine*, 26 Apr. 2024, www.skyatnightmagazine.com/space-science/when-stars-collapse-what-is-a-supernova.

“Anatomy of a Black Hole.” *Www.Eso.Org*, 10 Apr. 2019, www.eso.org/public/images/eso1907h/.

“Anatomy - NASA Science.” *NASA*, NASA, science.nasa.gov/universe/black-holes/anatomy/.

Crockett, Christopher. “What Do Redshifts Tell Astronomers?” *EarthSky*, 24 Jan. 2021, earthsky.org/astronomy-essentials/what-is-a-redshift/

“What Happens When Something Gets ‘Too Close’ to a Black Hole? - NASA Science.” *NASA*, NASA,science.nasa.gov/universe/what-happens-when-something-gets-too-close-to-a-black-hole/.

Dewar, Caitlyn, et al. “Phys. Rev. Lett. 133, 021401 (2024) - Self-Interacting Dark Matter Solves the Final Parsec Problem of Supermassive Black Hole Mergers.” *Self-Interacting Dark Matter Solves the Final Parsec Problem of Supermassive Black Hole Mergers*, 9 July 2024, journals.aps.org/prl/abstract/10.1103/PhysRevLett.133.021401.

Koo, Hyeonmo, et al. “Final Parsec Problem of Black Hole Mergers and Ultralight Dark Matter.” *Physics Letters B*, North-Holland, 29 July 2024, www.sciencedirect.com/science/article/pii/S0370269324004660.

O’Callaghan, Jonathan, et al. “How Do Merging Supermassive Black Holes Pass the Final Parsec?” *Quanta Magazine*, 23 Oct. 2024, www.quantamagazine.org/how-do-merging-supermassive-black-holes-pass-the-final-parsec-20241023/

“LISA--Laser Interferometer Space Antenna.” *Lisa--Laser Interferometer Space Antenna*, web.mit.edu/klmitch/classes/8.224/project/lisa.html.

Fatima, Aliza, and Sophia Kressy. *View of Effects of Black Holes on the Space-Time Continuum*, 2023, www.jsr.org/hs/index.php/path/article/view/5351/2501 .

What Happens When You Die? Life Beyond The Final Goodbye By Katherine Sheng

Introduction

Today, AI is expanding expeditiously with new learning models serving various purposes, including the latest versions of ChatGPT and Google Gemini. One of these innovations is AI-assisted bereavement support, also known as virtual avatars for people who lost loved ones or grief clones. These programs help people cope with grief by creating a virtual avatar mimicking how a familiar loved one would talk or interact. This innovation originates from the shared basic wish to continue seeing one's loved one again and feel their presence after they depart from the world. Some obvious benefits of having AI bereavement support include creating emotional closure through talking with the clones, alleviating grief, and helping to adapt an individual to the eventual life without that loved one.

Despite its innovativeness, this technology was not accepted by all who encountered it, with many stating that they found this machine "creepy." Admittedly, this new technology fulfills a need, but it is not without controversies. Many people find it unacceptable or even unethical. For example, one common argument is that the avatar is not real, and there is no genuine relationship between two interacting parties, making these interactions "fake" and "meaningless." These challenges may relate to the essential question of whether or not there's life after death. Because if life after death exists, this AI-assisted technology can be considered unethical and unhelpful for good philosophical reasons. This paper dives into the philosophical ideas behind life after death as an argument against using AI-assisted bereavement support technology.

Literature Review: The Philosophical Question of Life after Death

The advancements of AI have led to Chat GPT, IOS 18, and finally Deathbots. Nora Lindemann's *The Ethics of Deathbots* dives deep into the morals behind the so-called "Deathbots," an AI machine helping cope with grief from a lost loved one. Lindemann believes that AI-assisted bereavement support is both unethical and impractical.[1] By analyzing its usefulness and moral improvements, she concluded that the AI wouldn't seem believable to the person who lost a loved one and AI would be disrespecting the deceased person by replacing them. One negative aspect of these death bots is when their "owner" gets overly attached to them and relies on them too much. This will also eventually lead to the owner always looking for the deathbot when grieving anything, another form of overreliance. This overreliance will lead to the subject never actually healing from the grief because the death bot is a replacement for the person that passed, so always turning to it for comfort will have harmful impacts in the long run.

Through this over-attachment theory, Lindemann concludes that death bots don't work realistically even for healing from bereavement. For any bereavement support to work, the subject has to have a 'fulfilled' life again and resume living as one did before they passed,

[1] Nora Lindemann. 11-22-2022. "The Ethics of 'DeathBots." Science and Engineering Ethics.

adapting to this change over time. However, death bots just create the sense that they are still here, thus never fully living or adapting to that loss.

In addition, death bots will also negatively affect the grieving process. Sadness over a lost one would probably dissipate in a year or two, probably sooner. However, the virtual avatar would just remind the owner that the lost one has truly passed away every single day that the AI replacement is in their household. This would inevitably lead to much more grief, with the death bot doing the exact opposite of what it was 'trained' to do. This would only prolong the process of bereavement healing.

Death Bots also makes the atmosphere seem unnatural. Communicating with a loved one will appear very natural, everything said has compassion behind it. Compared to communicating with AI, all the dialogue will seem unnatural, as it has no real emotions behind it. Through too much of this type of communication, one may expect the same vibe in these interactions and everyday life. They might expect responses adapted for themselves and always the better attitude/option. They will adapt too much to this unnatural communication and lack real human communication. Over-embeddedness in such virtual reality may lead to other mental health challenges.

While Lindemann points out the negatives of the development of death bots, Hutson realizes its potential for our future and its significant perks. These skills that the death bots have 'include participating in dynamic dialogues, articulating diverse emotional states, and adapting to real-world occurrences'.[2]

He goes on to expand on the application of AI and robotics for the imitation or resurrection of dead people, a new turn in the way humans negotiate memory and loss. At the core of such studies come "Deathbots," which are digital avatars constructed to emulate the deceased through data that includes voice recordings, photos, and online profiles. Ameca and StoryFile are examples of how technologies like these can create lifelike interactions, muddling life and death. Comfort to the bereaved, death bots raise ethical issues, too: questions of consent, identity theft, and data ownership stretch already-framed notions of privacy and autonomy.

Much of this research paper references digital necromancy, involving the use of AI, data, and robotics to create a lifelike clone of deceased people.

With further improvements in necromancy, ethics will need to ensure that consent for pre- and post-death availability is considered in terms of ensuring people retain control over their bodies and personal data when they pass. It means, first of all, taking into consideration ownership and further probable issues related to ownership commercial or unauthorized re-creation, for example, this highly sensitive data used for creating these digital avatars.

This concept also poses a cultural and philosophical threat to notions of identity, memory, and mortality. The person making and consuming such technologies should consider the greater implications these technologies have in reconceptualizing life and death in virtual domains from an ethical standpoint beyond intent.

---

[2] James Hutson, Jay Ratican, Colleen Biri. 10-12-23. "Essence as Algorithm: Public Perceptions of AI-Powered Avatars of Real People". *Journal of Artificial Intelligence and Robotics*.

The balance required to understand this will come from such interdisciplinary collaboration-psychologists, ethicists, technologists, and jurists-who can together make sure that digital necromancy respects the singularity of persons and maintains the process of grieving without harming the social consequence. Such steps would ensure these novelties serve the purpose of remembering and connecting with humans without losing their essential ethical and emotional bottom line.

"Where Is My Loved One Now?": Rethinking Life after Death in the Age of AI

The general understanding of the definitions of terms such as "life," "death," "the soul," and "afterlife" varies. "Life" is one human being as conscious, breathing, and heart beating. "Death" is the opposite. However, the soul is always conscious. "Life after death" or "afterlife" involves humans as embodied beings, with another dimension beyond the physical bodies (the soul). Many religions and cultures believe that the soul will maintain its consciousness even after the body stops working. An individual's soul is conscious and exists through the life and death of a person. The undying question of "Where is my loved one now?" during bereavement cannot be denied.

Plato and Descartes provide fascinating theories on the idea of life after death. Plato's theory of the human soul mainly revolves around the idea that "All soul is immortal" [3] and that the soul is split into three parts- the rational, spirited, and appetitive dimensions.[4] He also believes that after an individual human being dies, the soul is continuously transported to a new body to begin life again, and the soul continues to exist even though the body has disintegrated. This occurs in a continuous cycle of birth, death, rebirth, etc.

The next philosopher who writes on the existence of the human soul beyond death is Descartes. His theory of dualism encompasses the idea that the human mind and body are interrelated and yet are still separate dimensions.[5] Think of it as an individual soul controlling your body like how code runs machines, and how that same code can be used on several different machines. In the same fashion, the soul can move to be separate from the body but together at the same time. Descartes suggests that there is life after death by emphasizing that the mind and the body are significantly different and separate, so your mind could keep on being conscious and working even after the body perishes.

These ideologies of the soul continuing after the body dies further the immorality of death bots as it will disrespect the passed soul's dignity and value. If the soul does continue as these theories say, the death bots will be viewed as their replacements, leading to less honor for them and less grieving for their accomplishments, etc. Deathbots would transition the human's attention away from the passed soul to the deathbot, disrespecting the actual soul. While this is

---

[3] Plato & Nichols, J. H. (tr. and ed.). Phaedrus. Cornell University Press. (1998).

[4] Plato. 1997. Complete Works. Edited by J.M. Cooper and D.S. Hutchinson. Indianapolis, Ind.: Hackett Pub

[5] "Mind Body Dualism." Wikipedia. Last modified September 13, 2021. https://en.wikipedia.org/wiki/Mind%E2%80%93body_dualism.

the death bots' purpose, it may still lead to negative consequences with the value or appreciation for the lost one.

Apart from philosophers, many scientists have also been questioning the morals of death bots on people after death. Admittedly, our understanding of death and our theories surrounding morals on death bots will vary with the rapidly advancing technology and increased perspectives. Numerous scientific studies have shown evidence of the disadvantages of surrounding ourselves with death bots too much. For example, researchers have found that too much contact with death bots will lead to one relying on them too much. "...it can lead to users over-relying on the bot to regulate their affects such that they constantly need it for emotion regulation (especially for regulating grief). This may lead to a situation in which users cannot regulate their effect without the bot"(Lindemann 4). This furthers the idea that the death bots are unethical because of the negative impacts on the humans using them. Overreliance will decrease one's capacity to grieve by oneself and only find comfort in the deathbot, which will inevitably lead to an unhealthy relationship. Another idea this study brought up was how easily the deathbot could manipulate the subject. This overreliance or trust would mean that the human would listen to anything that the deathbot says, which could have drastic consequences. In the worst case scenario, the human would be controlled by the deathbot as they have so much trust in them.

Religion plays a significant role in assessing life after death, especially for people who are culturally invested in a certain religion. This serves as its purpose for me as well. Religion paints life after death as a fundamental part of existence by engraving the belief that there is life after death for people who follow that religion. It uncovers the mystery around the afterlife and connects with people with the same inquiries. More specifically, three big religions have strong beliefs around life after death: Christianity, Islam, and Buddhism.

First, Christianity encompasses the tenacious belief that there is life after death, with God (Jesus) judging the individual souls that enter the afterlife to go to heaven, or hell. According to the Bible, on the day of judgment, Jesus personally judges souls and places them in either heaven or hell. Heaven is a place of eternal life where it's completely peaceful, while hell is a place of eternal punishment with suffering and torture involved.[6] Christianity also revolves around the idea that the soul is immortal and the body is mortal. So your soul lives on forever while your bodily existence can be replaced and done over.

Secondly, the religion of Islam involves the idea that the soul waits until judgment day when Allah questions each individual. On the one hand, if that individual correctly answers both questions, then they can avoid torture and live peacefully eternally. On the other hand, if they incorrectly answer either question, then they are subjected to punishment and torture.[10][7] The Quran 3.102 says, "Every soul shall taste death, and only on the Day of Judgment will you be

---

[6] "Death and the Afterlife: What Does Christianity Say about Life after Death?"BBC. https://www.bbc.co.uk/bitesize/guides/zdxdqhv/revision/5.

[7] "Key Beliefs in Islam." BBC. https://www.bbc.co.uk/bitesize/guides/zdxdqhv/revision/5.

paid your full recompense." This extends their religion in the afterlife to not only punishment or peace but also reincarnation.

Lastly, committed Buddhists believe that consciousness continues after death, with opportunities for reincarnation available for souls in the afterlife.[11][8] Not only do Buddhists teach their students about this, but they also teach them not to fear death. This implies that death is not the end and that the soul will live on forever/be reincarnated. These help the idea of life after death as each religion guides its followers with a sense of awareness.

Everyone has good deeds and bad deeds that earn them their place in the afterlife with a possibility of reincarnation (stated above). However, death bots completely erase their past accomplishments and honorings by standing in their place when they were alive. All the preparation for that person for the afterlife will be all for nothing as they won't be honored properly. This death bot has taken their place and makes the atmosphere worse as opposed to better. This constant reminder that the loved one has passed will make things significantly worse instead of alleviating it. And there will come a day when the owner realizes the loved one is no longer with them. This death bot is just pushing that day back.

Potential Risks of Over-reliance: The Concentration of AI Power

To further prove the futility of bereavement AI support, let's imagine a world where everybody has a death bot and constantly relies on it. Suppose every member of society believes that the deathbot will be able to help them 'overcome' their grief. This means that AI is given power over people who feel vulnerable after losing loved ones. These vulnerabilities may be exploited by tech companies. As a result, everyone's dependency on the deathbot will lead to a deathbot having some kind of central power.

There are many painful experiences in this world, and bereavement is no doubt one of them. When designers of new technology want to help people reduce or smooth these inevitable pains in life, they should be aware of some unintended consequences. Take the prevalence and overuse of pain medication in the US for example. Painkillers, especially opioids, have extremely devastating side effects including sedation, dizziness, nausea, vomiting, constipation, physical dependence, tolerance, and respiratory depression.[9] Some even say death is one of the consequences of overdosing on painkillers. However, even a small dose could lead to these negative effects. The overuse of these painkillers has turned into a global crisis with its highly addictive substances and overdosing. Most people who take such medications later turn to more powerful drugs like heroin or fentanyl to seek higher levels of euphoria. For example, in 2020

[8] "Death and the Afterlife: What Does Buddhism Teach about Life after Death?" BBC.https://www.bbc.co.uk/bitesize/guides/zyhmk2p/revision/1#:~:text=Buddhism%20teaches%20that%20there%20is,nirvana%2C%20an%20end%20to%20suffering.

[9] Ramsin Benyamin 1, Andrea M Trescot, Sukdeb Datta, Ricardo Buenaventura, Rajive Adlaka, Nalini Sehgal, Scott E Glaser, Ricardo Vallejo, 03-11-2008, "Opioid complications and side effects", PubMed, https://pubmed.ncbi.nlm.nih.gov/18443635/

alone, nearly 75% of the record 91,799 drug overdose deaths in the U.S. involved opioids.[10] This significant statistic shows just the extent of how impactful opioids are in the US. The annual number of opioid overdose deaths is projected to reach nearly 82 000 by 2025.[11] However, the consequences outweigh the positives as such predicaments are detrimental to one's mental health.

Overall, there are little to no advantages to having a death bot other than pushing the day that one truly overcomes the grieving stage back. The disadvantages far outweigh the advantages, from moral obstacles to well-being.

After examining all the philosophical, scientific, religious, and empirical evidence through a period of research, the conclusion was made that AI-assisted bereavement support is unethical for the deceased individual and the family in question.

[10] Participating Jurisdiction, "Drug Overdose Deaths: Facts and Figures", National Institute on Drug Abuse, https://nida.nih.gov/research-topics/trends-statistics/overdose-death-rates

[11] Paul Nelson, M.D., M.S., 02-01-2019, "Prescription Opioid Misuse and Projected Overdose Deaths in the United States", No Publication, https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2723405

Works Cited

1: Nora Lindemann. 11-22-2022. “The Ethics of ‘DeathBots.” Science and Engineering Ethics.

2: James Hutson, Jay Ratican, Colleen Biri. 10-12-23. “Essence as Algorithm: Public Perceptions of AI-Powered Avatars of Real People”. Journal of Artificial Intelligence and Robotics.

3: Whitten, Sarah. "'Soul' Is a 'Gorgeous Muddle' Filled with Artistry, Charm and a Tangled Plot." CNBC. Last modified December 22, 2020. https://www.cnbc.com/2020/12/22/soul-reviews-what-critics-are-saying-about-pixars-newest-film.html.

4: Plato & Nichols, J. H. (tr. and ed.). Phaedrus. Cornell University Press. (1998).

5: Plato. 1997. Complete Works. Edited by J.M. Cooper and D.S. Hutchinson. Indianapolis, Ind.: Hackett Pub

6: "Mind Body Dualism." Wikipedia. Last modified September 13, 2021. https://en.wikipedia.org/wiki/Mind%E2%80%93body_dualism.

7: "Death and the Afterlife: What Does Christianity Say about Life after Death?"BBC. https://www.bbc.co.uk/bitesize/guides/zdxdqhv/revision/5.

8: "Key Beliefs in Islam." BBC. https://www.bbc.co.uk/bitesize/guides/zdxdqhv/revision/5.

9: "Death and the Afterlife: What Does Buddhism Teach about Life after Death?" BBC.https://www.bbc.co.uk/bitesize/guides/zyhmk2p/revision/1#:~:text=Buddhism%20teaches%20that%20there%20is,nirvana%2C%20an%20end%20to%20suffering.

10: Ramsin Benyamin 1, Andrea M Trescot, Sukdeb Datta, Ricardo Buenaventura, Rajive Adlaka, Nalini Sehgal, Scott E Glaser, Ricardo Vallejo, 03-11-2008, "Opioid complications and side effects", PubMed, https://pubmed.ncbi.nlm.nih.gov/18443635/

11: Participating Jurisdiction, "Drug Overdose Deaths: Facts and Figures", National Institute on Drug Abuse, https://nida.nih.gov/research-topics/trends-statistics/overdose-death-rates

12: Paul Nelson, M.D., M.S., 02-01-2019, "Prescription Opioid Misuse and Projected Overdose Deaths in the United States", No Publication, https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2723405

The Determinant of Foreign Direct Investment: An Empirical Evidence from OECD Countries By Parth Sharma

Abstract

This study examines the determinants of foreign direct investment (FDI) in OECD nations by evaluating the impact of key macroeconomic variables. The study employs scatter plot diagrams, descriptive statistics, and regression analysis. The study identifies the positive relationship between FDI and GDP per capita growth, political stability, logistics, and rule of law and the negative relationship between FDI and inflation, thereby emphasizing macroeconomic stability. These insights present important implications for policymakers aiming to enhance FDI inflow by fostering favorable economic conditions and institutional frameworks.

Keywords Foreign Direct Investment, OECD, Scatter plot, regression

Introduction

Through the introduction of capital, technology, and knowledge, foreign direct investment (FDI) makes a substantial contribution to economic growth and development. Foreign direct investment is a crucial source of funding for OECD nations, as it is necessary to promote innovation and increase global competitiveness. Understanding the factors that influence foreign direct investment (FDI) is essential because it helps policymakers create an environment that encourages economic expansion and FDI.

This study employs regression analysis, descriptive statistics, and scatter plot diagrams to analyze the determinants of foreign direct investment in OECD countries. The empirical evidence highlights that political stability, logistics, GDP per capita growth, inflation, and the rule of law are statistically significant determinants of FDI. The study of these variables gives us useful information about how different factors affect foreign direct investment. It also gives policymakers suggestions on how to make their economy more attractive to investors.

The structure of this paper is as follows: section 2 reveals the relevant literature on FDI determinants. Section 3 describes the data and methodology used in the analysis. Section 4 represents the empirical findings. Finally, section 5 concludes with policy implications and suggestions for future research.

Literature Review

Foreign direct investment (FDI) significantly contributes to the economic development of host countries through cash inflows, knowledge transfer, and job creation. Over the years, a large number of studies have looked at the factors that influence foreign direct investment. This literature review presents the principal findings of many studies to provide a thorough grasp of the determinants influencing FDI. Faeth (2008) indicated that ownership advantages are a major factor of foreign direct investment (FDI). This paper indicates that horizontal FDI is motivated by consumer proximity, whereas vertical FDI is affected by variations in factor endowments.

Biswas (2002) examined both conventional and unconventional determinants affecting FDI through an analysis of 44 countries. Conventional factors such as infrastructure, wage levels, and market size strongly influence foreign direct investment (FDI), whereas non-conventional elements such as regime kinds, property rights, and governance equity are equally crucial. Schneider and Frey (1985) investigated foreign direct investment in less developed nations by evaluating economic and political factors. This analysis identified the primary determinant variables as real GNP per capita, GNP growth rate, inflation, wage costs, and the availability of skilled labor. Tatoglu and Glaister (1998) discovered that firm-specific advantages significantly influence foreign direct investment (FDI). The ownership structure of foreign direct investment varies according to transaction-specific costs. Blonigen and Piger (2011) conducted a global analysis on the determinants of FDI, identifying a comprehensive collection of 56 variables. This analysis indicates that regional trade agreements, GDP, distance, and relative labor endowments significantly influence FDI. Nevertheless, the attributes of the host country exerted a diminished influence on foreign direct investment (FDI). A study by Kumari and Sharma (2017) analyzed FDI inflows in 18 developing nations, revealing that FDI is significantly affected by market size, trade openness, and human capital. This analysis identified infrastructure and inflation as critical factors of FDI. Aqeel and Nishat (2004) conducted a study in Pakistan and found that reducing import tariffs and corporation tax rates favorably impacted FDI growth. Moreover, the appreciation of exchange rates was seen to stimulate foreign direct investment (FDI). This analysis highlighted that policy enhancements substantially facilitated the rise of FDI inflows. Kolstad and Villanger (2008) examined the influence of political and institutional determinants on foreign direct investment (FDI) in 57 countries, emphasizing the significance of democracy, institutional quality, and political risk. Their findings indicated that political and institutional variables greatly affect foreign direct investment (FDI), while also noting that FDI is mostly driven by local market size rather than trade openness. Mottaleb (2007) emphasizes the significance of modern infrastructure, particularly internet and telephone services, in attracting foreign direct investment (FDI). This study concludes that nations with a more business-friendly environment and reduced corruption levels are more likely to successfully attract foreign direct investment (FDI). Islam and Beloucif (2023) conducted a systematic literature analysis across 150 countries, analyzing the economic, political, institutional, and social aspects that affect foreign direct investment (FDI). This analysis concludes that market size is the primary driver, whereas trade openness, infrastructure quality, labor costs, and microeconomic stability also affect FDI. The authors observe that the influence of these elements differs according to the country's level of development.

Tocar (2018) analyzes several institutional, technological, infrastructural, and economic indicators influencing global foreign direct investment. This study indicates that infrastructure, cheap labor costs, and market size are significant determinants of foreign direct investment (FDI). In their 2019 study, Mahbub and Jongwanich examine the determinants of foreign direct investment (FDI) in Bangladesh, revealing that regulatory considerations are the most influential element. Furthermore, proficient labor, gas infrastructure, and economic growth significantly

contribute to attracting foreign direct investment (FDI). Asongu, Akpan, and Isihak (2018) examined foreign direct investment (FDI) in the BRICS and MINT nations, identifying trade liberalization, market size, and infrastructural accessibility as significant FDI determinants. This analysis indicates that foreign direct investment (FDI) is minimally influenced by institutional quality or natural resources. Dees (1990) examines foreign direct investment (FDI) in China during the 1990s and concludes that trade relations and market size positively influence FDI. Moreover, foreign direct investment in China significantly influences economic growth. Agiomirgianakis, Asteriou, and Papathoma (2003) analyzed foreign direct investment in OECD countries, emphasizing variables such as GDP per capita, GDP growth, trade openness, human capital, government spending, and infrastructure. This study determined that foreign direct investment in OECD nations is significantly affected by human capital, trade openness, infrastructure development, and GDP growth. Çevis and Çamurdan (2007) examined 17 emerging economies, revealing that FDI inflows are positively correlated with GDP growth, trade openness, and real interest rates, but negatively impacted by inflation. Singhania and Gupta (2011) demonstrated that inflation and research and development significantly affect foreign direct investment (FDI). Mottaleb and Kalirajan concluded that trade openness, GDP growth rates, and elevated GDP are significant determinants of FDI attraction after evaluating inflows in 68 countries. Yang, Groenewold, and Tcha (2000) examined foreign direct investment in Australia by analyzing parameters such as interest rates, GDP, trade openness, wages, industrial conflicts, and inflation. Their results show that higher interest rates have a positive effect on FDI.

Empirical Evidence

Data

Appendix 1 presents various economics and foreign direct investment variables from the World Bank World Development Indicator data set. The foreign direct investment variables include net current US$, logistics performance index as the ability to track and trace consignments (1=low to 5=high), real effective exchange rate index (2010 = 100), taxes on international trade as % of revenue, rule of law, political stability and absence of violence/terrorism, population, inflation as a consumer price as an annual %, labor force participation rate as % of total population ages 15+, and GDP per capita growth as an annual %.

Descriptive Statistics

Table 1. Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
FDI_net	37	166.8097	436.6027	-256.52	1558.719
Logistic	37	3.749	.401	3	4.3
Exchange Rate	32	95.446	14.9	61.006	128.511
Taxes	19	1.111	1.035	0	3.052
Rule of Law	37	1.113	.69	-.867	1.958

Political Stability	37	.511	.595	-1.285	1.312
Population	37	.3708	.60552	.0038	3.332
Inflation	37	10.844	11.152	2.498	72.309
LFP	37	61.872	5.028	49.1	74.603
GDP_percapita_g	37	3.122	2.243	-1.783	8.292

Table 1 provides descriptive statistics for various economic indicators used in the analysis. FDI_net has 37 observations with a mean of 166.809 and a standard deviation of 436.6027, ranging from a minimum of -256.52 to a maximum of 1558.719.
Logistic has 37 observations with a mean of 3.749 and a standard deviation of .401, ranging from a minimum of 3 to a maximum of 4.3.
Exchange Rate has 32 observations with a mean of 95.446 and a standard deviation of 14.9, ranging from a minimum of 61.006 to a maximum of 128.511.
Taxes have 19 observations with a mean of 1.111 and a standard deviation of 1.035, ranging from a minimum of 0 to a maximum of 3.052.
Rule of Law has 37 observations with a mean of 1.113 and a standard deviation of .69, ranging from a minimum of -.867 to a maximum of 1.958.
Political Stability has 37 observations with a mean of .511 and a standard deviation of .595, ranging from a minimum of -1.285 to a maximum of 1.312.
Population has 37 observations with a mean of .3708 and a standard deviation of .60552, ranging from a minimum of .0038 to a maximum of 3.332.
Inflation has 37 observations with a mean of 10.844 and a standard deviation of 11.152, ranging from a minimum of 2.498 to a maximum of 72.309.
LFP has 37 observations with a mean of 61.872 and a standard deviation of 5.028, ranging from a minimum of 49.1 to a maximum of 74.603.
GDP_percapita_g has 37 observations with a mean of 3.122 and a standard deviation of 2.243, ranging from a minimum of -1.783 to a maximum of 8.292.

Figure 1. Scatter plot graphics of FDI vs. GDPpercapitagrowth

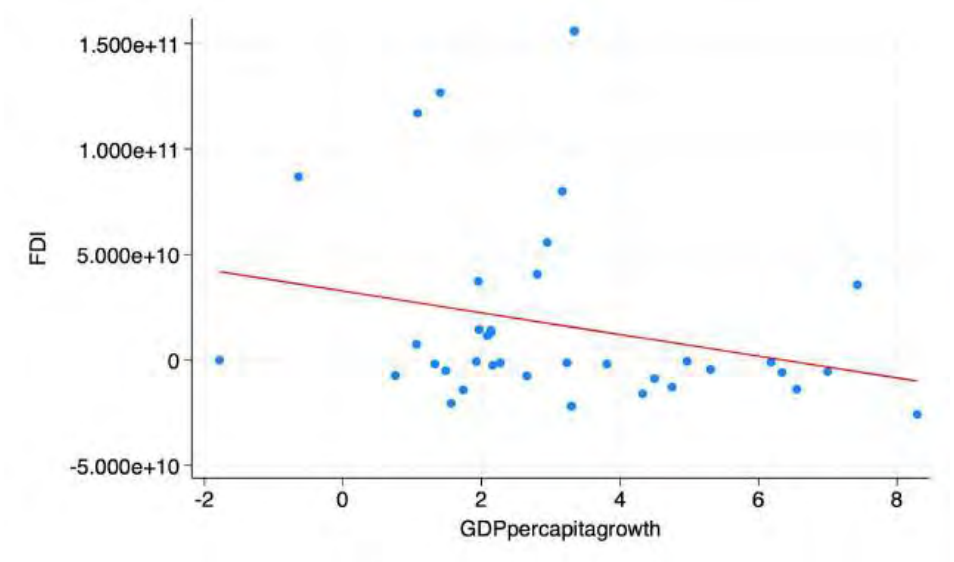


Figure 1 indicates a negative correlation between FDI_net and GDP_percapita, suggesting that a higher level of GDP_percapitagrowth might contribute to a lower level of FDI_net.

Figure 2. Scatter plot graphics of FDI vs. Labor Force Participation Rate

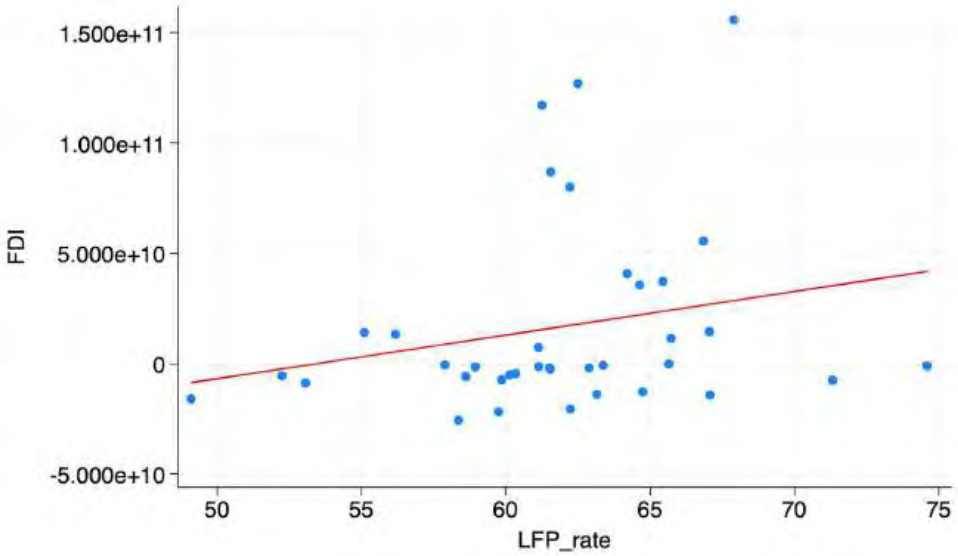


Figure 2 indicates a positive correlation between FDI_net and LFP_rate, suggesting that a higher level of LFP_rate might contribute to a higher level of FDI_net.

Figure 3. Scatter plot graphics of FDI vs. Taxes on International Trade

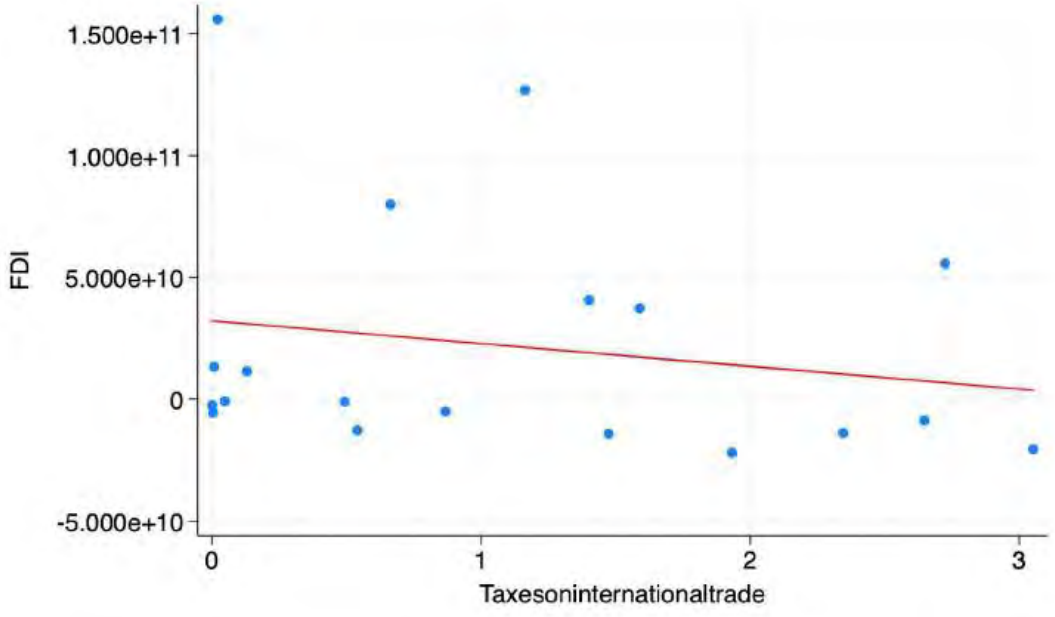


Figure 3 indicates a negative correlation between FDI_net and Taxes on international trade, suggesting that a higher level of Tax on international trade might contribute to a lower level of FDI_net.

Figure 4. Scatter plot graphics of FDI vs. Rule of Law

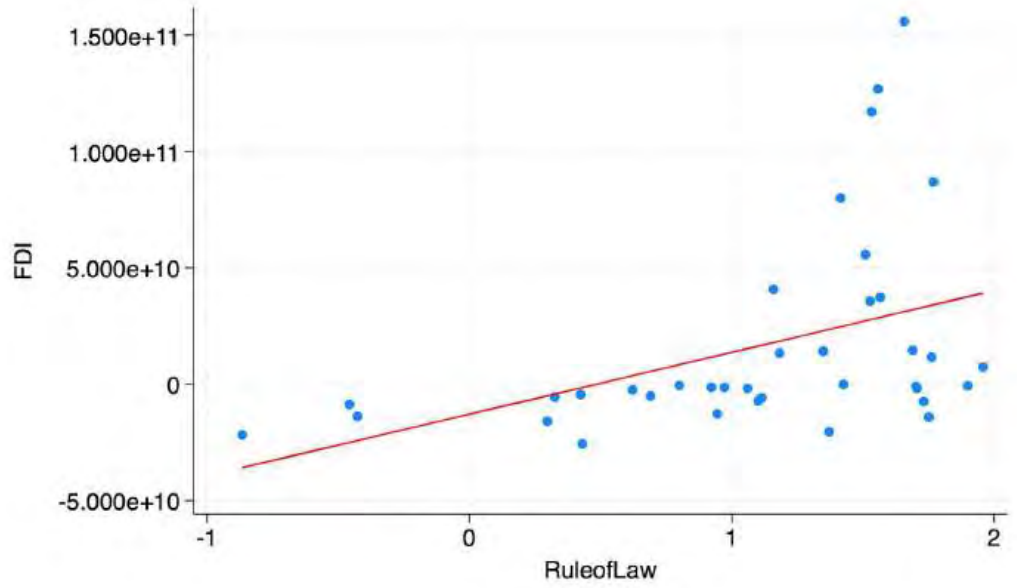


Figure 4 indicates a positive correlation between FDI_net and Rule of Law, suggesting that a higher level of Rule of Law might contribute to a higher level of FDI_net.

Figure 5. Scatter plot graphics of FDI vs. Logistics Performance

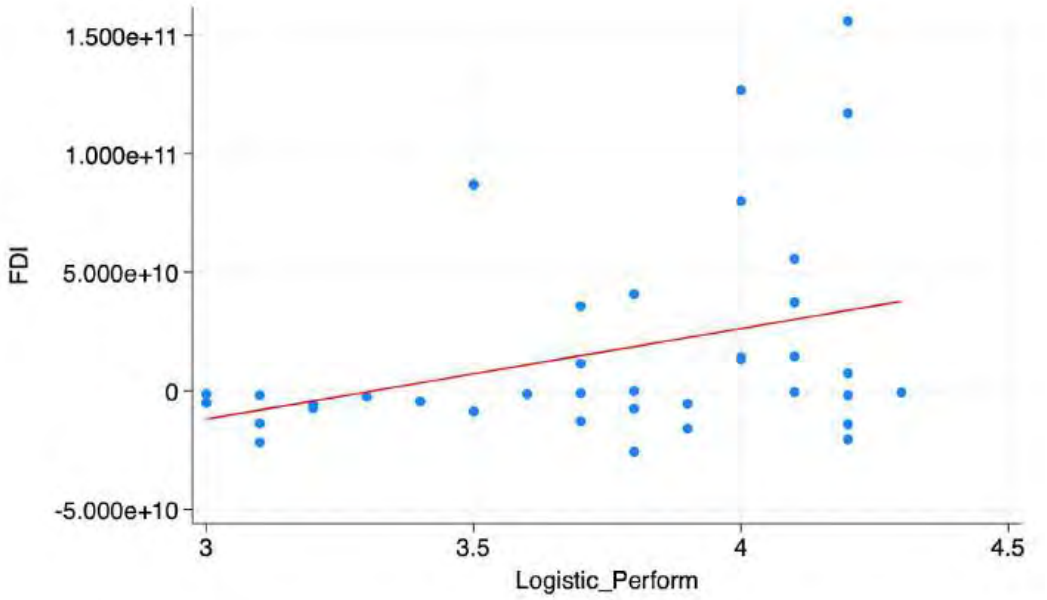


Figure 5 indicates a positive correlation between FDI_net and the Logistic performance index, suggesting that a higher level of logistic performance index might contribute to a higher level of FDI_net.

Figure 6. Scatter plot graphics of FDI vs. Exchange Rate

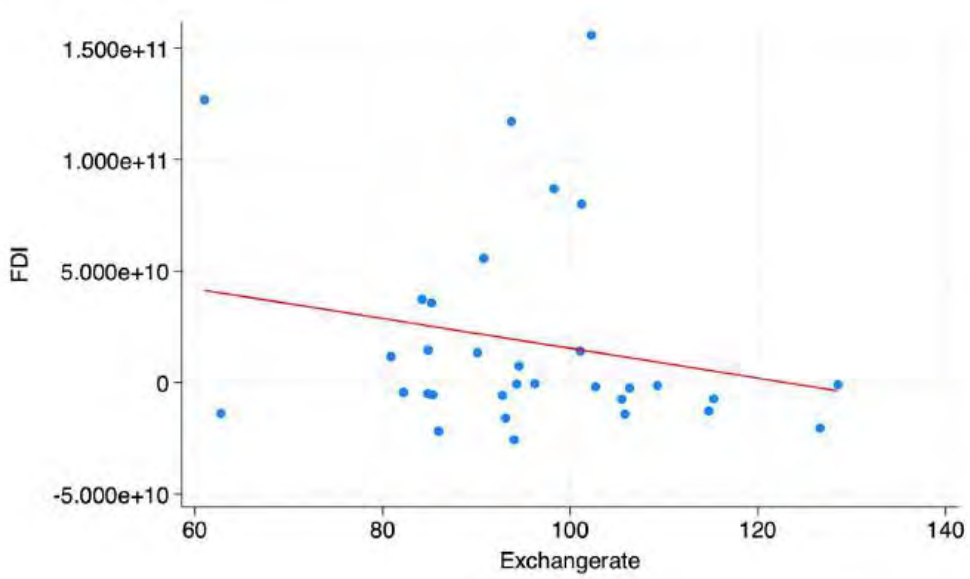


Figure 6 indicates a negative correlation between FDI_net and Exchange rate, suggesting that a higher level of exchange rate might contribute to a lower level of FDI_net.

Figure 7. Scatter plot graphics of FDI vs. Inflation

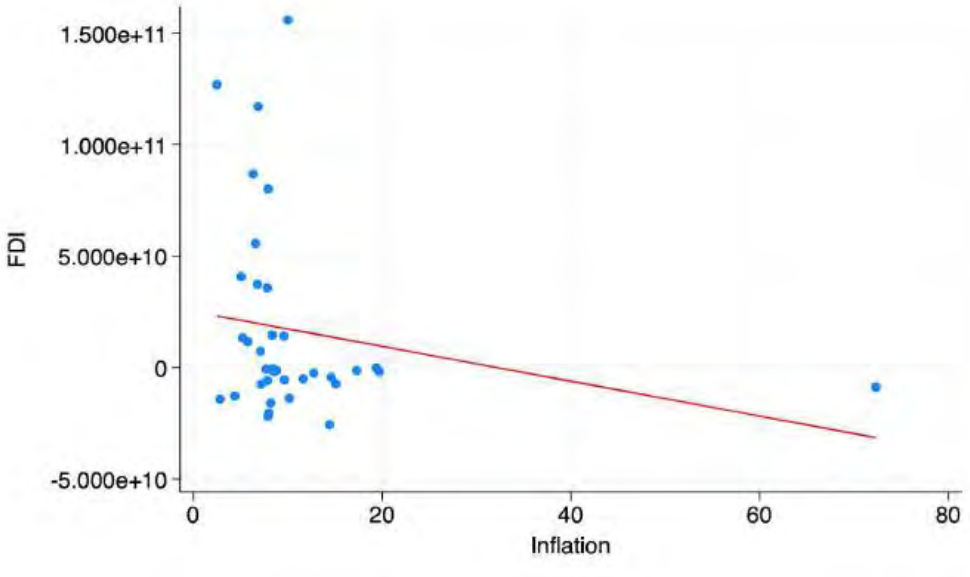


Figure 7 indicates a negative correlation between FDI_net and Inflation, suggesting that a higher level of Inflation might contribute to a lower level of FDI_net.

Model

$$Y_I = \beta_0 + \beta_1 X_{1I} + \beta_2 X_{2I} + \ldots + \beta_n X_{nI} + \varepsilon_I$$

Where Y_I is the dependent variable for observation i which refers to foreign direct investment net. β_0 is a constant term, representing the expected value of the dependent variable when dependent variables are zero. β_1 to β_n are the coefficient for the independent variables, which include GDP_percapita_g, LFP, Rule of Law, Inflation, Political Stability, Population, Taxes, Exchange Rate, and Logistic. ε_I is the error term, which represents the difference between the actual value and the predicted value from the model.

Findings

Table 1a. Regression Analysis

	(1)	(2)	(3)	(4)
VARIABLES	FDI_net	FDI_net	FDI_net	FDI_net
GDP_percapita_g	-51.507*			
	(26.178)			
LFP		19.760		
		(12.439)		
Rule of Law			265.607***	
			(76.380)	
Inflation				-7.822*
				(4.162)
Constant	327.616**	-1,055.797	-128.826**	251.635**
	(128.954)	(733.958)	(49.673)	(98.268)
Observations	37	37	37	37
R-squared	0.070	0.052	0.176	0.040

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

According to the regression analysis in Table 1, FDI_net is negatively and statistically significantly impacted by both inflation and GDP per capita growth. Both of these effects have p-values below 0.10. Conversely, FDI_net is positively and statistically significantly impacted by the rule of law, while Models 1 and 4 show that higher inflation slightly lowers FDI_net. FDI_net also goes down a little when GDP per head grows. Overall, the study shows that the Rule of Law has the biggest and most positive effect on foreign direct investment. On the other hand, GDP per capita growth and inflation have negative effects on FDI_net.

Table 1b. Regression Analysis

	(1)	(2)	(3)	(4)	(5)
VARIABLES	FDI_net	FDI_net	FDI_net	FDI_net	FDI_net
Political Stability	233.073***				
	(78.311)				
Population		29.64			
		(142.95)			
Taxes			-93.251		

			(103.948)		
Exchange Rate				-6.683	
				(5.678)	
Logistic					382.090**
					(159.986)
Constant	47.649	155.814**	322.020	821.746	-1,265.510**
	(51.279)	(73.719)	(194.371)	(566.782)	(551.087)
Observations	37	37	19	32	37
R-squared	0.101	0.002	0.037	0.046	0.123

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1b shows that Political Stability and Logistics are important factors that affect FDI_net, with p values less than 0.10, which means that higher FDI is linked to better political stability and logistics. Population, taxes, and the exchange rate don't have a statistically significant effect on FDI_net, which suggests they can't explain much in this case. This regression analysis highlights the importance of Political Stability and Logistics as key drivers of foreign direct investment.

Conclusion

This study explains the main determinants of foreign direct investment (FDI) in OECD nations, highlighting the significance of important macroeconomic factors in shaping the decisions about foreign investments. The analysis indicates that GDP per capita growth, political stability, logistics, and the rule of law are vital for attracting FDI inflows. Conversely, higher inflation results in reduced FDI inflows.

Policymakers in OECD countries should focus on improving economic stability, enhancing governance frameworks, and fostering an investment-friendly environment to encourage greater FDI inflows.

Future studies may investigate the influence of green investment on foreign direct investment patterns to yield more comprehensive insights.

Appendix

Appendix 1. Definition of Variable

Ticker	Variable	Definition	Source
FDI_net	Foreign direct investment, net (BoP, current US$)/$10^8$	Foreign direct investment are the net inflows of investment to acquire a lasting management interest (10 percent or more of voting stock) in an enterprise operating in an economy other than that of the investor. It is the	WDI: World Bank

		sum of equity capital, reinvestment of earnings, other long-term capital, and short-term capital as shown in the balance of payments.	
Logistic	Logistics performance index: Ability to track and trace consignments (1=low to 5=high)	Data are from the Logistics Performance Index survey conducted by the World Bank in partnership with academic and international institutions and private companies and individuals engaged in international logistics.	WDI: World Bank
Exchange Rate	Real effective exchange rate index (2010 = 100)	Real effective exchange rate is the nominal effective exchange rate (a measure of the value of a currency against a weighted average of several foreign currencies) divided by a price deflator or index of costs.	WDI: World Bank
Taxes	Taxes on international trade (% of revenue)	Taxes on international trade include import duties, export duties, profits of export or import monopolies, exchange profits, and exchange taxes.	WDI: World Bank
Rule of Law	Rule of Law: Estimate [RL.EST]	Rule of law captures perceptions of the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence.	WDI: World Bank
Political Stability	Political Stability and Absence of Violence/Terrorism: Estimate	Political Stability and Absence of Violence/Terrorism measures perceptions of the likelihood of political instability and/or politically-motivated violence, including terrorism. Estimate gives the country's score on the aggregate indicator, in units of a standard normal distribution, i.e. ranging from approximately -2.5 to 2.5.	WDI: World Bank
Population	Population, total/10^8	Total population is based on the de facto definition of population, which counts all residents regardless of legal status or	WDI: World Bank

		citizenship. The values shown are midyear estimates.	
Inflation	Inflation, consumer prices (annual %)	Inflation as measured by the consumer price index reflects the annual percentage change in the cost to the average consumer of acquiring a basket of goods and services that may be fixed or changed at specified intervals, such as yearly. The Laspeyres formula is generally used.	WDI: World Bank
LFP	Labor force participation rate, total (% of total population ages 15+) (modeled ILO estimate)	Labor force participation rate is the proportion of the population ages 15 and older that is economically active: all people who supply labor for the production of goods and services during a specified period.	WDI: World Bank
GDP_percapita_g	GDP per capita growth (annual %)	Annual percentage growth rate of GDP per capita based on constant local currency. GDP per capita is gross domestic product divided by midyear population.	WDI: World Bank

Works Cited

Agiomirgianakis, G. M., Asteriou, D., & Papathoma, K. (2003). The determinants of foreign direct investment: A panel data study for the OECD countries.

Asongu, S., Akpan, U. S., & Isihak, S. R. (2018). Determinants of foreign direct investment in fast-growing economies: evidence from the BRICS and MINT countries. Financial Innovation, 4(1), 1-17.

Biswas, R. (2002). Determinants of foreign direct investment. Review of development economics, 6(3), 492-504.

Blonigen, B. A., & Piger, J. (2011). Determinants of foreign direct investment (No. w16704). Foreign Direct Investment. Massachusetts Avenue.

Blonigen, B. A., & Piger, J. (2014). Determinants of foreign direct investment. Canadian Journal of Economics/Revue canadienne d'économique, 47(3), 775-812.

Çevis, İ., & Camurdan, B. (2007). The economic determinants of foreign direct investment in developing countries and transition economies. The Pakistan Development Review, 285-299.

Dees, S. (1998). Foreign direct investment in China: determinants and effects. Economics of planning, 31, 175-194.7

Faeth, I. (2009). Determinants of foreign direct investment–a tale of nine theoretical models. Journal of Economic surveys, 23(1), 165-196.

Ismail, N. W. (2009). The determinant of foreign direct investment in ASEAN: a semi-gravity approach. Transition Studies Review, 16, 710-722.

Kolstad, I., & Villanger, E. (2008). Determinants of foreign direct investment in services. European Journal of Political Economy, 24(2), 518-533.Mahbub, T., & Jongwanich, J. (2019).

Kumari, R., & Sharma, A. K. (2017). Determinants of foreign direct investment in developing countries: a panel data study. International Journal of Emerging Markets, 12(4), 658-682.

Mottaleb, K. A. (2007). Determinants of foreign direct investment and its impact on economic growth in developing countries.

Mottaleb, K. A., & Kalirajan, K. (2010). Determinants of foreign direct investment in developing countries: A comparative analysis. Margin: The Journal of Applied Economic Research, 4(4), 369-404.

Schneider, F., & Frey, B. S. (1985). Economic and political determinants of foreign direct investment. World development, 13(2), 161-175.

Singhania, M., & Gupta, A. (2011). Determinants of foreign direct investment in India. Journal of international trade law and policy, 10(1), 64-82.

Tatoglu, E., & Glaister, K. W. (1998). Determinants of foreign direct investment in Turkey. Thunderbird International Business Review, 40(3), 279-314.

Tocar, S. (2018). Determinants of foreign direct investment: A review. Review of Economic and Business Studies (REBS), (21), 165-196.

Yang, J. Y. Y., Groenewold, N., & Tcha, M. (2000). The determinants of foreign direct investment in Australia. *Economic Record*, *76*(232), 45-54.

https://databank.worldbank.org/

Syntactic Shifts in Contemporary American English: A Corpus-Based Study
By Arjun Raj

Abstract

This study analyzes syntactic shifts in contemporary American English using corpus-based data from the Brown, Frown, COCA, and NOW corpora. Key findings include the rise of the progressive aspect, the decline of traditional modals in favor of semi-modals, the increasing use of compact noun phrases, shifts in passive constructions, and simplified relative clauses. These trends reflect broader linguistic processes like colloquialization and grammaticalization, driven by digital communication and evolving cultural norms. The findings demonstrate how American English is becoming more streamlined and flexible, accelerated by the online platforms of modern communication.

Introduction

American English has evolved significantly over the past several decades. This paper outlines the grammatical changes that have reshaped how the English language is used in both its written and spoken forms.

Several factors have driven these developments: globalization, the increasing usage of digital communications, and shifting cultural norms toward informality and efficiency. Consequently, these developments are driving broader linguistic trends of colloquialization, grammaticalization, and simplification, which will be explained in further detail below.

By examining these trends over time through corpus-based analyses, we can better understand how American English has shifted from the more formal and structured forms prevalent in the mid-20th century to the more streamlined, flexible, and informal language used today.

In particular, this paper will explore key areas of syntactic change in American English, focusing on verb phrases, noun phrases, passive constructions, relativization, and non-finite clauses. By using data from the historical Brown and Frown corpora (1961–1992) as well as my proprietary analysis of contemporary corpora such as COCA and NOW, this paper aims to provide a comprehensive explanation of how these syntactic structures have shifted over time.

These broader linguistic and societal trends prioritizing brevity, directness, and flexibility are seen in this paper's five key findings: (i) the rise of the progressive aspect; (ii) the decline of core modals in favor of semi-modals; (iii) the increasing prevalence of compact noun phrases; (iv) the shift toward more direct passive constructions; and (v) the simplification of relative clauses.

These changes are not only reshaping formal written English but also becoming increasingly common in casual conversation and digital communications. This in turn accelerates shifts in the linguistic norms that govern American English, as spaces such as social media have greatly increased opportunities for informal communications both in America and globally.

Verb Phrase Changes

The most significant change in verb phrase usage in contemporary American English is the rise of the progressive aspect. Traditionally, the progressive aspect (e.g., "is running") was used to indicate ongoing actions, typically in the present tense. However, corpus data indicates that the progressive is now being used in a wider variety of contexts, including past, future, and even hypothetical situations. For instance, the combination of modal auxiliaries with the progressive (e.g., "should be going," "might be thinking") has become increasingly common in spoken English. This represents a major shift from earlier forms of English, where such combinations were less frequent and more restricted to formal registers.[12]

The Frown corpus, which provides data from the early 1990s, shows a marked increase in the use of the progressive aspect compared to earlier data from the Brown corpus (1961). The frequency of progressive forms in American English rose by approximately 30% between 1961 and 1992, particularly in informal genres such as fiction and journalism.[13] This trend has continued since, as usage has increased by 20% since 1992 (fig. 1). The rise of the progressive aspect can be attributed to several factors, including the increasing influence of spoken language on written English, as well as the growing need for nuance and flexibility in communication. The progressive allows speakers and writers to convey temporariness, politeness, as well as emotional involvement, making it a valuable tool for expressing a wide range of meanings.[14]

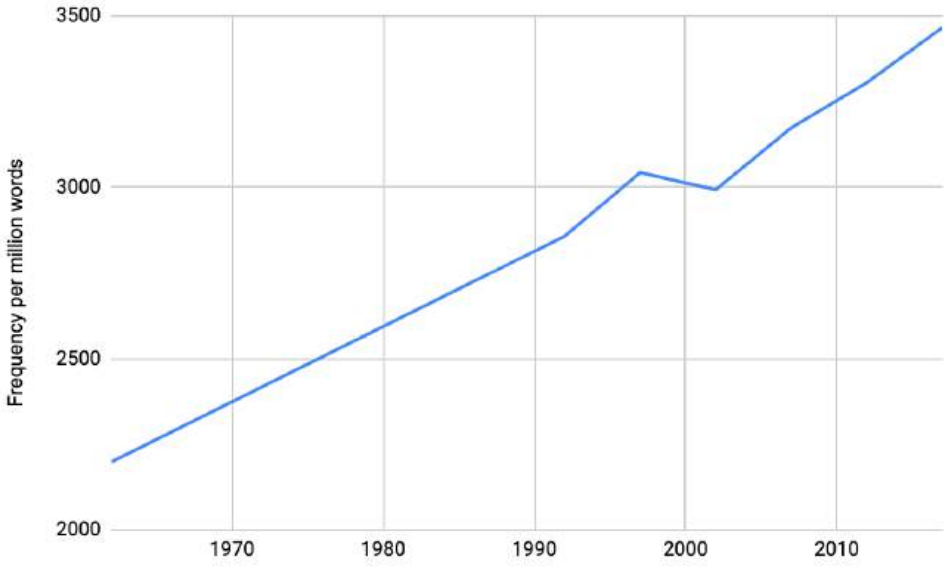


fig. 1

[12] Bas Aarts, Joanne Close, and Sean Wallis, *Recent Changes in the Use of the Progressive Construction in English*, accessed November 29, 2024, https://www.ucl.ac.uk/english-usage/projects/verb-phrase/RecentChangesProgressive.pdf.
[13] Geoffrey N. Leech, *Change in Contemporary English : a Grammatical Study* (Cambridge, UK: Cambridge University Press, 2009).
[14] Aarts, Close, and Wallis, *Recent Changes.*

In addition to its increased use, the progressive aspect is now being used in contexts that deviate from its traditional function of indicating ongoing action. For example, it is increasingly common to see the progressive used with stative verbs (verbs that indicate a state of being or perception – e.g., "I am loving this"), which traditionally do not take the progressive aspect. This expansion of the progressive's function reflects a broader trend in American English toward greater syntactic flexibility, as speakers and writers adapt existing grammatical structures to meet new communicative needs. In particular, the use of the progressive in metacommunicative contexts (when a secondary, implied meaning differs from the literal words being said – e.g., "He is always talking about work") has become more common, where it serves to express annoyance or emotional involvement rather than simply ongoing action.[15]

Another major change in the verb phrase category is the decline of traditional modal auxiliaries such as "shall," "must," and "may." These core modals, which were once common in formal and written English, have seen a sharp decline in frequency over the past few decades. The Brown and Frown corpora show that the frequency of "shall" dropped by nearly 70% between 1961 and 1992, while modals like "ought" and "may" declined by approximately 50%.[16] This trend has continued into the 21st century, as common English modal usage has declined by approximately 20% in the last three decades (fig. 2). This decline is particularly evident in spoken English, where modal verbs are often replaced by alternative structures such as semi-modals (e.g., "have to," "going to," "need to") to convey necessity, intention, or permission.

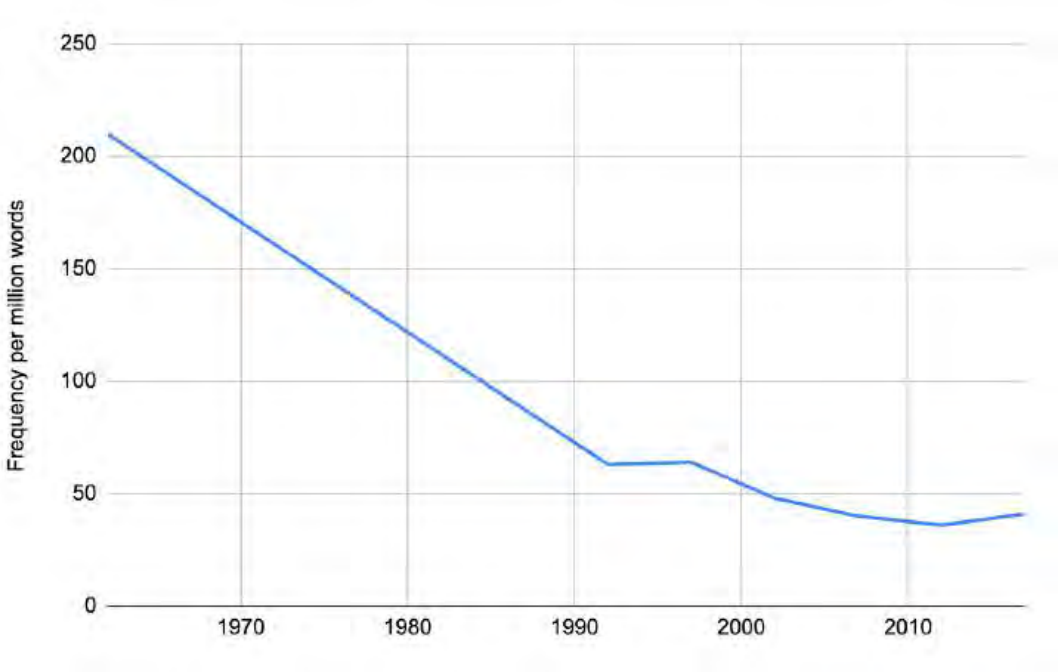


fig. 2

[15] Leech, *Change in Contemporary*.
[16] Leech, *Change in Contemporary.*

Semi-modals have become increasingly dominant in contemporary American English, particularly in informal registers. Forms like "gonna" and "wanna," which are reduced forms of "going to" and "want to," have become especially common in spoken language. These semi-modals are often used to express future actions or obligations in a more colloquial and relaxed manner than their core modal counterparts. The rise of semi-modals can be linked to the process of grammaticalization, in which conversational trends gradually take on grammatical functions.[17] In this case, semi-modals like "gonna" and "wanna" are increasingly being used as markers of future tense or intention, in much the same way that "will" and "shall" were used in earlier forms of English.[18]

The shift from core modals to semi-modals reflects broader sociolinguistic trends, including the move toward greater informality in American English. As digital spaces have created a uniquely large environment for casual and efficient communication, the language has adapted more quickly to reflect these changes. Semi-modals offer a more flexible way of expressing modality, which aligns with the broader trend toward syntactic simplification and efficiency in modern English.[19]

Noun Phrase Changes

Noun phrases have also undergone significant changes in contemporary American English, particularly in terms of their structure and complexity. One of the most notable trends is the increasing use of noun + noun sequences (e.g., "campaign manager," "weather report"). These more efficient constructions have become increasingly common across a wide range of genres, including journalism, academic writing, and scientific discourse. The Frown corpus shows a 20-30% increase in the frequency of noun + noun sequences between 1961 and 1992,[20] and a nearly 50% increase since then, according to my analysis of data from COCA (fig. 3).

[17] Christian Mair and Geoffrey Leech, "Current Changes in English Syntax," in *The Handbook of English Linguistics*, by Bas Aarts and April M. S. McMahon (Malden, MA: Blackwell Pub., 2006).
[18] Manfred G. Krug, *Emerging English Modals : a Corpus-based Study of Grammaticalization* (Berlin, Germany: Mouton de Gruyter, 2000).
[19] Krug, *Emerging English*.
[20] Leech, *Change in Contemporary*.

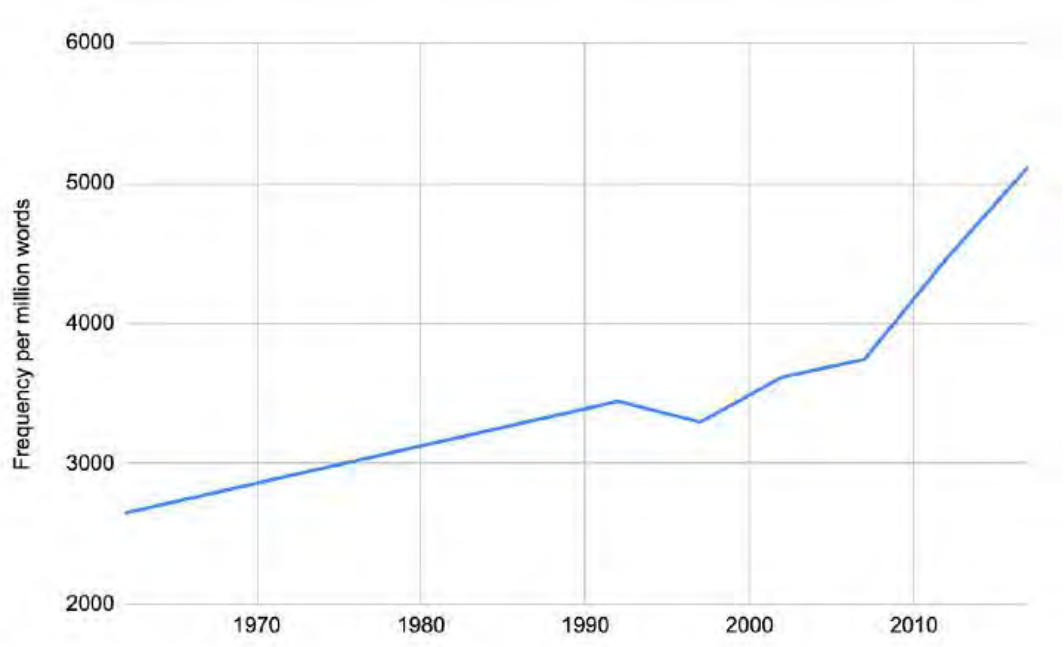


fig. 3

The rise of noun + noun sequences can be seen as part of a broader trend toward the densification of language. In academic and journalistic writing, where clarity and brevity are highly valued, these compact noun phrases allow writers to convey complex information in a more efficient manner. For example, a phrase like "campaign manager" conveys a specific role in just two words, whereas a more traditional construction might require additional prepositional phrases or modifiers (e.g., "the manager of the campaign"). This trend toward linguistic economy is not only limited to formal writing but is also becoming more common in spoken English, where speakers often seek to convey information as quickly and efficiently as possible.[21]

In addition to the rise of noun + noun phrases, there has also been a shift in the use of genitive constructions, or structures used to express possessive relationships, in American English. The s-genitive (e.g., "the book's cover") has become more prevalent, particularly in informal writing such as fiction and journalism, while the of-genitive (e.g., "the cover of the book") is more common in formal contexts, with the Frown corpus indicating a 15% rise in the use of s-genitives between 1961 and 1992.[22] This trend has continued into the 21st century, with COCA data indicating a 12.5% increase in the use of s-genitives since 1995 (fig. 4). This shift toward the s-genitive reflects the broader trend toward syntactic simplicity and informality in modern American English, as the s-genitive is typically shorter and more direct than the of-genitive.

[21] Martin Hilpert, *Constructional Change in English : Developments in Allomorphy, Word Formation, and Syntax* (Cambridge, UK: Cambridge University Press, 2013).
[22] Leech, *Change in Contemporary*.

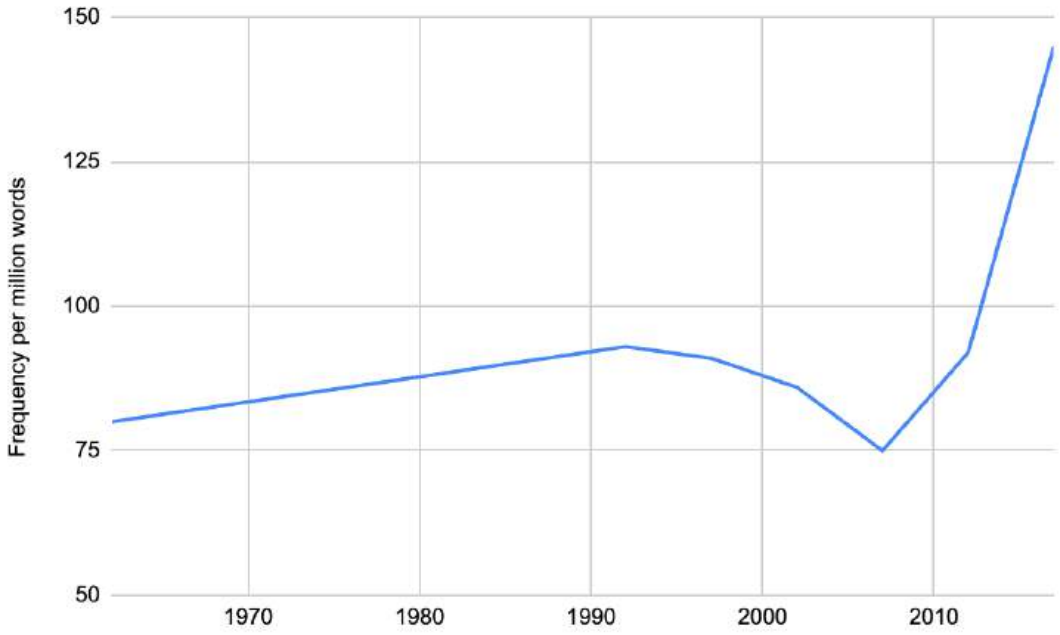


fig. 4

The increasing use of plural attributive nouns (e.g., "workers' rights," "students' union") is another notable change in the structure of noun phrases. These constructions, where a plural noun is used in an attributive position to modify another noun, have become more common in American English, particularly in contexts where relationships between entities need to be expressed concisely. The rise of plural attributive nouns suggests an ongoing trend of expressing complex relationships between entities in a more direct and efficient manner, further contributing to the overall densification of language in modern English.[23]

Passive Voice Constructions

The passive voice, particularly the be-passive (e.g., "The ball was kicked"), has traditionally been a staple of formal written English, allowing writers to focus on the action or the recipient of the action rather than the agent. However, corpus data shows a significant decline in the use of the be-passive in American English, particularly in informal registers such as journalism, fiction, and conversation. The Brown and Frown corpora reveal that the frequency of be-passive constructions dropped by approximately 30% between 1961 and 1992 (fig. 5).[24]

[23] Mair and Leech, "Current Changes".
[24] Leech, *Change in Contemporary*.

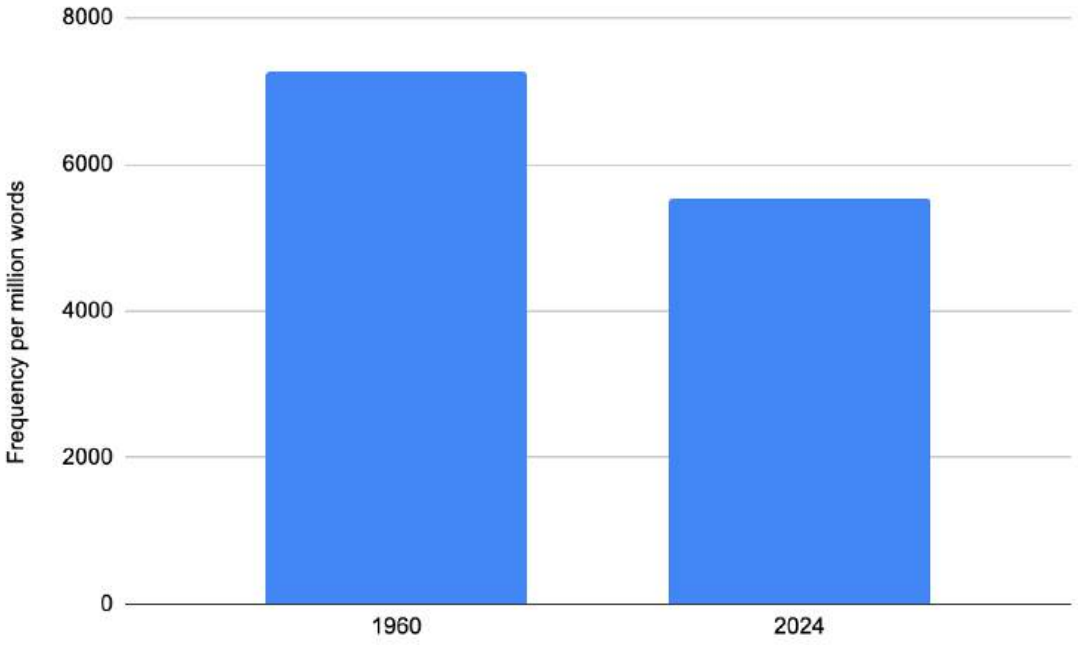


fig. 5

The decline of the be-passive can be attributed to several factors, including the growing preference for more active, direct forms of expression in American English. In genres like journalism and fiction, where clarity and immediacy are highly valued because they engage readers quickly and convey information efficiently, the passive voice is often seen as overly formal or cumbersome. Instead, writers are increasingly opting for active constructions, which allow them to convey information more directly and engage readers more effectively.[25] This shift toward the active voice is part of a broader move toward syntactic simplification in American English, as speakers and writers seek to streamline their language for greater efficiency.

At the same time, the get-passive (e.g., "He got fired") has seen a notable increase in frequency, particularly in informal spoken English. The get-passive, which is more dynamic and conversational than the be-passive, is often used to convey events that happen by chance or through external involvement. For example, a sentence like "He got hit by a car" implies an element of chance or misfortune, which is not as strongly conveyed in the be-passive equivalent, "He was hit by a car." The rise of the get-passive reflects the broader trend toward conversationalization and informalization in American English, as speakers increasingly favor more dynamic, agentless constructions in casual conversation.[26]

The mediopassive construction (a structure where the subject does not initiate the action described – e.g., "This book reads easily") offers another alternative to the traditional passive voice, particularly in specialized genres like advertising, instructional materials, and commercial writing. The mediopassive, which blurs the line between active and passive constructions, allows

[25] Christian Mair, *Twentieth-century English : History, Variation, and Standardization* (Cambridge, UK: Cambridge University Press, 2006)
[26] Mair, *Twentieth-century English.*

writers to highlight the subject's role in an action without explicitly mentioning the agent. This form has become more frequent in American English, particularly in contexts where efficiency and brevity are paramount. For example, in advertising copy, a phrase like "This product sells well" is more concise and engaging than the more traditional passive construction, "This product is sold well."

Relativization

Relative clauses, which allow speakers and writers to provide additional information about a noun, have also undergone significant changes in contemporary American English. One of the most notable shifts is the decline in the use of wh-relatives (e.g., "which," "who," "whom") in informal registers such as spoken English and journalism. The Brown and Frown corpora show a 10-20% decrease in the use of wh-relatives between 1961 and 1992 (fig. 6).[27]

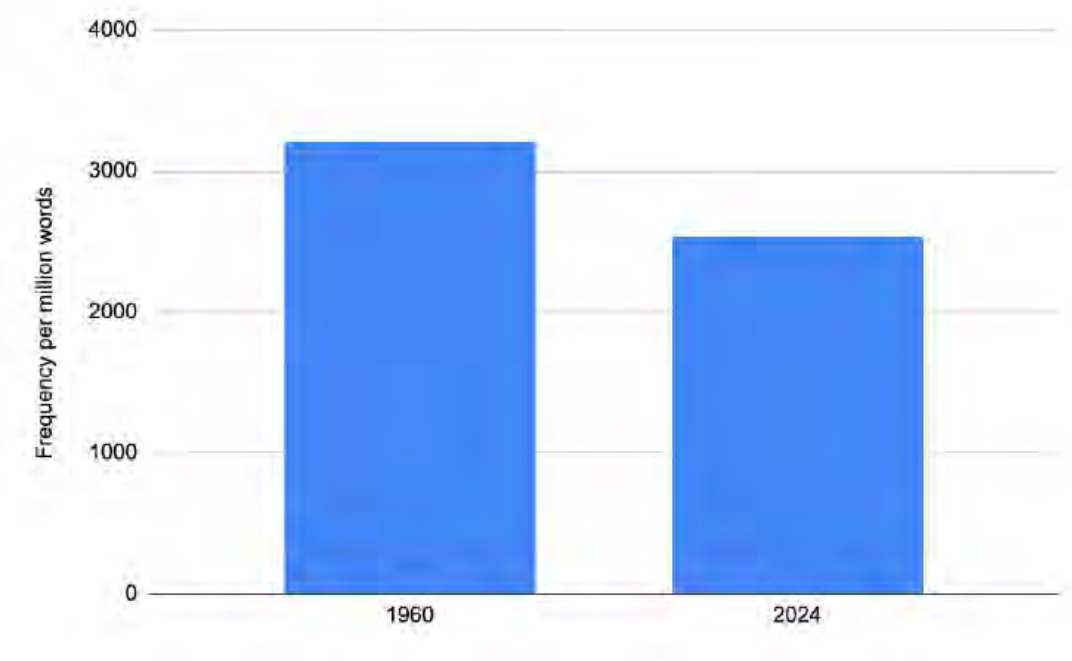


fig. 6

In place of wh-relatives, that-relatives (e.g., "the book that I read") and zero-relatives (e.g., "the book I read") have become more common, particularly in colloquial speech and informal writing. Zero-relatives, where the relative pronoun is omitted, are especially frequent in spoken English, contributing to a more streamlined and efficient way of linking clauses. For example, a sentence like "The project I'm working on" is more concise and direct than the equivalent construction with a wh-relative, "The project which I'm working on." This shift toward more straightforward relative clauses reflects the ongoing trend of reducing unnecessary grammatical elements in modern American English, making communication more direct and less formal.[28]

[27] Leech, *Change in Contemporary*.
[28] Mair and Leech, "Current Changes".

Another notable change in relativization is the shift away from "pied-piping" (in which prepositions are brought to the front of a clause along with a wh-expression – e.g., "To whom did you speak?") toward preposition stranding (e.g., "Who did you speak to?"). Preposition stranding, where the preposition is placed at the end of the clause, is more common in informal and spoken English, reflecting the preference for informality and conversational ease in contemporary American English. This shift away from "pied-piping" is particularly evident in spoken English, where preposition stranding is often used to make sentences sound more natural and less formal.[29]

Non-Finite Clauses

Non-finite clauses, particularly infinitival and gerundial constructions, have seen notable shifts in contemporary American English. Infinitival complements (e.g., "to run"), which are used after certain matrix verbs like "want" and "expect," have become more frequent in formal writing, particularly in academic and journalistic genres. The Brown and Frown corpora show a 15-20% increase in the use of infinitival clauses between 1961 and 1992,[30] with a 42% increase since (fig. 7).

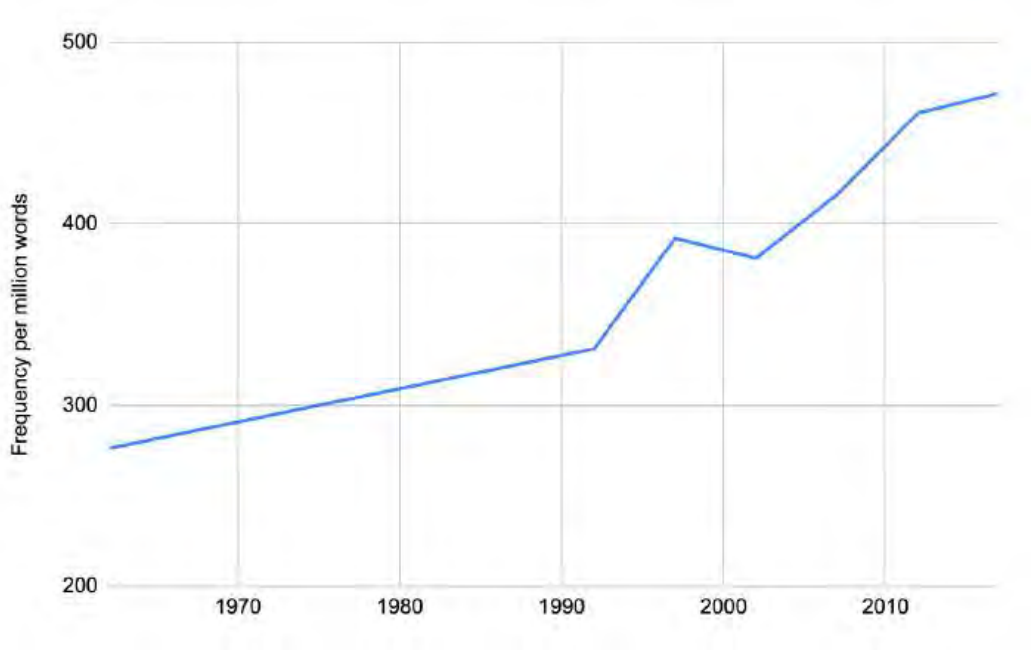


fig. 7

At the same time, gerund clauses (e.g., "running") continue to be widely used, particularly in spoken and informal contexts. Gerunds, which function as nouns but retain some verbal characteristics, are often used after verbs like "start," "stop," and "begin" (e.g., "He started running"). The competition between gerunds and infinitives in specific contexts (e.g., "started swimming" vs. "started to swim") highlights the flexibility of modern American

[29] Mair, *Twentieth-century English.*
[30] Leech, *Change in Contemporary.*

English, where both forms are used depending on the nuance the speaker or writer wishes to convey. In general, gerunds are more common in spoken English, while infinitives are more frequently used in formal writing.[31]

The rise of non-finite constructions reflects the broader trend toward syntactic simplification in American English, as speakers and writers increasingly prefer more compact and flexible sentence structures. Infinitives and gerunds allow for more efficient communication by reducing the need for additional clauses or modifiers, making them valuable tools for expressing future or ongoing actions in a streamlined manner.[32]

Conclusion

The syntactic shifts documented in this analysis highlight the ongoing evolution of American English, as the language adapts to shifting preferences as well as new technologies within modern communication. The rise of the progressive aspect, the decline of traditional modal auxiliaries, the increasing use of compact noun phrases, the shift from wh-relatives to that-relatives and zero-relatives, and the changing use of passive voice and non-finite clauses all point to a language that is becoming more streamlined, flexible, and informal. These changes are driven by broader sociolinguistic trends, including colloquialization, grammaticalization, and the influence of digital communication, which have reshaped the manner in which American English is used in both spoken and written forms.

The trends documented in this study suggest that American English is moving toward a more efficient and adaptable form, where brevity, directness, and informality are increasingly valued, due to the acceleration of linguistic evolution caused by factors such as globalization and online communications.

As American English continues to evolve, these syntactic changes will likely become even more pronounced, particularly as new forms of communication, such as social media and texting, continue to influence how people communicate. As the speed of change accelerates, perhaps even the grammar in a research paper such as this one will be unrecognizable in the coming decades.

[31] Mair and Leech, "Current Changes".
[32] Hilpert, *Constructional Change*.

Works Cited

Aarts, Bas. *The Verb Phrase in English : Investigating Recent Language Change with Corpora*. Cambridge, UK: Cambridge University Press, 2013.

Aarts, Bas, Joanne Close, and Sean Wallis. *Recent Changes in the Use of the Progressive Construction in English*. Accessed November 29, 2024. https://www.ucl.ac.uk/english-usage/projects/verb-phrase/RecentChangesProgressive.pdf.

Hilpert, Martin. *Constructional Change in English : Developments in Allomorphy, Word Formation, and Syntax*. Cambridge, UK: Cambridge University Press, 2013.

Krug, Manfred G. *Emerging English Modals : a Corpus-based Study of Grammaticalization*. Berlin, Germany: Mouton de Gruyter, 2000.

Leech, Geoffrey N. *Change in Contemporary English : a Grammatical Study*. Cambridge, UK: Cambridge University Press, 2009.

Mair, Christian. *Twentieth-century English : History, Variation, and Standardization*. Cambridge, UK: Cambridge University Press, 2006.

Mair, Christian, and Geoffrey Leech. "Current Changes in English Syntax." In *The Handbook of English Linguistics*, by Bas Aarts and April M. S. McMahon, 318-42. Malden, MA: Blackwell Pub., 2006.

UAPTA - Urban Air Pollution Treatment Apparatus By Júlia Silva Djahjah

Abstract

Intending to meet the Sustainable Development Goals (SDGs) outlined by the United Nations (UN) – with emphasis on SDG 11, which promotes the construction of sustainable cities and communities – this project studies the precariousness resulting from the exponential increase in population density in urban centers. Disorderly urbanization has exacerbated problems such as the inadequate occupation of space and the intensification of air pollutant emissions. The present study proposes a technological solution for the optimization of spatial use and the mitigation of atmospheric pollution levels in large metropolises via UAPTA (Urban Atmospheric Pollution Treatment Apparatus). This device was designed to mitigate the deleterious effects of pollutants in urban air, using microalgae as natural bio-filters. The technology exploits the photosynthetic and carbon capture capacity of microalgae to reduce pollutants. The underlying hypothesis of this initiative lies in the premise that with sustainable urban planning, disruptive technological innovation can be countered and ecological conservation and resilient and sustainable urban development fostered.

Keywords

Air pollution; Quality of life; Sustainable initiatives; Town planning; Global Health.

Introduction

The researchers believe that a careful analysis is required to understand phenomena related to air pollution and develop sustainable solutions to mitigate its impacts. Studies on air pollution and its relationship with human health, the environment, and the quality of life in urban centers can solve global problems of extreme importance (LUIZ, 2002).

Since the beginning of the Industrial Revolution, human activities have caused a transformation in the composition of the Earth's atmosphere. The burning of fossil fuels and other anthropogenic practices resulted in the emission of several hazardous pollutants, carbon dioxide (CO2), carbon monoxide (CO), sulfur dioxide (SO2), nitrogen oxides (NOx), volatile organic compounds (VOCs), ozone (O3), heavy metals and fine particles (PM2.5 and PM10), each with its chemical composition and variety of harmful effects.

Health professionals identify that air pollution has extensive impacts on human health. Medical research shows that the effects range from acute manifestations, such as irritation of the airways and aggravation of pre-existing diseases, to chronic consequences—respiratory and heart diseases, reduced life expectancy, and increased premature mortality. Epidemiological studies have associated chronic exposure to air pollutants with higher rates of morbidity and mortality in vulnerable populations.

Ecologists note that natural ecosystems also suffer the effects of air pollution, which contributes to soil acidification, vegetation degradation, and contamination of water bodies. Meteorologists explain that the dispersion of air pollutants by the wind amplifies these effects,

resulting in regional and global pollution problems that go beyond national borders (LUIZ, 2002).

In this context, the Urban Air Pollution Treatment Device (UAPTA) was developed, a multivalent solution designed to address the challenges of air pollution and preserve the quality of life and biodiversity in urban areas. The developers were inspired by biotechnological advances, such as the LIQUID 3 project at the University of Belgrade, with photosynthetic microalgae (2021).

The effectiveness of UAPTA lies in its ability to purify the air, capturing carbon dioxide and producing oxygen through photosynthesis, representing a sustainable and adaptable alternative to various urban environments.

The successful use of UAPTA uses technological advances to foster behavioral and political changes related to structural issues in society.

Accelerated urbanization in Brazilian metropolises has caused a deterioration in air quality, leading to the need for solutions that combine biotechnology and environmental engineering. The Urban Air Pollution Treatment Device (UAPTA) works with atmospheric bioremediation, using *Chlorella vulgaris* cultures in high-density vertical systems. As a biointegrated system, UAPTA comes from chemical principles that emphasize the desirability of the CO_2 sequestration characteristics of the microalgae *Chlorella vulgaris.*

The reaction stoichiometry shows that for each mole of CO_2 fixed, approximately 1.87 moles of O_2 are released, considering the ideal temperature and pressure conditions established by the system (LEGGIERI, 2015).

With mathematical modeling based on nonlinear differential equations, researchers have established that the function can express the efficiency of the UAPTA system:

$$E(t) = k\int(C_0\text{-}C(t))dt$$

K is the specific absorption coefficient of *Chlorella vulgaris*, C_0 the initial concentration of pollutants, and C(t) the concentration over time. This formulation helps in the interpretation of the operating parameters of the system.

Similarly, the chemical analysis of the metabolic processes of the microalgae shows that *Chlorella vulgaris*, when subjected to the controlled conditions of UAPTA, has a growth rate described by the equation:

$$\mu = (\ln X_2 - \ln X_1)/(t_2 - t_1)$$

The X is the biomass and t the time. This analysis maximizes the efficiency of the system in small urban spaces.

Additionally, the stoichiometric calculations show that each cubic meter of the UAPTA system, operating under optimal conditions, can process approximately 2.5 kg of CO_2 daily, converting it into algal biomass through photosynthetic reactions catalyzed by specific enzymes, notably RuBisCO (Ribulose-1,5-bisphosphate carboxylase oxygenase) (OHSE, 2008).

UAPTA in densely populated urban areas can result in a measurable reduction in air pollutant levels. The effectiveness of the system is remarkable in Brazilian metropolitan regions, where the concentration of particulate matter exceeds the limits established by the World Health Organization (ANVISA).

One aspect of the development of UAPTA is the optimization of the hydrodynamic parameters of the system, which influence the efficiency of the gas-liquid mass transfer. Through computational fluid dynamics (CFD) analysis, it was established that the controlled turbulent flow regime maximizes the air-liquid interface, increasing CO_2 absorption (WILZA, 2021).

In addition, the analysis of the growth kinetics of *Chlorella vulgaris* in the UAPTA system reveals a positive correlation between cell density and the rate of pollutant removal, following a modified Monod mathematical model. This model uses the light and nutrient limitations typical of the urban environment (BLAIR, 2014).

Theoretical Framework

The issue of air pollution goes beyond discussion about environmental phenomena and extends to controversial topics in the economy, ecology, quality of life, urban development and even historical issues.

In the economic sphere, the discussion about air pollution is often faced with industrial development, job creation and resource distribution. Developing countries, in their quest for rapid economic growth, often sacrifice environmental concerns for the sake of industrial progress, resulting in alarming levels of air pollution.

In a global context of accelerated economic growth, it is critical to recognize that environmental degradation is an unwanted by-product that becomes an increasing obstacle to long-term development.

Therefore, public policies should seek to internalize environmental costs, encouraging cleaner production practices and promoting the adoption of environmentally responsible technologies. This entails not only implementing stricter regulations and stricter emission standards, but also creating economic incentives, such as subsidies for investments in green technologies and carbon pricing mechanisms, that internalize environmental costs and promote innovation and environmental efficiency.

In many cases, the communities most affected by air pollution are those that already face significant economic and social inequalities, making it vital to formulate policies that address disparities in income and access to resources. This can include support and assistance programs for affected communities, as well as economic development initiatives that promote economic diversification and job creation in cleaner, more sustainable sectors. It is possible to perceive, in the socioeconomic sector, that the legacy of a slave society and the historical neglect of Brazilian native peoples left deep marks on the country's social structure, which are still visible today.

The slave system, which lasted for more than three centuries in Brazilian territory (from the arrival of the Portuguese in April 1500 to the proclamation of the Golden Law in May 1888),

exploited millions of Africans and shaped Brazil's social, economic, and political relations, creating an entrenched racial and social hierarchy.

The marginalization and exclusion faced by Afro-descendant communities in Brazil are reflections of a history marked by a hierarchical social structure system. Since colonial times, the arrival of enslaved Africans in Brazil to work on plantations and mines has contributed to the construction of the country's wealth, while these individuals were subjected to inhumane conditions and unequal treatment. The formal abolition of slavery in 1888, however, was not accompanied by effective measures of integration and reparation, leaving a gap in the search for equality.

Thus, racial segregation persisted in an unofficial way, being present in all spheres of Brazilian society, from access to education to employment opportunities. The lack of investment in infrastructure and basic services in areas inhabited predominantly by Afro-descendants perpetuated these inequalities, creating a vicious cycle of poverty and marginalization that persists today.

It is also necessary to mention police violence and structural racism, which are recurrent phenomena that contribute to a climate of mistrust between these communities and the authorities, further exacerbating their vulnerability and exclusion.

On the other hand, the discrimination faced by indigenous peoples in Brazil, a process analogous to racial boundaries, is the result of a history of exploitation and forced displacement, deeply rooted in the country's social and political structure.

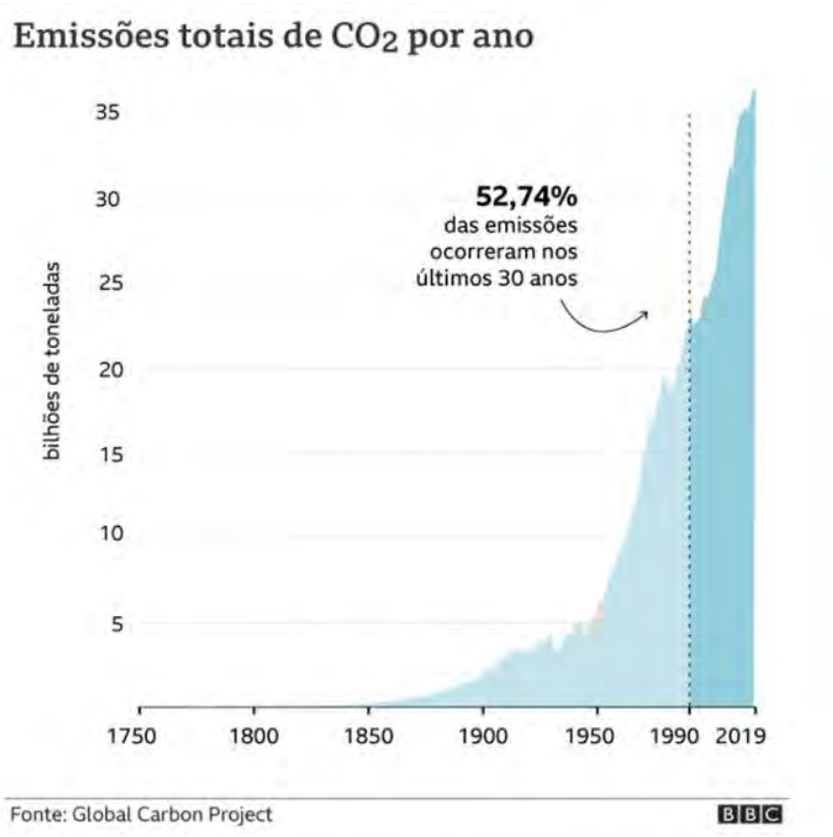


Fig 1: Total CO2 emissions per year
Source: https://www.globalcarbonproject.org/

Since the dawn of European colonization, indigenous peoples have been subjugated and their lands have been usurped for the sake of territorial and economic expansion. This process of territorial dispossession was accompanied by a policy of forced assimilation, which imposed Eurocentric standards of behavior and values that despised their millennial traditions.

Brazil's lack of recognition and effective protection of the territorial rights of indigenous peoples has been a constant source of insecurity for these communities.

The lack of adequate demarcation of their ancestral lands has allowed the invasion and illegal exploitation of natural resources by economic interests, exacerbating local tensions and conflicts.

Disorderly urbanization and uneven economic growth have been additional factors that make the conditions faced by indigenous peoples in Brazil even more alarming. Urban sprawl often comes at the expense of indigenous lands, which are often the target of real estate speculation, infrastructure projects, and industrial activities. This invasion of their traditional territories violates their fundamental rights and endangers their ways of life and livelihood, threatening their physical and cultural survival.

Thus, the exclusion faced by indigenous peoples in Brazil is the result of a set of historical, political, economic and social factors. To effectively analyze these problems, it is necessary to recognize the rights of social and economic inclusion of these communities in

Brazilian society. Only through a commitment to social justice and equal rights can a fairer and more sustainable future be built for all Brazilians.

The lack of access to basic resources, such as adequate housing, basic sanitation, and health services, is a reality for many Afro-Brazilian and indigenous communities, who continue to be marginalized and excluded from the country's social and economic progress. Disorderly urbanization has further exacerbated social inequalities. As urban areas have expanded rapidly without proper planning, these communities have often been pushed to the outskirts of cities, where they face precarious living conditions, a lack of basic infrastructure, and limited access to essential public services.

The lack of adequate housing policies and the neglect of land regularization have left many of these communities vulnerable to forced evictions and arbitrary removals, without guarantees of decent housing or fair compensation. As a result, many families end up in informal settlements, known as favelas, where they face extremely precarious living conditions, with limited access to clean water, basic sanitation, and health services.

The relationship between disorderly urbanization, social inequality and exposure to air pollution is an important step for the understanding of UAPTA. Rampant urbanization, often associated with lax urban planning and development policies, results in the concentration of populations in densely populated urban areas, where basic resources and services are often scarce and inadequate. This population agglomeration in urban spaces creates conditions conducive to the formation of pockets of poverty and social marginalization, in which historically oppressed communities, such as Afro-descendants and indigenous peoples, are disproportionately affected.

These groups, often relegated to peripheral urban areas devoid of adequate infrastructure, face exposure to air pollution. Proximity to sources of industrial pollution such as factories and refineries, coupled with a lack of effective environmental regulation in these areas, contributes to significantly higher levels of pollutants in the air, putting these communities at increased risk for respiratory, cardiovascular, and other health problems.

Air pollution infiltrates natural terrestrial and aquatic systems with devastating impacts on biota and vital cycles that are fundamental for sustaining life on the planet. The chemical analysis underlying ecology in this context reveals the interaction between air pollutants and the fundamental components of ecosystems. The presence of atmospheric contaminants, such as sulfur dioxide (SO2), nitrogen oxides (NOx), and particulate matter (PM), triggers chemical reactions that disrupt the natural balances of ecosystems.

These substances, in contact with the water present in the soil and water bodies, can react to form acids, leading to the acidification of these environments. This process compromises the availability of essential nutrients for the flora, directly affecting the growth and development of plants. The acidification of soils and surface and groundwater can also generate cascade effects, affecting the fauna associated with these environments and impacting the entire food chain. Acid deposition from air pollution is another factor that has direct effects on the viability of plant and animal species. For example, exposure to high concentrations of acids can cause damage to plant

cell membranes, interfere with nutrient uptake by the root system, and impair photosynthesis, thereby reducing biomass production and compromising plant health.

In addition, the bioaccumulation of air pollutants in the tissues of organisms and the subsequent biomagnification along the food chain amplify the effects of air pollution on ecosystems. Filter feeders, such as mollusks and crustaceans, are particularly susceptible to the absorption and accumulation of heavy metals and other air contaminants, compromising their health and that of the predators that feed on them, including mammals, birds, and even humans. The relationship, and to some extent dependence, between natural systems makes it visible that the impacts of air pollution extend far beyond urban and industrial areas, affecting entire ecosystems, from forests to oceans. This awareness of the interdependence of living things and the natural environment drives the search for a humanity that advocates for air pollution reduction measures to protect biological diversity and ensure a sustainable future for all forms of life on Earth.

In the context of air pollution, a dilemma inherent to human and economic development arises: reconciling economic growth and environmental preservation is difficult. Contemporary international dynamics are marked by an increasing interdependence between the economies of different countries, driven by globalization. In this context, in developing countries the need for economic growth to combat poverty and improve the living conditions of the population often conflicts with environmental preservation and long-term sustainability.

This interplay between economic development and environmental issues is most striking today by the increasing global competition for natural resources and markets, which often leads to resource overexploitation and environmental degradation. Developing countries, such as Brazil, find themselves in a position of vulnerability, being forced to prioritize immediate economic growth over environmental protection due to competitive pressure and the need to attract foreign investment. Thus, the challenge for these countries is to strike a balance between seizing the opportunities offered by globalization to drive economic development while adopting sustainable policies and practices that minimize negative impacts on the environment and promote a more equitable distribution of the benefits of growth.

The dependence of developing countries on more industrialized countries goes beyond the trade relationship. It is a global dynamic of power and exploitation, where the least developed countries find themselves in a position of structural disadvantage. This manifests itself not only in the export of raw materials and the supply of cheap labor, but also in technological dependence and the inability to control production and innovation processes. This dynamic perpetuates a kind of International Division of Labor (ILD) – a structure that has been adapting to technological advances since colonial times – where the least developed countries are relegated to peripheral roles in the global economy, while the more developed countries hold control over the higher value-added sectors. This divide limits opportunities for technological and sustainable growth in developing countries and contributes to the maintenance of unequal power structures.

This relationship of dependence and exploitation is reinforced by international institutions and agreements that favor the interests of the most developed countries. These include trade policies, tariff barriers, and regulations that make it difficult for developing countries to access global markets and limit their ability to compete on a level playing field. In contemporary times, society is immersed in individualism. This scenario, rooted in the incessant search for self-interest and immediate satisfaction, reveals itself as a psychosocial phenomenon of wide ramifications. Individualism, as a cultural ideology, promotes a disconnection between human beings and the environment that surrounds them, feeding a mentality of unbridled competition and social isolation.

Under the aegis of unbridled individualism, human beings are often compelled to act in disagreement with their natural instincts for cooperation and solidarity, plunging into a vicious cycle of greed and alienation. This disconnection of the self from the collective results in an inability to recognize and respond to the environmental and social challenges plaguing our contemporary world. In this context, the lust for resources and profits, without due consideration for ethical and environmental implications, becomes a hallmark of the individualistic mentality.

Unbridled consumption, another striking characteristic of today's society, and the incessant search for "having" eclipse the essence of "being", which resignifies the use of natural resources and validation through social relations as instruments of personal gratification. This reductionist logic perpetuates a cycle of environmental destruction and social inequality seen on a large scale in international dynamics (as in the DIT cited above).

In the globalized economy, the relentless pursuit of immediate profit supplants more effective ethical and environmental considerations. The current dynamics of production and consumption, based on the logic of the market and the maximization of financial returns, creates a substantial gap between what is feasible and what is morally correct. In this scenario, political and economic structures become complicit in the perpetuation of a cycle of exploitation and environmental degradation. Corporations, motivated by imperatives of profit and competition, tend to favor low-cost, high-yield production practices, often at the expense of environmental and social health.

This mindset of short-term, immediate profit imposes significant barriers to adopting more sustainable production and consumption practices. Even with the best intentions, economic agents often find themselves hostage to a system that values economic efficiency over long-term human and environmental well-being. The current dynamics of globalized society imposes a considerable challenge for the transition to a more sustainable development model. Even with the growing recognition of the need to preserve the environment, existing political and economic structures tend to perpetuate a cycle of exploitation and environmental degradation.

In addition, the fierce competition for natural resources and markets, which leads to over-exploitation and degradation of the environment. Developing countries, such as Brazil, often find themselves pressured to prioritize immediate economic growth over environmental protection due to the need to attract foreign investment and compete in the global market. Thus, to effectively address the challenges of air pollution and environmental degradation, it is

necessary to understand existing political and economic frameworks and adopt a realistic strategy for development.

In the graph below, it is possible to observe the substantial increase in carbon dioxide (CO2) emissions, especially after the year 1950. When analyzing this increase chronologically, globalization, the rural exodus and the consumption movements of modernity have a clear highlight in the justification of the phenomenon.

The graph – from the Global Carbon Project – fosters an important discussion about the concentration of the increase in emissions in recent decades, given that the period analyzed begins in the eighteenth century (GPC).

Since the dawn of human existence, the life and production habits of different peoples have been multifaceted and dependent on factors such as the culture, economic, social, natural and political situation of each region and period. Currently, factors such as urbanization and capital play a crucial role in understanding environmental problems such as the mission of greenhouse gases, pollution in cities and global warming.

The largest emitters of pollutants in the world today are China and the United States. Accordingly, the rates of production, consumption and population density are substantially high in these countries. The same occurs in Brazilian regions when analyzing pollutant emissions and urbanization and deforestation rates.

The Brazilian states with the most polluted air are in the Southeast region, as shown in the graph, which are classified as advanced in relation to globalization. An adaptable solution for sustainable development is necessary, so UAPTA was developed.

The most important issue to study in an attempt to promote sustainability in large cities is to understand what are the biggest impediments for this to occur. By observing the demographic density, it is possible to predict the impact that disorganized urbanization has had on green areas.

There is a need to adopt urban planning strategies that reconcile population growth with the preservation of green areas, aiming at quality of life and public health. In order to preserve nature through globalization, UAPTA is crucial. The device optimizes space and was formulated to facilitate urban planning because it is small and resilient to varied factors due to the project's design choices.

Methodology

The intentionality of the design of the LIQUID 3 of the University of Belgrade was the basis for the planning of UAPTA. The urban gardening initiative is intended to replace traditional trees. The main intention behind the design of LIQUID 3 is air purification (2021).

To achieve this goal, the project makes use of ecological technology. This strategy involves photosynthesis carried out by microalgae. With a total of 600 liters of water and a source of natural light, this urban photo-bioreactor is capable of capturing carbon from the atmosphere, working even in the coldest months – considering the European context in which it was developed.

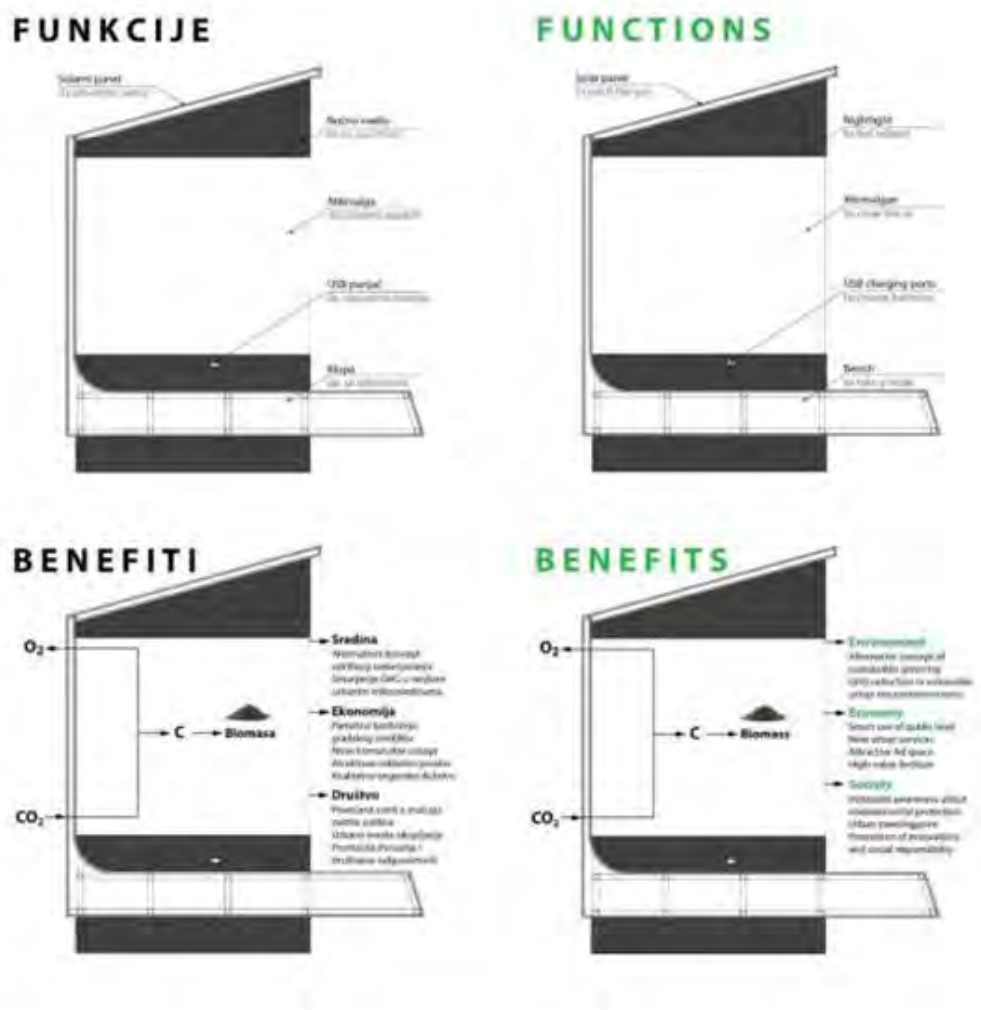


Fig 2: Liquid 3
Source: https://liquid3.rs/about/

The results are remarkable, as each LIQUID 3 can compensate for the presence of two 10-year-old trees or 200 square meters of green areas. The design of LIQUID 3 addresses the issue of air pollution, and promotes the efficient use of urban space and awareness of climate change, representing an innovative solution for smart and sustainable urban areas (UNIVERSITY OF BELGRADE, 2021).

UAPTA adapts the design concept of the LIQUID 3 and creates a small device with high productivity. The design of UAPTA – The Urban Atmospheric Pollution Treatment Device has its aspects of element arrangement very well thought out. The aquarium used has a cylindrical aspect to favor the circulation of microalgae and their proper oxygenation.

The cover has been cut with holes to ensure air inlet and out. The photovoltaic panel was placed above the lid to ensure the greatest solar incidence, therefore, the best use of light and thermal energy.

The air pump uses the energy from the photovoltaic panel for its operation. The pump ensures the circulation of microalgae and prevents live microalgae from settling on the bottom of the aquarium.



Fig 3: UAPTA prototype
Source Authorial

The algae are of the *Chlorella Vulgaris type*, since they have shown greater resistance to temperature and amount of water. In addition, algae of the *Chlorella* species usually have a lower cost, which facilitates the application of UAPTA financially.

Such microalgae are responsible for the photosynthesizing process, which aims to reduce air pollution. A hole was made that ensures the entry and exit of air. In the demonstrations, polluted air will be introduced to UAPTA, and the pollutant parameters will be measured. Soon, the air that is harmless to the environment or to human health will come out.

Finally, CO2 detection sensors in the environment were programmed, which were used for productivity tests and will be useful in demonstrating the operation of UAPTA.

The two sensors were programmed as Arduino on the arduino.cc/en/software platform.The Arduino MG811 used for the detection of carbon dioxide was used. The design of UAPTA, or Urban Air Pollution Treatment Device, effectively adapts and extends the innovative principles of the University of Belgrade's LIQUID 3.

In the subsequent stage of UAPTA's development, theories and features are implemented and tested to assess the effectiveness of the system. Before starting the experiment, an action

plan was developed to ensure that the steps were executed in a practical and methodical manner within the established timeframe.

The main strain of microalgae, *Chlorella Vulgaris*, was grown in a cylindrical aquarium. However, the use of a parallelepiped-shaped aquarium was tested, which did not work, because sediments accumulated in the corners of the aquarium. The other strain, *Chlorococcum Amblystomatis*, was cultivated for the purpose of comparison and validation of the superiority of resistance of *Chlorella Vulgaris*.

Two CO_2 sensors were installed in the system to measure the concentration of carbon dioxide before and after treatment via UAPTA. This allowed the quantification of reductions in CO_2 levels as a result of the photosynthetic activity of the microalgae.

During the experiment, the decantation of dead microalgae at the bottom of the aquarium was observed. This organic residue was later tested as a fertilizer for plants, showing efficacy in comparison with plants not fertilized with this material.

During the operation of UAPTA, continuous measurements of CO_2 concentration, oxygen levels, as well as microalgae growth and productivity were made. These measurements helped to assess the effectiveness of the system in terms of reducing pollutants and producing oxygen.

UAPTA has an average volume of 10 liters and through a function, the effectiveness of its application in the city of São Paulo was approximate.

Local da Medição	Concentração de CO_2 antes do ATUPA (ppm)	Concentração de CO_2 depois do ATUPA (ppm)	Redução estimada de CO_2 (ppm)
Centro da Cidade	450 ppm	400 ppm	50 ppm
Rodovia Principal	550 ppm	480 ppm	70 ppm
Bairro Residencial	420 ppm	380 ppm	40 ppm
Parque Urbano	400 ppm	360 ppm	40 ppm

Fig 4: Estimates for visualization of the effect of UAPTA
Source: Authorial

During the experiment, variations in UAPTA yield were observed due to climatic conditions. In response to this, a mechanism with thin sheets of aluminum and cork was developed to stabilize the temperature inside the UAPTA.

The next step for the improvement of UAPTA will be the adaptation of the device to meet specific needs in places such as industries, urban centers and bus stops. This will consider climatic factors and circumstances that may affect the survival of microalgae.

Results and discussion

Chlorella vulgaris, a single-celled microalgae belonging to the chlorophycea group, has attracted considerable interest due to its remarkable photosynthesis capacity and rapid growth rate (CHIA, 2013).

This chapter explores in detail the technical aspects of the cultivation of this algae, addressing the crucial factors that influence its productivity, as well as the methodologies used to optimize its growth under controlled conditions.

The high photosynthetic efficiency of this algae is one of the key factors for its rapid growth rate. Under ideal conditions, *Chlorella vulgaris* is able to capture an average of 7 mg/L of CO2 during the photosynthesis process (CHIA, 2013).

The productivity of *Chlorella vulgaris* can be significantly influenced by several environmental factors, which need to be carefully controlled to optimize cultivation. The water used in cultivation must be of high quality, free of contaminants that can inhibit algae growth. The presence of heavy metals, organic pollutants and pathogenic microorganisms can compromise the viability of *Chlorella crops*.

The availability of nutrients, particularly nitrogen, phosphorus, and potassium, is essential for the healthy growth of microalgae. Proper supplementation of micronutrients, such as iron, magnesium, and zinc, is also vital for maintaining high levels of productivity. The temperature of the growing medium must be strictly controlled. The ideal range for growing *Chlorella vulgaris* usually ranges between 20°C and 30°C. Temperatures outside this range can reduce photosynthetic efficiency and, consequently, the growth rate (BLAIR, 2014).

CO2 is a fundamental substrate for photosynthesis. Injecting CO2 into cropping systems can increase the growth rate of microalgae. The concentration of CO2 must be carefully monitored to avoid acidification of the growing medium, which can be harmful to algae.

Chlorella vulgaris *cultivation systems* can be classified into open and closed. Each type of system has specific advantages and disadvantages that influence the choice of cultivation method. Open systems, such as ponds and stirring tanks, are more cost-effective and easier to implement on a large scale. However, they are subject to contamination by other microorganisms and environmental variability, which can negatively impact productivity.

Closed systems include tubular and flat photo-bioreactors, which offer a controlled environment, minimizing the risk of contamination and allowing the optimization of growing conditions. These systems, although more expensive and complex, provide greater efficiency in biomass production.

The intensity and quality of light are adjusted to maximize photosynthetic efficiency. The use of LED light with specific spectrums can promote algae growth. Maintaining an optimal pH (usually between 7.0 and 8.0) is essential for the growth of microalgae. The pH can be adjusted through the controlled addition of acids or bases. Proper aeration ensures the even distribution of CO2 and nutrients in the growing medium. Mechanical or bubble mixing prevents sedimentation of the algae cells and improves light absorption. The use of sensors and real-time monitoring systems allows for immediate adjustment of growing conditions, ensuring process stability and efficiency.

Data	Temperatura	Coloração	Circulação
14/ago	20°C	Esverdeada	Média
15/ago	23°C	Esverdeada	Fraca
21/ago	22°C	Amarelada	Fraca
22/ago	22°C	Amarelada	Fraca
28/ago	25°C	Amarronzada	Média
29/ago	26°C	Amarelada	Boa
04/set	30°C	Esverdeada	Boa
05/set	29°C	Esverdeada	Boa
11/set	11°C	Amarronzada	Boa
12/set	12°C	Amarronzada	Boa
18/set	20°C	Amarelada	Boa
19/set	22°C	Esverdeada	Boa
25/set	26°C	Esverdeada	Boa
26/set	25°C	Esverdeada	Boa
02/out	29°C	Amarelada	Boa
03/out	28°C	Amarelada	Boa

Fig 5: UAPTA monitoring
Source: Authorial

Chlorella Vulgaris is known for its adaptability to temperature variations, but its ideal temperature range varies. In general, this microalgae can grow in a wide temperature range, from 15°C to 35°C (CHIA, 2013).

However, as can be seen in the table, its ideal growth temperature is usually around 20°C to 30°C. Chlorella *Vulgaris*' temperature tolerance is an important aspect in its cultivation, as extreme variations can hinder its growth and productivity.

Sudden temperature swings can stress the algae cells, leading to decreased growth rate and loss of beneficial properties.

To maintain the productivity and quality of *Chlorella Vulgaris*, it is essential to control the temperature in growing environments. Therefore, temperature insulating materials were used at UAPTA, in order to maintain the temperature and reduce possible oscillations. In addition to its environmental benefit, the filter aquarium also produces biomass rich in proteins, vitamins,

minerals and essential fatty acids. This biomass can be used as a resource to feed plants, contributing to soil fertilization and promoting vegetation growth in urban areas.

The analysis of the UAPTA system (Urban Atmospheric Pollution Treatment Apparatus) showed efficiency in the biofixation of atmospheric CO_2 through the biotechnological implementation of Chlorella vulgaris. Experimental data showing a carbon fixation capacity in the order of 7 mg CO_2/L, correlated with an average cell density of 1.2 x 10^6 cells/mL under ideal culture conditions (23°C ± 3°C).

The photosynthetic efficiency of C. vulgaris was mathematically expressed through the general equation of photosynthesis:

$$6CO_2 + 6H_2O + \text{light energy} \rightarrow C_6H_{12}O_6 + 6^o$$

The spectrophotometric analysis of the culture showed characteristic absorbance peaks at 435nm and 675nm, corresponding to chlorophyll-a, with variations correlated to the observed chromatic changes (greenish → yellowish → brownish), indicative of modulations in the photosynthetic apparatus.

Temperature monitoring (11-30°C) showed a second-degree polynomial correlation (R^2 = 0.89) between temperature and specific growth rate (μ), with optimum at 25°C, where μ = 0.42 day^{-1}.

The stoichiometric analysis of the biomass produced pointed to an average composition of:

$$CH_{1.83}O_{0.48}N_{0.11}P_{0.01}$$

The C:N:P ratio of 106:16:1, typical of the Redfield (AQUABASE) composition for microalgae. The fluid dynamics of the system demonstrated Reynolds (Re) numbers ranging between 2000 and 4000, characterizing a transition regime that uses mass transfer without affecting cell integrity. The CO_2 transfer efficiency (KLa) was calculated to be 0.025 s^{-1}, resulting in a CO_2 absorption rate of approximately 1.8 g CO_2/L/day under optimal conditions. The biochemical analysis of the biomass showed a protein content of 45-50% (w/w), with a complete amino acid profile and an essential amino acid index (EAAI) of 0.95. The lipid content ranged from 20-25% (w/w), with a predominance of polyunsaturated fatty acids (PUFAs), notably EPA (C20:5) and DHA (C22:6), representing potential added value to the system.

The mathematical modeling of growth followed the logistic model modified since the early inquisitions:

$$dX/dt = \mu max \cdot X \cdot (1-X/Xmax) \cdot (S/(Ks+S))$$

X is the cell concentration, μmax is the maximum specific growth rate (0.42 $days^{-1}$), Xmax is the maximum cell concentration (2.5 x 10^6 cells/mL), S is the limiting substrate concentration, and Ks is the saturation constant (15 mg/L).

The energy balance of the system showed an average consumption of 0.5 kWh/kg of biomass produced, resulting in a positive net energy efficiency when considering the CO_2

biofixation capacity. The thermodynamic analysis revealed a reaction enthalpy (ΔH) of -2870 kJ/mol fixed CO_2, with Gibbs free energy (ΔG) of -479 kJ/mol, confirming the energy viability of the process.

On an urban scale, UAPTA projects a CO_2 sequestration capacity equivalent to 10 adult trees per unit (considering an average tree with a fixation capacity of 22 kg CO_2/year), with advantages of lower spatial demand (1.5 m²) and greater adaptability to unfortunate urban environments.

Scalability calculations point to the prediction that a network of 1000 UAPTA units could sequester approximately 220 tons of CO_2 per year. Life cycle analysis (LCA) demonstrated a carbon payback of 3.5 months, considering the emissions associated with the manufacture and installation of the system.

Therefore, the feasibility of the UAPTA system as a biotechnological solution for the mitigation of urban atmospheric pollution is based, with foundations for the development of parameters for the future improvement of the system and perhaps direction for genetic modifications of the strain to increase photosynthetic efficiency and resistance to environmental stresses.

Works Cited

ANVISA. "Health Surveillance of Populations Exposed to Air Pollution (Vigiar)." *Ministry of Health*. Available at: https://www.gov.br/saude/pt-br/composicao/svsa/saude-ambiental/vigiar.

AQUABASE. "Redfield's Reason – General Clarifications – Aquabase." *Aquabase.com.br*. Available at: http://aquabase.com.br/wordpress/razao-de-redfield-esclarecimentos-gerais/. Accessed 1 Nov. 2024.

Blair, Matthew Forrest, Bahareh Kokabian, and Veera Gnaneswar Gude. "Light and Growth Medium Effect on *Chlorella vulgaris* Biomass Production." *Journal of Environmental Chemical Engineering*, vol. 2, no. 1, 2014, pp. 665–674. Available at: https://www.sciencedirect.com/science/article/pii/S2213343713002200.

Chia, Mathias A., Ana T. Lombardi, and Maria da Graça G. Melão. "Growth and Biochemical Composition of *Chlorella vulgaris* in Different Growth Media." *Annals of the Brazilian Academy of Sciences*, vol. 85, no. 4, 2013, pp. 1427–1438. Available at: http://www.scielo.br/pdf/aabc/v85n4/0001-3765-aabc-00-00-6613.pdf. Accessed 19 Apr. 2020.

Global Carbon Project. *Global Carbon Project (GCP).* Available at: https://www.globalcarbonproject.org/.

Leggieri, Leonardo R., and Nicolás A. Ferreiro. "Response of the Biomass and the C:N:P Composition of the First Trophic Level of a Pampean Stream to the Highest Drought in the Last 20 Years: A Case Study." *Ecología Austral*, vol. 25, no. 3, 2015. Available at: https://bibliotecadigital.exactas.uba.ar/greenstone3/exa/collection/ecologiaaustral/document/ecologiaaustral_v025_n03_p172. Accessed 1 Nov. 2024.

Luiz, Alfesio, Alberto Pereira, Paulo Hilário, et al. "Air Pollution and Its Effects on Human Health." *ComCiência*, 2002. Available at: https://www.comciencia.br/dossies-1-72/reportagens/cidades/paper_saldiva.pdf.

Ohse, Silvana, Roberto Bianchini Dérner, Renata Ávila Ozório, et al. "Growth of Microalgae in a Stationary Autotrophic System." *Biotemas*, vol. 21, no. 2, 2008.

University of Belgrade. "LIQUID3 – Urban Photo-Bioreactor." *LIQUID3*. Available at: https://liquid3.rs/.

Wilza, Carla. "Simulation of the Single-Phase Flow in One Stage of a Centrifugal Pump Using Computational Fluid Dynamics Techniques." *UFRN*, 2021. Available at: https://repositorio.ufrn.br/handle/123456789/13009. Accessed 1 Nov. 2024.

Acknowledgements

First, I would like to express my gratitude to my parents, my greatest motivation – Débora and Ricardo Djahjah.

I am immensely grateful to the colleagues who contributed to the realization of this study – Ana Luísa Cunha, Daniela Medeiros, Marcela di Fiore, Marina Carlésimo and Rafael Conde – and to Professor Tiago Naves Brigido, advisor of this research.

I extend my gratitude to the institution that provided us with the necessary infrastructure to conduct this research: Colégio Visconde de Porto Seguro – Morumbi.

To all those who, directly or indirectly, contributed to the realization of this study, my sincere recognition and thankfulness.

Analyzing the Failure Modes of Neuromodulation By Anirudh Nalluri

Abstract

Neural prosthesis provides a solution for a few of the most prominent and fatal diseases, such as Parkinson's disease, paralysis and quadriplegia. Prosthesis, such as deep brain stimulation and microelectrode arrays, either modulate neural activity or record it so that a patient may improve their quality of life. Failure modes for neuromodulation can occur through many mediums, including mechanical, chemical, electrical and magnetic. This article investigates two prominent mediums, electrical and magnetic stimulation, and explores the ways in which these devices can be improved.

KeyWords

micromagnetic simulation, epidural electrical stimulation, transcranial direct current stimulation, micro-coils, electric field, impedance

Introduction

Throughout the past millennia, prosthetics have had a significant impact on the lives of many individuals with physical disabilities. Prosthetics enable individuals with physical challenges to experience greater quality of life and enable them to perform many actions. There have been significant advancements in the field of prosthetics in the past decades. With the introduction of neural prosthetics, the possibility of recreating hand movement is getting closer to reality. Neural prosthetics enable the central nervous system to communicate with the peripheral nervous system. They differ from normal prosthetics because prosthetics usually focus on bringing movement back to limbs, whereas neural prosthetics aim to bring back control to either the spinal cord or the brain. Neural prosthetics are needed in the process of curing and palliating conditions like Parkinson's Disease, and a range of spinal cord injuries. Although there are a plethora of cases where they are useful, some of the most common instances include damage to the brain or spinal cord and neurological diseases.

Pathology

There are two types of spinal cord injury (SCI), complete SCI and incomplete SCI. Complete SCI is permanent damage to the area of the spinal cord that has been affected. Incomplete SCI is when sensory and motor function are partially preserved. Figure 1 shows the differences between complete and incomplete SCI. There are two phases to SCI. Primary phase is initial physical injury, disrupting the blood spinal cord barrier and blood vessels [1]. Secondary phase describes the cellular and chemical processes that further worsen the condition of the injury, due to the invasion of inflammatory cells. These inflammatory cells can either affect the spine through direct contact or fluid secretion [1]. Figure 2 shows the difference between primary and secondary phase [2]. SCI can occur at different sections of the spinal cord, some of them causing complete paralysis. This state of complete paralysis and SCI leads to quadriplegia.

Quadriplegia, also known as tetraplegia, is a dysfunction or a loss of motor and/or sensory function. Although such conditions are extremely detrimental, there are many ways through which they can either be lessened to improve the quality of life. Some of the current solutions for these diseases are shown in Figure 1 and Figure 2.

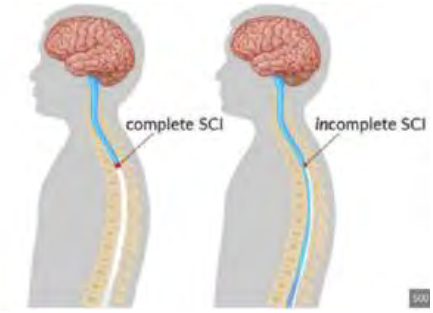


Figure 1

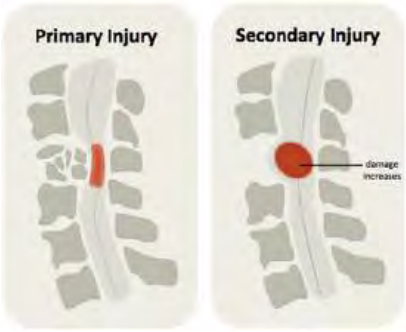


Figure 2

Current Neuromodulation Products on the market

An example of a neuromodulator is the Vercise genus deep brain stimulation (DBS) system by Boston Scientific Corporation. It performs magnetic stimulation to reduce tremors in the upper arms and is comparatively less invasive than other DBS devices.

A competitor for this device is the Abbott Laboratories' Infinity DBS System. It is aimed to be a personalized DBS therapy that can give patient's control of their Parkinson's disease or essential tremor system [10]. Not all neuromodulators have the same function, there are many different types of devices and varying mediums through which neuromodulators function. Some of the most widely used neuromodulators are transducers, epidural electrical stimulation and micromagnetic stimulation.

Transducers

Electric stimulation is the process where a device (typically a transducer or an electrode) gives an electric shock to a certain part of the nervous system, whether it be the brain or the spinal cord, in order to activate that part. Transducers play a pivotal role in neural prosthetics as they prove an electric current crucial to the activation of the brain. Electric stimulation transduction is extremely unstable because the channels respond to millivolt stimulus at a range of milliseconds. However, a beneficial avenue for future exploration would be the development

of a wireless neuromodulation, as these devices could respond to light, temperature, and magnetic waves. Each of these types of transducers have their own properties and advantages. For instance, mechanical stimulation communicates through the cell's membrane capacitance; it also does not require genetically modified cells [3].

An example of mechanical and electrical stimulation are piezoelectric transducers. The generation of mechanical force can be introduced by other forces like light and magnetic force. In order to test these characteristics, transducer- based polymer nanoparticles are used to stimulate neural activities, while nano-transducers induced acoustic waves are used to activate neurons. Upon light stimulation, the photoacoustic transducers were able to modulate brain activities. This was found in freely behaving mice. This technique exhibited a temporal resolution in the millisecond range, a spatial resolution in the sub-millimeter range, where no harmful temperature accumulation occurred. Chemical transducers work through peptide receptors, which works through the release of molecules to the brain [3]. It has high resolution and can be used in various molecules. Unlike electric, magnetic and ultrasound transducers, which are both non-invasive, can both go deep through the tissues.

EES

Epidural electric stimulation (EES) is a form of therapy to bring back control in the spinal cord, which in turn can bring back the volitional movement of limbs [1], which can be used to treat cardiovascular diseases. The stimulation of L4 and L5 levels of spinalized cats proved effective hindlimb motion [1]. However, the results of these animal studies could not be fully replicated in human studies. To bridge the connection between animal and clinical studies, a combination of neural recording techniques on the spinal cord were used in both animals and humans. EES and EMG (Electromyography). A series of three responses were observed;

1. Early Response: Direct result of motor neurons.
2. Monosynaptic Response: synapse(communication of neurons) without the interference of another neuron. (Reflexes - Nervous System Structure And Function - MCAT Content, n.d.)
3. Polysynaptic response: synapse is formed through the connection of multiple neurons. (Reflexes - Nervous System Structure And Function - MCAT Content, n.d.)

Different frequencies were recorded at different sections of the spinal cord to stimulate and innervate parts of the body. The following are the different outcomes based on different parameters:

Frequency (Hz)	Voltage (v)	Spinal cord Sector	Effect
25-60 Hz	1-9 v	C 7	enabled standing stepping, and voluntary leg controls [1]
25-60 Hz	1-9 v	T11-L1	enabled bearing of complete body weight [1]

Frequency of 2.2Hz showed potential signs of compound muscle action and frequencies of 25-50 Hz showed rhythmically alternating extension activity [1].

EES has proved to be effective such as improving weight bearing capability in exercise and daily activities after EES therapy [1]. also proven to enable verbal, auditory and visual function, although the effectiveness for patients varies [1]. Although EES is highly beneficial, it is still risky as implementing EES electrodes requires a laminectomy, which poses significant risks such as spinal cord instability and deformity, due to the biomechanical process [1]. EES also requires months of rehabilitation [1]. Many of the problems mentioned above, like size and comfort, can be solved by replacing the use of electrodes with microcoils, which uses magnetic force to create electric fields.

Magnetic Stimulation

In order to transfer electric signals, neurons have fluctuational membrane and dynamic changes, to produce action potentials [4]. Hence, this allows for the exogenous electric signals to modify these fluctuations, which is known as TMS (Transcranial Magnetic Stimulation). Image 3 illustrates the functioning of TMS. Transcranial Direct Current Stimulation uses weak but direct currents to do the same but in a non-invasive way [4]. Figure 3 shows a visual for the functioning of TMS [5].

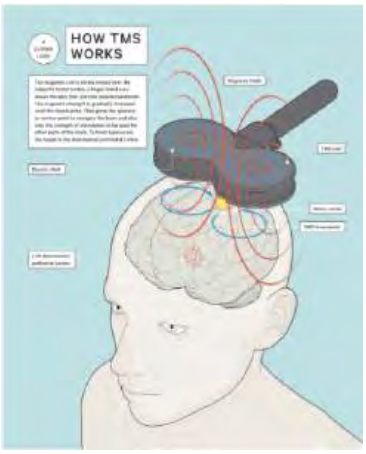
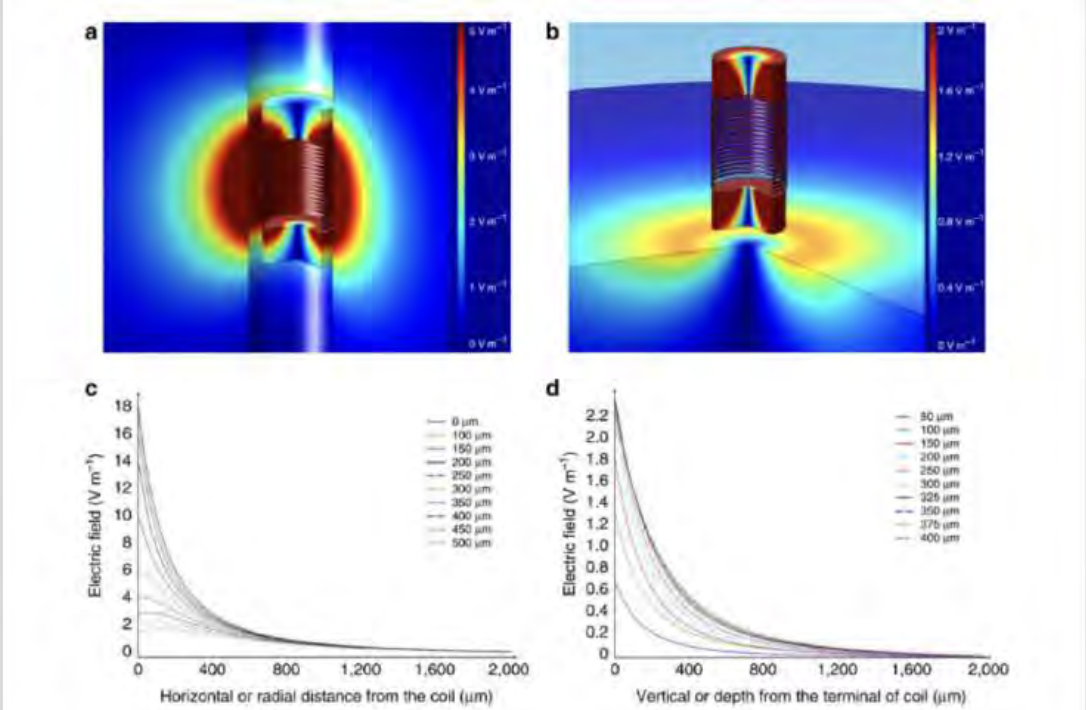


Figure 3

With current devices such as TMS and TDCS, there are many problems such as incompatibility with magnetic resonance imaging, less control of electric fields and also less effectiveness when encountered by tissue inflammation. When performing an MRI (Magnetic resonance imaging) the waves can cause neural damage when screening is done on a person with electrodes implemented in their body [4]. This is because the electrode tip becomes heated, as the MRI generated radio frequency waves interact with the conductive leads [6]. TMS creates long-time effects of current inducing magnetic fields. These coils are extremely costly and require the patient to stay in the clinic for long periods of time and also require large devices to

create the magnetic fields. Hence, a smaller device like electrodes can overcome the problems that TMS faces. These electrodes are still too big to not create a destructive footprint on the neural system. Another solution could be Micromagnetic Stimulation (μMS), as these coils can be turned off, making it suitable during an MRI scan. They are also isolated from the adjacent tissue, rather than being attached, ensuring further safety. In order to confirm that the microcoils can produce electric fields sufficient to stimulate neurons, an experiment was conducted, which explored the current field produced by the microcoils. In order to test the electrode, it was encased in a cylinder of 3mm radius and 3mm height, in which different materials are enclosed. creating a circulating electric field [6]. It was recognized that the field had a 500 μmeter radius. The strength of the field decreased with distance when the coil was oriented perpendicular to the retinal surface [14]. This decrease was non-linear. The following graphs show the electrode and its field:



(**a**) Three-dimensional distribution of the electric field around the coil starting at 100 μm from the edge of the device. (**b**) Three-dimensional distribution of the electric field starting at 100 μm below the terminal of the coil. (**c**) Plot of the radial distribution of the electric field from different elevations. Legend shows the different elevations from coil-center. (**d**) Plot of the axial distribution of the electric field from different radial distances from the axis of the device. Legend shows the different radial distances from coil-center.

Figure 4

Figure 4A. and Figure 4B. show the three-dimensional distribution of the electric field in different directions. Figure 4A shows the distribution of the electric field around the coil starting from 100 micrometers from the edge of the device, the color indicating the electric field (V m-1). Figure 4B. is similar to figure 4A, as it shows the distribution of the electric field starting at

below the coil. Figure 4C plots the horizontal distribution of the electric field from different elevations. Different colored lines represent the elevations from which the experiment was tested. Figure 4D. plots the vertical distribution of the electric field from different radial distances from the coil center. These studies were performed on retinal cells [6].

In this figure, retinal cells were stimulated first by direct activation through the μms resulting in a single action with latency =< 1ms. The second method is secondary activation which activates a neighboring cell which results in spike train signals. Although the exact reason is not known for the chain reaction, it is hypothesized that it is because the calcium gates are activated which ensure an increased release of synaptic transmitter. This availability of two methods is useful, as it can either be used to selectively target presynaptic terminals or prevent unwanted axon activation. μMS also shows similar effects as that of electric stimulation. During testing, waveforms were recorded at a sample of 20kHz. After processing this, it was recognized that the waveforms had an electric artifact arising from the ums pulse, which lasted for 20 milliseconds [6]. This was the same for all the trials with identical stimulus. Many responses also contained spikes, which were recognized as they were the same as those during light stimulation.

Inconsistencies and problems

Although there have been many wonderful advancements in the field of neuromodulation, there are many inconsistencies amongst their effectiveness which blocks many devices from being on the market. There is more to be done to uncover how neuromodulation affects the cellular level, which could be the key to establishing neuromodulation devices globally. One of the most significant reasons for inconsistency and varying results among different patients is tissue heterogeneity. Tissues are inhomogeneous as they consist of different biological elements throughout the structure, providing different resistances to electricity. Additionally, tissue swelling has been observed after intense neuronal activity, further varying the resistance. Due to this wide variety of resistances inside the tissue, there is a decrease in tissue conductivity and a smaller volume of tissue being stimulated by either microcoils or electrodes. Although this heterogeneity is disadvantageous, it is also beneficial as the heads of the neurons are anisotropic [4]. “Anisotropic” in this case refers to how the forces inside the tissue aren’t the same in all directions, which channels the electric current from the transcranial direct current stimulation (TDCS).

Another reason for inconsistencies amongst tissues is the impedance between the electrode and the contacting tissue. The impedance determines the current flow inside the tissue and is a measure of resistance of the flow. Impedance is also worsened by the insertion of electrodes. With the fluctuation of impedance, it is harder for the spread of electric current, which therefore makes it harder for the neurons to be stimulated. With the insertion of electrodes, the destroyed path leads to high impedance for the next 3-6 months [4]. During this time before stabilization, neuroinflammation can be caused by the invasive electrode. The introduction of electrodes can also lead to glial scarring, which prevents the readout of neural activity. The orientation of the cell to the electric current affects the pattern and intensity of the membrane

potential, as seen with Figure 1. Membrane potential is the gradual polarity of a membrane where one side is negative and the other is positive, which creates the flow of ions. Research on the orientation of the electric current showed that when an axon was parallel to the electrode, the activation threshold is lower than the angle between the electrode and axon when it is closer to perpendicular [10]. The reason for varying membrane potential is the size of the axonal diameter. A larger axon size is correlated with less resistance and lower threshold, whereas a smaller axon size generally means higher resistance and threshold. Neurons also contain a myelin sheath, which is not a great conductor of electricity, but rather is an insulator. Although it is an insulator, it is also a fast conductor at the nodes of Ranvier in between the myelinated sheaths. Inconsistencies are caused by the age of neurons, as the growth of neurons can lead to many changes. Electric current itself can alter the properties of neurons. High frequency can lead to a fatigued response in the axons of the spinal cord. Excess of potassium is released during electric stimulation, which leads to depolarization. Neurons use ephaptic effects to communicate. Ephaptic effect refers to electrical conduction. This field potential interferes with the electric stimulation, and vice versa [4].

An issue that interferes with the functioning of neuromodulation is the surgical implementation of the device. The insertion of devices into the brain immediately activates nearby microglia cells, which extend processes toward the implant surface [12]. The microglial cells start encapsulating the implanted device with lamellipodia [7]. This causes a thin cellular sheath to be formed that limits ionic exchange with the electrode surface. The impact of the immune response to foreign bodies on brain activity is not well understood, and may alter nervous function, especially within the reactive glial sheath area. Its impact may need to be considered in the interpretation of the collected data from both acute and longitudinal experiments, which shows how the function of dopamine in each cell varies. In a neuron, it regulates motion and emotions, whereas it induces calcium signaling in an astrocyte. In order to understand how the variability in injury impacts inconsistency of sensor performance, it is necessary to examine the inflammation biochemical cascades initiated by device insertion. The initial injury can lead to increased inflammation through a number of factors. Some of the factors include rupturing the blood brain barrier, reducing blood flow, oxygen perfusion, and neurotoxic waste removal, which in turn can cause ischemia and hypoxia [7]. Increasing mechanical strain from hemorrhage and the device volume are also causes.

Another reason for cell encapsulation and inflammation is the presence of metals polymers, silicon, and silica in the device; all of which are relatively hydrophobic. This quality makes their surfaces susceptible to protein absorption. One example of evidence that the electrode damages the blood brain barrier is that the drug carbidopa was observed to be present in the brain dialysates, although the drug molecule normally does not cross the blood brain barrier [7]. As this movement only occurred after the insertion of the device, it is evident that the implants open the blood brain barrier.

Another study that demonstrates the damage caused by the implant's implementation was done by using fluorescent nano-beads that were implemented into the blood flow. These beads

spill out of the blood vessels near implanted probes but not anywhere else in the brain, showing how the blood brain barrier is open. After the blood brain barrier disruption, inflammatory molecules can absorb onto the implant. The disruption of the blood brain barrier can lead to the deposition of plasma proteins foreign to the CNS, such as albumin and globulins. In particular, albumin has been shown to bind to transforming growth factor-β receptors in astrocytes. Inducing contractions and weakening cell−cell adhesion, in turn propagates additional BBB leakage [7]. In addition, albumin has been shown to activate astrocytes and microglia through the mitogen-activated protein kinase pathway resulting in increased levels of nitric oxide in astrocytes.

There are also significant risks that are associated with penetration of brain implants, which can lead to stress upon the device. Pericytes are critical components of the BBB and are tightly coupled to the capillary endothelial cells. They reduce BBB permeability from periphery-derived harmful substances while allowing influx of nutrients. During the surgical implantation of the device, the insertion injury causes pericyte loss. This can lead to many problems, such low brain microcirculation, diminished brain capillary perfusion and secondary neural degeneration [7]. The increase in swelling caused by the insertion injury can also lead to the pinching of nearby blood vessels and capillaries, which in turn can result in secondary injury beyond the initial injury site. The insertion damage can be reduced by decreasing the size of the electrode, as the smaller the implanted device, the less pressure that will be applied to the surrounding brain tissue. For example, a small carbon fiber electrode has much better outcomes than the large microdialysis probes. Although smaller device size can lead to better results, it does not eliminate glial activation. Instead, proteins can be used in devices in order to encourage cell attachment and growth. In the future, targeting neuron survival may be more beneficial than just reducing glial activation [7].

As shown before, there are a plethora of different failure modes in the field of neuromodulation, such as unreliable signals, glial scar reactions and cell death. In order to create a standard procedure for identifying failure modes, the researchers at New Jersey Medical School have conducted a retrospective study on an experiment that was conducted in 1996 [8]. In this experiment, silicon electrode arrays were implanted in 27 monkeys, and all the electrode arrays were checked for failure. One of the most common reasons for failure is physical trauma during the insertion of the device [13]. The presence of multiple foreign bodies caused by insertion trauma can lead to numerous infections. As a result, a general effort has been made to mimic neurosurgical protocols and procedures, with an intent to reduce the inclusion of foreign materials.

In order to create categorization of different failures, a definition for “failure” must be established. Failure is simply the absence of extractable action potentials for all electrodes on an array [8]. Another way is to classify them into biological, material, mechanical and unknown. Biological failures refer to the foreign body response initiated by the tissue against the sensor. Whereas material failures describe design flaws and material degradation. Some instances of mechanical failure include physical factors that move the sensor from its desired location or

damage the hardware to a point where it stops recording. Unknown failures, on the other hand, are from signal loss not directly related to one of these methods. Another way to categorize failures is by either defining them as acute or chronic. Acute refers to the rapid loss of signals within the span of 7 days [8], whereas chronic failure is the slow progressive loss of neural signals over a longer period, ranging from a month to years. Acute failures are often caused by physical damage to the hardware or subject, like wire bundle disconnection or destructive removal of skull-based connectors. Another cause for acute failures can be rapid infectious diseases.

Of the 78 arrays used in this analysis, 62 arrays (79%) completely failed by the operational definition, 9 (12%) array experiments were electively terminated while the arrays were functional, and 7 (9%) arrays remained active at the time this study was closed. Acute failures were the most common as they comprised 45 array failures. Two of were biological failures, which were caused during the implantation procedure when a piece of gauze caught by the drill impacted the surface of the brain. The acute material failures were the most apparent to recognize, as they never recorded, from the start. This issue was likely caused by manufacturing defects and the shirting of connectors. The majority of acute failures was made up of acute mechanical failures, taking up 48.4% of the entirety of failures [8]. One cause for acute mechanical failures is surgical issues, which includes wire bundle disconnection, connector removal and internalization of the array. Another reason is trauma, caused by irreparable damage inflicted upon connector pins. The trauma subsection of acute mechanical failure also includes monkeys manipulating the skull mounted connector in some way. There were significantly less chronic failures than acute failures, as they only made up 27.4%. All of the chronic biological failures were caused by the meningeal encapsulation of the array and the extrusion from the cortex [8]. Chronic material failure was shown to be the progressive loss of signal by the shorting of connector pads. There were no chronic failures that were mechanical [8].

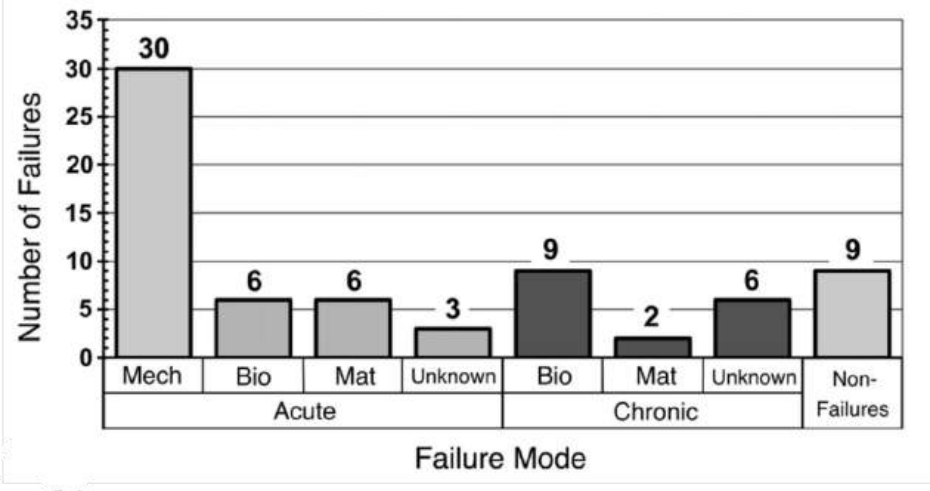


Figure 5 (Mat = material)

Figure 5 shows the number of arrays that failed, organized by failure modes. Acute mechanical failures were the most common cause. Another important aspect is that only 9 of the

71 (12.68%) arrays were unknown, reflecting the efficiency of this categorization. Impedance data was also recorded over the span of the experiment. The pre implant average impedance across all the electrodes was 305 kΩ. The results after the first two weeks showed that the average impedance more than doubled(2.48 ×) with a mean value of 755 kΩ. Across all arrays, the mean spike amplitude (PTP) for viable channels started at approximately 69 μV in the first week and rose to approximately 104 μV over the first month. The PTP amplitude trended downward over time, and reached a mean of 48 μV by the final recording session. Although impedance plays a role in predicting the capability of an electrode to record neural signals, many other features such as proximity to cells, cell size and geometry, tip geometry, and tissue reactions may affect recording [8]. In addition, the relationship between signal quality and impedance is weak.

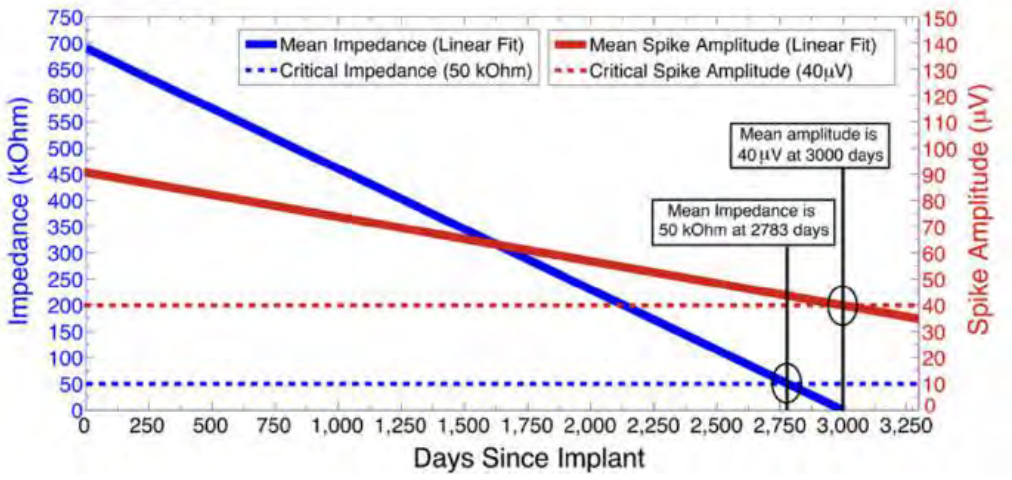


Figure 6. Figure 6 shows the rates of signal decay. This graph overlaps the models for impedance and spike amplitude (PTP), in order to compare their intercepts. The instances where the device has failed are marked by dashed lines. Mean impedance values that degrade to <50 kΩ suggest that signals are largely shunted to ground; spikes with amplitudes <40 μV are effectively impossible to separate from background noise. The different y intercepts could suggest multiple, interacting failure modes but the data suggest that the technology could function for up to eight years [8].

Future efforts

Companies need to work on improving the compatibility of their magnetic stimulators. Most companies focus on reducing the size of the electrodes, reducing invasiveness and other issues. However, there are many different configurations of neurons, all of them have different thresholds, resistances, and potential. A possible solution for this is the development of micromagnetic stimulation, which uses coils and can control neuron activity, block unwanted conductance as well as hyperactivity [4].

In order to progress further in this field, a more developed understanding of neural anatomy is key. The different organelles of neurons all have different functions and properties,

most of which either interfere or aid EMS. Therefore, it is crucial to understand how they work to make neuromodulation more effective. The dynamic state of neurons can be used to boost the efficiency of modulation. Preconditioning the neuronal tissues can prove to be very effective in seeing results of stimulation, like proving random noise stimulation to precondition the primary motor cortex, showing lower thresholds [4].

The current simulations of neuromodulators should start including neural dynamics. Most of the current simulations include the electrode, neural model and the electric field produced from the electrode. This framework assumes that the electric field is the same in every model, which is not the case [6]. This model also ignores the current produced by the membrane. There is an increasing significance in developing closed loop stimuli and making real time changes to the electrodes. New technology should provide real-time, closed-loop measurement of the brain's state in order to actively change and adjust the dynamic neutrons [3]. These can enhance recovery from neural injuries such as SCI. As μms has a similar effect to that of electric stimulation, this might lead to further research on expanding its electric field to be considerably larger than that of electric stimulation, as the number of spikes between the coil and cell increased. This shows the possibility of it working amongst glial scars, due to its enhanced range of electric field.

An important tool to study and improve neurotransmitters is microdialysis, which samples extracellular brain fluid. This operates on the technique of diffusion, where small molecules move across a concentration gradient from the extracellular base into the dialysis fluid inside the probe. A solution to counter cell encapsulation and improving signal sensitivity is neurovascular mapping. By creating a three-dimensional map using a two-photon microscopy, it can be used to guide the device insertion in a way that won't disrupt major vascular structures. As discussed earlier, mechanical strain on the tissue happens when the material softness of the device and tissue don't match, which leads to inflammation. Flexible devices are key to solving this problem, as they apply lower pressure on the chronic tissue. This in turn reduces inflammation, thereby improving the implant's integration. However, this might still pose problems, as during implantation, flexible devices are going to deflect while being directed to the target area.

Another solution to combat implantation injury is the shape of the device. While sharper tapered tips may be more mechanically fragile, they can reduce insertion related damage by allowing implantation without tissue compression during pial penetration. Alternatively, "open-architecture" or "lattice style" probes with reduced volume and surface area also reduce the tissue response even when the initial insertion injury is identical to a solid substrate design. While focusing on the integration of the device, one solution is to use soluble and surface bound biological molecules or pharmaceutical drugs to blur the line between biology and machine technology. Overall, there are many biomolecular approaches for improving tissue integration.

One method to improve tissue integration is releasing soluble cues to reduce inflammatory tissue response or attract neural processes [7]. Another solution in order to prevent inflammation and glial scars is to ensure that the glial cells do not disrupt the implantation

process. For this to take place, the microglia cells must stop functioning for the period of time it takes for the insertion to take place. This can be achieved through hematopoietic stem cell transplantation, which is a form of chemotherapy carried out in order to reset the human immune system. As microglia cells are a part of the immune system, this would be successful in temporarily shutting them off. The reset of the immune system through the hematopoietic stem cell transplantation is already known to be successful, as it is used in sclerosis. A drug that can be used to silence or reduce the functioning of glial cells for a period of time is Low dose Naltrexone.

Although there are multiple research labs and companies for neuromodulation, many researchers around the world agree on many aspects, like future points of interest, current choices for treatment and many.

Similarities amongst the experiments and methods of articles

Both the articles, Neuron Matters and Nanotransducers for wireless neuromodulation emphasize the invasiveness of implementing nanotransducers and electrodes into the body. This is because bulky devices can lead to glial scars in the brain. These glial scars can also provide impedance against the electrodes. The invasive electrodes also cause neuroinflammation. Both articles put forward the solution of minimizing the space that the device occupies in the brain, i.e. making it smaller, as this decreases the amount of neural tissue destroyed while inserting the electrode [3, 4].

Another possible solution is to implement more of neural dynamics, which is a form of computer simulation that includes a number of aspects that are important in neuromodulation such as the electric field, and the current produced. One such application is the FEM (Finite Element Method) [6].

There is also a need to understand and develop energy sources required for the device to function. This includes acoustic devices, light sources and magnetic systems, which are relatively large and need to decrease in size. [3]. The struggle of bypassing the human nervous and immune system in order to implant the device is a major issue that can lead to numerous problems like inflammation, glial scars and many more, which has been emphasized by many researchers who aim to introduce less destructive methods of implementation, in order to prevent the immune system from attacking the device [7, 9].

The discussion about the functions and advantages of magnetic stimulation is prevalent in many articles, as it is possibly the best medium for neuromodulation. There are many advantages of magnetic stimulation over others. They are comparatively non-invasive, as the microcoils are much smaller than the electrodes in electric stimulation. They provide much better comfort to the patient too. Stimulators that use electric stimulation interfere with other medical equipment in use, such as MRI scanners as the electric field interferes with the scanner. But with neuromodulators that make use of magnetic devices can be turned off, as the electric field they use can be disabled by deactivating the magnets. The articles also declare Transcranial Magnetic Stimulation (TMS) much more inefficient than the methods they propose; which is the use of

micro-coils. TMS uses large handheld devices that require the patient to stay in the clinic for 2-3 days. In addition, it is also extremely expensive. In contrast, the micro-coils are implanted into the patient, making it much smaller. They also can give an opportunity to create closed loop systems, where the device takes data from the body and makes adjustments based on that, which is not possible with TMS [3, 4, 6].

Conclusion

Neuromodulation is a growing field of medical treatment and provides a solution for many neurological diseases such as Parkinson's disease, Alzheimer's disease and physical trauma such as spinal cord injury. Neuromodulation can be life changing for many;. It provides aid to some of the most quintessential components of the body: the spinal cord and the brain, either through SCI, micromagnetic stimulation, epidural electric stimulation, or any other form of neuro transducers. As it is still a growing field with new discoveries made all the time, it is inevitable for downsides to be present. This includes inflammation during implantation, formation of glial scars, lack of signal quality. With advancements – such as the use of microcoils, smaller devices and the use of pharmaceutical drugs – some of the problems have been solved, such as MRI incompatibility and less damage to the blood brain barrier. However, more work needs to be done to solve additional issues, and this can only be achieved through expanding on the knowledge of the interaction between the cells and the implant, the effect of the electric fields and many more, requiring the implementation of not only biology but also physics, biochemistry, computer science and many more subjects into the field of neuromodulation.

Work Cited

[1] Choi, Elliot et al. (15/12/2021). Epidural Electrical Stimulation for Spinal Cord Injury. https://doi.org/10.4103/1673-5374.313017. Retrieved: 04/05/2024

[2] O'Higgins, Madeleine et al. (09/05/2017). What Is Spinal Cord Injury? https://kids.frontiersin.org/articles/10.3389/frym.2017.00017/full Retrieved: 06/07/2024

[3] Li, Xiuying et al. (05/05/2021). Nanotransducers for Wireless Neuromodulation. https://doi.org/10.1016/j.matt.2021.02.012. Retrieved: 15/04/2024

[4] Hui Ye et al. (03/11/2022). Neuron Matters: Neuromodulation with Electromagnetic Stimulation Must Consider Neurons as Dynamic Identities. https://doi.org/10.1186/s12984-022-01094-4. Retrieved: 28/04/2024

[5] Unknown. N/A. Transcranial Magnetic Stimulation (TMS) to Treat Symptoms of Depression https://www.americastms.com/blog/transcranial-magnetic-stimulation-tms-to-treat-symptoms-of-depression. Retrieved: 10/07/2024

[6] Bonmassar, Giorgio et al. (26/06/2012). Microscopic Magnetic Stimulation of Neural Tissue. https://doi.org/10.1038/ncomms1914. Retrieved: 15/06/2024

[8] Barrese, James C et al. (12/11/2013). Failure Mode Analysis of Silicon-Based Intracortical Microelectrode Arrays in Non-Human Primates. https://doi.org/10.1088/1741-2560/10/6/066014. Retrieved: 01/06/2024

[9] Polikov, Vadim S. et al. (27/09/2024). Response of brain tissue to chronically implanted neural electrodes. https://doi.org/10.1016/j.jneumeth.2005.08.015. Retrieved: 28/05/2024

[10] Angeli CA et al. (07/04/2024). Altering spinal cord excitability enables voluntary movements after chronic complete paralysis in humans. https://doi.org/10.1093/brain/awu038. Retrieved: 15/04/2024

[11] Montgomery, E. B. Jr, Gale, John T. (27/06/2007). Mechanisms of action of deep brain stimulation(DBS). https://doi.org/10.1016/j.neubiorev.2007.06.003. Retrieved: 27/06/2024

[12] Szarowski, D. H et al. (05/09/2003) Brain responses to micro-machined silicon devices. https://doi.org/10.1016/S0006-8993(03)03023-3. Retrieved: 20/05/2024

[13] Bjornsson CS et al. (21/06/2006). Effects of insertion conditions on tissue strain and vascular damage during neuroprosthetic device insertion. https://doi.org/10.1088/1741-2560/3/3/002. Retrieved: 30/04/2024

[14] Slopsema Julia P. et al. (05/09/2018). Clinical deep brain stimulation strategies for orientation-selective pathway activation. https://iopscience.iop.org/article/10.1088/1741-2552/aad978/meta. Retrieved: 07/06/2024

The Psychology and Effects of Workplace Ageism By Abhimanyu Shridhar

Abstract

Workplace ageism is a major issue in the workplace that affects people of all ages and is mostly unrecognized in its effect on different workplace factors. This research examined how ageism impacts workplace culture, productivity, and intergenerational perceptions. A 29-question survey was conducted with 31 respondents aged between 31-60+ from California-based companies and LinkedIn networks to understand their experiences with ageism and work environment preferences.

The statistical methodologies employed, encompassing Chi-square tests and sentiment analysis alongside Google Forms data, indicated that 16% of participants encountered ageism within their work environments, and 26% acknowledged its effect on allocating tasks among team members, thereby influencing overall productivity. Negative perceptions were further shown by stereotypes regarding the technological competencies of older employees; younger workers assessed their older counterparts as less efficient, although a significant number showed strong interest in gaining knowledge from them. Although more neutral to learning from younger colleagues, older workers also showed interest. Concerns regarding ageism also varied across age groups; 70% felt comfortable voicing their concerns, but those aged 41-50 demonstrated greater uncertainty, while the 6% who were uncomfortable were in the younger age range of less than 50 years. Experiences of ageism were associated with mixed preferences for hybrid versus in-person work arrangements. Notably, 93% of respondents felt that the workplace was open to all ages, and there was less certainty about whether ageism affected prospects for promotion, showing that some workplace culture factors may not be as influenced by age discrimination. Although the hypothesis that ageism negatively impacts perceptions and workplace interactions was only partially validated, the results indicate that it affects the allocation of tasks and relationships between different generations. Limitations included a small sample size, which points out the need for further study with a larger population.

Introduction

Among the various forms of discrimination in our society, ageism—which involves stereotyping people based on their age—remains one of the most overlooked issues today (United Nations, 2021). Particularly in the workplace, age discrimination is prevalent yet has not been researched as much as other recognized forms of bias, such as racism and sexism. Firzly et al. (2021) stated that "…research on ageism lags behind when compared with research conducted on other types of prejudice." Tragically, ageism "is tolerated or — even worse — unrecognized for what it truly is: discrimination, plain and simple" (Kita, 2019).

This issue affects individuals across all generations, but it particularly impacts older workers. An AARP survey from 2019 revealed that "nearly 1 in 4 workers age 45 and older have been subjected to negative comments about their age from supervisors or coworkers," with "76 percent of these older workers viewing age discrimination as a hurdle to finding a new job…"

(Kita, 2019). Choi-Allum (2022) found that "about two in three adults ages 50-plus in the labor force (64%) think older workers face age discrimination in the workplace today." In a more recent survey by SHRM in 2023, "nearly a third (30 percent) of U.S. workers report feeling unfairly treated due to their age at some point in their careers" (SHRM, 2023).

Such findings show the need to address age discrimination, as the workforce is essential for our society. The Bureau of Labor Statistics reported that in 2023, civilian labor participation rates for those aged 45-54 and 55+ were 82.1% and 65.8%, respectively, with projected rates for 2033 at 82.0% and 69.1%. Meanwhile, participation rates for younger generations aged 25-34 and 35-44 were 83.8%, expected to decrease only slightly to 82.8% and 82.0% in 2033 (Bureau of Labor Statistics, 2023). This data shows the important role all generations play in the workforce, yet age discrimination can stifle the contributions they make. Additionally, current research highlights that societies that value their elders experience higher life expectancies (Weir, 2023).

In response to ageism in multiple environments, intergenerational programming has been utilized, encouraging meaningful interactions between younger and older generations. Recent trends show a growing separation between these groups, particularly in living arrangements and recreational activities (Kaplan, 2001). Intergenerational programming has been shown to "diminish ageist stereotypes, strengthen communities, and lead to improved services for children, youth, and older adults" (Kaplan, 2001). Firzly et al. (2021) surveyed 612 undergraduates and found that "perception of the workplace intergenerational climate as well as knowledge-sharing practices directly and positively influence job satisfaction." Similarly, research by Pillemer et al. (2022) has demonstrated that intergenerational learning in high schools results in positive attitudes toward elders among youth.

Despite notable research on the impacts of ageism and the effectiveness of intergenerational learning, there are multiple gaps. Many current studies primarily focus on the discrimination faced by older people, often placing less focus on the perceptions of younger generations. There is little exploration of how different age groups view one another, with most research centered on the experiences of older adults, who are generally categorized as 50 or older. Additionally, systemic ageism, which involves policies inspired by age discrimination, has been thoroughly investigated, but less attention has been placed on the impacts of age discrimination on individual behaviors and preferences. Insights for these behavioral aspects could improve work environments and tailor intergenerational programs more effectively to combat ageism.

The outcomes of intergenerational programs have been observed, yet research on their implementation within workplaces is little. A critical question remains: do diversity programs or intergenerational programming effectively mitigate the impacts of ageism in the workplace?

Given this context, the purpose of this research is to examine ageism in terms of its systemic impacts while analyzing the psychology of workplace age discrimination across various age groups. This study aims to discover the perspectives of multiple generations concerning ageism, assess its prevalence, and evaluate the effectiveness of intergenerational learning and

diversity programs in mitigating age discrimination. Based on current research on age discrimination, it is hypothesized that survey data will reveal that ageism negatively impacts productivity, work culture, and intergenerational perceptions in the workplace.

Methods

This study aimed to collect insights from a professional network of employees on workplace ageism through a survey distributed via email and LinkedIn. As a resident of California, I made a list of 450 companies within California for accessibility, and they were compiled using Generative AI. This list was narrowed to 100 companies based on their likelihood of responding. Emails were sent to human resources departments or general information addresses, requesting employees to participate. Additionally, the survey was shared through a public LinkedIn post, leveraging the platform's professional network to enhance response diversity. These dual approaches ensured a broader demographic reach. Surveys were distributed in early August, with a completion deadline of November 15.

The final sample consisted of 31 participants: 18 males and 13 females. Age group distribution was as follows: 6 respondents aged 31–40, 14 aged 41–50, 9 aged 51–60, and 2 aged 60+. The majority worked in industries such as technology, communications, and biotechnology. The 29-question survey included 21 multiple-choice questions and 8 open-ended responses. Likert scales were used to measure ageism's effects, diversity program effectiveness, and perceptions of generations, while linear scales assessed job satisfaction, perception scores, and intergenerational learning interest. To maintain ethics, all respondents' or companies' identities were made anonymous when filling out the Google Form.

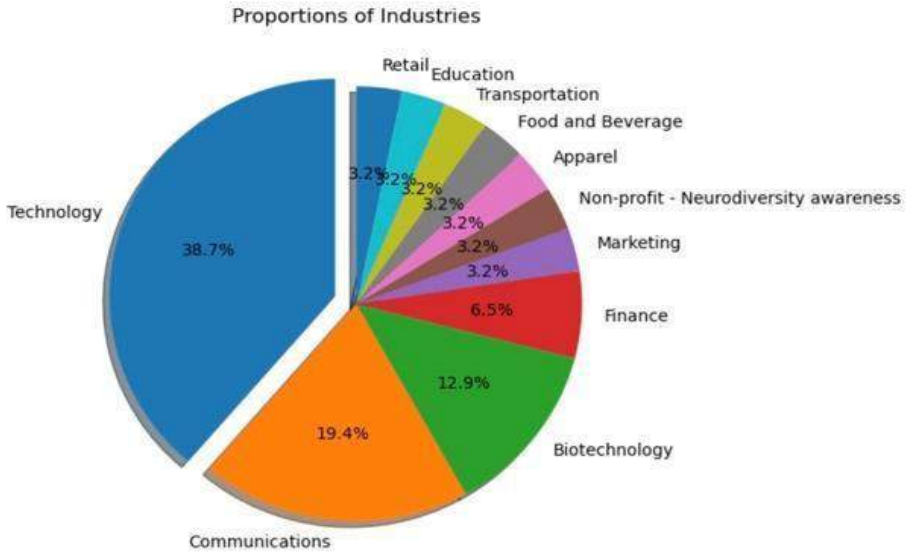


Figure 1. Proportions of the participants' industries they work in

This survey was influenced by prior studies: Pillemer et al. (2022) analyzed intergenerational learning interventions, Firzly et al. (2021) observed workplace ageism and its effects on job satisfaction and ageism awareness, and Choi-Allum (2022) explored the

prevalence of age discrimination among older employees. Unlike these studies, this project assessed ageism across generations, examining correlations with preferences and perceptions. Components from previous research mentioned, including ageism awareness and job satisfaction, were implemented in the survey too.

Google Forms data was first processed to calculate proportions and trends. Further analysis was conducted using Python to find correlations. TextBlob, a natural language processing library, was used to quantify sentiment in open-ended responses for interest in learning and perception, generating scores from -1 (negative) to 1 (positive). Responses marked "not applicable" were excluded to ensure accuracy. To explore the strong relationships in the dataset, Python's Itertools module generated all possible combinations of categorical variables in the dataset. This approach allowed for the identification of combinations by applying Chi-square tests from Python's SciPy module to the combinations to find the notable correlations ($p < 0.05$), such as whether a respondent experienced ageism versus their industry.

Significant correlations relevant to the hypothesis were visualized using Matplotlib and Seaborn to understand the associations better. These visualizations, alongside statistical analysis, helped to analyze the effects of age discrimination on various factors, like workplace dynamics.

Results

The average job satisfaction score was 8, and almost all respondents said that ageism didn't affect their job satisfaction if it was low (<7). That said, first, respondents aged 50+ were asked to rate their perception of respondents aged <50, and vice versa. Then, they were asked to rate based on their interest in learning from one another. Figures 2 and 3 show the perception score distributions and figures 4 and 5 show the interest score distribution.

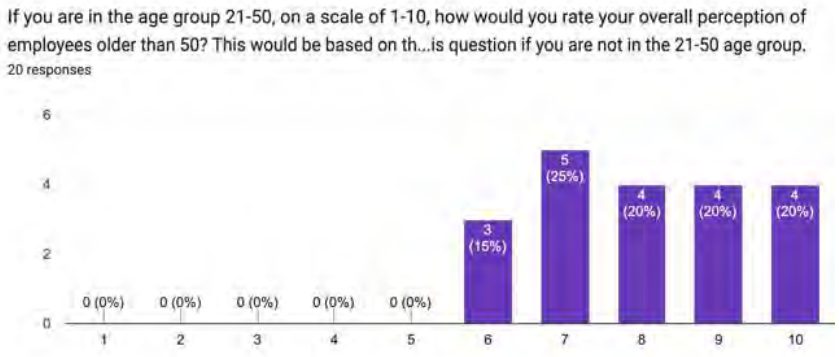


Figure 2. Perception score of respondents aged <50 years old toward people aged 50+ years old

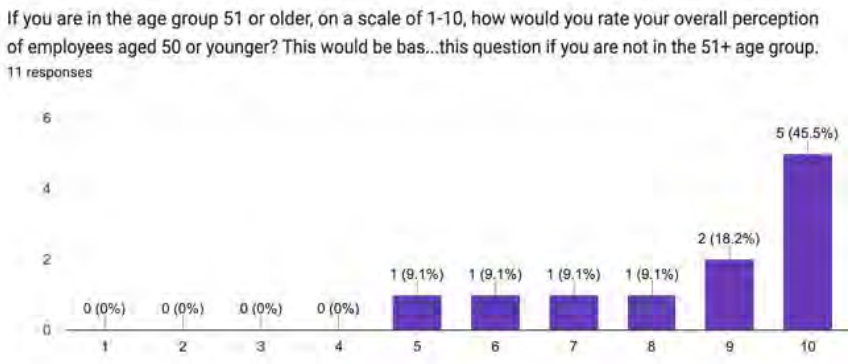


Figure 3. Perception scores of respondents aged 50+ years old toward people aged <50 years old



Figure 4. Distribution of interest scores from respondents aged 50+ years old toward people aged <50 years old

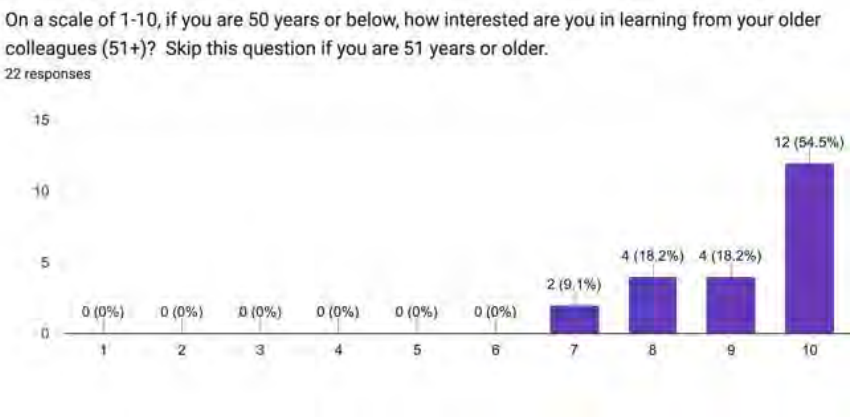


Figure 5. Distribution of interest scores from respondents aged <50 years old toward people aged 50+ years old

Open-ended responses for the questions representing figures 2-5 showed a trend where older generations thought younger generations needed more guidance, and younger generations thought that older workers' methods were sometimes outdated. Though, both respected each other for their novel methods and expertise.

Afterward, respondents were questioned as to whether they experienced ageism (Figure

6), are they comfortable with voicing concerns about ageism (Figure 7), and whether ageism affects the distribution of tasks (Figure 8) and career advancement (Figure 9). Also, they were asked whether they received diversity training for ageism (Figure 10), how effective it was (Figure 11), and whether all age groups were welcomed in the workplace (Figure 12).

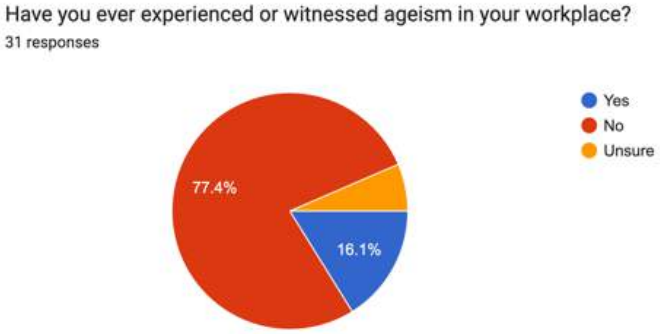


Figure 6. Distribution of respondents who witnessed or experienced ageism

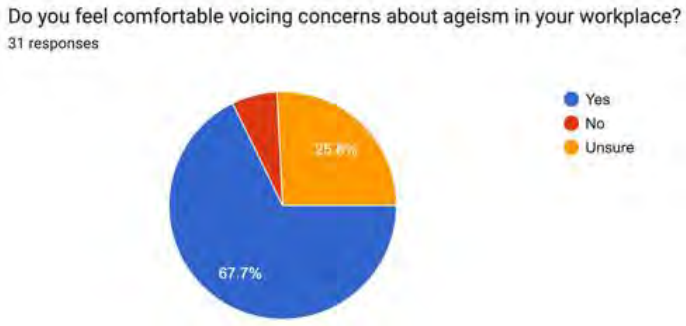


Figure 7. Distribution of respondents based on how comfortable they were about voicing concerns about ageism

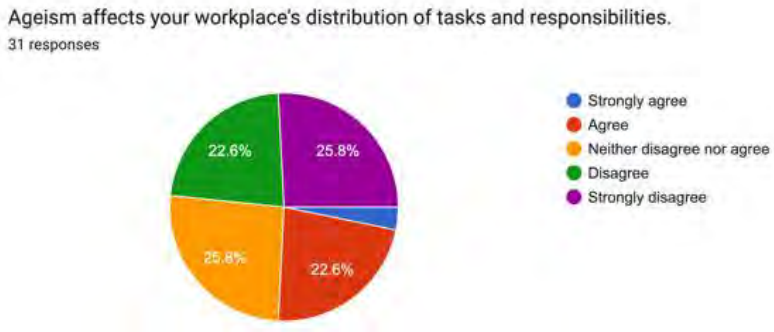


Figure 8. Distribution of respondents based on whether they think ageism affects the distribution of tasks

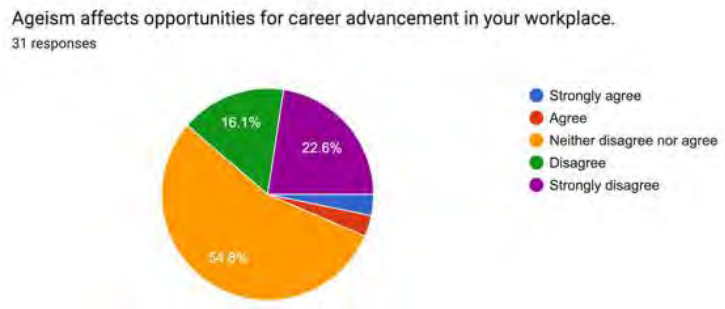


Figure 9. Distribution of respondents based on whether they think ageism impacts career advancement



Figure 10. Distribution of respondents on whether they've received training related to ageism

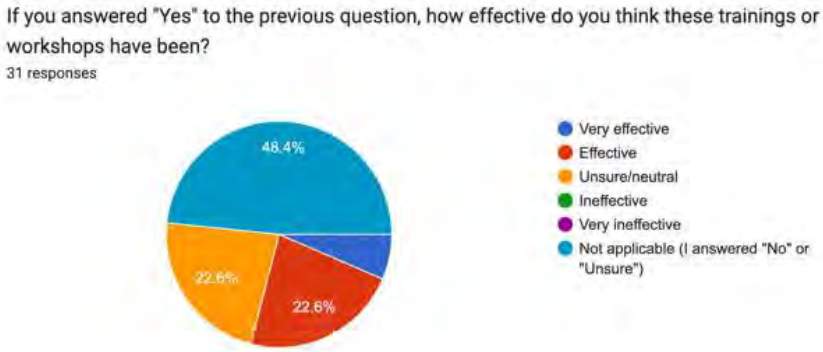


Figure 11. Distribution of respondents based on whether they thought the diversity programs were effective

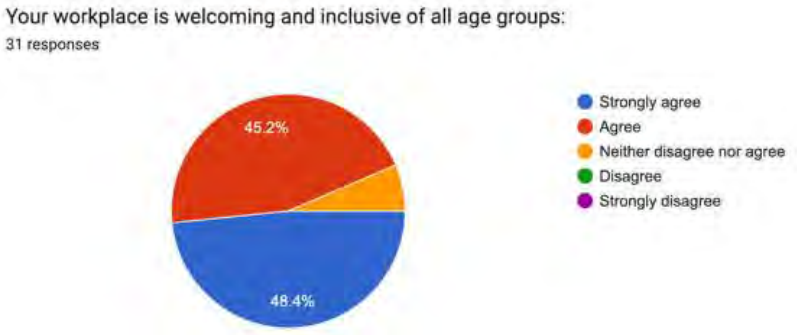


Figure 12. Distribution of respondents on whether they think their workplace is welcoming and inclusive of all age groups

Then, Textblob was used to quantify the open-ended justifications by sentiment score as to why respondents selected their scores from Figures 2-5, as shown in Figure 13.

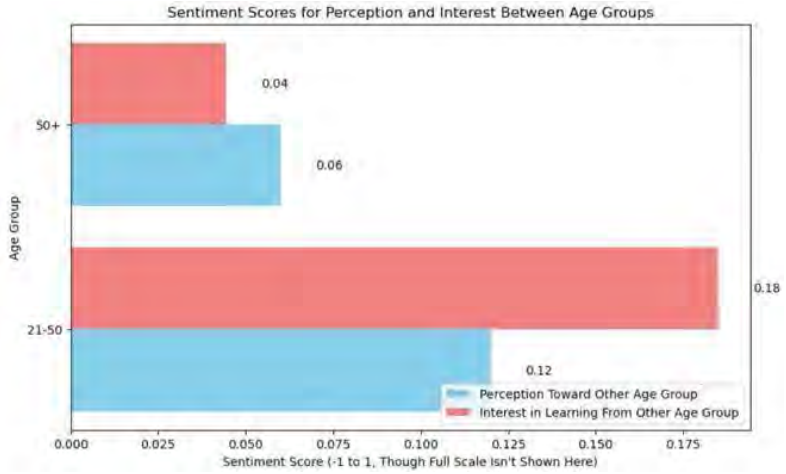


Figure 13: Sentiment scores of each age group's open-ended responses in justifying answers

Finally, a program utilizing SciPy iterated through the combinations made by Itertools and extracted the significant correlations ($p < 0.05$). The correlations relevant to the hypothesis were then extracted. There was a significant correlation between whether a respondent experienced ageism and their workplace preference (Figure 14). Additionally, age group and comfort in voicing concerns had a notable correlation (Figure 15).

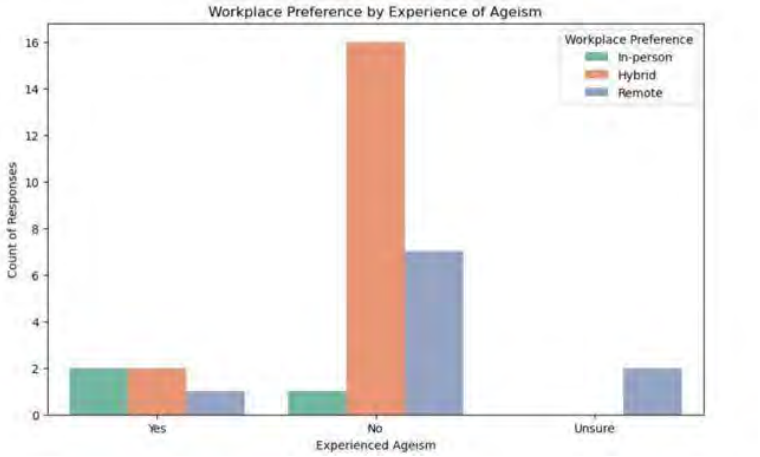


Figure 14: Bar graph visualizing the correlation between whether a respondent experienced ageism and their workplace preference

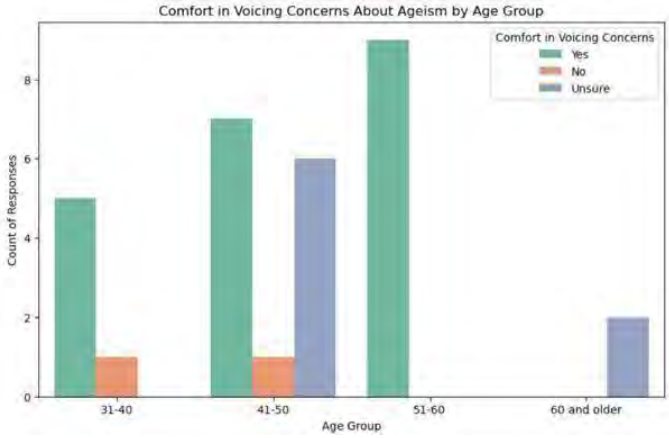


Figure 15: Bar graph visualizing the correlation between whether a respondent experienced ageism and their workplace preference

Discussion

The findings indicate that while ageism exists and significantly influences workplace culture and perceptions, its impact may not be as major as initially hypothesized. As shown in Figure 2, younger workers had a mean view score of approximately 8 regarding older colleagues' efficiency and ability, with a mode of 7. In comparison, older employees rated younger staff with a mean score of approximately 8.5 and a mode of 10 (Figure 3). This difference shows that some younger staff may perceive older workers as less efficient, potentially due to concerns about outdated ideas, as expressed in open-ended responses. However, both age groups expressed a strong interest in learning from one another, as proven by modes of 10 in Figures 4 and 5. Additionally, some older workers noted in open-ended responses that younger colleagues may need some workplace guidance. This view about older employees struggling with new technology may influence overall perceptions of age-related competence, although younger workers also experience discrimination, albeit to a lesser extent. Notably, responses underscored

the importance of younger workers' new ideas alongside older employees' expertise. Pillemer et al. (2022) found that perceptions improved following intergenerational education, suggesting that tailored training programs could help mitigate age-related stereotypes.

Figure 6 revealed that about 16% of participants either experienced or witnessed ageism, while nearly 25% felt that it influenced task distribution (Figure 8). Choi-Allum (2022) also found a significant number of employees affected by ageism, though notably more than the finding in this study. Furthermore, about 50% neither disagreed nor agreed as to whether ageism impacted career advancement, with almost 40% disagreeing or strongly disagreeing (Figure 9). These findings highlight the prevalence of ageism in the workplace, which may hinder productivity, particularly if task distribution is affected, though career advancement may not be as impacted.

Approximately 70% of participants reported being comfortable voicing issues about ageism, and about 25% remained uncertain (Figure 7), suggesting that a notable number may either not encounter ageism in their environments or are bound by organizational barriers, which could affect work culture. Firzly et al. (2021) similarly noted that awareness of ageism remained relatively unchanged despite initiatives for intergenerational education, potentially due to a lack of exposure to age-related biases. Similarly, the 25% who were unsure could be linked to a lack of exposure to ageist behavior.

Regarding intergenerational learning, Figure 10 conveyed that respondents were almost evenly split on the efficacy of diversity programs for addressing age discrimination. Figure 11 showed that approximately 50% of those who received these programs (25% of all participants) indicated some positive impact. These findings suggest that ageism may not be recognized enough, as shown by the lack of exposure to diversity programs. The uncertainty of efficacy may suggest that ageism is not common in certain workplaces and therefore does not require special attention despite the diversity programs executed. Figure 12 supported this idea, showing that about 93% of participants agreed that individuals of all ages are welcome in the workplace. Sentiment analysis indicated that younger workers showed a subtly higher interest in learning from older colleagues compared to vice versa, with similar perception sentiment scores across age groups (Figure 13). This trend highlights enthusiasm among younger employees for the expertise offered by experienced colleagues, while older staff maintained a more neutral but interested stance toward learning from younger workers. Firzly et al. (2021) found that knowledge sharing promotes a positive work culture, a sentiment echoed by Pillemer et al. (2022). This feeling may be prevalent among all age groups of employees, as shown by the keen interest in learning from each other.

Correlation analysis showed a strong link between experiences of ageism and workplace preferences (Figure 14). Those who did not face ageism preferred remote or hybrid work, likely due to its flexibility. Conversely, those who experienced ageism displayed mixed preferences, possibly seeking a balance of hybrid flexibility and in-person visibility to support career advancement even with discrimination. This indicates that certain ageist work cultures may affect preferences for workplace type.

Another notable link was identified between age group and comfort in voicing concerns (Figure 15). Participants aged 60 and older, along with many in the 41-50 age group, expressed uncertainty about voicing issues, which may suggest that ageism is either infrequent or inadequately addressed in their environments, regardless of age. Although only approximately 6% of all participants expressed discomfort, all were under 50, indicating that ageism could also affect younger workers.

The study's limitations include a small sample size of 31, impacting the ability to generalize findings. Additionally, there was an uneven balance among age groups and industries, with no participants in the age range of 21-30, which may have affected the findings. Furthermore, the cross-sectional survey design limits the ability to imply causation over the correlations. Future research should focus on larger, more balanced samples and longitudinal survey designs to observe ageism's influence on workplace culture, views, and preferences. Additionally, Python modules, particularly Textblob, shouldn't be relied upon fully to generalize the results of feelings from respondents. Implementing intergenerational learning programs across various settings and measuring the impact, similar to the approaches taken by Pillemer et al. (2022) and Firzly et al. (2021), could improve understanding and reduce age-related stereotypes.

Conclusion

This study addressed the need to recognize age discrimination's negative impact on workplace culture, productivity, and intergenerational perceptions. The findings partially supported this hypothesis, showing that ageism impacts task distribution and perceptions of efficiency, notably for older individuals. Stereotypes about outdated methods contributed to ageist views, though mutual respect between generations was present, with both younger and older respondents eager to learn from each other.

Ageism was observed to influence task distribution by nearly 25% of respondents, supporting part of the hypothesis that it affects productivity. However, its impact on career advancement was shown to be less prominent, with 50% of respondents uncertain about its role in ageism. Encouragingly, 93% of participants agreed that individuals of all ages feel welcome in their workplaces, suggesting that while ageism exists, it does not fully mean there is no inclusivity. These results demonstrate that age discrimination impacts workplace culture in subtle ways but also shows opportunities to promote intergenerational understanding and collaboration.

These findings have practical implications, such as designing intergenerational learning sessions and raising awareness about ageism. To further combat ageism, organizations should implement more policies that promote inclusivity and mentorship. While limited by a small sample size, this study offers valuable insights for future research into ageism's impacts in various sectors. Ultimately, addressing age discrimination is essential for enhancing workplace culture and supporting diversity and inclusion efforts.

Acknowledgments

To start, I would like to show gratitude to my father for his assistance in gathering data for this study, particularly through his aid in dispersing the questionnaire on LinkedIn. His support was essential in compiling an adequate dataset that provided valuable insights.

I am also thankful to Mr. Dworzak, my science teacher, for his guidance on data collection methods during challenging moments.

Additionally, I would like to express my appreciation to the representatives of the companies I contacted through email who took the time to participate in the surveys. Your contributions significantly contributed to this research. I extend my gratitude to the respondents from my LinkedIn post and those from the companies who provided critical insights regarding age discrimination.

Lastly, a special thank you to Mrs. Bui, my English teacher, for her assistance in ensuring clarity in the text and for providing helpful suggestions that enhanced the overall quality of this paper.

Works Cited

Choi-Allum, L. (2022). Age discrimination among the 50+: Annotated questionnaire. https://doi.org/10.26419/res.00545.001

Firzly, N., Van de Beeck, L., & Lagacé, M. (2021). Let's Work Together: Assessing the Impact of Intergenerational Dynamics on Young Workers' Ageism Awareness and Job Satisfaction. Canadian Journal on Aging / La Revue Canadienne Du Vieillissement, 40(3), 489–499. https://doi.org/10.1017/S0714980820000173

Kaplan, M. (2001, January 5). Some what's and why's of Intergenerational Programming. Department of Agricultural Economics, Sociology, and Education. https://aese.psu.edu/outreach/intergenerational/curricula-and-activities/handouts/factsheets/some-whats-and-whys-of-intergenerational-programming

Kita, J. (2019, December 30). Workplace age discrimination still flourishes in America. AARP. https://www.aarp.org/work/age-discrimination/still-thrives-in-america/

Pillemer, K., Nolte, J., Schultz, L., Yau, H., Henderson, C. R., Jr., Cope, M. T., & Baschiera, B. (2022). The Benefits of Intergenerational Wisdom-Sharing: A Randomized Controlled Study. International Journal of Environmental Research and Public Health, 19(7), 4010. https://doi.org/10.3390/ijerph19074010

SHRM. (2023, December 12). New SHRM research details age discrimination in the workplace. SHRM. https://www.shrm.org/about/press-room/new-shrm-research-details-age-discrimination-workplace

United Nations. (2021, March 18). World losing billions annually to age-based prejudice and discrimination: UN report | UN News. https://news.un.org/en/story/2021/03/1087672

U.S. Bureau of Labor Statistics. (2023). Civilian labor force participation rate by age, sex, race, and ethnicity. https://www.bls.gov/emp/tables/civilian-labor-force-participation-rate.htm

Weir, K. (2023, March 1). Ageism is one of the last socially acceptable prejudices. Psychologists are working to change that. Monitor on Psychology, 54(2). https://www.apa.org/monitor/2023/03/cover-new-concept-of-aging

Neural Mechanisms of Trauma and Anxiety Disorders: Insights from DSM-5 Classifications for Advancing Treatment Strategies By Euhyun Brooke Jahng

Introduction

Anxiety disorders and trauma, which are classified in the DSM-5, are supported by distinct neural mechanisms (American Psychiatric Association, 2022, p. 265). These prevalent mental illnesses in the United States are often diagnosed based on symptom clusters rather than a comprehensive understanding of their underlying pathophysiology and individual neurobiological variations (Brehl et al., 2020).

Post-traumatic stress disorder (PTSD), which is classified as trauma in DSM-5, is a chronic condition with significant physiological and psychological manifestations and is frequently assessed through structured clinician interviews or self-reports. Self-reports contain a range of symptoms and behaviors as patients cope with their traumatic experiences; common actions include: "anhedonic and dysphoric symptoms, externalizing angry and aggressive symptoms, or dissociative symptoms" (American Psychiatric Association, 2022, p. 265). These assessments consistently predict various cognitive, behavioral, and emotional health outcomes. Research suggests that immune system dysregulation, particularly the role of inflammation and its impact on brain function, may play a significant role in the pathophysiology of PTSD (Kim D. et al., 2020). This has been linked to anatomical changes in the brain, such as reduced cortical complexity in frontal, parietal, and temporal regions and increased glial activation in those with more severe PTSD symptoms (Invernizzi et al., 2023). For instance, the long-term effects of PTSD following the 9/11 attacks are evident, with approximately 23% of World Trade Center (WTC) responders still experiencing chronic PTSD twenty years after the incident. In the field of PTSD, "Posttraumatic stress disorder (PTSD) with delayed expression (also known as delayed PTSD or delayed-onset PTSD) is a diagnostic category that applies to people who first meet the criteria for a PTSD diagnosis at least six months following exposure to a traumatic event" (Smid et al., 2022).

Anxiety disorders, on the other hand, encompass disorders characterized by excessive anxiety and are accompanied by behavioral problems (American Psychiatric Association, 2022, p.189). Unlike fear, anxiety is an anticipation of future threats. Anxiety disorders differ in terms of the objects or situations that cause dread, anxiety, or avoidance behavior, as well as the related cognitive conceptualization. Thus, while anxiety disorders are often highly comorbid, they can be distinguished by a detailed study of the types of circumstances dreaded or avoided, as well as the content of the accompanying thoughts or beliefs. Anxiety disorders reach out to encompass other branches, including selective mutism and phobias, and repetitive symptoms can result in responses such as panic attacks. Anxiety disorders are characterized by increased inherent amygdala hypersensitivity, low prefrontal control (pFC), and high locus coerulus (LC) drive. Among these factors, the amygdala is specifically responsible for emotions and activating feelings of fear, which is a human instinct to protect oneself. Hyperactivity is commonly shown in individuals with anxiety disorders. Crucially, regulatory core hubs, PFC and LC, play a

significant role in regulating this response; "patients with PTSD revealed increased activation in the dorsolateral and ventromedial PFC" (Brehl et al., 2020, p. 729).

This paper is a compilation of different studies to aid the understanding of the research question: What are the neural mechanisms underlying trauma and anxiety disorders as classified in the DSM-5, and how can understanding these mechanisms inform the development of effective treatments and interventions? As PTSD and anxiety disorders are relatively common mental disorders, several research studies using neuroscientific approaches such as neuroimaging, neurophysiology, and neurogenetics have been conducted to explore interventions. This essay will also use the literature and keywords from various research papers to further expand on this topic and answer the research question.

Neural Mechanisms Underlying Trauma (Trauma and Stressor related disorders)

Post-traumatic stress disorder, also known as PTSD, is a prevalent trauma-related disorder throughout the world (American Psychiatric Association, 2022). The diagnosis criteria for PTSD include having experienced a traumatic event, intrusive symptoms, avoidance, negative cognitive and affective changes, alterations in arousal and reactivity, and the duration of symptoms that impede functioning. In the recent article "Labels Matter: Is it Stress or Is it Trauma?" Ritcher and Sandi discuss the importance of precise terminology in the field of stress and trauma research (2021). The article proposes a conceptual framework that differentiates between stress and trauma based on the consequences of exposure to challenging experiences. The authors argue that the terms "stress" and "trauma" are often used interchangeably, which can lead to confusion in understanding the distinct neural mechanisms and outcomes associated with each.

Resilience and stress are linked; most people who are exposed to challenging situations do not go on to develop mental health disorders (Ritcher and Sandi, 2021). Most stress-related reactions are adaptive, assisting people in managing difficult situations. Contrarily, trauma describes events that have pathological effects and impair a person's capacity to deal with difficulties in the future. Understanding this difference is essential to comprehending the complex neural mechanisms at play.

The piece contends that the brain processes engaged by traumatic events that result in psychopathology, such as PTSD, vary from those activated by stressful situations that do not result in such results as well (Ritcher and Sandi, 2021). This dichotomy is critical for understanding how certain events promote resilience while others result in long-term pathological alterations. The term "pathological metaplasticity" (Sandi, 2021, p.3) is used to characterize the long-term neurological alterations that occur in the brains of people who acquire trauma-related diseases. These alterations contrast with the adaptive plasticity in those who successfully deal with stress. By differentiating between stress and trauma in research, scientists may better identify and examine the exact brain pathways that contribute to resistance or sensitivity to trauma. This strategy is required to provide more tailored and effective therapies for trauma-related illnesses.

Moreover, understanding the neurological aspects of mental health disorders is essential. The article "Genetic and Neuroimaging Approaches to Understanding Post-Traumatic Stress

Disorder" is a comprehensive discussion of the neurobiological mechanisms underlying PTSD, with an emphasis on genetic and neuroimaging discoveries (Nisar et al., 2020). There are various networks of brain regions: amygdala, hippocampus, and prefrontal cortex. These regions are crucial for memory, emotional control, and decision-making, and anomalies in them have been associated with PTSD symptoms.

The hippocampus, in particular, takes a significant role in regulating anxiety– especially in emotional behaviors and PTSD (Ghasemi, M. et al., 2022). In a more profound factor, structural brain changes, especially in the hippocampus and amygdala, have been regularly seen in PTSD patients. These alterations are considered to contribute to the difficulty in-memory processing and heightened fear reactions associated with PTSD. Furthermore, PTSD patients exhibit structural alterations in the dorsolateral prefrontal cortex (DLPFC) and anterior cingulate cortex (ACC), which are associated with compromised emotional control and cognitive performance.

On the other hand, functional brain changes are also shown. Studies on functional neuroimaging in PTSD patients show changes in brain activity patterns and connections, especially in the amygdala, which is connected to elevated fear and anxiety reactions (Ghasemi, M. et al., 2022). A disturbance in the regular control of fear and stress reactions is suggested by an unusual connection between the amygdala and brain areas involved in emotion regulation, such as the prefrontal cortex, as shown by resting-state fMRI (Functional MRI is a type of MRI scan that can show which areas of your brain are most active) studies. Task-based fMRI studies further demonstrate that when trauma-related stimuli are displayed to people with PTSD, there is hypoactivation in brain regions linked to cognitive control and hyperactivation in brain areas connected to fear processing.

PTSD is highly vulnerable to specific genetic factors (Jiang et al., 2019). Certain genetic variations may heighten PTSD vulnerability after trauma exposure. For example, changes in the stress hormone-regulating FKBP5 gene have been connected to changes in the structure and function of the brain in PTSD. The brain's structural integrity and functional response to stress and fear are both impacted by other genes that are linked to PTSD, including those that encode serotonin transporters and dopamine receptors. The complicated interplay between genes and trauma in influencing the brain's stress response is highlighted by the connection between these genetic elements and environmental stresses, which leads to the neurobiological abnormalities observed in PTSD.

When PTSD is triggered, it involves complex interactions between genetic, neurotransmitter, and neuroendocrine factors (Jiang et al., 2019). The disruption of neurotransmitters in PTSD leads to feelings of anxiety, hypervigilance, and emotional dysregulation. Neurotransmitters are essential for mood regulation, arousal, and stress responses. By altering brain structure and function, particularly in reaction to stress and terror, some genetic variations, such as those in the FKBP5 gene, serotonin transporters, and dopamine receptors, enhance vulnerability to post-traumatic stress disorder (PTSD). The control of mood, arousal, and stress responses is significantly impacted by the dysregulation of neurotransmitter systems, such as those involving serotonin, dopamine, and GABA. These neurotransmitters are altered in PTSD, which can lead to symptoms including anxiety, hypervigilance, and emotional

dysregulation. For example, variations in dopamine signaling are connected to reduced motivation and attention, whereas changes in serotonin receptors and transporter genes are linked to increased anxiety and poor fear extinction.

Neuroendocrine dysregulation, namely of the hypothalamic-pituitary-adrenal (HPA) axis, is another feature of PTSD (Nisar et al., 2020). This exacerbates the neurobiological alterations in the brain and adds to the symptoms of post-traumatic stress disorder by causing aberrant cortisol levels and a disturbed stress response. The intricate interplay of elements that determine the brain's reaction to trauma and stress in post-traumatic stress disorder (PTSD) is highlighted by the connection among genetic predisposition, neurotransmitter dysregulation, and neuroendocrine abnormalities.

The paper "Neurobiological mechanisms underlying delayed expression of posttraumatic stress disorder: A scoping review" delves into the complex neurobiological elements that lead to delayed-onset PTSD, in which symptoms appear at least half a year after a traumatic incident (Smid et al., 2022).

Several interrelated systems, including neural, neuroendocrine, neuroinflammatory, and neurobehavioral/contextual mechanisms, are identified in the review as being involved in this delayed expression. The hippocampus, amygdala, and prefrontal cortex (PFC) are part of the neurocircuitry of fear, which is central to PTSD. Studies show that in PTSD patients, fear learning, memory, and emotional regulation are impacted by elevated amygdala responsiveness, decreased hippocampus volume, and aberrant PFC function. Neural network models indicate that recurrent traumatic events may modify these networks, causing processes such as kindling and sensitization to intensify PTSD symptoms progressively.

The hypothalamic-pituitary-adrenocortical (HPA) axis and the autonomic nervous system are essential to the body's stress response, and their dysregulation characterizes PTSD (Nisar et al., 2020). The hallmarks of this disorder include sensitization of the HPA axis and autonomic hyperarousal, which result in increased stress reactivity. Genetic and epigenetic factors further modulate these neuroendocrine responses, and changes in gene expression over time contribute to the development of PTSD symptoms.

Proinflammatory cytokines, in particular, have an influential role in the immune system's involvement in PTSD (Smid et al., 2022). These cytokines can potentially affect brain activity, impairing learning and memory. The idea that PTSD and traumatic brain injury (TBI) are two completely different disorders is called into question because TBI can exacerbate PTSD by inducing immunological activation and neuroinflammation.

Exposure to supplementary stressors after the initial trauma can hasten the onset of post-traumatic stress disorder (PTSD) symptoms (Nisar et al., 2020). Comorbid conditions such as depression, dissociation, and other mental or physical illnesses can also affect how PTSD develops; delayed-onset PTSD is ubiquitous in people who already have pre-existing vulnerabilities. The clinical picture is further complicated by depression, which often co-occurs with PTSD and has a significant impact on cognitive function and employment status. Hence, the amygdala, hippocampus, and prefrontal cortex are three important brain regions where PTSD is linked to anatomical and functional abnormalities. These areas are crucial in-memory processing, decision-making, and emotional control.

There is a common understanding of how trauma affects certain brain areas, as evidenced by the consistent reporting of structural alterations such as decreased hippocampal and amygdala volumes being linked to PTSD (Smid et al., 2022). The neurotransmitter systems in PTSD are essential (particularly serotonin, dopamine, and GABA) since they control critical facets of mood, arousal, stress reactions, and emotional regulation—all of which are significantly impacted by this illness. Dysregulation in these systems is associated with the symptoms of PTSD, such as heightened fear, anxiety, and impaired stress responses. There is agreement that genetic variations affecting these neurotransmitter systems can influence PTSD development and symptom severity. There is a connection between the FKBP5 gene and a higher chance of developing PTSD after trauma. The FKBP5 gene restrains the hypothalamic-pituitary-adrenal (HPA) axis, which governs the release of stress hormones such as cortisol and, thus, the body's stress response. Changes in this gene can result in a modified stress response, which increases a person's vulnerability to the impacts of traumatic experiences. HPA axis dysregulation plays a significant role in PTSD, leading to abnormal cortisol levels and a disrupted stress response, which exacerbates the neurobiological changes in the brain.

Neural Mechanisms Underlying Anxiety

Anxiety is an essential defensive behavior controlled by a network of brain regions, including the hippocampus, prefrontal cortex, amygdala, hypothalamus, and nucleus accumbens (Ghasemi M. et al., 2022). The hippocampus, in particular, takes a significant role in regulating anxiety– especially in emotional behaviors and anxiety states in response to stress.

Anxiety disorders are among the most prevalent mental health conditions, with a lifetime prevalence of 33.7% in the U.S. (Kessler et al., 2012). These disorders are typically diagnosed based on clusters of symptoms rather than an understanding of the underlying neurobiological mechanisms. Exposure-based cognitive behavioral therapy (CBT) is the first-line treatment for anxiety disorders, though its efficacy varies, with response rates ranging from 45% to 65% (Bandelow et al., 2014; Loerinc et al., 2015). Research indicates that anxiety is linked to structural and functional changes in brain regions like the amygdala, known for its role in emotional processing. Increased amygdala hyperactivity is often observed, alongside deficits in prefrontal control (PFC) and elevated drive from the locus coeruleus (LC), both of which regulate the fear response (Etkin & Wager, 2007). These changes underscore the complex neurobiology underlying anxiety disorders.

Recent findings have emphasized the importance of targeting specific neurobiological markers in the treatment of anxiety disorders (Brehl et al., 2020). Amygdala hypersensitivity, low PFC regulation, and noradrenergic dysregulation in the LC are key factors contributing to heightened anxiety responses. Treatments that downregulate amygdala responsiveness, such as enhancing GABAergic mechanisms, have shown promise in reducing anxiety symptoms, including post-traumatic stress disorder (PTSD). Moreover, exposure-based CBT has effectively strengthened PFC functioning to mitigate persistent symptoms. Future treatments may increasingly focus on addressing these underlying biomarkers, offering more tailored and practical approaches to managing anxiety and PTSD.

The hippocampus plays a crucial role in modulating anxiety by interacting with various brain regions, including the prefrontal cortex, amygdala, hypothalamus, and nucleus accumbens (Ghasemi et al., 2021). These regions regulate emotional and stress responses, adjusting anxiety levels in response to different stressors. Research suggests that heightened neuronal excitability in the hippocampus is linked to anxiety, with disruptions in local regulation of this excitability contributing to behavioral patterns seen in certain anxiety disorders. Some anxiolytic treatments alter the excitability and plasticity of hippocampal neurons, highlighting the potential for targeted interventions that focus on the cellular and molecular mechanisms governing hippocampal function. This is particularly relevant for treatment-resistant anxiety disorders, where traditional therapies may fail to address the underlying neural circuits involved.
Both innate and learned fear and anxiety responses in rodents are closely associated with hippocampal activity (Ghasemi et al., 2021). Conversely, reducing fear responses has lowered hippocampal excitability (Zhang et al., 2017). Neuronal excitability in the hippocampus is influenced by the conductance of voltage-gated ion channels, which modulate synaptic inputs and action potential generation (Shim et al., 2018). Changes in hippocampal excitability may arise from fear of learning, stress exposure, or biological causes unrelated to stress. Understanding the mechanisms underlying ventral hippocampal hyperactivity in anxiety and stress responses could lead to more tailored treatments for anxiety disorders by addressing the specific neural circuits involved (Ghasemi et al., 2021).

Cutting-edge techniques for treating anxiety have evolved, and mindful-based interventions (MBI) have shown promise in regulating PFC activity and enhancing emotional control (Gerber et al., 2023). Neuroimaging methods such as fMRI and electroencephalogram (EEG) have shown positive patterns indicating a neurological foundation for their efficacy. One research demonstrated that non-pharmaceutical treatments–therapeutic methods–are better than cramming traditional anti-depressions, which increases happy hormones such as oxytocin, as non-pharmaceutical treatments have been found to alter anterior cingulate cortex (ACC) activities. This activity is particularly relevant considering that neurogenesis, or the formation of new neurons, can be harmful in anxiety circumstances.

Adolescence is a critical developmental period during which anxiety often emerges, with over 31% of U.S. adolescents reporting symptoms (Merikangas et al., 2010). Anxiety in youth is frequently linked to behavioral avoidance, mirroring patterns seen in adults (Galván & Peris, 2014; Maner & Schmidt, 2006). Interestingly, this period is also marked by increased risk-taking, characterized by approach behaviors (Baker et al., 2020). This paradox raises the question of why adolescence, a time of heightened anxiety and avoidance, also promotes behaviors that involve risk. Neuroimaging studies suggest that anxiety influences neural functioning through atypical modulation of attention, particularly toward threatening or fearful stimuli. Key brain regions such as the amygdala and the prefrontal cortex (PFC), which comprise the "salience network," are implicated in controlling responses to threats, with anxiety leading to unusual activation in these areas (Baker, 2020, p. 2).

Recent findings have also highlighted the role of the striatum in adolescent anxiety (Lago et al., 2017). As this brain region becomes sensitized around the same time anxiety symptoms emerge, it plays a crucial role in fear processing. The striatum’s connections with the

amygdala, hippocampus, and ventromedial PFC further suggest its involvement in adolescent anxiety, particularly in motivation and attention. Future research could benefit from techniques like Dynamic Causal Modeling (DCM), which would help examine the causal architecture between brain regions linked to anxiety, such as the amygdala, PFC, and ventral striatum. Longitudinal studies tracking risk-taking, approach-avoidance motivations, and mental health through puberty are essential for understanding the underlying mechanisms of risk and resilience in adolescents at risk for anxiety (Baker et al., 2020).

In psychoanalysis, anxiety is viewed as a signal of threat whenever a forbidden emotion arises (Grecucci et al., 2021). This anxiety activates defenses and maladaptive behaviors, which can lead to clinical issues. As a result, managing anxiety is central to psychodynamic-oriented treatments, aiming to help clients regulate their emotional responses. This theoretical paper explores anxiety generation and dysregulation from both a neural and psychodynamic perspective, drawing on neuroimaging and psychophysiological research. By comparing these findings with psychodynamic theories of anxiety, the paper sheds light on the psychological mechanisms and neural bases of emotion regulation, particularly within the therapeutic setting. Two fundamental approaches to emotion and anxiety regulation are discussed: one grounded in the cognitive model of emotion regulation and the other rooted in psychodynamic principles and affective neuroscience.

According to James J. Gross' emotion regulation framework, based on appraisal theory, emotions arise from how individuals interpret a stimulus, and cognitive-behavioral therapies (CBT) focus on managing emotions through cognitive processes (Grecucci et al., 2021). From this perspective, dysregulated emotions result from ineffective mental, behavioral, or attentive strategies. In contrast, dynamic experiential treatments distinguish between primary emotions—those stemming from real-world stimuli—and dysregulated affect states (DAS), which are tied to excessive conditioned anxiety. DAS, which results from pairing emotions with unregulated anxiety, requires specialized regulation techniques. Experiential dynamic therapists work to reduce stress to a level where patients can fully experience their emotions without triggering defenses. To prevent relapse, therapists go a step further, helping patients tolerate higher levels of emotion without triggering anxiety, allowing them to break free from the defenses that lead to symptoms and problems.

Social anxiety disorder (SAD), formerly known as social phobia, is characterized by persistent fear, nervousness, and avoidance in social situations, especially when interacting with strangers or being observed by others (Yu et al., 2021). SAD often begins in adolescence and, if left untreated, can lead to comorbid conditions such as depression, substance abuse, and other anxiety disorders. As research progresses, more scientists turn to neuroimaging techniques like magnetic resonance imaging (MRI) to uncover the cognitive neural mechanisms underlying SAD. Understanding these mechanisms could eventually improve the diagnosis and treatment of the disorder, offering hope for more effective interventions in the future. A recent meta-analysis using activation likelihood estimation (ALE) for whole-brain analysis revealed key neurological differences between individuals with SAD and healthy controls. The analysis showed reduced activation in the cingulate gyrus in people with SAD, a region crucial for attention control. Additionally, the research highlighted that brain activation patterns varied

based on the type of task used in experimental paradigms, suggesting that the results should be interpreted cautiously. This variability underscores the importance of considering task differences when studying the neurological underpinnings of social anxiety, as different experimental designs can influence brain activation patterns.

Discussion

This paper went over the definition of both trauma and anxiety disorders as classified in the DSM-5 (American Psychiatric Association, 2022). PTSD, one of the most prevalent trauma disorders, shows symptoms such as negative cognitive and affective changes, alterations in arousal and reactivity, and intrusive symptoms after experiencing a traumatic event. On the other hand, anxiety is a disorder characterized by fear, which is an anticipation of future threats. It shows symptoms such as dread, avoidance behavior, as well as related cognitive conceptualization. Unlike PTSD, anxiety is a bigger picture of a disorder that branches out to other fear responses–such as SAD (Yu et al., 2021). The research question is answered by going through different neural mechanisms to cure, such as Exposure-based cognitive behavioral therapy, activation likelihood estimation, EEG, and fMRI (Bandelow et al., 2014; Gerber et al., 2023).

Understanding the neural mechanisms underlying trauma and anxiety disorders can help develop focused and effective treatments. For example, trauma disorders tend to cause structural brain changes; regulating hormones has shown positive outcomes in changing the structure and function of the brain rather than giving physical changes in the brain (Jiang et al., 2019). Especially changes in the FKBP5 gene have been shown to control the HPA axis, which governs the release of stress hormones such as cortisol and, thus, the body's stress response.

For anxiety disorders, both two specific techniques, Cutting-edge techniques (CBT) and mindful-based interventions (MBI), have been shown to enhance emotional control in regulating the pre-frontal cortex area of the brain (Brehl et al., 2020; Gerber et al., 2023). MBI shows the understanding of how specific brain regions, such as the amygdala, hippocampus, and prefrontal cortex, contribute to trauma and anxiety symptoms, allowing for therapies that target their malfunction. Both trauma and anxiety disorders are affected by specific brain regions–hippocampus, prefrontal cortex, amygdala, and hypothalamus. Several treatments address both shared anxiety and trauma disorders. For instance, fMRI studies indicated the foundations of where the problem occurs by displaying the hypoactivation in brain regions. Specific mechanisms such as enhancing GABAergic have been shown to regulate amygdala responsiveness, which is responsible for emotions and activating feelings of fear, showing reduced symptoms of anxiety and PTSD (Brehl et al., 2020).

Studying the neural mechanisms can improve the psychological treatment of disorders in individuals. Personalized therapies such as biomarker identifications can guide treatment selections according to individual neural profiles. Especially the neuroimaging method (fMRI and EEG), as mentioned, gives an objective measure that helps figure out biological aspects to improve further treatments and generate predictions of behavioral changes (Sohail et al., 2024).

Limitation

Although focusing on the neural mechanisms might overlook some aspects of trauma and anxiety disorders, such as the lived experience, it is crucial to understand the underlying causes rather than focusing on the symptoms to offer a more individualized and successful therapy paradigm. Another limitation of the review was time. Hence, it was a brief review rather than a systematic or scoping review.

Conclusion

In summary, the review explored the neural mechanisms underlying trauma and anxiety disorders. By developing a better understanding of these mechanisms, we can create more focused and efficient treatments and interventions that deal with the underlying causes of the conditions. Moreover, this information holds great promise for enhancing patient outcomes and public health in addition to advancing science. In order to develop new therapeutic strategies and improve current ones, more research in this field is necessary. This will eventually lead to more effective and individualized medical interventions.

Works Cited

American Psychiatric Association. (2022). Diagnostic and Statistical Manual of Mental Disorders. American Psychiatric Association Publishing. https://doi.org/10.1176/appi.books.9780890425787

Baker, A. E., & Galván, A. (2020). Threat or thrill? The neural mechanisms underlying the development of anxiety and risk taking in adolescence. Developmental Cognitive Neuroscience, 45, 100841.

Bandelow, B., Lichte, T., Rudolf, S., Wiltink, J., & Beutel, M. E. (2014). The diagnosis of and treatment recommendations for anxiety disorders. Deutsches Arzteblatt International, 111(27-28), 473–480. https://doi.org/10.3238/arztebl.2014.0473

Brehl, A.-K., Kohn, N., Schene, A. H., & Fernández, G. (2020). A mechanistic model for individualised treatment of anxiety disorders based on predictive neural biomarkers. Psychological Medicine, 50(5), 727–736. https://doi.org/10.1017/S0033291720000410

Etkin, A., & Wager, T. D. (2007). Functional neuroimaging of anxiety: A meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. The American Journal of Psychiatry, 164(10), 1476–1488. https://doi.org/10.1176/appi.ajp.2007.07030504

Gerber, C., & Matuschek, P. (2023). Neural mechanisms of mindfulness-based interventions in anxiety disorders: A systematic review. Archives of Clinical Psychiatry, 50(6).

Ghasemi, M., Navidhamidi, M., Rezaei, F., Azizikia, A., & Mehranfard, N. (2022). Anxiety and hippocampal neuronal activity: Relationship and potential mechanisms. Cognitive, Affective, & Behavioral Neuroscience, 22(3), 431–449.

Grecucci, A., Sığırcı, H., Lapomarda, G., Amodeo, L., Messina, I., & Frederickson, J. (2020). Anxiety regulation: From affective neuroscience to clinical practice. Brain Sciences, 10(11), 846.

Invernizzi, A., Rechtman, E., Curtin, P., et al. (2023). Functional changes in neural mechanisms underlying post-traumatic stress disorder in World Trade Center responders. Translational Psychiatry, 13, 239. https://doi.org/10.1038/s41398-023-02526-y

Jiang, S., Postovit, L., Cattaneo, A., Binder, E. B., & Aitchison, K. J. (2019). Epigenetic modifications in stress response genes associated with childhood trauma. Frontiers in Psychiatry, 10, 808. https://doi.org/10.3389/fpsyt.2019.00808

Kessler, R. C., Petukhova, M., Sampson, N. A., Zaslavsky, A. M., & Wittchen, U. (2012). Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. International Journal of Methods in Psychiatric Research, 21(3), 169–184. https://doi.org/10.1002/mpr.1359

Kim, T. D., Lee, S., & Yoon, S. (2020). Inflammation in post-traumatic stress disorder (PTSD): A review of potential correlates of PTSD with a neurological perspective. Antioxidants, 9(2), 107. https://doi.org/10.3390/antiox9020107

Lago, T., Davis, A., Grillon, C., & Ernst, M. (2017). Striatum on the anxiety map: Small detours into adolescence. Brain Research, 1654(Pt B), 177–184. https://doi.org/10.1016/j.brainres.2016.06.006

Loerinc, A. G., Meuret, A. E., Twohig, M. P., Rosenfield, D., Bluett, E. J., & Craske, M. G. (2015). Response rates for CBT for anxiety disorders: Need for standardized criteria. Clinical Psychology Review, 42, 72–82. https://doi.org/10.1016/j.cpr.2015.08.004

Maner, J. K., & Schmidt, N. B. (2006). The role of risk avoidance in anxiety. Behavior Therapy, 37(2), 181–189. https://doi.org/10.1016/j.beth.2005.11.003

Merikangas, K., He, J., Burstein, M., Swanson, S., Avenevoli, S., Cui, L., & Swendsen, J. (2010). Lifetime prevalence of mental disorders in US adolescents: Results from the national comorbidity study-adolescent supplement (NCS-A). Journal of the American Academy of Child & Adolescent Psychiatry, 49(10), 980–989. https://doi.org/10.1016/j.jaac.2010.05.017

Nisar, S., Bhat, A. A., Hashem, S., Syed, N., Yadav, S. K., Uddin, S., Fakhro, K., Bagga, P., Thompson, P., Reddy, R., Frenneaux, M. P., & Haris, M. (2020). Genetic and neuroimaging approaches to understanding post-traumatic stress disorder. International Journal of Molecular Sciences, 21(12), 4503. https://doi.org/10.3390/ijms21124503

Richter-Levin, G., & Sandi, C. (2021). Labels matter: Is it stress or is it trauma? Translational Psychiatry, 11(1), 385. https://doi.org/10.1038/s41398-021-01514-4

Smid, G. E., Lind, J., & Bonde, J. P. (2022). Neurobiological mechanisms underlying delayed expression of posttraumatic stress disorder: A scoping review. World Journal of Psychiatry, 12(1), 151–168. https://doi.org/10.5498/wjp.v12.i1.151

Sohail, A., & Zhang, L. (2024). Informing the treatment of social anxiety disorder with computational and neuroimaging data. Psychoradiology, 4, kkae010. https://doi.org/10.1093/psyrad/kkae010

Yu, X., Ruan, Y., Zhang, Y., Wang, J., Liu, Y., Zhang, J., & Zhang, L. (2021). Cognitive neural mechanism of social anxiety disorder: A meta-analysis based on fMRI studies. International Journal of Environmental Research and Public Health, 18(11), 5556.

miR-223 Regulating Neutrophils Via the NF-kB Pathway in Chronic Obstructive Pulmonary Disease By Jaelyn Chow

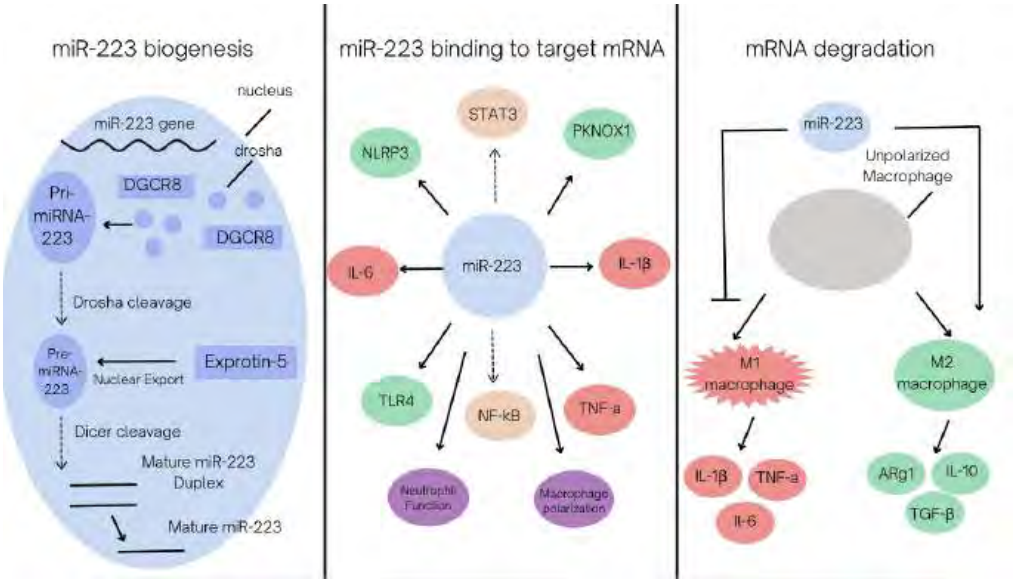
Abstract

Micro-RNA-223(miR-223) serves as a key regulator of inflammatory processes in Chronic Obstructive Pulmonary Disease (COPD) (1, 2). This research evaluates the multifaceted role of miR-223 in COPD pathogenesis, focusing specifically on the impact on neutrophil function and count, macrophage polarization, and the Nf-κB pathway (1, 3). In this review, I first introduced the pathogenesis and inflammatory mechanisms of chronic obstructive pulmonary disease (COPD), the biogenesis and expression of miR-223 in inflammation, macrophage polarization, T-cell responses, metabolic reprogramming, and signaling pathways: NF-κB, STAT/JAK, cancer and DNA damage with oxidative stress, in miR-223's role in balancing inflammatory responses (2, 3). Afterward, I discuss the potential role of miR-223 in COPD, atherosclerosis, neuroinflammation, and potential therapeutic causes (4, 5).

Introduction: The molecular basis of Chronic inflammation in COPD

At the molecular level, chronic inflammation is caused by the prolonged inflammatory response that is further caused by the ongoing activation of monocytes and lymphocytes that release proinflammatory mediators and damage tissues (6, 7, 8). MicroRNAs are small RNAs that regulate gene expression (9). MicroRNA-223(miR-223) also significantly regulates this process by influencing different parts of the inflammatory response (10).

Figure 1.



Overview of miR-223 biogenesis, its role in binding to target mRNA, and mRNA degradation (9).

Chronic Obstructive Pulmonary Diseases (COPD): Pathogenesis and Inflammatory Mechanisms

Chronic inflammation plays a huge role in the pathogenesis of Chronic Obstructive Pulmonary Disease(COPD) (11). COPD affects millions of people around the world, with approximately 215 million people diagnosed (94). COPD is most commonly found in 40-year-olds, specifically men more than women (94). COPD is often described as a persistent inflammation in the airways and lung tissues (12). The abnormalities of airflow limitation, air trapping, and reduced gas exchange all contribute to the abnormal functions that lead to the diagnosis of COPD (13). This happens because inflammatory cells keep getting activated and releasing pro-inflammatory mediators that cause more inflammation. In COPD, the balance between the macrophages in M1 and M2 is disrupted, and an increase in M1 macrophages produces lots of TNF-α, IL-1β, and IL-8, which contributes to and affects the tissue remodeling and airway (14). Neutrophils play the biggest part in COPD, by releasing enzymes that are capable of breaking down proteins such as matrix metalloproteinases (MPPs) which disrupt the structural components of the tissue causing emphysema (15, 16).

In COPD patients, T lymphocytes, specifically CD8+T cells are increased in the airways and lung parenchyma; the part of the lung that is responsible for gas exchange (17, 18). These CD8+T cells release perforins and granzyme (19). The release of these mediators causes alveolar cell apoptosis and contributes to the slow progression of emphysema. Leading to COPD, the Nf-κB pathway is activated due to the increased production of the inflammatory cytokines that chemokines (20, 21). Chronic activation plays a huge part in oxidative stress that is caused by smoking and environmental pollutants (22). Similarly, the STAT/JAK pathway, specifically the STAT3 activation in the COPD pathogenesis promotes inflammation and mucus hypersecretion (23, 24).

Similar to the Warburg effect in cancer cells, metabolic reprogramming occurs in COPD with the shift of glycolysis in immune and structural cells (25). The metabolic shift supports the high demand for energy for chronic inflammation that contributes to the high inflammatory responses in the lungs (26, 27). The airways that have been seen in COPD patients are increased such as pattern recognition receptors (28). These receptors can recognize the pathogen-associated molecular patterns and damage-associated patterns of (DAMP), which leads to the activation of immune responses and prolonged inflammation (29, 30).

In addition, the extracellular matrix (ECM) in a COPD patient's lungs goes through a significant amount of remodeling (31, 32). Higher activity of MMPS, specifically MMP-9 and MMP-12 can lead to the degradation of the ECM substances that lead to lung emphysema (33). The increase of abnormal ECM proteins leads to the thickening of the airway wall and fibrosis in the smaller airways of the lung (34). Oxidative stress plays a huge role in COPD, as it causes a huge imbalance in oxidants and antioxidants (36). Activated inflammatory cells produce high levels of ROS and RNS that cause direct damage to the lung tissues, activate pro-inflammatory receptors, and cause more tissue damage (37, 38). The presence of antibodies against elastins and other self-antigens has been seen in COPD patients, leading to the suggestion that these mechanisms may also contribute to prolonged inflammation and tissue damage (39, 40). These autoantibodies cause ongoing inflammation by sticking to the body's tissue and activating immune cells (41). This triggers a chain reaction of inflammatory responses, causing continuous damage and creating a cycle of more inflammation (42). The persistent and ongoing

inflammation in the lungs ultimately leads to airway wall thickening, mucus production, and the destruction of alveolar walls (43).

Role of miR-223 in inflammation

MiR-223 is a small, noncoding RNA that regulates gene expressions and controls the processes that include inflammation and immune cell differentiation (44). miR-223 has been shown to target specific genes like CDK2, TGF-β, IGF-1R, Mef2c, and p53, that are involved in inflammation (45, 2). For example, miR-223 can inhibit the production of inflammatory mediators or block inflammation through direct inhibition in the signaling pathways, therefore protecting the tissues from prolonged inflammation damage (46). This regulatory function of miR-223 has been seen in various immune conditions such as acute lung injury, rheumatoid arthritis, and central nervous system inflammation (46, 10). Overall, miR-223 and its ability to modulate immune responses and inflammation make it a potential candidate when developing therapeutic strategies aimed at treating chronic inflammatory disorders (10).

Macrophage Polarization: miR-223's Role in Balancing Inflammatory Responses

Macrophages' ability to take on different phenotypes is a factor in chronic inflammation, with M1 and M2 being the two main phenotypic classifications serving as proinflammatory and anti-inflammatory phenotypes, respectively (47, 48). M1 macrophages are known for producing cytokines like TNF-α, IL-1β, IL-6, and reactive oxygen and nitrogen species that intensify tissue damage, while M2 macrophages are meant to repair tissue damage and fibrosis, releasing anti-inflammatory cytokines like IL-10 and TGF-β (14, 49). An imbalance between M1 and M2 macrophages can lead to prolonged tissue damage and chronic inflammation (50). miR-223 plays a crucial role in regulating macrophage polarization by influencing the expression of specific genes involved in inflammation and inhibiting proinflammatory activation by targeting pathways such as NLRP3, which is important in M1 macrophage activation (46, 51). By promoting the anti-inflammatory M2 phenotype that inhibits the pro-inflammatory M1 phenotype, miR-223 helps balance the inflammatory response and prevent excessive tissue damage (45, 51). The overexpression of miR-223 has been shown to reduce M1 polarization markers such as TNF-α and IL-6 while enhancing M2 markers such as IL-10 and Arg1, resulting in a shift towards the M2 phenotype that is important for resolving inflammation and promoting tissue repair (52, 51). In COPD, the balance between M1 and M2 macrophages is severely disrupted and contributes to the prolonged inflammatory state of this disease (53). The airways of COPD patients show a marked increase in M1 macrophages, which release high levels of proinflammatory mediators like TNF-α, IL-1β, and IL-6, exacerbating tissue damage and airway remodeling (54, 55). The abnormal level of the M1 phenotype is accompanied by a deficiency of M2 macrophages, which limits the lung's ability to resolve inflammation and repair the damaged tissue (56, 14). The abnormality in macrophage polarization in COPD is further compounded by exposure to cigarette smoke and other environmental pollutants that can directly influence the macrophage phenotype and function (57, 58). Targeting macrophage polarization, potentially through modulation of miR-223 expression, represents a promising therapeutic avenue for COPD, aiming

to restore the balance between proinflammatory and anti-inflammatory macrophage populations and control the chronic inflammatory processes driving disease progression (47, 45).

T Cell Responses: miR-223's Role in Balancing Inflammatory Responses

Similarly, lymphocytes play a key role in chronic inflammation. TH1 cells, otherwise known as T helper one cells, are a subset of CD4+ T cells that produce IFN-y, which is responsible for activating macrophages, while Th17 cells release IL-17, which attracts neutrophils (59, 60). However, T reg cells also play a role by releasing anti-inflammatory cytokines, but their effects are often lowered in inflammatory disorders that last for a long period (61). miR-223 can modulate these T cell responses, potentially helping to regulate the balance between proinflammatory Th1/Th17 cells and anti-inflammatory Treg cells, its exact role however, is still in need of further study (46). In COPD, the T cell response is altered, contributing to the chronic inflammatory state (62). CD8+T cells are particularly increased in many of the airways and lung parenchyma of COPD patients. These cells release cytotoxic mediators like perforins and granzymes, leading to alveolar cell apoptosis and contributing to the slow progression of emphysema (63, 17). The balance between different T cell subsets is disrupted in COPD, with an increase in proinflammatory Th1 and Th17 cells and a relative decrease in anti-inflammatory Treg cells (64). This imbalance perpetuates the inflammatory cycle and impairs the resolution of the inflammation in COPD lungs (65). The role of miR-223 in modulating these T cell responses in the context of COPD is an area of ongoing research, with potential implications for developing targeted therapies to restore T cell homeostasis in the diseased lung (2, 45).

Signaling Pathways: miR-223's modulation of NF-kB and STAT/JAK in COPD

In the immune system, several pathways contribute to chronic inflammation. The NF-κB pathway is a transcription factor and central component of pro-inflammatory programs in which it is activated by cytokines and pathogen-associated molecular patterns (21). miR-223 is often involved in regulating this pathway, particularly by inhibiting NLRP3 inflammasome activation to suppress IL-1β secretion, which is part of the Nf-κB feed-forward loop (66). The action of miR-223 helps limit chronic inflammation. The long activation creates prolonged inflammation that leads to chronic inflammation, such as autoimmune disorders, and often involves compensatory feedback mechanisms (67, 68). There are many other pathways, such as the STAT/JAK pathway, that can be activated via cytokine IL-6 and IFN-γ. In this pathway, miR-223 targets STAT3, which plays a role in inflammatory signaling (78). In COPD, the Nf-κB is activated, leading it to play a central role in the chronic inflammatory response. This pathway tends to be sensitive to oxidative stress that is often caused by cigarette smoke and environmental pollutants, leading to increased production of proinflammatory cytokines and chemokines in the lungs. The persistent activation of NF-κB in COPD contributes to the ongoing inflammatory cells that are being released, tissue remodeling, and mucus hypersecretion (70) Additionally, the STAT/JAK pathway, especially STAT3 activation, is implicated in COPD pathogenesis, which further promotes inflammation and excessive mucus production in the airways (71). These signaling pathways and miR-223 in COPD are in areas that are active for

research, leading to potential targeted therapies that can modulate and balance the inflammatory cytokines.

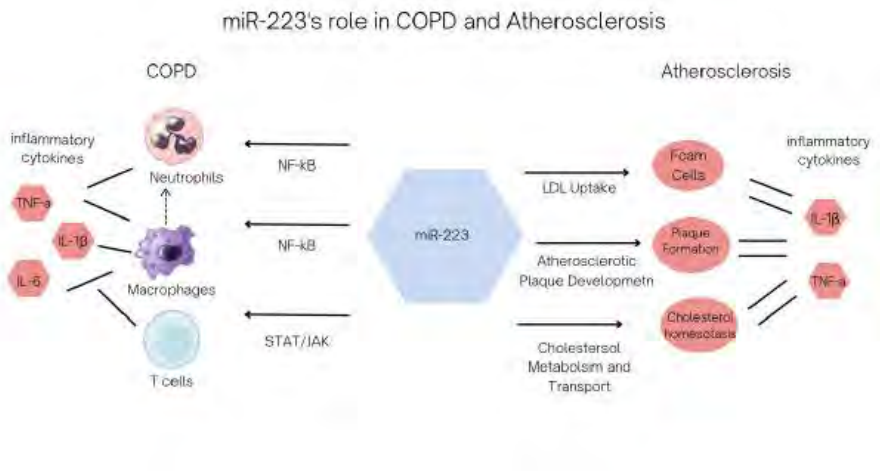
Metabolic Reprogramming: miR-223's Impact on Immune Cell Metabolism in COPD

Chronic inflammation can also be related to and affected by the metabolic reprogramming of cells in our immune system. The cells change how they make energy, and instead of using oxygen like most cells prefer, these cells start breaking down sugar for energy, even if there's oxygen. This change in metabolism is capable of meeting the high energy needs of activating the cells in our immune system and supporting the production of inflammatory mediators that cause inflammation. In macrophages, miR-223 has been implicated in this metabolic reprogramming, specifically in changing the expression levels of glycolytic machinery, such as PDF and PDK-1, as well as HIF-1α, an upstream regulator of metabolism (72). Some glycolytic metabolites like citrate and succinate also work as signaling molecules that increase our body's inflammatory response (45). In COPD, metabolic reprogramming plays a significant role in chronic inflammation. The immune cells in the lungs of COPD patients, specifically macrophages and neutrophils, often undergo a shift towards glycolysis, even with the presence of oxygen (73). This metabolic shift supports the high energy demands of chronic inflammation, contributing to the heightened inflammatory responses in COPD lungs. The altered metabolism in immune cells also leads to an increase in the production of inflammatory mediators and reactive oxygen species, further damaging the tissue and disrupting airway remodeling (73). miR-223s are also involved in metabolic reprogramming, specifically in macrophages, which have important implications for disease pathogenesis and potential therapeutic interventions that are targeted toward reducing inflammation (92).

Atherosclerosis and COPD: Inflammatory mechanisms influenced by miR-223

Chronic inflammation is often tightly linked with many long-term diseases. In a process called atherosclerosis, fat accumulates and builds inside of your arteries causing them to harden and narrow down. Foamy macrophages contribute to this process, and miR-223 plays a significant role in regulating cholesterol homeostasis and foam cell formation (51, 76). This suggests that miR-223 could influence plaque stability, as certain proteins like IL-1β and TNF-α can destabilize these plaques, potentially leading to fatal heart attacks due to reduced blood flow (51). While atherosclerosis is not directly related to COPD, they both share a very similar inflammatory mechanism that is influenced by miR-223. In COPD, chronic inflammation in the lungs contributes to cardiovascular complications. The role of miR-223 is to regulate homeostasis and foam cell formation, which are extremely crucial in atherosclerosis and may also have implications for COPD patients (51). The ability of miR-223 to influence plaque stability could be particularly relevant in COPD since these patients often have high levels of inflammatory markers like IL-1β and TNF-α, which can destabilize plaques (2).

Figure 2.



Outlines the foam cells, plaque formation, and cholesterol homeostasis with inflammatory cytokines IL-1β and TNF-α, and shows the role of miR-223 in chronic inflammation and contributing to disease progression. (51).

Cancer and DNA Damage: miR-223's role and oxidative stress

In cancer, inflammation creates a space that allows cancer cells to grow and spread to other tissues. miR-223 has been seen in this process, showing pro- and anti-tumorigenic roles depending on the specific cancer type and context of each case (77). miR-223 can influence tumor-associated inflammation, potentially making it a target for certain cancer therapies (79). Inflammation is also capable of changing and damaging DNA via the production of reactive oxygen and nitrogen species (RONS) by immune cells (79, 80). These reactive compounds were meant to destroy pathogens but can harm the DNA around them and cause mutations (81). Consequently, if the DNA is not repaired then the genetic mutations can lead to the early stages of cancer. mirR-223 has been shown to regulate oxidative stress responses that potentially influence this aspect of inflammation-induced DNA damage. In COPD patients, there is an increased risk of lung cancer, which may be partly due to the chronic inflammatory environment in the lungs (82). The role of miR-223 in this context is very complex since they can both exhibit both pro- and anti-tumorigenic effects depending on the cellular environment (83). In COPD-associated lung cancer, miR-223 might influence tumor-associated inflammation, potentially affecting cancer cell growth and metastasis (45). Chronic oxidative stress that's present in COPD lungs leads to an increase in the production of reactive oxygen and nitrogen species (RONS), which can cause DNA damage and mutations (84). miR-223's role in regulating oxidative stress responses can be relevant in COPD and potentially influences the risk of inflammation-induced DNA damage and subsequent cancer development in the lungs of COPD patients (2).

Neuroinflammation: Potential Systemic Effects of COPD on Brain Health

In neuroinflammation, some brain cells such as astrocytes and microglia are activated due to abnormal proteins and signals that have been released from damaged neurons (85). These signals attract microglia, the brain's immune cells, which causes them to be more active (86). Once these cells are active, then microglia will release inflammatory substances and bring in

immune cells that will continue the cycle of inflammation (87). This activation of the brain cells caused by chronic inflammation damages the neurons in the brain and is a feature of Alzheimer's and Parkinson's diseases (88). While COPD only affects the lungs, there is potentially more evidence of systematic effects that can change the impact on the central nervous system (89). In COPD patients, chronic hypoxia and systemic inflammation can also lead to neuroinflammation, potentially contributing to cognitive decline and other neurological complications (90). Similar to the processes observed, COPD-related neuroinflammation may involve the activation of microglia in the brains of COPD patients (90). These activated glial cells can release proinflammatory mediators that can potentially damage neurons and disrupt normal brain function and overall brain health in COPD patients (91).

Therapeutic Approaches: miR-223 for COPD Treatment

miR-223 is potentially a therapeutic solution for chronic inflammatory disease, specifically COPD, because of its many functions in regulating inflammatory processes (46). miR-223 influences many inflammatory processes relevant to COPD pathogenesis, making it a promising target for developing treatments. One approach that is currently being researched is mimicking the miR-223, which is used to replicate miR-223's role. By increasing miR-223 levels, these mimics could potentially increase their anti-inflammatory effects in COPD patients (46). This could reduce proinflammatory cytokine production and help shift macrophage polarization towards the anti-inflammatory M2 phenotype, potentially reducing tissue damage and airway inflammation. (47). Nanoparticle-based delivery systems for miR-223 or its modulators are another topic that's being researched (93). These systems could improve the stability of miR-223-based therapeutics and enhance their targeted delivery to the lungs, increasing efficacy while also minimizing off-target effects (93). The therapeutic potential of miR-223 in COPD is particularly important due to its role in regulating neutrophil function. By modulating neutrophil activity, miR-223-based therapies could help reduce the severity of COPD and potentially slow the disease progression (45).

Works Cited

1. Roffel, Mirjam P., et al. “MiR-223 Is Increased in Lungs of Patients With COPD and Modulates Cigarette Smoke-induced Pulmonary Inflammation.” *AJP Lung Cellular and Molecular Physiology*, vol. 321, no. 6, Oct. 2021, pp. L1091–104. https://doi.org/10.1152/ajplung.00252.2021.
2. Roffel, Mirjam P., Ken R. Bracke, et al. “miR-223: A Key Regulator in the Innate Immune Response in Asthma and COPD.” *Frontiers in Medicine*, vol. 7, May 2020, https://doi.org/10.3389/fmed.2020.00196.
3. Shi, Mingyu, et al. “miR-223: A Key Regulator of Pulmonary Inflammation.” *Frontiers in Medicine*, vol. 10, July 2023, https://doi.org/10.3389/fmed.2023.1187557.
4. Zeng, Yan, et al. “MicroRNA-223 Attenuates Hypoxia-induced Vascular Remodeling by Targeting RhoB/MLC2 in Pulmonary Arterial Smooth Muscle Cells.” *Scientific Reports*, vol. 6, no. 1, Apr. 2016, https://doi.org/10.1038/srep24900.
5. *MicroRNAs as Therapeutic Targets in Lung Disease: Prospects and Challenges*. journal.copdfoundation.org/jcopdf/id/1085/MicroRNAs-as-Therapeutic-Targets-in-Lung-Disease-Prospects-and-Challenges.
6. Hannoodee, Sally, and Dian N. Nasuruddin. “Acute Inflammatory Response.” *StatPearls - NCBI Bookshelf*, 8 June 2024, www.ncbi.nlm.nih.gov/books/NBK556083.
7. Gardner, T. W., Antonetti, D. A., Barber, A. J., LaNoue, K. F., & Levison, S. W. (2002). Diabetic retinopathy. *Survey of Ophthalmology*, *47*, S253–S262. https://doi.org/10.1016/s0039-6257(02)00387-9
8. Pahwa, Roma, et al. “Chronic Inflammation.” *StatPearls - NCBI Bookshelf*, 7 Aug. 2023, www.ncbi.nlm.nih.gov/books/NBK493173.
9. “MicroRNAs: Understanding Their Role in Gene Expression and Cancer.” *Einstein (São Paulo)*, vol. 19, Jan. 2021, https://doi.org/10.31744/einstein_journal/2021rb5996.
10. Yuan, Shun, et al. “miR-223: An Immune Regulator in Infectious Disorders.” *Frontiers in Immunology*, vol. 12, Dec. 2021, https://doi.org/10.3389/fimmu.2021.781815.
11. Xu, Jiao, et al. “Inflammation Mechanism and Research Progress of COPD.” *Frontiers in Immunology*, vol. 15, Aug. 2024, https://doi.org/10.3389/fimmu.2024.1404615.
12. “COPD - Symptoms and Causes.” *Mayo Clinic*, www.mayoclinic.org/diseases-conditions/copd/symptoms-causes/syc-20353679#:~:text=Chronic%20obstructive%20pulmonary%20disease%20.
13. Agarwal, A. K., Raja, A., & Brown, B. D. (2023, August 7). *Chronic obstructive pulmonary disease*. StatPearls - NCBI Bookshelf. https://www.ncbi.nlm.nih.gov/books/NBK559281/
14. Strizova, Zuzana, et al. “M1/M2 Macrophages and Their Overlaps – Myth or Reality?” *Clinical Science*, vol. 137, no. 15, Aug. 2023, pp. 1067–93. https://doi.org/10.1042/cs20220531.
15. Demkow, U., and F. J. Van Overveld. “Role of Elastases in the Pathogenesis of Chronic Obstructive Pulmonary Disease: Implications for Treatment.” *European Journal of Medical Research*, vol. 15, no. S2, Nov. 2010, https://doi.org/10.1186/2047-783x-15-s2-27.
16. Demkow, U., and F. J. Van Overveld. “Role of Elastases in the Pathogenesis of Chronic Obstructive Pulmonary Disease: Implications for Treatment.” *European Journal of Medical Research*, vol. 15, no. S2, Nov. 2010, https://doi.org/10.1186/2047-783x-15-s2-27.

17. Williams, Maya, et al. “The Role of CD8 + T Lymphocytes in Chronic Obstructive Pulmonary Disease: A Systematic Review.” *Inflammation Research*, vol. 70, no. 1, Oct. 2020, pp. 11–18. https://doi.org/10.1007/s00011-020-01408-z.
18. Barnes, Peter J., and Manuel G. Cosio. “Characterization of T Lymphocytes in Chronic Obstructive Pulmonary Disease.” *PLoS Medicine*, vol. 1, no. 1, Oct. 2004, p. e20. https://doi.org/10.1371/journal.pmed.0010020.
19. Janeway, Charles A., Jr, et al. “T Cell-mediated Cytotoxicity.” *Immunobiology - NCBI Bookshelf*, 2001, www.ncbi.nlm.nih.gov/books/NBK27101.
20. Edwards, Michael R., et al. “Targeting the NF-κB Pathway in Asthma and Chronic Obstructive Pulmonary Disease.” *Pharmacology & Therapeutics*, vol. 121, no. 1, Oct. 2008, pp. 1–13. https://doi.org/10.1016/j.pharmthera.2008.09.003.
21. Liu, Ting, et al. “NF-κB Signaling in Inflammation.” *Signal Transduction and Targeted Therapy*, vol. 2, no. 1, July 2017, https://doi.org/10.1038/sigtrans.2017.23.
22. “Oxidative Stress: Harms and Benefits for Human Health.” *Oxidative Medicine and Cellular Longevity*, vol. 2017, no. 1, Jan. 2017, https://doi.org/10.1155/2017/8416763.
23. “Epigenetics of Mucus Hypersecretion in Chronic Respiratory Diseases.” *American Journal of Respiratory Cell and Molecular Biology*, vol. 58, no. 3, Nov. 2017, pp. 299–309. https://doi.org/10.1165/rcmb.2017-0072tr.
24. Verres, Yann, et al. “Impact of JAK/STAT Inhibitors on Human Monocyte‐derived‐macrophages Stimulated by Cigarette Smoke Extract and Lipopolysaccharide.” *Clinical and Experimental Pharmacology and Physiology*, vol. 49, no. 11, July 2022, pp. 1187–96. https://doi.org/10.1111/1440-1681.13705
25. Navarro, Carla, et al. “Metabolic Reprogramming in Cancer Cells: Emerging Molecular Mechanisms and Novel Therapeutic Approaches.” *Pharmaceutics*, vol. 14, no. 6, June 2022, p. 1303. https://doi.org/10.3390/pharmaceutics14061303.
26. Kominsky, Douglas J., et al. “Metabolic Shifts in Immunity and Inflammation.” *The Journal of Immunology*, vol. 184, no. 8, Apr. 2010, pp. 4062–68. https://doi.org/10.4049/jimmunol.0903002.
27. Wang, Hui, and Jianping Ye. “Regulation of Energy Balance by Inflammation: Common Theme in Physiology and Pathology.” *Reviews in Endocrine and Metabolic Disorders*, vol. 16, no. 1, Dec. 2014, pp. 47–54. https://doi.org/10.1007/s11154-014-9306-8.
28. Kotlyarov, S. (2022). Involvement of the innate immune system in the pathogenesis of chronic obstructive pulmonary disease. *International Journal of Molecular Sciences*, *23*(2), 985. https://doi.org/10.3390/ijms23020985
29. Ma, M., Jiang, W., & Zhou, R. (2024). DAMPs and DAMP-sensing receptors in inflammation and diseases. *Immunity*, *57*(4), 752–771. https://doi.org/10.1016/j.immuni.2024.03.002
30. Roh, Jong Seong, and Dong Hyun Sohn. “Damage-Associated Molecular Patterns in Inflammatory Diseases.” *Immune Network*, vol. 18, no. 4, Jan. 2018, https://doi.org/10.4110/in.2018.18.e27.
31. “Extracellular Matrix Remodelling in COPD.” *European Respiratory Review*, vol. 29, no. 158, Nov. 2020, p. 190124. https://doi.org/10.1183/16000617.0124-2019.

32. Karakioulaki, M., Papakonstantinou, E., & Stolz, D. (2020). Extracellular matrix remodeling in COPD. *European Respiratory Review*, *29*(158), 190124. https://doi.org/10.1183/16000617.0124-2019
33. Gharib, Sina A., et al. "Matrix Metalloproteinases in Emphysema." *Matrix Biology*, vol. 73, Mar. 2018, pp. 34–51. https://doi.org/10.1016/j.matbio.2018.01.018.
34. Ito, Juliana T., et al. "Extracellular Matrix Component Remodeling in Respiratory Diseases: What Has Been Found in Clinical and Experimental Studies?" *Cells*, vol. 8, no. 4, Apr. 2019, p. 342. https://doi.org/10.3390/cells8040342.
35. "Nonspecific Interstitial Pneumonia (NSIP)." *Cleveland Clinic*, 1 May 2024, my.clevelandclinic.org/health/diseases/14804-nonspecific-interstitial-pneumonia-nsip.
36. Barnes, Peter J. "Oxidative Stress in Chronic Obstructive Pulmonary Disease." *Antioxidants*, vol. 11, no. 5, May 2022, p. 965. https://doi.org/10.3390/antiox11050965.
37. Mittal, M., Siddiqui, M. R., Tran, K., Reddy, S. P., & Malik, A. B. (2013). Reactive oxygen species in inflammation and tissue injury. *Antioxidants and Redox Signaling*, *20*(7), 1126–1167. https://doi.org/10.1089/ars.2012.5149
38. Yu, Yunfei, et al. "Roles of Reactive Oxygen Species in Inflammation and Cancer." *MedComm*, vol. 5, no. 4, Apr. 2024, https://doi.org/10.1002/mco2.519.
39. "The Persistent Inflammation in COPD: Is Autoimmunity the Core Mechanism?" *European Respiratory Review*, vol. 33, no. 171, Jan. 2024, p. 230137. https://doi.org/10.1183/16000617.0137-2023.
40. De Oliveira Rodrigues, Sarah, et al. "Mechanisms, Pathophysiology and Currently Proposed Treatments of Chronic Obstructive Pulmonary Disease." *Pharmaceuticals*, vol. 14, no. 10, Sept. 2021, p. 979. https://doi.org/10.3390/ph14100979.
41. Janeway, Charles A., Jr, et al. "Autoimmune Responses Are Directed Against Self Antigens." *Immunobiology - NCBI Bookshelf*, 2001, www.ncbi.nlm.nih.gov/books/NBK27155.
42. Chen, Linlin, et al. "Inflammatory Responses and Inflammation-associated Diseases in Organs." *Oncotarget*, vol. 9, no. 6, Dec. 2017, pp. 7204–18. https://doi.org/10.18632/oncotarget.23208.
43. Bird, Kim, and Jawedulhadi Memon. "Bronchiectasis." *StatPearls - NCBI Bookshelf*, 22 May 2023, www.ncbi.nlm.nih.gov/books/NBK430810.
44. Yuan, Xiaoyi, et al. "MicroRNA miR-223 as Regulator of Innate Immunity." *Journal of Leukocyte Biology*, vol. 104, no. 3, July 2018, pp. 515–24. https://doi.org/10.1002/jlb.3mr0218-079r.
45. "miR-223: A Key Regulator of Pulmonary Inflammation." *Frontiers in Medicine*, vol. 10, July 2023, https://doi.org/10.3389/fmed.2023.1187557.
46. Jiao, Peng, et al. "miR-223: An Effective Regulator of Immune Cell Differentiation and Inflammation." *International Journal of Biological Sciences*, vol. 17, no. 9, Jan. 2021, pp. 2308–22. https://doi.org/10.7150/ijbs.59876.
47. Atri, Chiraz, et al. "Role of Human Macrophage Polarization in Inflammation During Infectious Diseases." *International Journal of Molecular Sciences*, vol. 19, no. 6, June 2018, p. 1801. https://doi.org/10.3390/ijms19061801.

48. Parisi, Luca, et al. “Macrophage Polarization in Chronic Inflammatory Diseases: Killers or Builders?” *Journal of Immunology Research*, vol. 2018, Jan. 2018, pp. 1–25. https://doi.org/10.1155/2018/8917804.
49. Mamilos, Andreas, et al. “Macrophages: From Simple Phagocyte to an Integrative Regulatory Cell for Inflammation and Tissue Regeneration—A Review of the Literature.” *Cells*, vol. 12, no. 2, Jan. 2023, p. 276. https://doi.org/10.3390/cells12020276.
50. Li, Meirong, et al. “Macrophage Related Chronic Inflammation in Non-Healing Wounds.” *Frontiers in Immunology*, vol. 12, June 2021, https://doi.org/10.3389/fimmu.2021.681710.
51. “miR-223 Exerts Translational Control of Proatherogenic Genes in Macrophages.” *Circulation Research*, vol. 131, no. 1, May 2022, pp. 42–58. https://doi.org/10.1161/circresaha.121.319120.
52. Dang, Cong Phi, and Asada Leelahavanichkul. “Over-expression of miR-223 Induces M2 Macrophage Through Glycolysis Alteration and Attenuates LPS-induced Sepsis Mouse Model, the Cell-based Therapy in Sepsis.” *PLoS ONE*, vol. 15, no. 7, July 2020, p. e0236038. https://doi.org/10.1371/journal.pone.0236038.
53. Vlahos, Ross, and Steven eBozinovski. “Role of Alveolar Macrophages in Chronic Obstructive Pulmonary Disease.” *Frontiers in Immunology*, vol. 5, Jan. 2014, https://doi.org/10.3389/fimmu.2014.00435.
54. Hu, Tingting, et al. “The Activation of M1 Macrophages Is Associated With the JNK-m6A-p38 Axis in Chronic Obstructive Pulmonary Disease.” *International Journal of COPD*, vol. Volume 18, Oct. 2023, pp. 2195–206. https://doi.org/10.2147/copd.s420471.
55. King, Paul T. “Inflammation in Chronic Obstructive Pulmonary Disease and Its Role in Cardiovascular Disease and Lung Cancer.” *Clinical and Translational Medicine*, vol. 4, no. 1, July 2015, https://doi.org/10.1186/s40169-015-0068-z.
56. “Macrophages: Their Role, Activation and Polarization in Pulmonary Diseases.” *Immunobiology*, vol. 223, no. 4–5, Nov. 2017, pp. 383–96. https://doi.org/10.1016/j.imbio.2017.11.001.
57. Kim, Gun-Dong, et al. “Macrophage Polarization and Functions in Pathogenesis of Chronic Obstructive Pulmonary Disease.” *International Journal of Molecular Sciences*, vol. 25, no. 11, May 2024, p. 5631. https://doi.org/10.3390/ijms25115631.
58. “Smoking-Dependent Reprogramming of Alveolar Macrophage Polarization: Implication for Pathogenesis of Chronic Obstructive Pulmonary Disease.” *The Journal of Immunology*, vol. 183, no. 4, July 2009, pp. 2867–83. https://doi.org/10.4049/jimmunol.0900473.
59. Rezaiemanesh, A., Abdolmaleki, M., Abdolmohammadi, K., Aghaei, H., Pakdel, F. D., Fatahi, Y., Soleimanifar, N., Zavvar, M., & Nicknam, M. H. (2018). Immune cells involved in the pathogenesis of ankylosing spondylitis. *Biomedicine & Pharmacotherapy*, *100*, 198–204. https://doi.org/10.1016/j.biopha.2018.01.108
60. Tesmer, Laura A., et al. “Th17 Cells in Human Disease.” *Immunological Reviews*, vol. 223, no. 1, June 2008, pp. 87–113. https://doi.org/10.1111/j.1600-065x.2008.00628.x.
61. Vignali, Dario a. A., et al. “How Regulatory T Cells Work.” *Nature Reviews. Immunology*, vol. 8, no. 7, June 2008, pp. 523–32. https://doi.org/10.1038/nri2343.

62. Qi, Yanan, et al. “Inflammatory and Immune Mechanisms in COPD: Current Status and Therapeutic Prospects.” *Journal of Inflammation Research*, vol. Volume 17, Sept. 2024, pp. 6603–18. https://doi.org/10.2147/jir.s478568.
63. Freeman, Christine M., et al. “Cytotoxic Potential of Lung CD8+ T Cells Increases With Chronic Obstructive Pulmonary Disease Severity and With in Vitro Stimulation by IL-18 or IL-15.” *The Journal of Immunology*, vol. 184, no. 11, Apr. 2010, pp. 6504–13. https://doi.org/10.4049/jimmunol.1000006.
64. Lourenço, Juliana Dias, et al. “Th17/Treg Imbalance in Chronic Obstructive Pulmonary Disease: Clinical and Experimental Evidence.” *Frontiers in Immunology*, vol. 12, Dec. 2021, https://doi.org/10.3389/fimmu.2021.804919.
65. Barnes, Peter J. “Inflammatory Mechanisms in Patients With Chronic Obstructive Pulmonary Disease.” *Journal of Allergy and Clinical Immunology*, vol. 138, no. 1, May 2016, pp. 16–27. https://doi.org/10.1016/j.jaci.2016.05.011.
66. Xu, Wenjuan, et al. “MiR-223 Plays a Protecting Role in Neutrophilic Asthmatic Mice Through the Inhibition of NLRP3 Inflammasome.” *Respiratory Research*, vol. 21, no. 1, May 2020, https://doi.org/10.1186/s12931-020-01374-4.
67. “Chronic Inflammation.” *StatPearls - NCBI Bookshelf*, 7 Aug. 2023, www.ncbi.nlm.nih.gov/books/NBK493173.
68. Jarczak, Dominik, and Axel Nierhaus. “Cytokine Storm—Definition, Causes, and Implications.” *International Journal of Molecular Sciences*, vol. 23, no. 19, Oct. 2022, p. 11740. https://doi.org/10.3390/ijms231911740.
69. Jarczak, Dominik, and Axel Nierhaus. “Cytokine Storm—Definition, Causes, and Implications.” *International Journal of Molecular Sciences*, vol. 23, no. 19, Oct. 2022, p. 11740. https://doi.org/10.3390/ijms231911740.
70. Colarusso, C., Terlizzi, M., Molino, A., Pinto, A., & Sorrentino, R. (2017). Role of the inflammasome in chronic obstructive pulmonary disease (COPD). *Oncotarget*, *8*(47), 81813–81824. https://doi.org/10.18632/oncotarget.17850
71. Kiszałkiewicz, J. M., et al. “Evaluation of Selected IL6/STAT3 Pathway Molecules and miRNA Expression in Chronic Obstructive Pulmonary Disease.” *Scientific Reports*, vol. 11, no. 1, Nov. 2021, https://doi.org/10.1038/s41598-021-01950-8.
72. “Over-expression of miR-223 Induces M2 Macrophage Through Glycolysis Alteration and Attenuates LPS-induced Sepsis Mouse Model, the Cell-based Therapy in Sepsis.” *PLoS ONE*, vol. 15, no. 7, July 2020, p. e0236038. https://doi.org/10.1371/journal.pone.0236038.
73. Chan, Stanley M. H., et al. “Pathobiological Mechanisms Underlying Metabolic Syndrome (MetS) in Chronic Obstructive Pulmonary Disease (COPD): Clinical Significance and Therapeutic Strategies.” *Pharmacology & Therapeutics*, vol. 198, Feb. 2019, pp. 160–88. https://doi.org/10.1016/j.pharmthera.2019.02.013.
74. Hough, Kenneth P., et al. “Airway Remodeling in Asthma.” *Frontiers in Medicine*, vol. 7, May 2020, https://doi.org/10.3389/fmed.2020.00191.
75. Pahwa, Roma, and Ishwarlal Jialal. “Atherosclerosis.” *StatPearls - NCBI Bookshelf*, 8 Aug. 2023,

www.ncbi.nlm.nih.gov/books/NBK507799/#:~:text=Atherosclerosis%20is%20a%20chronic%20inflammatory,(ASCVD)%20as%20a%20result.
76. Lightbody, Richard James, et al. “MicroRNA Sequences Modulating Inflammation and Lipid Accumulation in Macrophage ‘Foam’ Cells: Implications for Atherosclerosis.” *World Journal of Cardiology*, vol. 12, no. 7, July 2020, pp. 303–33. https://doi.org/10.4330/wjc.v12.i7.303.
77. Aziz, Faisal, et al. “Relevance of miR-223 as Potential Diagnostic and Prognostic Markers in Cancer.” *Biology*, vol. 11, no. 2, Feb. 2022, p. 249. https://doi.org/10.3390/biology11020249.
78. Haneklaus, M., et al. “miR‐223: Infection, Inflammation and Cancer.” *Journal of Internal Medicine*, vol. 274, no. 3, June 2013, pp. 215–26. https://doi.org/10.1111/joim.12099.
79. Kay, Jennifer, et al. “Inflammation-induced DNA Damage, Mutations and Cancer.” *DNA Repair*, vol. 83, July 2019, p. 102673. https://doi.org/10.1016/j.dnarep.2019.102673.
80. Kidane, Dawit, et al. “Interplay Between DNA Repair and Inflammation, and the Link to Cancer.” *Critical Reviews in Biochemistry and Molecular Biology*, vol. 49, no. 2, Jan. 2014, pp. 116–39. https://doi.org/10.3109/10409238.2013.875514.
81. Kay, J., Thadhani, E., Samson, L., & Engelward, B. (2019). Inflammation-induced DNA damage, mutations and cancer. *DNA Repair*, *83*, 102673. https://doi.org/10.1016/j.dnarep.2019.102673
82. *Linking COPD and Lung Cancer | Sierra Hematology & Oncology*. www.sierrahemonc.com/blog/copd-and-lung-cancer-navigating-the-connection-for-a-better-tomorrow/#:~:text=Some%20experts%20believe%20that%20both,airways%20more%20susceptible%20to%20carcinogens.
83. Favero, Andrea, et al. “The Many Facets of miR‐223 in Cancer: Oncosuppressor, Oncogenic Driver, Therapeutic Target, and Biomarker of Response.” *Wiley Interdisciplinary Reviews - RNA*, vol. 12, no. 6, May 2021, https://doi.org/10.1002/wrna.1659.
84. Boukhenouna, Samia, et al. “Reactive Oxygen Species in Chronic Obstructive Pulmonary Disease.” *Oxidative Medicine and Cellular Longevity*, vol. 2018, no. 1, Jan. 2018, https://doi.org/10.1155/2018/5730395.
85. Zhang, W., Xiao, D., Mao, Q., & Xia, H. (2023). Role of neuroinflammation in neurodegeneration development. *Signal Transduction and Targeted Therapy*, *8*(1). https://doi.org/10.1038/s41392-023-01486-5
86. Colonna, Marco, and Oleg Butovsky. “Microglia Function in the Central Nervous System During Health and Neurodege
87. neration.” *Annual Review of Immunology*, vol. 35, no. 1, Feb. 2017, pp. 441–68. https://doi.org/10.1146/annurev-immunol-051116-052358.
88. Qin, Jun, et al. “Microglia Activation in Central Nervous System Disorders: A Review of Recent Mechanistic Investigations and Development Efforts.” *Frontiers in Neurology*, vol. 14, Mar. 2023, https://doi.org/10.3389/fneur.2023.1103416.
89. Kinney, Jefferson W., et al. “Inflammation as a Central Mechanism in Alzheimer’s Disease.” *Alzheimer S & Dementia Translational Research & Clinical Interventions*, vol. 4, no. 1, Jan. 2018, pp. 575–90. https://doi.org/10.1016/j.trci.2018.06.014.

90. Agusti, A. G. N. “Systemic Effects of Chronic Obstructive Pulmonary Disease.” *Proceedings of the American Thoracic Society*, vol. 2, no. 4, Nov. 2005, pp. 367–70. https://doi.org/10.1513/pats.200504-026sr.
91. Torres-Sánchez, Irene, et al. “Cognitive Impairment in COPD: A Systematic Review.” *Jornal Brasileiro De Pneumologia*, vol. 41, no. 2, Apr. 2015, pp. 182–90. https://doi.org/10.1590/s1806-37132015000004424.
92. Larrea, Ane, et al. “Neuroinflammation in the Evolution of Motor Function in Stroke and Trauma Patients: Treatment and Potential Biomarkers.” *Current Issues in Molecular Biology*, vol. 45, no. 11, Oct. 2023, pp. 8552–85. https://doi.org/10.3390/cimb45110539.
93. Moore, L. B., Sawyer, A. J., Saucier-Sawyer, J., Saltzman, W. M., & Kyriakides, T. R. (2016). Nanoparticle delivery of miR-223 to attenuate macrophage fusion. *Biomaterials*, *89*, 127–135. https://doi.org/10.1016/j.biomaterials.2016.02.036
94. Seyhan, Attila A. “Trials and Tribulations of MicroRNA Therapeutics.” *International Journal of Molecular Sciences*, vol. 25, no. 3, Jan. 2024, p. 1469. https://doi.org/10.3390/ijms25031469
95. Adeloye, D., Song, P., Zhu, Y., Campbell, H., Sheikh, A., & Rudan, I. (2022). Global, regional, and national prevalence of, and risk factors for, chronic obstructive pulmonary disease (COPD) in 2019: a systematic review and modelling analysis. *The Lancet Respiratory Medicine*, *10*(5), 447–458. https://doi.org/10.1016/s2213-2600(21)00511-7

The Industrial Revolution: From Artisanal Uniqueness and Craftsmanship to Reproducibility and Affordability By Taryn Hwang

Prior to industrialization, the majority of people lived in rural areas and either worked as farmers or artisans. Robert Marks describes how "rural families were free to decide what and how much to grow and how to allocate family labor on the farm" (Marks 109). Individual families generated what they needed for everyday survival and traded for what they could not create themselves. A farmer might have traded wheat with a blacksmith for horseshoes. But with the onset of industrialization during the 18th century, everything changed. The Industrial Revolution marked a notable shift in the value that consumers placed on goods: a transition from an emphasis on material quality and artisanship to a focus on perfection, reproducibility, and affordability. However, it also inspired counter-efforts that both emphasized financial struggles that workers experienced and sought to highlight the value and beauty of handmade goods that were lost to industrialization.

As manufacturers mechanized production, consumers began valuing perfection, reproducibility, and affordability over artisanal uniqueness. For instance, the implementation of the steam loom in the 1820s allowed for the production of uniform textiles, which meant that they could be repeatedly produced to exact specifications at a large scale. According to Richard Guest, "The best hand Weavers seldom produce a piece of uniform evenness…In Steam Looms, the lathe gives a steady, certain blow, and when once regulated by the engineer, moves with the greatest precision from the beginning to the end of the piece. Cloth made by these Looms, when seen by those manufacturers who employ hand Weavers, at once excites admiration and a consciousness that their own workmen cannot equal it. The increasing number of Steam Looms is a certain proof of their superiority over the Hand Looms" (Guest). In the first part of the passage, Guest distinguishes between the imperfection of the cloth produced by hand weavers and the perfection of the cloth produced by the steam loom. The fact that steam looms had apparently overcome hand looms in production illustrates the value that consumers now placed on perfection, as producers would not have continued to invest in more machines if their initial sales hadn't been successful. But most telling was the reaction from "manufacturers" that Guest described. The admiration and consciousness that the manufacturers felt toward cloth produced by the steam loom illustrates how this perfection was strongly viewed as positive. Furthermore, Guest's comparison to workmen not being able to equal this perfection highlights how perfection was valued *over* uniqueness, which also attaches a negative connotation to the uniqueness of products produced by hand. The ability to produce identical goods on a large scale reinforced the value placed on consistency and reproducibility in the minds of consumers, because they found comfort in knowing that they could expect a certain standard of quality and performance from their purchases every single time. Mechanization also allowed producers to sell their items at more affordable prices, which resulted in consumers expecting affordable goods. In terms of production, affordability was achieved through efficient and timely mass production. The ratio that "each machine will do as much work in twelve hours, as ten men can in that time do by hand" (Blackburn, Lobley) shows how much manual labor is required to fill the work of one

machine. This newfound efficiency brought about through mechanization allowed for goods to be sold at a more affordable price, and consumers sought out these affordable prices.

However, striving towards perfection, reproducibility, and affordability was not a uniform attitude. While they may have constituted a minority, there was a sufficient number of people who accounted for the backlash and counter-reformation towards mechanization and the Industrial Revolution. There were two main motivations for these counter-efforts. The first was that mechanization stripped away countless jobs of those "who did that business before machines were invented" (Blackburn, Lobley). While affordability was achieved through the use of machinery, this often came at the cost of individual laborers and their ability to find jobs. A petition written by Joseph Hepworth Thomas Lobley, Robert Wood Thos. Blackburn, and others in 1786 was addressed to those in support of the continuation of mechanization, and it reads that "Men of common sense must know, that so many machines in use, take the work from the hands employed in Scribbling…How are those men, thus thrown out of employ to provide for their families…Some say, Begin and learn some other business. - Suppose we do; who will maintain our families, whilst we undertake the arduous task" (Blackburn, Lobley). After losing their previous jobs to machines, people could no longer provide for their families. They had to learn new work, but while doing so, who would look after their children and family? Ultimately, mechanization took away the pre-industrial jobs of countless people who now needed to adapt by drastically changing their lifestyle. These new struggles received pushback, as evidenced by this petition. The description of these new industrial tasks as "arduous" also shows how the group of people who were now pushed out of their jobs (most of which were previously craftsmen) viewed these tasks. The pre-industrial process behind producing a vase, a chair, a piece of fabric, had often been passed down through family lineage, and held a sense of richness and meaning to each new generation of craftsmen. This sense of purpose in craft was now lost to mechanization, which leads to the second reason for pushback against industrialization: the belief that in seeking perfection and reproducibility, a lack of uniqueness and individuality meant that products had lost their "soul." Not only was there a loss of richness and lineage to each handmade piece, but each piece had also lost its own uniqueness and imperfections. Many people, especially craftsmen, strove to bring these aspects of handmade goods to light, as they believed they were what gave each product its "soul." This effort was exemplified by the Arts and Crafts movement, which was an international movement in the decorative and fine arts. The cultural significance of this movement lies in the fact that it introduced a different perspective for determining the value of goods: rather than the mechanical reproducibility and affordability of the objects themselves, people began to value *how* things were made and the individuality of items rather than simply the objects themselves.

While the Industrial Revolution marked a shift in the value that consumers placed on goods—a transition from an emphasis on material quality and artisanship to a focus on reproducibility and affordability—it also received pushback that brought to light the financial struggles that workers experienced and highlighted the richness of handmade goods that was lost to mechanization. The emphasis on reproducibility and affordability influenced the habits of consumers to revolve around being able to obtain these goods. It marked a change in people's

entire lifestyle. Where people had previously labored for survival, in agriculture or craft, for instance, laborers of industrial tasks now worked *in order* to purchase goods.

Works Cited

Guest, Richard. The Steam Loom. 1823. The Cotton Manufacture of Great Britain, edited by John Smith, London, 1825, pp. 45-48.

Hepworth, Joseph, et al. The Humble Address and Petition of Thousands, Who Labour in the Cloth Manufactory. 1786. Society and Politics in England, 1780-1960, edited by J. F. C. Harrison, Harper & Row, 1965, pp. 71-72.

Marks, Robert B. The Origins of the Modern World: Fate and Fortune in the Rise of the West. Rowman & Littlefield, 2007, p. 109.

The Double Duty Defender: Vitamin D's Multifaceted Role in Skin Cancer Progression and Prevention By Isabelle Yang

Abstract

Vitamin D is a fat-soluble hormone essential to many organisms. In humans, it is primarily synthesized through ultraviolet B (UVB) radiation and activated in the epidermis. Vitamin D levels are tightly regulated to prevent toxicity from excess or deficiency from insufficient levels. Widely recognized for its role in calcium and phosphate homeostasis, recent research has also uncovered vitamin D's potential as a tumor suppressor in skin cancer prevention and inhibition. Vitamin D interacts with the vitamin D receptor (VDR) to execute its functions, influencing various cancer-related pathways and genes. This review paper explores the interactions between vitamin D and VDR with two major skin cancer pathways—the Wnt-β-catenin pathway and the p53 gene pathway. These interactions promote unique anti-cancer mechanisms, including autophagy, DNA repair, anti-proliferation, and cell differentiation. However, this hypothesis is under debate, as some evidence suggests that vitamin D may have cancer-promoting functions under certain conditions. The complex relationship between vitamin D and skin cancer is highlighted by the paradox that while UV radiation increases the risk for skin cancer, it also drives the production of vitamin D, a potential inhibitor of the disease.

I. Introduction

A sculptor delicately carves into an untouched block of marble. With each meticulous tap of his chisel, he uncovers the concealed figure beneath. Gradually, his work reveals delicate, expressive, human-like features. As one moves closer to the work, small accidental chips and other imperfections can be observed. Despite this, he continues to refine his masterpiece.

Similarly, ever since the formation of the planet, ultraviolet (UV) radiation has been one of the major sculptors of biological life. Approximately 3.45 billion years ago, the earliest forms of life appeared on Earth as microscopic single-celled prokaryotes, known as bacteria and archaea [1]. These were the first organisms to experience UV radiation through sunlight which gave them the ability to mutate and diversify. 1.3 billion years later, cyanobacteria in the ocean were the first to harness solar energy through photosynthesis, a process that paved the way for higher forms of life to develop on Earth [2]. As prokaryotes gave rise to more complex eukaryotes, these organisms continued to evolve into distinct species [3], likewise to how a sculptor constantly refines a sculpture. Along the road of evolution, phytoplankton and zooplankton arose and became the first organisms to produce vitamin D catalyzed by absorbing UV radiation which aided in protecting their DNA, RNA, and proteins from UV damage [4]. However, UV radiation’s ability to mutate DNA has not always been beneficial, as cancer is a potential consequence. Nonetheless, UV radiation has been a powerful force in shaping biological systems, influencing their biochemical, physiological, and ecological traits throughout evolutionary history [5].

The Paradox Involving Skin Cancer and Vitamin D

While UV radiation significantly increases the risk of skin cancer, this double edged sword is also necessary for our bodies, as it drives the synthesis of vitamin D in vivo. Interestingly, remarkable research over the past 40 years has uncovered vitamin D's anti-cancer effects, not only in preventing the formation of skin cancer but also in inhibiting its progression. The interconnection of these two processes, the formation and progression of skin cancer and the synthesis of vitamin D, presents an intriguing paradox:

Knowing that exposure to UV allows for more vitamin D production, it is believed that with more UV exposure, the body will become more resistant to cancer. However, this is also contradictory since more exposure to UV increases the risk of skin cancer. How does vitamin D interact with cancer cells? Why is vitamin D unable to inhibit cancers completely if it has anti-cancer effects? Is there a balanced amount of UV exposure to ensure sufficient vitamin D levels and put a person at a lesser risk of skin cancer?

II. Hallmarks of Skin Cancer

Overview of Skin Cancer

One of the most critical biological phenomena related to UV radiation is skin cancer. There is strong epidemiological and molecular data support that UV exposure is linked to all skin cancers [6]. Nearly 90% of nonmelanoma skin cancers and 65% of melanoma skin cancer cases are directly linked to UV exposure, and it is the most prevalent carcinoma in the United States [6]. However, it's important to emphasize that skin cancer is also one of the most preventable cancers. By implementing proper sun protection and safety measures, the risk of contracting skin cancer can be greatly diminished [7].

Hallmarks of Cancer

Cancerous cells and tissues are characterized by six fundamental traits that sustain tumor growth and facilitate metastatic dissemination: sustaining proliferative signaling, evading growth suppressors, enabling replicative immortality, resisting cell death, activating invasion and metastasis, and inducing angiogenesis. Sustaining proliferative signaling allows cancer cells to bypass regulatory mitosis checkpoints and reproduce rapidly. Evading growth suppressors involves inactivating tumor suppressor genes and other growth-limiting factors. By enabling replicative immortality, cancer cells can replicate indefinitely, exceeding normal cellular aging and apoptosis. These cells resist programmed cell death by deactivating apoptosis pathways, ensuring their continued survival. Their ability to invade and metastasize allows for the spread of malignant cells to distant tissues. Additionally, inducing angiogenesis ensures tumors obtain the necessary nutrients and oxygen while effectively removing waste products, supporting continued growth. Each of these hallmarks demonstrates the complexity and resilience of cancer, making it a formidable challenge to combat [8].

Etiology

Skin cancer is caused by both genetic risk factors and environmental risk factors.

Genetic factors strongly affect one's likelihood of inheriting skin cancer. Those with physical traits like fair skin, blonde or red hair, many moles, freckles, or skin that reddens and burns easily under the sun are more inclined to have skin cancer [9]. Melanin production in skin cells is also a trait that plays a large role in determining one's skin's sensitivity and skin cancer risk, since melanin is a pigment that absorbs UV radiation and protects epidermal cells from damage [10]. Additionally, family history or previous diagnosis of skin cancer also increases one's cancer risk [11]. Inherited gene diseases or mutations may also increase the risk of skin cancers, especially during childhood [12]. Immunosuppressed individuals will also be more susceptible to cancer as their immune system cannot recognize and destroy cancerous cells [13].

Most skin cancers are related to UV exposure. Chronic and/or high intensity exposure to UVB solar radiation may lead to tanning which indicates damage to the skin tissue, skin cells, and DNA [14], while sunburn indicates cell death [15]. UV exposure induces the formation of cancer cells through damaging DNA and creating mutations that produce altered proteins that dysregulate cellular processes, such as the cell cycle. Chronic and intense exposure to UV and any other carcinogen will induce the initiation, promotion, and progression of cancer [16].

There are three types of UV rays: UVA, UVB, and UVC. UVA has the longest wavelength (315-400 nm) and penetrates through the ozone layer in the atmosphere, past the epidermis, into the dermis. UVA contributes to premature skin aging. UVB has a shorter wavelength (280-315 nm), and some of it penetrates through the ozone into the epidermis, less deep than UVA. UVB stimulates the production of melanin in skin cells, resulting in a tanned appearance and damaged skin. Overexposure will result in erythema, inflammation, and pain in the form of a sunburn. Both UVA and UVB contribute to carcinogenesis. UVC rays have the shortest wavelength (100-280 nm) and are completely absorbed by the ozone layer and atmosphere, therefore they do not contribute to skin cancer [17].

Three Major Types of Skin Cancer

Skin cancer has three major types: basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and melanoma.

Basal cell carcinoma is the most common skin cancer as it accounts for approximately 80% of all skin cancer cases [18]. It has a slow growth rate and limited metastasis, usually only invading the dermis layer, making it the least dangerous skin carcinoma detected early. It begins in basal keratinocyte stem cells which are located in the basal cell layer or the bottom layer of the epidermis. The basal cell layer regenerates the upper parts of epidermis by pushing new cells up while the uppermost dead skin cells sheds away. BCC is often seen in areas that are chronically exposed to direct sunlight, such as the head, neck, and trunk [19].

Squamous cell carcinoma is the second most common non-melanoma skin cancer, constituting approximately 20% of skin cancer cases [20]. SCC is more aggressive and invasive than BCC, since it can metastasis into other tissues, most commonly lymph nodes, making it more lethal if not detected early. It originates in the squamous keratinocyte cells in the epidermis, which are constantly being pushed up and shed away. A common precursor for SCC is actinic keratosis which are skin lesions caused by UV damage [21]. SCC commonly occurs in sun

damaged areas, similar to BCC, including the backs of the hands, head, neck, trunk, and the arms [20].

Melanoma is the most lethal skin cancer as it tends to spread to lymph nodes and metastasize quickly, however it is very uncommon as it represents less than 5% of skin cancer cases [22]. It originates in the melanocytes embedded in the basal cell layer in the epidermis. Melanocytes are responsible for producing melanin. The amount of melanin produced increases with the exposure of UVA and UVB rays as a defense mechanism to absorb harmful radiation [10]. Melanoma can originate in areas not prone to sun exposure, like in the palms of hands, but they are most commonly found on the backs and shoulders of men, legs of women, and the head and neck for both men and women [22].

Pathophysiology and Major Signaling Pathways

Most SCC cases are caused by mutations in the p53 tumor suppressor gene in the p53 pathway that inactivates the inhibition of proliferation in keratinocytes and DNA repair [23]. BCC is most commonly associated with mutations in the PTCH1 gene, a tumor suppressor in the hedgehog signaling pathway that regulates proliferation in keratinocytes, as well as the P53 gene [24]. Melanoma includes mutations in the CDKN2A gene, a tumor suppressor gene in the p53 and RB1 pathways that regulates proliferation, MCR1, a gene in the cAMP pathway that regulates the amount of melanin is produced by melanocytes, B-Raf, a gene in the MAPK pathway that regulates proliferation, and DNA repair genes that may result in xeroderma pigmentosum [25] [26] [27] [28].

Current Treatments

Current treatments for skin cancer primarily focus on the removal and/or destruction of cancer cells. These methods often include:

1. Surgical Removal: This involves excising the tumor and some surrounding healthy tissue to ensure complete removal of cancerous cells. Surgical options can range from simple excisions for small tumors to more complex procedures removing large or recurring tumors [29]. As seen in Mohs micrographic surgery, which is usually used for high risk nonmelanoma skin cancers, involves removing the cancer layer by layer and examining each layer under a microscope until no abnormal cells remain [29]. Curettage and electrodesiccation is another commonly used method for small tumors, where tumors are scraped off with a curette and cauterized by an electric current to prevent further growth and bleeding [29].
2. Phototherapy: This therapy is minimally invasive and includes photodynamic and photothermal therapy. Photodynamic therapy involves using a certain wavelength of light to activate topical medicine containing photosensitizers, resulting in the production of reactive singlet oxygen and free radicals that damage the cell's critical functions and ultimately destroy surrounding cancer cells [30]. Similarly, photothermal therapy involves shining a certain wavelength of light, however the light is converted into thermal

energy by the applied medicine containing photothermal transduction agents and used to induce rapid cell death in the area [30].

3. Immunotherapy: Immunotherapy uses antigen-antibody interactions to eradicate cancer cancer cells. Substances made in the body or a laboratory, like interferon, imiquimod, and immune checkpoint inhibitors, are injected, consumed, or absorbed topically. They function to support or restore the immune system's ability to target cancer cells more effectively. Often used to treat SCC, the immune checkpoint inhibitors known as the PD-1 inhibitors and PD-L1 inhibitors can block PD-1 proteins on T cells from binding to PD-L1 proteins on certain cancer cells. This allows the T cell to identify and destroy the cancer cell [29].
4. Chemotherapy: These drugs, often referred to as chemotherapeutic agents, work by killing, slowing, and/or stopping the division of rapidly dividing cells including cancer cells [29]. Some drugs are systemic, affecting cells throughout the body, while others are topical, applied directly to the skin. BCC and SCC are often treated with topical medicine [29]. Since chemotherapy targets all rapidly dividing cells, there are many side effects including hair loss and damage to organs [31] [32].
5. Radiotherapy: Radiotherapy is a successful and widely used treatment for skin cancer. This treatment uses x-ray, gamma-rays, protons, and other forms of high-energy radiation to kill cancer cells. Radiation therapy is particularly useful for tumors that are difficult or not recommended to be removed surgically or for patients who cannot undergo surgery as it ensures tissue preservation [29].
6. Targeted Therapy: Through the use of medical drugs in targeted therapy, the specific genetic mutations or proteins in cancerous cells are targeted and cell proliferation is able to be stopped [33]. By targeting only the harmful mutated genes in cancer cells, surrounding healthy tissue will be protected and result in less side effects [33]. However, targeted therapy is specific to only mutations in the B-Raf gene, MEK gene, C-KIT gene currently, so it is not applicable to all cancer patients as of now [34].

These treatments can be tailored based on the type of skin cancer, its stage, and the overall health of the patient. Surgical procedures for skin cancer are most often paired with adjuvant therapies to ensure that the cancer does not return. Early detection and treatment are crucial for improving outcomes and reducing the risk of recurrence [35].

III. Metabolism and Regulation of Vitamin D Levels

Overview of Vitamin D

Over a century ago, vitamin D was first discovered and became known as the "sunshine vitamin" after it was revealed to be produced in the skin in response to sunlight. This discovery was monumental as it identified the underlying cause of the vitamin D deficiency endemic that spread throughout the world during the Industrial Revolution as populations moved from farms to cities. Its exact mechanism and biological roles were understood over the next several years through various studies [36].

Vitamin D is a fat-soluble hormone that primarily exists in two chemical forms: vitamin D2 (ergocalciferol) and vitamin D3 (cholecalciferol). Vitamin D2 is derived from plants and fungi, while Vitamin D3 is sourced naturally from animal products and synthesized in the human skin upon exposure to sunlight [37].

Vitamin D functions biologically to regulate calcium and phosphorus homeostasis in the blood, which is crucial for bone health, muscle function, balance, and other crucial metabolic processes. Immunomodulatory functions include regulating the immune system and protecting against MS, type 1 diabetes, and psoriasis. Cardiovascular effects include renin-angiotensin regulation and decreased the risk for hypertension, type 2 diabetes, and heart failure [38].

Source and Synthesis

Sufficient levels of vitamin D are mainly obtained through dietary intake of vitamin D and endogenous synthesis of vitamin D triggered by UVB radiation. Dietary intake contributes approximately 10% of vitamin D, while synthesis in the skin accounts for 90% [39]. The two inactive chemical forms of vitamin D, cholecalciferol and ergocalciferol, can both be obtained through consumption. Fatty fish, egg yolks, milk, beef liver, reindeer lichen, fish liver oil, and UV activated mushrooms are strong sources of Vitamin D. Fortified foods, like milk and orange juice, and dietary supplements are also viable sources of vitamin D [40].

In vitamin D3 synthesis, 7-dehydrocholesterol from the cholesterol biosynthesis pathway goes through a photochemical reaction in the epidermis using UVB radiation where it is converted to pre-vitamin D3 and then inactive form of vitamin D3 (cholecalciferol) [37].

Metabolism and Regulation

Cholecalciferol and ergocholecalciferol are both biologically inert molecules, so they require a conversion to an active form. After cholecalciferol is synthesized in the skin or ingested, it travels into the bloodstream where it circulates by binding to the vitamin D binding protein (DBP), eventually reaching the liver. In the liver, the enzyme 25-hydroxylase encoded from the gene CYP2R1 converts vitamin D, in a hydroxylation reaction, into 25-hydroxycholecalciferol (calcifediol). Calcifediol is the most abundant circulating form of vitamin D and is used as a biomarker to test vitamin D levels [37]. It then travels to the proximal tubular epithelial cells of the kidneys where the enzyme 1α-hydroxylase, encoded by the gene CYP27B1, converts it through hydroxylation into 1,25-dihydroxycholecalciferol (calcitriol), the active form of vitamin D [37].

A similar process happens for vitamin D2, but its active form is not as abundant or as potent as vitamin D3 [41].

This entire metabolic process has also been discovered to occur cutaneously in the epidermis as well as in other tissues besides the liver and kidneys [37]. Besides being the site of vitamin D synthesis, keratinocytes also have the same enzyme machinery needed to activate cholecalciferol into calcitriol and have vitamin D receptors (VDR) to respond to calcitriol. Many cutaneous functions are mediated this way, including inhibition of proliferation, stimulation of differentiation, promotion of innate immunity, and the promotion of the hair follicle cycle. [42].

The production of calcitriol is regulated by the 1α-hydroxylase enzyme (CYP27B1) and must be activated by low phosphate levels and the parathyroid hormone (PTH), which is triggered by low calcium levels. It is also inhibited by fibroblast growth factor 23 (FGF23), a hormone that is produced and secreted by the osteocytes and osteoblasts in bones, which is triggered by high levels of calcitriol, calcium, and phosphate [43].

Calcitriol's most notable role is to increase calcium and phosphate levels in the blood when these levels are too low. After binding to the vitamin D receptor (VDR), it increases absorption of calcium and phosphate in the gastrointestinal tract and reabsorption in the kidneys, returning calcium and phosphate into the blood. It also increases osteoclast activity, which breaks down bone, allowing for calcium and phosphate to return into the blood. The breaking down of bone allows calcium and phosphate in the blood to mineralize and deposit into the bone, using osteoblasts to rebuild and strengthen new bone tissue [44].

Degradation and Inactivation

Besides regulating the production of vitamin D, degradation and inactivation of vitamin D is also used to ensure normal amounts in the body and avoid vitamin D toxicity through a negative feedback pathway. The CYP24A1 gene produces two different enzymes that initiate two different degradation pathways: 24-hydroxylase and 23-hydroxylase. Different species have differing 24-hydroxylase to 23-hydroxylase activity ratios. In humans, 24-hydroxylase activity predominates over 23-hydroxylase [37]. The CYP24A1 gene is induced by high levels of FGF23, calcitriol, and/or calcifediol. Meanwhile, CYP24A1 is inhibited by PTH and low calcium levels [43]. The enzyme 24-hydroxylase will catabolize active calcitriol and inactive calcifediol into 1,24,25-trihydroxycholecalciferol (calcitetrol) and 24,25-dihydroxycholecalciferol, respectively, and finally convert the metabolites into calcitroic acid to be excreted. Meanwhile, 23-hydroxylase degradates calcitriol and calcifediol into 1α,25-dihydroxycholecalciferol-26,23-lactone (calcitriol lactone) that will be excreted as well [43].

Vitamin D Deficiency and Toxicity

Vitamin D deficiency affects people of all ages and ethnicities. Around 1 billion of the world population is vitamin D deficient. To add, 50% of the worldwide population is vitamin D insufficient. This is mainly caused by lack of exposure to UV radiation or sunlight. The composition of people's skin also determines how much UV is absorbed and can be used to synthesize vitamin D. People with darker, melanin-rich skin tend to produce less vitamin D because the melanin absorbs the UVB available, compared to those with lighter skin and less melanin. Vitamin D deficiencies may also be related to medication, malabsorption syndromes, chronic liver disease, or insufficient consumption of vitamin D in the diet [45]. A rare mutation in the CYP27B1 gene that compromises the conversion to active calcitriol was recently discovered to contribute to vitamin D deficiency as well [46].

In an individual with deficient levels of vitamin D, there is decreased intestinal calcium and phosphorus absorption which causes hypocalcemia. To combat low calcium, excess amounts of PTH is released and causes secondary hyperparathyroidism. Secondary hyperparathyroidism

increases osteoclast activity and breaks down the bone to release calcium and phosphate. However, calcium builds up in the blood due to decreased calcium absorption, while phosphate gets excreted, resulting in dropping levels of phosphate, leading to phosphaturia. With no phosphate to mineralize with calcium, bone gets broken down without rebuilding, causing the bone to become weak [45]. These bone issues lead to diseases known as rickets in children and osteomalacia in adults. In rickets, children's collagen rich bones fail to mineralize, resulting in stunted bone growth and bendy or deformed bones [47]. In osteomalacia, adult bone density decreases, and bones become weak. Osteomalacia can potentially develop into osteoporosis where bones become brittle over time [48].

To increase vitamin D levels in one's body, receiving plentiful sun exposure, taking vitamin D supplements, and eating fortified foods are the most common and most efficient ways [49].

Vitamin D toxicity or hypervitaminosis is relatively rare. Sun overexposure usually will not result in vitamin D toxicity as the body will regulate the vitamin D pathway. Eating foods with naturally occurring vitamin D is also unlikely to cause vitamin D toxicity as these foods contain only trace amounts of vitamin D. However, overdosing vitamin D supplements or consuming fortified foods with excess amounts of vitamin D are common causes [50]. Meanwhile, mutations in the vitamin D pathway, such as an abnormal CYP24A1 gene which results in inhibited degradation of excess vitamin D, may lead to vitamin D toxicity [51].

Vitamin D toxicity causes hypercalcemia through inducing increased bone resorption leading to abnormally high amounts of calcium in the bloodstream and increased intestinal calcium absorption. Calcinosis, when calcium builds up in soft tissue, may occur and affect the kidney, bones, gastrointestinal system, cardiac system, and central nervous system. Symptoms include calcium based kidney stones, polyuria, polydipsia, bone pain and weakness, abdominal pain, peptic ulcers, arrhythmia, and depression in the nervous system [52].

According to the Endocrine Society, ideal vitamin D levels in the body should be approximately 40-60 ng/mL. Infants, children and adolescents, and adults should have daily intakes of 400-1000 IU, 600-1000 IU, and 1500-2000 IU, respectively [53].

IV. Mechanisms of Vitamin D's Inhibitory Effects on Skin Cancer

Vitamin D Receptor (VDR) and Gene Regulation

The vitamin D receptor (VDR) is a nuclear receptor and transcription factor that mediates the diverse functions of vitamin D [37]. While VDR is typically associated with target tissues in the liver and kidneys, it is also present in various other tissues, including the epidermis [42]. The VDR plays a pivotal role in modulating numerous biological processes, such as calcium and phosphate homeostasis, immune response, and cell proliferation and differentiation [38].

When calcitriol (the active form of vitamin D) reaches its target tissue, it binds to VDR in the cytoplasm, inducing a conformational change. This activated VDR heterodimerizes with a nuclear hormone receptor from the retinoid X receptor (RXR) family. The VDR-RXR complex then translocates to the nucleus. Inside the nucleus, it binds to specific sequences in the DNA known as vitamin D response elements (VDREs) that are located near the promoter regions of

VDR-regulated genes. This binding initiates a complex cascade of interactions with additional coregulators [37].

Coregulators, including coactivators and corepressors, are essential for modulating the transcriptional activity of the VDR-RXR complex. Coactivators promote, while corepressors inhibit the transcription of target genes [37]. Furthermore, other transcription factors can also interact with the VDR-RXR complex, facilitating the assembly of a multi-protein transcriptional complex. This precise and dynamic regulation ensures that the appropriate gene expression is achieved, which is crucial for maintaining cellular homeostasis and responding to physiological signals [37]. By affecting the transcription of various genes, the VDR-mediated signaling pathway influences numerous physiological processes, including bone formation, immune function, and skin health [54].

The coactivator factors, including steroid receptor coactivator (SRC) 1, SRC2, SRC3, and CREB-binding protein p3000, work to reconfigure the chromatin structure, making it more accessible and allowing the VDR complex to bind efficiently to the transcription start site of the target gene [37]. This chromatin remodeling process is crucial for facilitating the recruitment of RNA polymerase II, which is carried out by SWI-SNF ATP dependent chromatin remodeling complex, methyltransferases, and the DRIP complex as well as the general transcription machinery necessary for initiating transcription [37]. In addition to chromatin remodeling, coactivators also interact directly with RNA polymerases and other transcription factors, enhancing the efficiency and specificity of transcriptional activation. For example, the expression of certain coactivators in immune cells can enhance the role of vitamin D in modulating immune responses, while in osteoblasts, other coactivators may promote bone formation [54].

Conversely, corepressors function to repress transcription by recruiting histone deacetylases (HDACs), methyl transferases (MT), or other chromatin-modifying enzymes that condense the chromatin structure, making it less accessible to the transcriptional complex. This action prevents the binding of RNA polymerase II and inhibits gene transcription [37].

The interaction between VDR, its coregulators, and the transcriptional machinery demonstrates the complexity and versatility of vitamin D signaling. This network of interactions allows VDR to modulate a wide range of biological processes, from calcium homeostasis and bone metabolism to immune function and cellular differentiation [38]. Understanding these interactions provides significant insight into the potential therapeutic applications of vitamin D in treating various diseases and conditions [55].

Studies have uncovered an inverse correlation between VDR expression and the progression and prognosis of melanoma. In an experiment utilizing CRISPR gene-editing technology, researchers precisely deleted the VDR gene in human melanoma cells, ensuring the complete absence of VDR proteins. These VDR knockout (KO) cells exhibited enhanced proliferation in monolayer culture, increased spheroid and colony formation, and significantly greater migration compared to the scramble control group (which retained VDR expression). When activated forms of vitamin D were introduced into both the VDR KO culture and the scramble melanoma cell culture, distinct differences emerged. In the VDR KO cells, there was no change in proliferation or migration rates and only a slight decrease in colony and spheroid formation. Conversely, the scramble control cells exhibited a minor yet significant inhibition in

proliferation, along with a marked reduction in colony and spheroid formation and migration rates. These findings suggest that vitamin D functions as a melanoma tumor suppressor, with its effects being mediated through the VDR. The absence of VDR in KO cells resulted in resistance to the anti-proliferative and anti-migratory effects of vitamin D, highlighting the critical role of VDR in modulating these processes. The presence of VDR correlates with the inhibition of cancer growth and suppression, so it may potentially be able to be used as a therapeutic target in melanoma treatment [56].

In addition, VDR's role as a tumor suppressor is further supported by a study using a group of VDR null mice with a mutated APC gene (Apc1638N/+Vdr−/−). The APC gene is a tumor suppressor gene that when altered often results in colorectal cancers. These mice did not develop GI or extra-colonic tumors, but the loss of VDR decreased the intestinal tumor latency period, increased the incidence of extra-colonic lesions, and accelerated colon cancer progression to adenocarcinoma compared to wild type (WT) control group mice. When these mice were treated with the carcinogen azoxymethane (AOM), they developed squamous cell carcinoma of the anus (anal cancer) [57].

However, VDR is not always a tumor suppressor. Pancreatic adenocarcinoma patients with elevated VDR levels exhibited a significantly reduced overall survival since pancreatic adenocarcinoma cells with overexpression of VDR promoted macrophage polarization towards the M2 phenotype (pro-tumor) and recruitment rather than the M1 phenotype (anti-tumor) in vitro and vivo, which ultimately led to increased M2 macrophage infiltration and worsened clinical outcomes [58]. Furthermore, in the largest-ever randomized clinical trial testing vitamin D for cancer prevention, VITamin D and OmegA-3 TriaL (VITAL), the results showed that the vitamin D supplement significantly reduced cancer mortality, but did not aid in reducing cancer incidence [59].

Some studies have reported positive anticancer effects of vitamin D and VDR, while others have shown no effect or even increased cancer risk. The available evidence is debatable.
Crosstalk Between VDR and the Wnt-β-catenin Pathway

The Wnt/β-catenin signaling pathway is a complex mechanism that regulates various biological processes, playing pivotal roles in development, tissue homeostasis, bone and teeth maintenance, cancer, and other diseases.

The Wnt/β-catenin signaling pathway is crucial in embryonic development, particularly in the morphogenesis of oral organs like teeth, taste buds, and taste papillae. Its dynamic activation is necessary for normal development. Stage-dependent roles in tooth development are important since transient inhibition or overactivation can disrupt normal morphogenesis [60]. The Wnt/β-catenin pathway is also significant for bone mass maintenance, predominantly by regulating the activity of osteoblasts and osteoclasts, which are essential for skeletal development, formation, and health [61].

In tissue homeostasis, the Wnt/β-catenin signaling pathway is vital for regulating adult stem cell homeostasis through promoting proliferation, differentiation, and self renewal abilities in tissue specific stem cells, including the epidermis, liver, stomach, mammary gland, and intestine epithelium [62]. Hyperactivation of this pathway is observed in almost all colorectal cancers and is also relevant in breast, liver, lung, and prostate cancers. This pathway is heavily

implicated in cancer development and progression, and modulation of this pathway offers potential strategies for cancer therapy [63].

This pathway also plays a significant role in skin cancer development. Studies have shown many epidermal genes induced by β-catenin, including the stem cell marker keratin 15, contain vitamin D response elements (VDREs), meaning that the VDR signaling pathway can regulate the Wnt-β-catenin signaling pathway as well as the same genes. Hair follicle formation in the adult epidermis is induced by β-catenin which requires VDR. However, human trichofolliculomas, or hair follicle tumors, are characterized by high nuclear β-catenin and subsequent high VDR expression, while in high beta-catenin and low VDR conditions, BCC formations occur. EB1089, a Vitamin D analogue, has therapeutic potential by synergizing with β-catenin to stimulate hair follicle differentiation while preventing trichofolliculoma formation in the skin [64].

VDR is central to the regulation of β-catenin activity in the epidermis as it interacts with β-catenin to modulate gene expression, including genes involved in cell proliferation and differentiation. Vitamin D, through its receptor VDR, negatively regulates the Wnt/β-catenin pathway, meaning that reduced VDR expression can lead to increased pathway activation. Studies showed dysregulation of the vitamin D pathway, characterized by suppression of VDR expression in keratinocytes, enhances β-catenin-mediated transcription of pro-proliferation genes, including cyclin D1 and cMyc, resulting in excessive cell proliferation. Abnormal regulation also decreases the formation of β-catenin-E-cadherin cell-cell junctions, impairing keratinocyte differentiation. This increased proliferation and hindered differentiation promotes tumor growth, significantly impacting keratinocyte transformation and increasing susceptibility to BCC and other skin carcinomas in VDR-null mice [64].

It has been further confirmed that VDR is a TCF/Lef-independent transcriptional effector of the Wnt-pathway β-catenin. Genomic and non-genomic mechanisms involving the vitamin D, VDR, and vitamin D derivatives inhibiting β-catenin signaling in BCC offer additional evidence for Vitamin D's protective role against skin cancer [64].

The regulatory role of vitamin D through VDR is essential for maintaining normal activity of the Wnt/β-catenin pathway and is crucial for keratinocyte proliferation and differentiation. Dysregulation of this pathway contributes significantly to the onset and progression of skin carcinomas [64].

Crosstalk Between Vitamin D and p53 Signaling in Cellular Functions

The interactions between vitamin D and p53 signaling is crucial in regulating autophagy in skin cancer. This process involves the formation of autophagosomes, which engulfs portions of the cytoplasm, including pathogens, old organelles, and proteins that require degradation. To maintain cellular homeostasis and promote organelle regeneration, autophagosomes merge with lysosomes to create autophagolysosomes, also known as autolysosomes, to degrade cellular material [65].

Autophagy's role in cancer is dual-faceted, influencing both tumor suppression and cancer progression depending on the stage. In the early stages, autophagy acts as a tumor suppressor by digesting damaged organelles and potentially carcinogenic substances, thus

preventing the spread of damage, including DNA alterations. Conversely, in advanced stages, autophagy can facilitate tumor growth by providing nutrients and other resources for cell metabolism and growth and aiding angiogenesis, enabling tumor cells to endure stressful conditions and cancer treatments [66]. Notably, patients with tumors exhibiting high levels of autophagy often have poor clinical outcomes [67].

p53, a well-established tumor suppressor, is frequently mutated in skin cancers such as squamous cell carcinoma, basal cell carcinoma, and melanoma [68]. It plays a crucial role as a transcription factor in regulating many pathways, including DNA repair, apoptosis, autophagy, and the cell cycle [69]. The relationship between p53 and autophagy is complex, involving both activation and inhibition depending on the subcellular localization of the p53 protein. Nuclear p53 can act as a transcription factor and activate autophagy, apoptosis, and cell cycle arrest in response to stress, such as genotoxic stress. This activation can aid cells in surviving nutrient deprivation. Conversely, cytoplasmic p53 can operate at the mitochondria to inhibit autophagy and activate apoptosis [70].

How vitamin D and p53 interact is crucial in determining whether autophagy will induce apoptosis in damaged cells or provide cancer cells with survival. VDR as well as independent vitamin D can regulate the expression of the p53 family and its downstream gene expressions, such as MDM2, XPC, and DDB2, which are involved in cell growth, apoptosis, DNA repair, and autophagy [71] [72] [73] [74].

In healthy skin with functional p53, vitamin D and p53 may synergistically trigger autophagy to clear damaged organelles and cells, thus preventing tumorigenesis. Western blot analysis has confirmed the direct binding of VDR to p53 [75]. Additionally, VDR binds to the MDM2 P2 promoter, a key negative regulator of p53 activity, upregulating MDM2 expression levels in the presence of the p53 protein which activates the promoter [71]. Moreover, VDR also binds to VDREs in nucleotide excision repair (NER) genes as a transcription factor and regulates the expression of genes such as XPC and DDB2 [72] [76]. Studies have shown that vitamin D can induce autophagy in skin cells as a protective response against oxidative stress and DNA damage [77] [78]. These interactions may play a crucial role in vitamin D's ability to enhance the survival of skin cells following UV irradiation.

In a study involving TERT-immortalized human keratinocytes, it was found that the knockdown of p63, a gene that is a part of the p53 family, significantly reduced VDR and XPC expression, indicating a role for p63 in regulating these genes. Interestingly, the knockdown of VDR alone moderately reduced XPC mRNA levels while p63 and XPC protein expression were not impacted. The combined depletion of both p63 and VDR led to considerably slower removal of UV-induced DNA damage. This suggests that VDR can partially account for p63's effects on the nucleotide excision repair process [72].

In normal keratinocytes, XPC accumulates at DNA damage foci and actively participates in the nucleotide excision repair (NER) process, peaking within 15 minutes and gradually decreasing over 90 minutes. However, in keratinocytes depleted of either p63 or VDR, XPC over-accumulates at the sites of damage and exhibits slower dissociation compared to control cells, indicating a disruption in the repair process. The concurrent knockdown of both VDR and p63 results in even slower XPC release from DNA damage sites, thereby further impairing the

repair process. This suggests a complex interaction between VDR and p63 in regulating the efficacy of NER in response to UV-induced DNA damage [72].

In skin cancer cells, approximately half of all skin cancers in individuals without DNA repair deficiencies harbor p53 mutations. In contrast, up to 90% of skin cancers in individuals with xeroderma pigmentosum (XP), a genetic condition with decreased ability to repair DNA after UV damage, exhibit p53 mutations. Therefore, the presence of mutated p53 compromises the regulation of crucial cellular processes, leading to impaired cellular function and tumor progression [79].

The p63 protein as a transcription factor plays a vital role in the formation and proliferation of the epidermis and other stratified epithelia. Notably, the isoform ΔNp63α is found at significantly higher levels in non-melanoma skin cancers, including BCC and SCC, and their precursors, compared to normal, non-cancerous skin [80]. Research has demonstrated a positive correlation between expression levels of p63 and VDR where p63 and VDR levels are both highest in BCC, followed by SCC and SCC precursors. This correlation suggests a potential association between p63 and VDR in skin carcinogenesis [81]. Moreover, it has been observed that the vitamin D-VDR-mediated increase in ΔNp63α protein levels can be reversed by inhibiting both Akt and p38 activation pathways [81].

Conversely, p53 mutations can convert vitamin D into an antiapoptotic agent. It has been shown that mutant p53 can interact both functionally and physically with VDR and increase the nuclear accumulation of VDR, modulating the expression of VDR-regulated genes [82]. ΔNp63 can also regulate VDR signaling through competitive binding to target genes such as p21Waf1/Cip1 [83]. The status of the p53 gene, whether it is mutated or wild type, determines how it interacts with VDR and vitamin D, making it a possible tumor suppressor or tumor promoter influencing the biological impact of vitamin D on tumor cells [82].

These findings reveal the complexity of the molecular interactions involved and the potential for targeting these pathways in therapeutic strategies for skin cancer. While UV exposure is the most common cause of skin cancer, it is also necessary for the production of vitamin D3 in the skin, which plays a crucial role in the molecular interactions for various cellular processes, adding another facet of complexity to skin cancer prevention and treatment strategies.

V. Discussion

Vitamin D and its receptor (VDR) are integral to numerous biological functions, including cancer development. Current research presents a complex picture of vitamin D's role in skin cancer. On one hand, vitamin D acts as a tumor suppressor through mechanisms like pro-differentiation, anti-proliferative effects, pro-apoptotic pathways, immunomodulatory activities, anti-migration, DNA repair functions, and autophagy during early cancer stages. On the contrary, other studies have shown that vitamin D can promote cancer through similar mechanisms but under different conditions, particularly in late-stage cancers. For instance, its pro-angiogenic activities, anti-apoptotic effects, and inhibition of DNA repair mechanisms can facilitate tumor growth and progression. This suggests that the role of vitamin D in cancer is

highly context-dependent and varies greatly depending on the stage and type of cancer being studied.

Despite these contradictions, there is considerable evidence suggesting the potential of vitamin D as a tumor suppressor. To harness this potential effectively, further research is needed to understand how to manipulate vitamin D's mechanisms favorably for cancer inhibition. Investigations should focus on other pathways and genes involving VDR and explore additional functions of vitamin D in controlling cancer cells. Moreover, exploring the roles of significant receptors responsive to vitamin D could reveal broader biological functions beyond VDR. Research into advanced technologies might also help redirect UV radiation from damaging cells and DNA to beneficial uses like vitamin D synthesis.

Skin cancer currently is a very prevalent yet preventable carcinoma in the United States. In a 2019 study where occurrence of 30 different cancer types were studied, data showed UV radiation induced skin cancer was the second largest contributor to total cancer cases in men and the fifth largest contributor in women. In addition, more than 80% of melanoma cases and deaths were potentially preventable through modifiable risk factors, and 93% of melanoma cases and deaths were linked with UV radiation exposure. Although non-melanoma skin cancer cases were less lethal than melanomas, they were substantially more common with an estimated 4.3 million new cases in 2019 [84]. It is especially crucial for people to be aware of their sun exposure and take proper sun protection and skin cancer precautionary measures, including using sunscreen, wearing protective clothing, and avoiding peak sun hours are necessary to prevent skin damage, to lower one's risk of skin cancer [85].

Finally, it is crucial to emphasize the importance of UV radiation in biological functions despite concerns about skin cancer. Insufficient sun exposure also poses a significant health crisis currently worldwide as studies have reported that insufficient sun exposure may be responsible for 340,000 deaths in the United States per year. It may also increase one's risk of breast cancer, colorectal cancer, hypertension, cardiovascular disease, and other various health problems [86]. Sensible sun exposure is vital as it helps maintain adequate vitamin D levels, which is essential for numerous physiological processes required to keep the body healthy, including cancer prevention mechanisms. While UV radiation exposure poses significant health risks and must not be underestimated, it is essential to life and should not be avoided entirely.

VI. Acknowledgements

I would like to thank Snigdha Arun for being a valuable mentor to me throughout the process of creating this paper.

Works Cited

Alegado, Rosanna A, and Nicole King. “Bacterial influences on animal origins.” *Cold Spring Harbor perspectives in biology* vol. 6,11 a016162. 3 Oct. 2014, doi:10.1101/cshperspect.a016162

Demoulin, Catherine F et al. “Cyanobacteria evolution: Insight from the fossil record.” *Free radical biology & medicine* vol. 140 (2019): 206-223. doi:10.1016/j.freeradbiomed.2019.05.007

Martin, William F et al. “Endosymbiotic theories for eukaryote origin.” *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* vol. 370,1678 (2015): 20140330. doi:10.1098/rstb.2014.0330

Keegan, Raphael-John H et al. “Photobiology of vitamin D in mushrooms and its bioavailability in humans.” *Dermato-endocrinology* vol. 5,1 (2013): 165-76. doi:10.4161/derm.23321

Rai, Sandhya et al. “Eco-evolutionary impact of ultraviolet radiation (UVR) exposure on microorganisms, with a special focus on our skin microbiome.” *Microbiological research* vol. 260 (2022): 127044. doi:10.1016/j.micres.2022.127044

D'Orazio, John et al. “UV radiation and the skin.” *International journal of molecular sciences* vol. 14,6 12222-48. 7 Jun. 2013, doi:10.3390/ijms140612222

US Department of Health and Human Services. *The Surgeon General's Call to Action to Prevent Skin Cancer*. Office of the Surgeon General (US), 2014.

Hanahan, Douglas, and Robert A Weinberg. “Hallmarks of cancer: the next generation.” *Cell* vol. 144,5 (2011): 646-74. doi:10.1016/j.cell.2011.02.013

Beral, V et al. “Cutaneous factors related to the risk of malignant melanoma.” *The British journal of dermatology* vol. 109,2 (1983): 165-72. doi:10.1111/j.1365-2133.1983.tb07077.x

Brenner, Michaela, and Vincent J Hearing. “The protective role of melanin against UV damage in human skin.” *Photochemistry and photobiology* vol. 84,3 (2008): 539-49. doi:10.1111/j.1751-1097.2007.00226.x

Wei, Erin X et al. “Having a first-degree relative with melanoma increases lifetime risk of melanoma, squamous cell carcinoma, and basal cell carcinoma.” *Journal of the American Academy of Dermatology* vol. 81,2 (2019): 489-499. doi:10.1016/j.jaad.2019.04.044

Fogel, Alexander L et al. “Genetic diseases associated with an increased risk of skin cancer development in childhood.” *Current opinion in pediatrics* vol. 29,4 (2017): 426-433. doi:10.1097/MOP.0000000000000514

Collins, Lindsey et al. “Skin Cancer and Immunosuppression.” *Dermatologic clinics* vol. 37,1 (2019): 83-94. doi:10.1016/j.det.2018.07.009

Schulman, Joshua M, and David E Fisher. “Indoor ultraviolet tanning and skin cancer: health risks and opportunities.” *Current opinion in oncology* vol. 21,2 (2009): 144-9. doi:10.1097/CCO.0b013e3283252fc5

Van Laethem, An et al. “The sunburn cell: regulation of death and survival of the keratinocyte.” *The international journal of biochemistry & cell biology* vol. 37,8 (2005): 1547-53. doi:10.1016/j.biocel.2005.02.015

Basu, Ashis K. “DNA Damage, Mutagenesis and Cancer.” *International journal of molecular sciences* vol. 19,4 970. 23 Mar. 2018, doi:10.3390/ijms19040970

Watson, Meg et al. “Ultraviolet Radiation Exposure and Its Impact on Skin Cancer Risk.” *Seminars in oncology nursing* vol. 32,3 (2016): 241-54. doi:10.1016/j.soncn.2016.05.005

Berking, Carola et al. “Basal cell carcinoma-treatments for the commonest skin cancer.” *Deutsches Arzteblatt international* vol. 111,22 (2014): 389-95. doi:10.3238/arztebl.2014.0389

Fagan, Jake, et al. “Basal Cell Cancer.” *StatPearls*, StatPearls Publishing, 14 April 2023.

Fania, Luca et al. “Cutaneous Squamous Cell Carcinoma: From Pathophysiology to Novel Therapeutic Approaches.” *Biomedicines* vol. 9,2 171. 9 Feb. 2021, doi:10.3390/biomedicines9020171

Hadian, Yasmin, et al. “Cutaneous Squamous Cell Carcinoma.” *StatPearls*, StatPearls Publishing, 2 July 2024.

Matthews, Natalie H., et al. “Epidemiology of Melanoma.” *Cutaneous Melanoma: Etiology and Therapy*, edited by William H. Ward et. al., Codon Publications, 21 December 2017. doi:10.15586/codon.cutaneousmelanoma.2017.ch1

Black, A P B, and G S Ogg. “The role of p53 in the immunobiology of cutaneous squamous cell carcinoma.” *Clinical and experimental immunology* vol. 132,3 (2003): 379-84. doi:10.1046/j.1365-2249.2003.02159.x

Kim, Mi-Yeon et al. “Mutations of the p53 and PTCH gene in basal cell carcinomas: UV mutation signature and strand bias.” *Journal of dermatological science* vol. 29,1 (2002): 1-9. doi:10.1016/s0923-1811(01)00170-0

Helgadottir, Hildur et al. “Phenocopies in melanoma-prone families with germ-line CDKN2A mutations.” *Genetics in medicine : official journal of the American College of Medical Genetics* vol. 20,9 (2018): 1087-1090. doi:10.1038/gim.2017.216

Pasquali, Elena et al. “MC1R variants increased the risk of sporadic cutaneous melanoma in darker-pigmented Caucasians: a pooled-analysis from the M-SKIP project.” *International journal of cancer* vol. 136,3 (2015): 618-31. doi:10.1002/ijc.29018

Ascierto, Paolo A et al. “The role of BRAF V600 mutation in melanoma.” *Journal of translational medicine* vol. 10 85. 9 Jul. 2012, doi:10.1186/1479-5876-10-85

Cleaver, James E. “Diagnosis of Xeroderma Pigmentosum and Related DNA Repair-Deficient Cutaneous Diseases.” *Current medical literature. Dermatology* vol. 13,2 (2008): 41-48.

PDQ Adult Treatment Editorial Board. “Skin Cancer Treatment (PDQ®): Health Professional Version.” *PDQ Cancer Information Summaries*, National Cancer Institute (US), 11 March 2024.

Overchuk, Marta et al. “Photodynamic and Photothermal Therapies: Synergy Opportunities for Nanomedicine.” *ACS nano* vol. 17,9 (2023): 7979-8003. doi:10.1021/acsnano.3c00891

Botchkarev, Vladimir A. “Molecular mechanisms of chemotherapy-induced hair loss.” *The journal of investigative dermatology. Symposium proceedings* vol. 8,1 (2003): 72-5. doi:10.1046/j.1523-1747.2003.12175.x

Zeien, Justin et al. “Clinical implications of chemotherapeutic agent organ toxicity on perioperative care.” *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie* vol. 146 (2022): 112503. doi:10.1016/j.biopha.2021.112503

Padma, Viswanadha Vijaya. “An overview of targeted cancer therapy.” *BioMedicine* vol. 5,4 (2015): 19. doi:10.7603/s40681-015-0019-4

Davar, Diwakar, and John Munn Kirkwood. “New therapies in the treatment of melanoma.” *Expert opinion on investigational drugs* vol. 21,11 (2012): 1643-59. doi:10.1517/13543784.2012.713938

Jerant, A F et al. “Early detection and treatment of skin cancer.” *American family physician* vol. 62,2 (2000): 357-68, 375-6, 381-2.

Holick, Michael F. “The One-Hundred-Year Anniversary of the Discovery of the Sunshine Vitamin D3: Historical, Personal Experience and Evidence-Based Perspectives.” *Nutrients* vol. 15,3 593. 23 Jan. 2023, doi:10.3390/nu15030593

Bikle, Daniel D. “Vitamin D: Production, Metabolism and Mechanisms of Action.” *Endotext*, edited by Kenneth R Feingold et. al., MDText.com, Inc., 31 December 2021.

Burns, Erin M et al. “Vitamin D and skin cancer.” *Photochemistry and photobiology* vol. 91,1 (2015): 201-9. doi:10.1111/php.12382

Tukaj, Cecylia. “Właściwy poziom witaminy D warunkiem zachowania zdrowia” [Adequate level of vitamin D is essential for maintaining good health]. *Postepy higieny i medycyny doswiadczalnej (Online)* vol. 62 502-10. 9 Oct. 2008

Benedik, Evgen. “Sources of vitamin D for humans.” *International journal for vitamin and nutrition research. Internationale Zeitschrift fur Vitamin- und Ernahrungsforschung. Journal international de vitaminologie et de nutrition* vol. 92,2 (2022): 118-125. doi:10.1024/0300-9831/a000733

Armas, Laura A G et al. “Vitamin D2 is much less effective than vitamin D3 in humans.” *The Journal of clinical endocrinology and metabolism* vol. 89,11 (2004): 5387-91. doi:10.1210/jc.2004-0360

Bikle, Daniel D. “Vitamin D metabolism and function in the skin.” *Molecular and cellular endocrinology* vol. 347,1-2 (2011): 80-9. doi:10.1016/j.mce.2011.05.017

Saponaro, Federica et al. “An Update on Vitamin D Metabolism.” *International journal of molecular sciences* vol. 21,18 6573. 8 Sep. 2020, doi:10.3390/ijms21186573

Lung, Brandon E., et al. “Calcitriol.” *StatPearls*, StatPearls Publishing, 17 July 2023.

Sizar, Omeed, et al. “Vitamin D Deficiency.” *StatPearls*, StatPearls Publishing, 17 July 2023.

Al Homyani, Doua Khalid, and Shahad Khalid Alhemaiani. “Novel Homozygous CYP27B1 Gene Mutation in Vitamin D-Dependent Rickets Type 1A (VDDR1A) Disorder: A Case Report.” *Frontiers in endocrinology* vol. 13 862022. 18 May. 2022, doi:10.3389/fendo.2022.862022

Chanchlani, Rahul et al. “An Overview of Rickets in Children.” *Kidney international reports* vol. 5,7 980-990. 11 Apr. 2020, doi:10.1016/j.ekir.2020.03.025

Zimmerman, Lance, et al. “Osteomalacia.” *StatPearls*, StatPearls Publishing, 2 September 2024.

Srivastava, Sneha Baxi. “Vitamin D: Do We Need More Than Sunshine?.” *American journal of lifestyle medicine* vol. 15,4 397-401. 3 Apr. 2021, doi:10.1177/15598276211005689

Asif, Anum. and Nauman Farooq. “Vitamin D Toxicity.” *StatPearls*, StatPearls Publishing, 24 May 2023.

Collins, Lucy et al. “Hypervitaminosis D Secondary to a *CYP24A1* Loss-of-Function Mutation: An Unusual Cause of Hypercalcemia in Two Siblings.” *JBMR plus* vol. 7,9 e10788. 8 Aug. 2023, doi:10.1002/jbm4.10788

Sadiq, Nazia M., et al. “Hypercalcemia.” *StatPearls*, StatPearls Publishing, 7 May 2024.

Chauhan, Krati, et al. “Vitamin D.” *StatPearls*, StatPearls Publishing, 9 April 2023.

Li, Yanqi et al. “Modulation of the vitamin D/vitamin D receptor system in osteoporosis pathogenesis: insights and therapeutic approaches.” *Journal of orthopaedic surgery and research* vol. 18,1 860. 13 Nov. 2023, doi:10.1186/s13018-023-04320-4

Plum, Lori A, and Hector F DeLuca. “Vitamin D, disease and therapeutic opportunities.” *Nature reviews. Drug discovery* vol. 9,12 (2010): 941-55. doi:10.1038/nrd3318

Podgorska, Ewa et al. “Knocking out the Vitamin D Receptor Enhances Malignancy and Decreases Responsiveness to Vitamin D3 Hydroxyderivatives in Human Melanoma Cells.” *Cancers* vol. 13,13 3111. 22 Jun. 2021, doi:10.3390/cancers13133111

Bong, Yong-Sik et al. “A role for the vitamin D pathway in non-intestinal lesions in genetic and carcinogen models of colorectal cancer and in familial adenomatous polyposis.” *Oncotarget* vol. 7,49 (2016): 80508-80520. doi:10.18632/oncotarget.12768

Li, Hengzhen et al. “VDR promotes pancreatic cancer progression in vivo by activating CCL20-mediated M2 polarization of tumor associated macrophage.” *Cell communication and signaling : CCS* vol. 22,1 224. 10 Apr. 2024, doi:10.1186/s12964-024-01578-x

Manson, JoAnn E et al. “Principal results of the VITamin D and OmegA-3 TriaL (VITAL) and updated meta-analyses of relevant vitamin D trials.” *The Journal of steroid biochemistry and molecular biology* vol. 198 (2020): 105522. doi:10.1016/j.jsbmb.2019.105522

Liu, F, and S E Millar. “Wnt/beta-catenin signaling in oral tissue development and disease.” *Journal of dental research* vol. 89,4 (2010): 318-30. doi:10.1177/0022034510363373

Duan, Peipei, and L F Bonewald. “The role of the wnt/β-catenin signaling pathway in formation and maintenance of bone and teeth.” *The international journal of biochemistry & cell biology* vol. 77,Pt A (2016): 23-29. doi:10.1016/j.biocel.2016.05.015

Steinhart, Zachary, and Stephane Angers. “Wnt signaling in development and tissue homeostasis.” *Development (Cambridge, England)* vol. 145,11 dev146589. 8 Jun. 2018, doi:10.1242/dev.146589

Song, Pan et al. “Wnt/β-catenin signaling pathway in carcinogenesis and cancer therapy.” *Journal of hematology & oncology* vol. 17,1 46. 18 Jun. 2024, doi:10.1186/s13045-024-01563-4

Pálmer, Héctor G et al. “The vitamin D receptor is a Wnt effector that controls hair follicle differentiation and specifies tumor type in adult epidermis.” *PloS one* vol. 3,1 e1483. 23 Jan. 2008, doi:10.1371/journal.pone.0001483

Zhen, Yan, and Harald Stenmark. “Autophagosome Biogenesis.” *Cells* vol. 12,4 668. 20 Feb. 2023, doi:10.3390/cells12040668

Chavez-Dominguez, Rodolfo et al. “The Double-Edge Sword of Autophagy in Cancer: From Tumor Suppression to Pro-tumor Activity.” *Frontiers in oncology* vol. 10 578418. 7 Oct. 2020, doi:10.3389/fonc.2020.578418

Ma, Xiao-Hong et al. “Measurements of tumor cell autophagy predict invasiveness, resistance to chemotherapy, and survival in melanoma.” *Clinical cancer research : an official journal of the American Association for Cancer Research* vol. 17,10 (2011): 3478-89. doi:10.1158/1078-0432.CCR-10-2372

Benjamin, Cara L, and Honnavara N Ananthaswamy. “p53 and the pathogenesis of skin cancer.” *Toxicology and applied pharmacology* vol. 224,3 (2007): 241-8. doi:10.1016/j.taap.2006.12.006

Wang, Haolan et al. “Targeting p53 pathways: mechanisms, structures, and advances in therapy.” *Signal transduction and targeted therapy* vol. 8,1 92. 1 Mar. 2023, doi:10.1038/s41392-023-01347-1

Maiuri, Maria Chiara et al. “Autophagy regulation by p53.” *Current opinion in cell biology* vol. 22,2 (2010): 181-5. doi:10.1016/j.ceb.2009.12.001

Chen, Hankui et al. “Vitamin D directly regulates Mdm2 gene expression in osteoblasts.” *Biochemical and biophysical research communications* vol. 430,1 (2013): 370-4. doi:10.1016/j.bbrc.2012.11.003

Wong, Christian T et al. “Regulation of XPC Binding Dynamics and Global Nucleotide Excision Repair by p63 and Vitamin D Receptor.” *The journal of physical chemistry. B* vol. 127,10 (2023): 2121-2127. doi:10.1021/acs.jpcb.2c07257

Melis, Joost P M et al. “The role of XPC: implications in cancer and oxidative DNA damage.” *Mutation research* vol. 728,3 (2011): 107-17. doi:10.1016/j.mrrev.2011.07.001

Stoyanova, Tanya et al. “DDB2 decides cell fate following DNA damage.” *Proceedings of the National Academy of Sciences of the United States of America* vol. 106,26 (2009): 10690-5. doi:10.1073/pnas.0812254106

Pemsel, Anja et al. “Tandem Affinity Purification and Nano HPLC-ESI-MS/MS Reveal Binding of Vitamin D Receptor to p53 and other New Interaction Partners in HEK 293T Cells.” *Anticancer research* vol. 38,2 (2018): 1209-1216. doi:10.21873/anticanres.12341

Demetriou, Stephanie K et al. “Vitamin D receptor mediates DNA repair and is UV inducible in intact epidermis but not in cultured keratinocytes.” *The Journal of investigative dermatology* vol. 132,8 (2012): 2097-100. doi:10.1038/jid.2012.107

Das, Lopa M et al. “Vitamin D improves sunburns by increasing autophagy in M2 macrophages.” *Autophagy* vol. 15,5 (2019): 813-826. doi:10.1080/15548627.2019.1569298

Bhutia, Sujit Kumar. “Vitamin D in autophagy signaling for health and diseases: Insights on potential mechanisms and future perspectives.” *The Journal of nutritional biochemistry* vol. 99 (2022): 108841. doi:10.1016/j.jnutbio.2021.108841

Giglia-Mari, Giuseppina, and Alain Sarasin. “TP53 mutations in human skin cancers.” *Human mutation* vol. 21,3 (2003): 217-28. doi:10.1002/humu.10179

Alshammari, Eid S et al. “ERK3 is transcriptionally upregulated by ΔNp63α and mediates the role of ΔNp63α in suppressing cell migration in non-melanoma skin cancers.” *BMC cancer* vol. 21,1 155. 12 Feb. 2021, doi:10.1186/s12885-021-07866-w

Hill, N T et al. “1α, 25-Dihydroxyvitamin D_3 and the vitamin D receptor regulates ΔNp63α levels and keratinocyte proliferation.” *Cell death & disease* vol. 6,6 e1781. 11 Jun. 2015, doi:10.1038/cddis.2015.148

Stambolsky, Perry et al. “Modulation of the vitamin D3 response by cancer-associated mutant p53.” *Cancer cell* vol. 17,3 (2010): 273-85. doi:10.1016/j.ccr.2009.11.025

Reichrath, Jörg et al. “Tumor suppression in skin and other tissues via cross-talk between vitamin D- and p53-signaling.” *Frontiers in physiology* vol. 5 166. 3 Jun. 2014, doi:10.3389/fphys.2014.00166

Islami, Farhad et al. “Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States, 2019.” *CA: a cancer journal for clinicians* vol. 74,5 (2024): 405-432. doi:10.3322/caac.21858

Sultana, Nahid. “Sun Awareness and Sun Protection Practices.” *Clinical, cosmetic and investigational dermatology* vol. 13 717-730. 29 Sep. 2020, doi:10.2147/CCID.S265477

Alfredsson, Lars et al. “Insufficient Sun Exposure Has Become a Real Public Health Problem.” *International journal of environmental research and public health* vol. 17,14 5014. 13 Jul. 2020, doi:10.3390/ijerph17145014

How Does Autopilot Control A Plane and How Can We Further Improve This Technological Concept? By Ethan Tang & Jerry Liu

I.Abstract

Ever since aviation was first introduced as a new form of transportation, it was a revolutionary innovation that allowed people to travel from country to country, over vast oceans and historic landmarks, in just several hours. The components that make modern aviation possible, such as flight control systems, primary and secondary control surfaces, and autopilot and auto landing, come out of sophisticated design and complicated calculations. This paper will focus on the evolution of autopilot systems, including the history, composition, and type, as well as its advantages and future directions.

II.Introduction

The autopilot may be one of the greatest aviation inventions. It helps to relieve pilots' efforts and is always a reliable assistant in the cockpit (*The Evolution of Autopilot*). However, autopilot was not always as advanced as it is now in modern times. Although autopilot has changed so much technologically, there are still similarities between modern autopilot and when it was a new concept. For example, both modern and old autopilot controlled flight surfaces such as the rudder, elevator, and ailerons, as seen in Figure 1. In this figure, outside forces are detected by the attitude sensor/rate gyro which is connected to the transducer. The transducer transmits the information through electrical signals to the signal processing. Finally, the output signal will control the servo motor to adjust the control surface in order to achieve the optimal position. However, there are drastic differences between the new and old systems. The old autopilot was a gyroscopic design. It incorporated heading and altitude indicators to hydraulically operate elevators and rudders. On the contrary, the modern autopilot mainly uses electronic devices and signals to control flight. In the future, advancing this complicated system even further would be a massive milestone for all of air travel. This paper will explain the basics of flight controls, which will help provide context to understand the autopilot apparatus and improvements that can be made to it in the future.

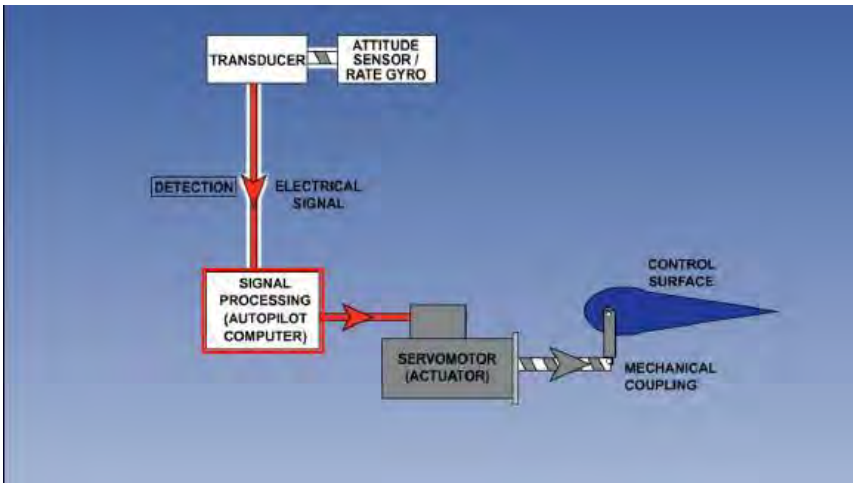


Figure 1: Demonstrates the process of the autopilot control system (*Auto Pilot Working And Auto Pilot Operations*)

III. The History of Autopilot and Autopilot Systems

Autopilot systems have come a long way since their inception in the early 20th century. The design was initially created to relieve pilots of the monotonous and physically demanding tasks involved in flight, specifically tasks that could lead to a higher chance of pilot error. The autopilot back then could only control the aircraft's elevator and rudder. However, today's autopilot system encompasses a wide range of functions, from basic course holding to sophisticated flight management systems that can control pretty much every single aspect of flight, such as takeoff, cruise, and landing. At the very core of these advancements lie significant improvement in sensor technology, computing power, and algorithms. These improvements help make the overall experience of air travel smoother with processes such as aerodynamic feedback, as shown in Figure 2. Aerodynamic feedback is where the aircraft's control systems respond to aerodynamic forces in real-time in order to maintain stable and efficient flight. In this feedback, there are two main components that help to ensure stable and precise control of the aircraft: outer and inner loop. The outer loop is responsible for higher-level control tasks, which include maintaining the aircraft's altitude, position, and heading. The inner loop handles the lower-level control tasks, such as the aircraft's attitude (pitch, roll, yaw) and stability. The inner loop takes commands from the outer loop to adjust the control surfaces and achieve the desired attitude.

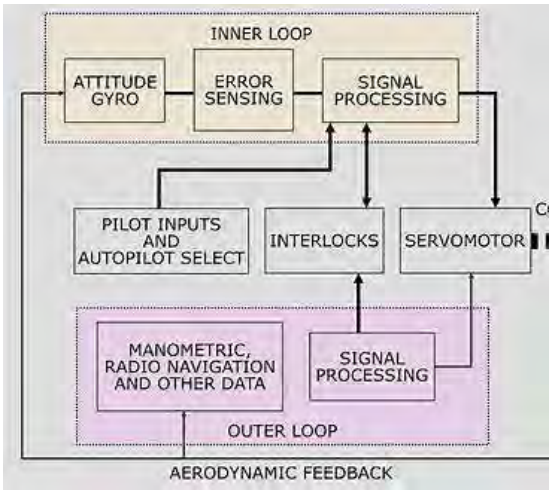


Figure 2: Aerodynamic feedback with correlation between outer and inner loops ("Auto Pilot")

Currently, autopilots utilize advanced technology including precise GPS as well as sophisticated sensors and gyroscopes to monitor the aircraft's position and attitude. There are four main components of an autopilot system: the sensing unit, autopilot computer, autopilot servo, and command unit (admin). The sensing unit is a device that generates an electrical signal in response to an aircraft movement, which is taken as the input for the system. The attitude and directional gyros, turn coordinator, and altitude control are examples of autopilot sensing units. These gyros are located in the cockpit mounted instructions. The autopilot computer is a controller that receives and processes signals. The amplifier helps the autopilot computer strengthen the signal for processing. By taking input information from the gyroscope and other systems, the autopilot computer can then send the information to the autopilot servo as input for the system that controls the flight. Third, the autopilot servo is the output element that causes actuation of the flight control surfaces. These servos are designed on the basis of the method of actuation of flight controls. In short, the servo motor spins the trim wheel up and down, allowing the aircraft to adjust itself according to environmental factors. Different types of flight controls will have different types of autopilot servos (*How Autopilot Controls the Aircraft - AviationHunt*). Lastly, the command unit is the human interface of the autopilot. The pilot tells the autopilot what to do. For example, the pilot is able to adjust the heading at which the plane should fly or the altitude it should maintain. With all of these autopilot systems in mind, it is important to note that different types of aircrafts have different autopilot systems. For example, the Cessna-172 and Boeing-737 have different autopilot systems. Because the Cessna-172 is an older and simpler plane, there is no autopilot implemented into the cockpit of the Cessna. However, the Boeing-737 is a commercial aviation jet and must have an autopilot system in order to ensure that the flight is smooth and efficient. It is also newer and has more complex machinery implemented.

IV. Flight Control Systems (FCS)

The way an aircraft is controlled has evolved throughout the years. This is called the flight control system (FCS) of an aircraft. As more FCSs came into play, autopilot advanced just as fast. There are three main flight control systems: mechanical, hydraulic, and fly by wire (AeroGuardAdmin). Figure 3 shows the evolution of the different flight control systems throughout the years.

Generation	Name of aircraft	Control System of Aircraft
First Generation Aircraft	De Havilland Comet -1(1949) Boeing 707 (1954)	Hydro mechanical
Second Generation Aircraft	Boeing 727 (1963) Boeing737(1967) Boeing 747(1969) Lockheed L1011(1970) McDonnell Douglas(1967) DC9 (1965) DC10 (1970) Airbus A300(1972)	Hydro mechanical
Current Generation Aircraft	A319(1995) AirbusA320(1987) AirbusA330(1992) AirbusA340(1991) AirbusA380(2005) Boeing777(1994) Boeing787(2009)	Fly-By-Wire

Figure 3: Flight control system evolution from hydromechanical to fly-by-wire (Islam)

The mechanical flight control system is the most traditional out of the three FCSs. It uses a series of levers, rods, cables, pulleys, and more, which pilots move to adjust control surfaces of the aircraft as shown in Figure 4 (*Mechanical Control Systems*). This gives the pilots a more accurate depiction of how the aircraft is handling aerodynamic forces as they fly. To ensure that everything is working properly, there are regular inspections and lubrication of the cables. The mechanical flight control system is the most economical type of FCS, which makes it abundant and efficient (*Mechanical Control Systems*). However, mechanical systems are very bulky and heavy, limiting passengers and cargo that could be loaded onto the aircraft. An example of a plane using this mechanical system would be light and training aircraft, such as the Cessna 172 and Piper Cub.(*Mechanical Control Systems*)

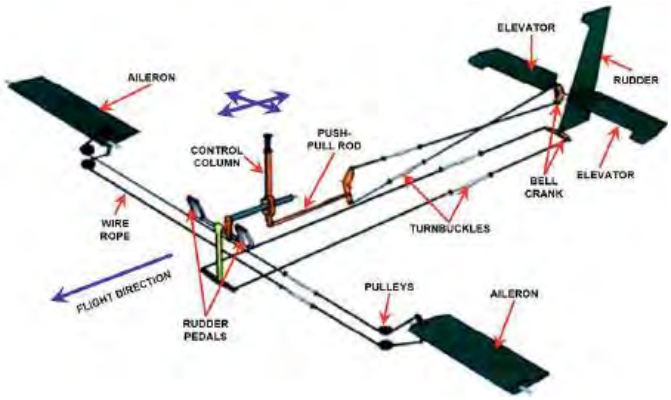


Figure 4: Mechanical FCS diagram with relationships between control surfaces (Nae et al.)

Hydraulic flight control systems are quite modern as they were first invented during World War II. These aircraft systems transfer power needed to move essential components such as brakes, landing gear and other flight control surfaces through hydraulic power (*Aircraft Hydraulic Systems (Authoritative Guide) | AviationMatters.Co*). This power is created by pumps that are either electrically driven or powered by the aircraft's engine.That power is then transferred where it is needed through a series of pipes and valves. Figure 5 shows the hydraulic system redundancies in a large commercial aircraft, highlighting three hydraulic systems (orange, teal, blue) and their connections to components like landing gear, brakes, and flight control surfaces. Basic aircraft hydraulic systems contain four main components: the hydraulic reservoir, the pump, the selector, and accumulators. Reservoirs are where the hydraulic fluid is stored. They allow the systems to be replenished and for air to be bled from the system. The pump gets power from the rotation of the engines. The accumulator is a pressurized container that stores fluid under pressure. Finally, the selectors allow pilots to select the direction of fluid flow (*Aircraft Hydraulic Systems (Authoritative Guide) | AviationMatters.Co*). However, disadvantages of this flight system include its heavy weight and the corrosiveness of the hydraulic fluid. Hydraulic systems are becoming less common in commercial aviation, but the Boeing 747 is a commercial aircraft that still uses this system (“Large Aircraft Hydraulic Systems - Boeing 737 and Boeing 777”). These planes have not been equipped with newer control systems as transitioning from hydraulics into other different systems would be very costly. Hydraulics are still very reliable and efficient control systems, so swapping them out would not provide much benefit.

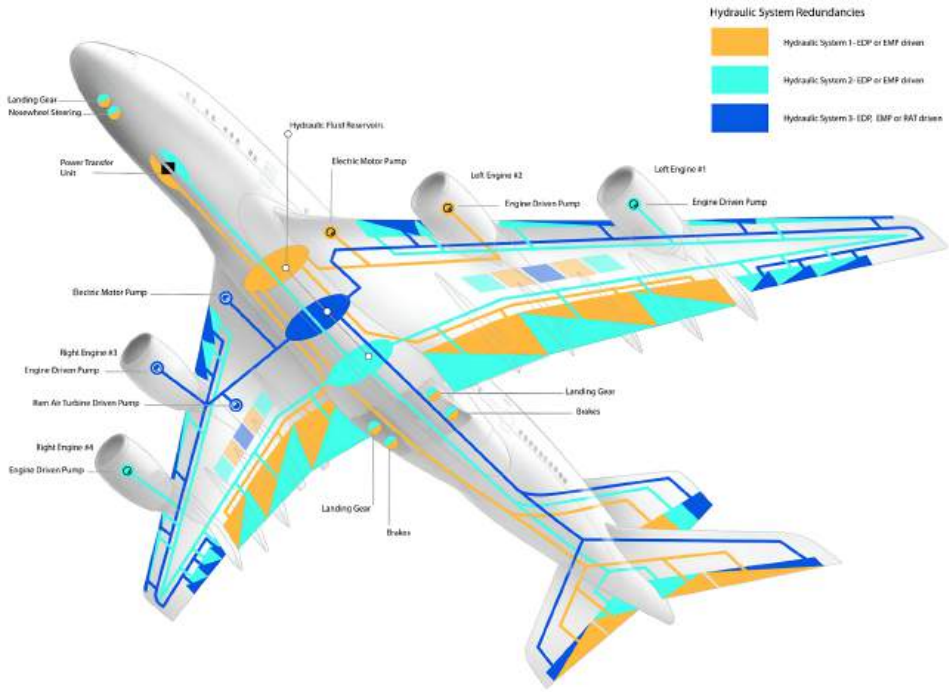


Figure 5: Structure of an aircraft hydraulic system (Shen and Zhao)

The fly-by-wire flight control system is the most used in commercial aviation today. In short, this type of FCS is a semi-automatic, computer-regulated aircraft flight control system that replaces mechanical flight controls with electronic interface. When a pilot moves flight controls, the movements are then converted into electrical signals, which are computed by the aircraft's flight control computers to adjust actuators that move the primary and secondary flight control surfaces. Figure 6 shows the connections of networks that transfer the electrical signals to each control surface. The fly-by-wire control system is purely electric, making it a lot lighter and less bulky than traditional mechanical control systems ("What Are Fly-by-Wire Systems?"). However, fly-by-wire planes also use hydraulic actuators to move the control surfaces, but the yoke and actuator are not physically connected with one another. When an airplane encounters situations such as sudden gusts, the fly-by-wire system will send signals to different control systems to act accordingly. This purely electric system allows an increase in fuel efficiency and aircraft design flexibility. In order to ensure that flight critical failure chances are very low, there are triple or quadruple redundancy backups built into the fly by wire systems ("What Are Fly-by-Wire Systems?"). Though there are many advantages to this flight system, flaws are inevitable. These include cost, complexity, and the need for constant monitoring. Though there are cons regarding the fly-by-wire control system, this flexible, fuel efficient, and reliable system is still the best choice and most widely used for all commercial aircrafts modernly ("Boeing vs Lockheed Martin").

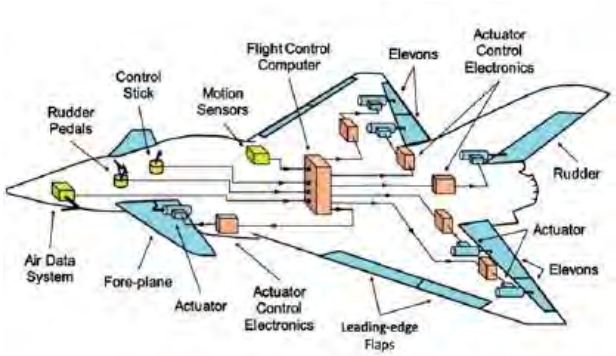


Figure 6: Fly-by-wire FCS diagram displays the electric signal networks ("Fly-by-Wire")

V. Primary and Secondary Control Surfaces

Fully understanding how autopilot controls the aircraft requires familiarity with its different control surfaces. There are two types of control surfaces: primary and secondary. Primary control surfaces are required on every aircraft, while secondary control surfaces are not required, but will assist in making flying a much less tedious task.(Ahlgren)

Primary control surfaces consist of ailerons, elevators, and the rudder, as shown in Figure 7. Ailerons control roll about the longitudinal axis and are attached to the outboard of the trailing edge of each wing. Ailerons move inversely with one another; if the right wing aileron is pitched up, the left wing aileron will be pitched down, which in turn disrupts the airflow and causes the aircraft to roll (*How It Works*). The elevator controls pitch about the lateral axis and is usually located on the horizontal tail stabilizer. The way the elevator is positioned is also determined by the tail shape (Joshi and Memon). Third, the rudder controls the yaw of the aircraft based on its vertical axis. The rudder is fixed to the vertical tail stabilizer and is controlled by rudder pedals which are located under the control panel in the cockpit. Below, Figure 8 depicts a diagram showing what the different axes represent when an aircraft is in flight.

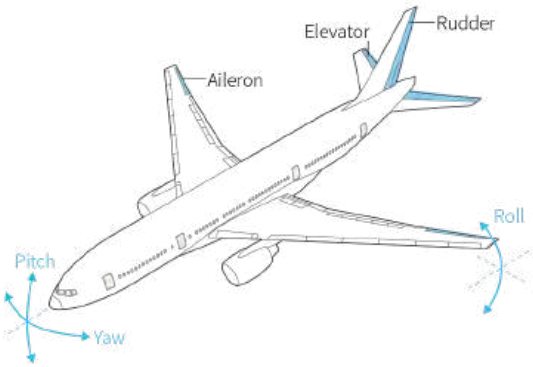


Figure 7: Primary control surface locations on an aircraft ("Determining the Final Moments of MH370")

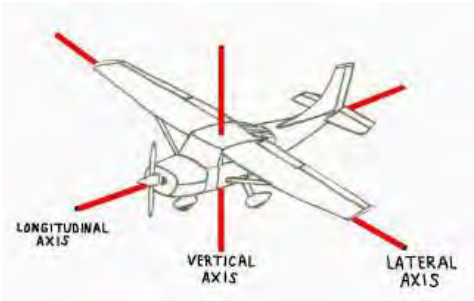


Figure 8: The axes on an aircraft (*Pergerakan Pesawat Terbang*)

Secondary control surfaces include spoilers, flaps, and trim tabs. Spoilers are panels on the wings of an airplane which deploy mainly during the landing portion of the flight (Maaz, "How Do Aircraft Spoilers & Speed Brakes Work?"). This control surface helps to decrease lift and increase drag, which helps the braking efficiency. When spoilers are extended in the air, they act as speed brakes to increase the rate of descent as well as slow the aircraft down (Maaz, "How Do Aircraft Spoilers & Speed Brakes Work?"). Flaps are used during takeoff and landing, which manipulate air flow by changing the shape of the wing (Casinader). This helps to create more lift, subsequently allowing the airplane to fly at lower speeds (Casinader). There are five main types of flaps installed on aircrafts: plain flaps, split flaps, slotted flaps, fowler flaps, and slotted fowler flaps, as shown in Figure 9 (Institute, "The 5 Types of Flaps Explained"). Plain flaps are the simplest type of flaps. They extend from the trailing edge of the wing and angle downward when deployed, which increases the wing's camber. Split flaps are attached to the underside of the wing and when they're deployed, they also drop down from the wing's trailing edge. Split flaps are more efficient at generating lift than plain flaps. Slotted flaps have one or more slots that allow high-pressure air from the underside of the wing to pass through the top where the flaps are deployed. This delays the airflow separation that leads to a stall, therefore allowing for a higher angle of attack and hence more lift. Fowler flaps slide backward on tracks before hinging downward. This extends the wing's surface area and increases the camber as well. Because of this, Fowler flaps have a significant lift increase. Slotted Fowler flaps slide backward and angle downwards when deployed. Because there are slots in these kinds of flaps, this enables high-pressure air from beneath the wing to move to the surface of the wing and produces significantly more lifting power than other flap designs (Institute, "The 5 Types of Flaps Explained"). Flaps are also utilized in the autopilot system. The autopilot will use the input from the aileron and flap channel and "mix" them to calculate how the flaps on the aircraft should be moved.

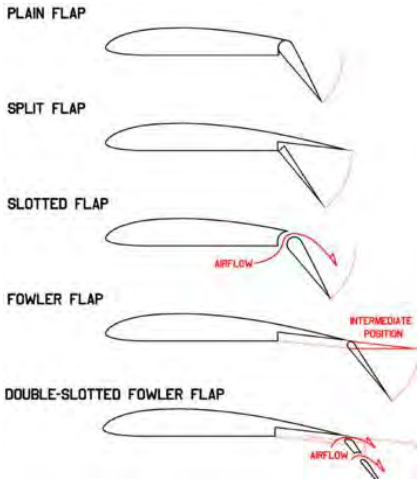


Figure 9: The 5 different types of flaps and how they are controlled (*Flaps | SKYbrary Aviation Safety*)

Finally, trim tabs are small surfaces on the trailing edges of primary control surfaces and they help to maintain a control surface's given position. These tabs significantly reduce the amount of work the pilot needs to input into the aircraft, which makes flying less tedious. An example of trim tabs in action is when the pilot wants the aircraft to climb to a specific altitude. Instead of having to pull the yoke for the whole climb duration, the trim tab can be adjusted accordingly with the elevator, thereby maintaining the climb without needing the pilot's input (*What Are Trim Tabs on Airplanes?*). With aircrafts equipped with stabilators, the trim moves the entire stabilator instead of a separate trim tab. This then correlates to the autopilot to help maintain a steady path. In bigger and more advanced aircrafts, auto trim is implemented into the cockpit. This system allows the plane to trim itself without the pilot having to adjust the trim wheel. Instead of a mechanical trim wheel, auto trim uses an electric trim wheel. A servo motor spins the trim wheel up and down, depending on the autopilot's command. With auto trim, the load of the pilots decreases even further, therefore assisting in a smoother and even more efficient flight.(Institute, "Aircraft Trim Explained")

VI. Different Levels of Autopilot

The autopilot consists of three levels of control in modern autopilot. The levels include a single axis autopilot, double axis autopilot, and three axis autopilot. The single axis autopilot is a single channel and only controls the roll axis of the aircraft. This type of autopilot is also known as a wing leveler system. Aircrafts that use the single axis autopilot are small and light aircrafts, such as the Cessna 150 and Piper Cub. The two axis autopilot controls the aircraft both in pitch and roll axes. As one autopilot channel controls the elevators in pitch, the other controls the ailerons in the roll. Medium sized general aviation aircrafts and older commercial aircrafts use the two axis autopilot, such as the Cessna 172, Cirrus SR22, and Diamond DA40. Lastly, three axis autopilots control the aircraft in roll, pitch, and yaw. Essentially, the autopilot controls all the primary control surfaces, which include the elevators, ailerons, and rudder. Almost all modern aircrafts and fighter jets have the three axis autopilot implemented. Some aircraft examples include the Boeing 737 & 777, Airbus A320, Gulfstream G650, Cirrus Vision Jet, and

Bombardier Global 7500. Airplanes with this kind of autopilot can perform auto landings as well as auto rollout (Maaz, “Autopilot”). Although three-axis autopilots certainly are the most advanced, there are scenarios where a two-axis autopilot is preferred over a three-axis one. Cost and complexity, weight considerations, and pilot preference are some big factors where a two-axis autopilot may be more fitting for an aircraft than a three-axis autopilot.

VII. Autoland

The autoland system was built on the existence of other useful cockpit aids, with the Instrument Landing System (ILS) being a crucial foundation of this system. There are three main categories of the ILS system: Category (CAT) I, II, and III. Which category pilots use depends on the pilot’s level of training and the type of equipment the aircraft has. Each category has different weather minimums that must be followed. For CAT I, the decision height (DH) can not be lower than 200 ft, and the runway visual range (RVR) cannot be less than 800m. For CAT II, the DH can be lower than 200 ft but not lower than 100 ft, and the RVR can not be less than 300m. Finally, for CAT III, the DH can be lower than 100ft, and the RVR cannot be less than 175m. Autolanding is usually in use during low visibility or foggy situations. The aircraft uses ILS and autothrust for self-landing. For instance, ILS is needed for the localizer and glide slope signals, and a radio altimeter is needed for height determination. This enables the aircraft to establish and maintain the correct approach, as shown in Figure 10. As the aircraft nears the runway, autothrust comes into play as well. At the appropriate altitude, autoland will initiate the flare by pitching up and reducing thrust via autothrust (Hayward and Hardiman). Even though the autoland system sounds like magic, there are still limitations. For example, the autoland system is not able to be executed if there is a strong crosswind, and the pilots will have to manually land the airplane in this circumstance. Although autoland is available to pilots, it is actually rarely used. It has been estimated that only 1% of all commercial flight landings used autoland. Additionally, pilots may disengage autoland when there is severe turbulence in order to have a smoother approach upon landing.

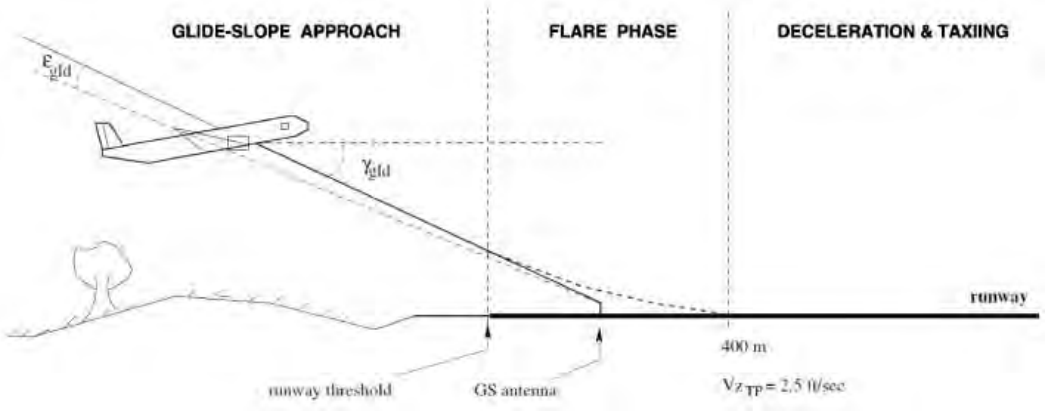


Figure 10: Autoland glideslope diagram which helps direct pilots to the runway at a controlled descent (Basic Segments of the Approach and Landing Maneuver)

VIII. Safety Concerns and Future Directions

As stated before, the modern autopilot is very advanced. It reduces the mental and physical fatigue of flying an aircraft, lets the pilot focus on other tasks such as air traffic control (ATC) conversations, and can give the passengers a much smoother ride as the autopilot reacts faster than human pilots. Even though autopilot is a great relief and assistant in the cockpit, it can also be dangerous at times if not properly maintained. An example of this would be Lion Air Flight 610. It crashed into the Java Sea thirteen minutes after takeoff because of the errors in its Maneuvering Characteristics Augmentation System (MCAS). The MCAS system is commanded by the Flight Control Computer, which in turn is correlated to autopilot. The MCAS system pushes the nose down whenever the angle-of-attack (AoA) sensor calibrates a too steep angle of attack. Unfortunately, the AoA sensor miscalibrated and sent the erroneous data to MCAS. Therefore, the plane nose dived into the body of water, killing all 189 people on board (Baker).

Continued efforts are needed to further optimize autopilot to help air travel be even more efficient and safe. Airbus is working on a project called "Optimate" that incorporates a more advanced and safe autopilot system to reduce the number of human pilots needed in the cockpit (*New Pilot Assistance Technologies Take to the Road with Airbus' Optimate Demonstrator | Airbus*). Airbus says that this project will take roughly three years and will combine a wide range of cutting-edge technologies including advanced automation, computer vision, data fusion, and machine learning (*New Pilot Assistance Technologies Take to the Road with Airbus' Optimate Demonstrator | Airbus*). This project will also develop and test automatic taxiing based on more accurate and reliable position calculation, as well as assessing the potential of quantum sensing in order to improve position availability and navigation system robustness. Another goal for the project is to investigate the capabilities of a collaborative map and virtual flight assistant that will support the pilot's strategic decisions and interactions with ATC and airline operations centers. Furthermore, MIT researchers are developing a new technique to solve the complex stabilize-avoid problem ("A Step toward Safe and Reliable Autopilots for Flying"). This will ensure that the autopilot will avoid obstacles while maintaining a desirable flight path ("A Remedy for the Spread of False News?"). Autoland is another area that can be greatly improved. For example, an advanced autoland system that can handle emergencies and relay constant updates to passengers is being worked on. It is expected that this system will not only make the pilot's job easier, but also be able to inform passengers on what is going on in the flight deck.

IX. Conclusion

All in all, the autopilot system has truly been one of the most important breakthroughs in the aviation industry. It controls a plane by manipulating the various mechanics of an aircraft, such as the primary & secondary flight control systems. All of the autopilot components ensure that the different apparatuses of an aircraft are working together properly and efficiently. Further improving autopilot is a challenging but exciting process. Utilization of a wide range of cutting-edge technologies is necessary to advance the field, so that a more safe, efficient and pleasant air travel experience can be achieved for all passengers.

Works Cited

"A Remedy for the Spread of False News?" MIT News | Massachusetts Institute of Technology, 17 Mar. 2021, https://news.mit.edu/2021/social-media-false-news-reminders-0317.

"A Step toward Safe and Reliable Autopilots for Flying." MIT News | Massachusetts Institute of Technology, 12 June 2023, https://news.mit.edu/2023/safe-and-reliable-autopilots-flying-0612.

admin. "Autopilot System." Aeroclass.Org, 23 May 2022, https://www.aeroclass.org/autopilot-system/.

AeroGuardAdmin. "Cable Systems vs. Hydraulic Systems vs. Fly by Wire in Aircraft." AeroGuard, 20 July 2020, https://www.flyaeroguard.com/blog/cable-systems-vs-hydraulic-systems-vs-fly-by-wire-in-aircraft/.

Ahlgren, Linnea. "How Do Aircraft Flight Control Surfaces Work?" Simple Flying, 12 Mar. 2021, https://simpleflying.com/aircraft-flight-control-surfaces/.

Aircraft Hydraulic Systems (Authoritative Guide) | AviationMatters.Co. 5 Feb. 2021, https://www.aviationmatters.co/aircraft-hydraulic-systems/.

"Auto Pilot." Study Aircrafts, https://www.studyaircrafts.com/auto-pilot. Accessed 10 Nov. 2024.

Auto Pilot Working And Auto Pilot Operations | Video 03. Directed by Airplane Tech Talk, 2021. YouTube, https://www.youtube.com/watch?v=w_2_-CET5LY.

Baker, Sinéad. "This Timeline Shows Exactly What Happened on Board the Lion Air Boeing 737 Max That Crashed in Less than 13 Minutes, Killing 189 People." Business Insider, https://www.businessinsider.com/lion-air-crash-timeline-boeing-737-max-disaster-killed-189-2019-10. Accessed 3 Nov. 2024.

"Boeing vs Lockheed Martin: A Comparison of Aviation Giants." Aero Corner, 31 Oct. 2023, https://aerocorner.com/blog/boeing-vs-lockheed-martin/.

Casinader, Trevin Christian. "How Aircraft Flaps Work: A Brief Guide." Simple Flying, 19 May 2024, https://simpleflying.com/how-aircraft-flaps-work/.

"Determining the Final Moments of MH370." Reuters, http://fingfx.thomsonreuters.com/gfx/rngs/MALAYSIA-AIRLINES-MH370/010020W524Y//index.html. Accessed 10 Nov. 2024.

"Figure 1. Basic Segments of the Approach and Landing Maneuver in a..." ResearchGate, https://www.researchgate.net/figure/Basic-segments-of-the-approach-and-landing-maneuver-in-a-vertical-plane_fig1_323098449. Accessed 10 Nov. 2024.

Flaps | SKYbrary Aviation Safety. https://skybrary.aero/articles/flaps. Accessed 10 Nov. 2024.

"Fly-by-Wire." Elaine's Idle Mind, 13 Feb. 2017, https://elaineou.com/2017/02/12/byzantine-fault-tolerant-airplanes/fly-by-wire/.

Hayward, Justin, and Jake Hardiman. "How Do Autoland Systems Work?" Simple Flying, 15 Aug. 2022, https://simpleflying.com/how-do-autoland-systems-work/.

How Autopilot Controls the Aircraft - AviationHunt. 20 May 2020, https://www.aviationhunt.com/aircraft-autopilot-system/.

How It Works: Ailerons. 9 Jan. 2019, https://www.aopa.org/news-and-media/all-news/2019/september/flight-training-magazine

/how-it-works-ailerons.
Institute, Pilot. “Aircraft Trim Explained (With Video).” Pilot Institute, 19 Jan. 2024, https://pilotinstitute.com/aircraft-trim-explained/.
---. “The 5 Types of Flaps Explained.” Pilot Institute, 26 June 2023, https://pilotinstitute.com/flaps-types-explained/.
Islam, Rezawana. Evolution of Aircraft Flight Control System and Fly-by-Light Flight Control System. Dec. 2013, pp. 2250–459.
Joshi, Gaurav, and Dr Omar Memon. “Ailerons And Elevators – What Are They And How Do They Work?” Simple Flying, 11 July 2020, https://simpleflying.com/ailerons-and-elevators/.
“Large Aircraft Hydraulic Systems - Boeing 737 and Boeing 777.” Aircraft Systems, http://www.aircraftsystemstech.com/p/large-aircraft-hydraulic-systems.html. Accessed 29 Sept. 2024.
Maaz, Anas. “Autopilot: A Complete Guide.” Simple Flying, 29 Oct. 2022, https://simpleflying.com/autopilot-complete-guide/.
---. “How Do Aircraft Spoilers & Speed Brakes Work?” Simple Flying, 6 Nov. 2022, https://simpleflying.com/how-aircraft-spoilers-speed-brakes-work/.
Mechanical Control Systems: Flight Control Systems - Aviation Gurukul, GOLN. 21 Aug. 2018, https://aviationgoln.com/mechanical-control-systems/.
Nae, Cătălin, et al. “Military Aircraft Flight Control.” Aeronautics - New Advances, IntechOpen, 2022. www.intechopen.com, https://doi.org/10.5772/intechopen.105491.
New Pilot Assistance Technologies Take to the Road with Airbus’ Optimate Demonstrator | Airbus. 17 May 2024, https://www.airbus.com/en/newsroom/press-releases/2024-05-new-pilot-assistance-technologies-take-to-the-road-with-airbus.
Pergerakan Pesawat Terbang. http://boniepku.blogspot.com/2013/09/pergerakan-pesawat-terbang_4.html. Accessed 10 Nov. 2024.
Shen, Kenan, and Dongbiao Zhao. “An EMD-LSTM Deep Learning Method for Aircraft Hydraulic System Fault Diagnosis under Different Environmental Noises.” Aerospace, vol. 10, no. 1, 1, Jan. 2023, p. 55. www.mdpi.com, https://doi.org/10.3390/aerospace10010055.
The Evolution of Autopilot. 21 Aug. 2020, https://www.key.aero/article/evolution-autopilot.
“What Are Fly-by-Wire Systems?” BAE Systems | United States, https://www.baesystems.com/en-us/definition/what-are-fly-by-wire-systems. Accessed 29 Sept. 2024.
What Are Trim Tabs on Airplanes? | Blog- Monroe Aerospace. https://monroeaerospace.com/blog/what-are-trim-tabs-on-airplanes/. Accessed 29 Sept. 2024.

Artistic Currents and Commercial Tides: South-East Asian Art Expansion Through Trans-Oceanic Trade Networks in the First Millennium CE By Riya Rao

Abstract

Many historians have noted the flourishing of trade between India, the Middle East, and China, in which Southeast Asian communities played a central role in the first millennium CE. As the historian Geoff Wade argues, in the four centuries from circa 900 to 1300 CE, commercial and financial changes in China, South Asia, the Middle East, and the Southeast Asian region increased maritime trade and links between societies. In this early Age of Commerce, these networks allowed not only for the exchange of goods, but also of ideas and beliefs. They introduced new economic opportunities as well as cultural frameworks in the region.

While considerable research on the overland Silk Road trade exists, scholars have yet to map the full extent to which the expansion of this maritime trade affected cultural exchanges across these regions and, in particular, how they affected the establishment of political states in Southeast Asia. Southeast Asian material culture reveals how maritime activities redefined political authority in the region, showing the importance of Buddhism in this transformation. By exploring the interconnection between religious and political dynamics within the framework of transoceanic trade, it is clear that artistic and archaeological findings show the pivotal role of Buddhism and how it shaped political power as Southeast Asian kingdoms navigated the shifting power dynamics influenced by maritime commerce in the Early Age of Commerce. By using the region's advantageous location and resources, these kingdoms became intricate states, as Buddhism was a fundamental element in the stability of their government. Thus, Southeast Asian kingdoms emerged in the first millennium CE as active participants in transoceanic commercial and cultural exchanges.

Introduction

The geography of Southeast Asia, and in particular the abundance of trade routes and waterways, makes this area an essential center from a historical and religious perspective. In 2013, the Phanom Surin shipwreck was found in the Samut Sakhon Province in Thailand. Dating back to the eighth century CE, this shipwreck manifests the stitched construction typical of the Arab dhow tradition that dominated the Western Indian Ocean. Its ceramic contents give us a glimpse of the heterogeneous cross-cultural exchange that was created by the circulation of people and commodities in this era, linking the Middle East, Southeast Asia, and China:

> Excavated local and overseas artifacts were found in the Phanom Surin shipwreck indicating it was involved in long-distance trade. The ceramics like glazed and unglazed jars were located along the Pear River delta. "One jar has incised Chinese characters on the outer surface, which have been identified as 'Chen' (陳) (Jumprom 2019), which may be the name of the Chinese owner of the cargo." There is an incised inscription on the exterior surface of the torpedo jar that has been interpreted as a personal name written in the Pahlavi script. This indicates ownership of the contents

in the jar (Guy 2019a, 2019b).[33]

Thus, the ship's construction in the western Indian Ocean tradition as well as the ceramic typologies found on the Phanom Surin shipwreck originating in China and the Persian Gulf, signal its complex global trade route across oceanic communities. In addition, the five torpedo jars found onboard contain two types of Iranian bitumen, further demonstrating the links between central Iraq/southwest Iran and Thailand during this period.[34] Further, while considerable scholarship exists on the spread of Islam from the Middle East to Asia during this period, there is relatively less research about the spread and impact of Buddhism. Examining art and archaeological evidence in Southeast Asia from this period can illuminate the changing political impact of this maritime trade, and the role of Buddhism and material culture in this history.

Many historians have noted the flourishing of trade between India, the Middle East, and China, in which Southeast Asian communities played a central role. As Geoff Wade, a historian of Sino-Southeast Asian interactions, has argued, "the four centuries from circa 900 to 1300 CE can be seen as an Early Age of Commerce in Southeast Asia."[35] According to Wade, commercial and financial changes in China, South Asia, the Middle East, and the Southeast Asian region increased maritime trade and links between societies. These networks allowed not only for the exchange of goods, but also of ideas and beliefs. They introduced new economic opportunities as well as cultural frameworks in the region. The first century CE also saw the rise and spread of Buddhism across Asia. Buddhism originated in India in the teachings of the Buddha in the fifth to sixth century BCE. Steven Kossak, the MET's curator of the Asian art department for 20 years, explains that the Buddha taught that "Through recognition of the illusory nature of the phenomenal world and through the transcendence of ego one could escape from human suffering and be released from the endless cycle of birth and rebirth."[36] With time, these Buddhist ideals spread across Asia; as several scholars, including Kenneth Hall, have observed, the expansion of maritime trade in the region was accompanied by the dissemination of Buddhist ideals and institutions.

As scholars have yet to map the full extent to which the expansion of this maritime trade affected cultural exchanges across these regions and their effect on the establishment of states in Southeast Asia, this study explores the following questions: how did transoceanic trade networks shape the cultural transformations of religion and art in Southeast Asia? What can art or archeological evidence from Southeast Asia tell us about how maritime trade transformed political power in the region and Buddhism's role therein? Analyzing how religious and political influence came to become intertwined in Southeast Asia in the context of trans-oceanic trade,

---

[33] Kimura, Jun. "Archaeological Evidence of Shipping and Shipbuilding Along The Maritime Silk Road." *The Maritime Silk Road: Global Connectivities, Regional Nodes, Localities*, edited by Franck Billé et al., Amsterdam UP, 2022, pp. 97-127. JSTOR, https://doi.org/10.2307/j.ctv2x00w7b.8.

[34] Connan, Jacques, et al. "Geochemical Analysis of Bitumen from West Asian Torpedo Jars from the c. 8th Century Phanom-Surin Shipwreck in Thailand." *Journal of Archaeological Science*, vol. 117, 2020, p. 105111. Science Direct, https://doi.org/10.1016/j.jas.2020.105111.

[35] Wade, Geoff. "An Early Age of Commerce in Southeast Asia, 900-1300 CE." *Journal of Southeast Asian Studies*, vol. 40, no. 2, 2009, pp. 221-65. Cambridge UP, https://doi.org/10.1017/S0022463409000149.

[36] Kossak, Steven., et al. *The Arts of South and Southeast Asia*. Metropolitan Museum of Art, 1994. P. 8.

this study proposes that the artistic and archaeological evidence analysis demonstrates the centrality of Buddhism's political influence as Southeast Asian kingdoms negotiated changing power relations affected by maritime trade in the Early Age of Commerce. By capitalizing on the region's strategic location and natural resources, these kingdoms transformed into complex states, with Buddhism playing a critical role in integrating and sustaining their regimes.[37] Southeast Asian kingdoms thus emerged in the first millennium CE through their participation in trans-oceanic commercial and cultural networks.

A Brief History of Maritime Trade and Its Transformations in the Early Age of Commerce

While maritime trade relations between South East Asia, China, India, and the Middle East had existed for 2000 years, as Jan Christie, a historian of Southeast Asian archaeology, argues, there was a unique expansion of maritime trade in South East Asia that led to profound social and political change, as "between the early tenth and mid-thirteenth centuries, a boom occurred in the trade linking the seas of maritime Southeast Asia to the Indian Ocean and the South China Sea."[38] The ensuing emergence of new ports and urban centers, the movement of administrative capitals toward the coast, population growth, the expansion of Buddhism and Islam, and the development of new industries, forms of consumption, and mercantile organizations were some of the key social, economic, and political transformations generated by these new maritime trade networks. The establishment of Southeast Asian kingdoms was thus deeply intertwined with their participation in expansive commercial networks that spanned the Indian Ocean and extended to the Mediterranean and East Asia beginning in the first centuries CE.

Artistic and archaeological evidence from this period indicates the cultural exchanges that emerged from these trade networks, as merchants, traders, and priests traveled these routes across Asia. For example, as the noted historian of early India, Romila Thapar observed, "Voyages to south-east Asian ports were encouraged by the search for spices. Graffiti in Brahmi on sherds, carnelian beads and rouletted ware surface in south-east Asia at places accessible to ships from the Bay of Bengal: the Irrawady Delta; the Malay peninsula; and even as far as Oc-eo in the Mekong Delta, and the island of Bali."[39] The kingdoms of Pyu in modern Myanmar and Oc Eo in modern Vietnam were among the earliest to emerge as economic centers, engaging in a vibrant exchange of goods and culture.[40] Archaeological findings at Oc Eo in southern Vietnam, for example, reveal a thriving manufacturing economy producing pottery, beads, and metal items intended for export. Artifacts such as colorful glass beads, Indian Rouletted Ware, Roman coins, and Mediterranean-style ornaments further highlight the city's extensive trade connections.

---

[37]Laliberté, André. "Religious Traditions in Politics: Buddhism." *Oxford Research Encyclopedia of Politics*, Oxford UP, 25 Feb. 2019, https://oxfordre.com/politics/view/10.1093/acrefore/9780190228637.001.0001/acrefore-9780190228637-e-691.

[38]Christie, Jan Wisseman. "Javanese Markets and the Asian Sea Trade Boom of the Tenth to Thirteenth Centuries A.D." *Journal of the Social and Economic History of the Orient*, vol. 41, no. 3, 1998, pp. 344-81.

[39] Thapar, Romila. *The Penguin History of Early India*. Penguin Books, pp. 200; 256.

[40]Aung-Thwin, Michael A. "The Pyū Millennium." *The Mists of Ramanna: The Legend That Was Lower Burma*, U of Hawai'i P, 2005, pp. 13-42. JSTOR, http://www.jstor.org/stable/j.ctt1wn0qs1.5.

Notably, the glass beads recovered at Oc Eo resemble those manufactured in the Middle East, indicating long-distance exchanges that bridged vast geographies.[41] The findings of unusual ornaments and potash glass vessels from South East and East Asia at the monastic site of Tissahamara in Sri Lanka further indicate robust maritime long-distance networks between the South China Sea and the Gulf of Bengal.[42] The discovery of artifacts with Mediterranean and Middle Eastern motifs, including Roman gold medallions dated to the reigns of Antoninus Pius (r. 138–161 CE) and Marcus Aurelius (r. 161–180 CE), underscores Oc Eo's integration into wider trade networks during the 2nd century CE. These early kingdoms were able to generate considerable revenue by facilitating the trade of precious metals, ceramics, and other goods, providing the material foundations necessary to support the emergence of complex social and political structures.[43]

The Tenets of Buddhism and their Impact on State Formation

Buddhism's significance in China grew substantially by the sixth century, with Southeast Asia serving as a crucial bridge between China and Buddhism's South Asian origins. As Hall demonstrates, religious exchange flourished through multiple routes: Buddhist monks traveled along maritime paths and through Burma by land, Chinese monks made pilgrimages to India to deepen their understanding, and Indian monks journeyed to China to teach their Chinese followers. A thriving Buddhist community had already emerged in Vietnam's Red River Delta by the 3rd century, hosting more than five hundred monks across twenty temples. By the time the Chinese pilgrim I-ching made his journey in the 7th century, this settlement had evolved into both a major trading hub and an essential Buddhist center where pilgrims stopped before entering China.[44]

The integration of Buddhism into Southeast Asia was both a spiritual transformation and a strategic development that strengthened political authority across emerging kingdoms. As Indian merchants and monks traveled to Southeast Asia, they brought Buddhist doctrines that resonated with local elites and facilitated alliances between newly established states and powerful Buddhist institutions. These states found that supporting Buddhism not only enriched their cultural landscapes but also legitimized their authority, framing rulers as righteous and benevolent figures aligned with the teachings of the Buddha.

It is important to note that unlike Hinduism and Jainism, Buddhism as a faith was not opposed to the economic world of trade and profit. In fact, it facilitated the expansion of

---

[41]Liebner, Horst Hubertus. "South-East Asian Maritime Spaces and Networks: Spatial History and Social Dynamics in the Maritime Regions of Southeast Asia." Scholars' Bank, U of Oregon, scholarsbank.uoregon.edu.

[42]Borell, Brigitte, and Laure Dussubieux. "Exceptional Potash Glass Artifacts Excavated at Tissamaharama (Sri Lanka)." *Journal of Glass Studies*, vol. 64, 2022, pp. 33-57.

[43] Hall, Kenneth R. "The 'Indianization' of Funan: An Economic History of Southeast Asia's First State." *Journal of Southeast Asian Studies*, vol. 13, no. 1, 1982, pp. 81–106. *JSTOR*, http://www.jstor.org/stable/20070472.

[44] Hall, Kenneth R. *Maritime Trade and State Development In Early Southeast Asia.* E-book, Honolulu: University of Hawai'i Press, 1985, pp. 37. https://hdl.handle.net/2027/heb31047.0001.001.

maritime trade in this region. As Osmund Bopearachchi observes, since its early history, Buddhism "was closely linked with the merchant class, which came to support Buddhist monasteries and other institutions. Buddhist communities were established along key trade routes that connected important political and economic centers, providing hospitality for merchants and other travelers."[45] This strategic positioning allowed Buddhist practice to spread widely, as merchants within the Buddhist tradition helped expand the geographical reach of their commercial activities. In addition to hundreds of inscriptions and Buddhist texts highlighting the contributions of wealthy merchants to Buddhist complexes, the study of Buddhist iconography, including paintings and ex-votos, offers valuable insights into the dynamics of trade. These artistic depictions help reveal trade routes' chronological and geographical orientations, further illustrating how Buddhism and commerce were mutually reinforcing forces.

These historical developments meant that the connection between Buddhism and political power was not merely ideological but had practical implications for governance. By endorsing Buddhism, rulers were able to harness the religion's ethical teachings to promote social cohesion and justify taxation systems that sustained both religious and state institutions. Buddhist concepts, such as the ideal of the chakravartin, or "universal ruler," became central to royal imagery, depicting kings as protectors of the Dharma and stewards of prosperity.[46] This fusion of religious and political authority solidified rulers' legitimacy and established a stable framework for managing diverse territories, and Buddhist monasteries and stupas became focal points for both spiritual practice and political power.

As a result, Buddhism during this period experienced a remarkable expansion throughout Southeast Asia. It became a powerful cultural and political force, which can be discerned through its imprint in the art, material culture, and architecture of this region. The Khmer Empire, in modern day Cambodia, was a significant Buddhist civilization. It's home to Angkor Wat, the largest religious monument in the world, showing the religious and architectural achievements of the time.[47] Other Buddhist kingdoms also stretched across the region, from Myanmar to Srivijaya. These created a network of cultural and religious exchange that shaped the landscape of Southeast Asia.[48]

Further, the city of Angkor[49], has urban complexity that contrasts the traditional historical accounts of cities at the time.[50] It was extremely expansive, and the population may have

---

[45] Bopearachchi, Osmund. "Indian Ocean Trade through Buddhist Iconographies." *The Maritime Silk Road: Global Connectivities, Regional Nodes, Localities*, edited by Franck Billé et al., Amsterdam UP, 2022, pp. 243-66. JSTOR, https://doi.org/10.2307/j.ctv2x00w7b.13.

[46] Sen, Tansen. "Maritime Southeast Asia Between South Asia and China to the Sixteenth Century." *TRaNS: Trans-Regional and National Studies of Southeast Asia*, vol. 2, no. 1, 2014, pp. 31-59.
[47] Cœdès, G. ed. Walter F. Vella, trans. Susan Brown Cowing. *The Indianized States of Southeast Asia*. University of Hawaii Press, 1970.
[48] Pierre-Yves Manguin, "The Archaeology of the Early Maritime Trading Polities of Southeast Asia," in *Southeast Asia: From Prehistory to History*. Eds. Peter Bellwood, Dr Ian Glover, Routledge, 2004.
[49] Fletcher, R. "From Angkor to the Future: Heritage and Archaeology." 2010.
[50] Stark, M. T. "Early Mainland Southeast Asian Landscapes in the First Millennium CE." 2006.

exceeded one million. This surpasses even Rome during the height of the Roman Empire.[51] This urban center was not merely a religious site, but a complex metropolitan area with sophisticated water management systems, intricate canal networks, and a robust agricultural infrastructure that supported its massive population. The economic model of Angkor and similar Southeast Asian kingdoms demonstrated a remarkable ability to integrate religious, agricultural, and commercial activities, creating a holistic system of urban development that was centuries ahead of its time in complexity and scale.

As Southeast Asia's commercial landscape evolved, certain regions began to shift from passive participation in trade to actively controlling major trade routes, leading to a further consolidation of power[52]. The kingdom of Srivijaya, centered in Sumatra, exemplifies this transition. Srivijaya became a key maritime force in the 7th century CE, as it took control of trade routes in the Strait of Malacca and turned the capital, Palembang, into somewhat of a "hub" of commerce. Srivijaya imposed taxes on merchants traveling through and created an advanced agricultural system, which helped to further increase its wealth and power. This transition from just participating in trade to dominating it represented a significant change in the growth of Southeast Asian kingdoms.

These kingdoms were able to put money into improving agriculture from the riches gained from the trade, which helped in forming and stabilizing their states. The extra food produced also created a steady income for these early governments. A good example of this is the Pyu cities in Myanmar, which were positioned along key trade routes.[53]As commerce deepened their ties with distant polities, these emerging Southeast Asian kingdoms solidified their identities as states capable of sustaining complex economies and centralized government. The alignment between commercial interests and political authority laid a strong foundation for the rise of Southeast Asia's early states.

As Tansen Sen, a scholar specializing in maritime connections between China, India, and Southeast Asia, has shown, while Buddhist ideas circulated along with goods through trade networks across Asia, they had made limited inroads in the maritime regions of Southeast Asia before the fifth century AD[15] Thus, as many historians like Sen demonstrate, it was not only the interactions with merchants and traders but also with monks and priests from across the Bay of Bengal, from present-day India and Sri Lanka, that shaped emerging Southeast Asian politics as they began to consolidate power and develop state structures. This formative period saw the birth of significant commercial epicenters and the adoption of South Asian cultural and religious elements, particularly Buddhism, which became instrumental in justifying political authority.[54]

---

[51] Buckley, B. M., et al. "Climate as a Contributing Factor in the Demise of Angkor, Cambodia." *Proceedings of the National Academy of Sciences* (*PNAS*), 2010.

[52] Department of Ancient Near Eastern Art. "Trade Routes between Europe and Asia during Antiquity." *Heilbrunn Timeline of Art History*, Metropolitan Museum of Art, Oct. 2000, www.metmuseum.org/toah/hd/trade/hd_trade.htm.

[53] Alamgir, Jalal. "Myanmar's Foreign Trade and Its Political Consequences." *Asian Survey*, vol. 48, no. 6, 2008, pp. 977-96. JSTOR, https://doi.org/10.1525/as.2008.48.6.977.

[54] Sen, Tansen. "The Spread of Buddhism." *The Cambridge World History*. Ed. Benjamin Z. Kedar and Merry E. Wiesner-Hanks. Cambridge: Cambridge University Press, 2015. 447–480. Print. The Cambridge World History.

The Rise of Buddhism: Cultural Integration and Political Power

A significant aspect of Buddhism's rise in these states involved the construction of monasteries, stupas, and other religious infrastructure that became focal points for both spiritual practice and governmental power. Rulers in the Srivijaya empire, for instance, were active patrons of Buddhism, supporting monastic communities and establishing their kingdom as a prominent center of Buddhist learning and pilgrimage. Srivijaya's rulers leveraged these religious affiliations to cultivate relationships with influential Buddhist centers such as Nalanda in India, enhancing both their spiritual prestige and political reach. The close relationship between the government and Buddhist institutions played a crucial role in enhancing Srivijaya's influence over regional trade, as the support for religious practices helped gain the trust of local communities and made governance more effective. Dharma, which represents the moral and social order in Buddhism, was actively promoted by rulers through the construction of temples, funding for monasteries, and the endorsement of Buddhist principles. By embracing dharma, the leaders of Srivijaya not only validated their authority but also created a unified cultural and ethical system that improved their governance and expanded their regional power.

Monks from the Srivijaya kingdom traveled widely, sharing Buddhist teachings and building diplomatic ties. Their journeys not only attracted pilgrims but also boosted the local economy, establishing Srivijaya as a key hub for religious interactions. This strategic blend of political and religious leadership allowed the regime to solidify its strength, with Buddhist monasteries acting as vital centers for education and cultural exchange, while also functioning as important parts of the state's administrative framework. By backing these religious institutions, the rulers of Srivijaya showcased their legitimacy as Buddhist monarchs and developed a bureaucratic system capable of managing their maritime empire effectively. The monasteries' roles as both spiritual and administrative hubs helped reinforce state authority and established lasting structures that maintained the regime's control over trade and cultural interactions throughout the region.

Maritime routes were critical hubs of commerce and cultural exchange for these Southeast Asian kingdoms. Archaeological findings from Oc Eo, located in the Mekong delta of southern Vietnam, show that these urban areas were advanced "industrial" hubs focused on creating and exporting high-quality items both locally and internationally. This included a variety of products such as pottery, beads, and detailed ornaments made from gold, tin, and bronze.

In the Śrīvijaya kingdom, Buddhism emerged as a powerful cultural and economic force, with archaeological sites in Jambi and Palembang revealing "Mahāyāna Buddhism prospered"[55] through an abundance of sophisticated statues from the seventh and eighth centuries. While temple construction and elaborate religious artifacts were primarily funded by ruling elites and wealthy merchants, Buddhism maintained a remarkably inclusive character. Monasteries served as crucial social institutions that provided education, community support, and spiritual guidance across social boundaries. Monks from various social backgrounds could be found in these

[55] Guy, John, et al. *Lost Kingdoms: Hindu-Buddhist Sculpture of Early Southeast Asia*. Metropolitan Museum of Art, 2014.

temples, and Buddhist practices remained accessible through communal rituals and the fundamental principle of compassion that transcended strict social hierarchies.[56]

Material Culture, Archaeological Sites, and Traces of History [insert relevant map here]

What traces of this complex, ambulatory maritime flow of people, ideas, and power remain today? What survives, from the ocean and the land, to disclose to us this history of shifting political authority and cross-cultural exchange? Lars Fogelin, a scholar who studies the archaeology of Buddhism in South Asia, posits that archaeology is a useful method "for directly investigating the broader social context of Buddhism" in ways that can enhance our understanding of the historical context.[57] Archaeological findings across Southeast Asia illuminate the powerful interplay between trade, religion, and political authority that shaped these early kingdoms. It is instructive to map these shifts through two sites which were also centers for Buddhist monasteries and trade: Pyu and Sri Vijaya.

The Pyu cities of Myanmar, particularly Sri Ksetra, illustrate the intricate interplay between Buddhism, trade, and governance. Sri Ksetra's city layout, marked by numerous monastic complexes and stupas such as the prominent Bawbawgyi stupa, reflects the kingdom's deep commitment to Buddhism. By carving royal decrees and religious writings into long-lasting materials like stone tablets and parts of buildings, the rulers of Sri Ksetra created a powerful story of divine rule. The leaders cleverly used inscriptions to build and strengthen their political power. They were placed in spots like the bases of temples and entrances to the city, thus turning ordinary places into significant political and religious symbols. This allowed the rulers to connect their authority with spiritual matters, effectively weaving their power into the very fabric of the city. The examples from the Pyu and Srivijaya civilizations show how these inscriptions highlight how rulers used Buddhism as a tool for legitimizing political power.

Also in Pyu, bilingual inscriptions in Pyu and Pali languages mention royal support for Buddhist institutions, with kings and high-ranking officials often donating to monasteries. These inscriptions suggest that Pyu leaders viewed supporting Buddhism as both a religious duty and a way to legitimize their authority. By backing Buddhist institutions, the rulers wove religious support into the community, securing loyalty and strengthening the kingdom's stability.
The larger picture of trade and cultural exchange highlights how advanced these civilizations were. Evidence of manufacturing and the presence of goods from far-off places show that Funan, especially in its central regions, was well-connected to long-distance trade networks. The creation and export of crafted items, along with the influence of Mediterranean and Middle Eastern designs, point to a bustling economy linked to the wider Indian Ocean trade systems.

In the capital of Srivijaya, Palembang, archaeological discoveries also reveal a strong connection between Buddhism and political power. Researchers found clay tablets inscribed with

---

[56] Guy, John, et al. *Lost Kingdoms: Hindu-Buddhist Sculpture of Early Southeast Asia*. Metropolitan Museum of Art, 2014.
[57] Fogelin, Lars. "Sacred Architecture, Sacred Landscape: Early Buddhism in North Coastal Andhra Pradesh." *Archaeology as History in Early South Asia*, edited by Himanshu Prabha Ray and Carla M. Sinopoli, Aryan Books International, 2005, pp. 376-91.

Buddhist texts alongside crafted statues, showing the kingdom's commitment to supporting the religion. Also, these tablets were not just for worship; they symbolized the rulers' authority, as the writings declared them as "champions" of Buddhism. The leaders of Srivijaya understood how Buddhism could boost their political influence, which is evident from their strategic use of religious symbols in governance. By creating these clay tablets and commissioning detailed statues, they established themselves as both spiritual leaders and political figures. This approach helped legitimize their rule, presenting them as guardians of Buddhism and its followers. Their close ties to Buddhism not only strengthened their control at home but also improved their image all over the world.

An understanding of the relationship between archaeological evidence and socio-economic relations in this history would be incomplete without a brief look at the case of Rouletted Ware, a type of fine gray or black pottery with a smooth surface, featuring concentric, wheel-like grooves or patterns. In explaining how material culture and artifacts disclose the long history of cross-cultural encounters and trade to us, Hans-Joachim Weisshaar notes, "Connections between South Asia and the West (Red Sea, Arabian Peninsular, Persian Gulf) go way back to before the Periplus as indicated by common Indian pottery in Oman (Pavan and Schenk, 2012) or imported fine cutlery like the Rouletted Ware or Wheeler types 10 and 18 in southern Arabia and the Red Sea (Schenk, 2006: 131, fig. 4)."[58] In the case of the history of inter-Asia maritime trade networks, Himanshu Ray explains, "pottery that has often been used in secondary writings to discuss early trade contacts across the Indian Ocean is termed Rouletted Ware."[59] According to Ray, "Rouletted Ware first emerged in the 3rd and 2nd centuries BCE and belongs to a group of pottery that has been found at archaeological sites along the east coast of India from Bengal to the Tamil coast and farther south to Sri Lanka."[60] This pottery has also been discovered at sites in coastal Malaysia, Thailand, Java, and Bali, as well as Vietnam.[61] Ray posits that Rouletted Ware arrived at sites in Southeast Asia between the 3rd and 2nd centuries BCE, according to the results of painstaking archeological excavations at the revered monastic site of Tissamaharama in southeastern Sri Lanka. Some were shallow dishes, and others were inscribed by donors, with inscriptions like "Dining Plate of the Noble Devi."[62] The existence of these Rouletted Ware sherds, and "[T]he find-spots of Rouletted Ware fragments along the east coast of India reflect a maritime network extending across the seas to Vietnam and the South China Sea to the east and Berenike on the Red Sea coast to the west," per Ray.[63]

Buddhist Developments in India and Their Influence: The Andhra Pradesh Connection

---

[58] Weisshaar, Hans-Joachim. "Ancient Tissamaharama: The Formation of Urban Structures and Growing Commerce." *Maritime Contacts of the Past: Deciphering Connections amongst Communities*, ed. Sila Tripati, Delta Book World, 2015, p. 208.
[59] Ray, Himanshu Prabha. *Coastal Shrines and Transnational Maritime Networks across India and Southeast Asia*. Routledge, 2021. 26.
[60] Ray, ibid., 26.
[61] Ray, ibid., 28.
[62] Ray, ibid., 40.
[63] Ray, ibid., 46.

For those interested, it is worth alluding briefly to archaeological findings related to Buddhism in the coastal centers of maritime trade in premodern South Asia. After all, the spread of Buddhism from India into Southeast Asia enumerated in the above discussion was rooted in the dynamic Buddhist developments occurring on the Indian subcontinent, particularly in regions like Andhra Pradesh. By the 1st century BCE, Andhra Pradesh had become a thriving center of Buddhist activity, producing influential texts, artwork, and architectural styles that would later resonate across Southeast Asia. The Sada dynasty, which ruled parts of the Krishna-Godavari region, played a pivotal role in the establishment and flourishing of Buddhist institutions. In analyzing not only Buddha images but also the opulent narrative reliefs of Buddhist monasteries in South India, Monika Zin argues that the narrative representations of the Aparaśaila/Aparamahāvinaśailas, the only school in Nagarjunakonda whose stūpas were covered with narrative reliefs, suggest that these narrative reliefs played a crucial role in Buddhist art of ancient Āndhradeśa.[64]Archaeological discoveries from Andhra Pradesh, such as the monastic complexes at Amaravati and Guntupalli, reveal a vibrant culture of Buddhist patronage where rulers and elites supported the construction of stupas and monasteries, elevating Buddhism's societal status.

Inscriptions from sites like Amaravati depict donations made by high-ranking officials and court patrons, illustrating how Buddhism was intricately tied to political authority in the region. This model of royal patronage—where rulers legitimized their power through acts of religious support—became a template that Southeast Asian kingdoms later emulated. For instance, the elaborate stone carvings on the coping stones of Amaravati's stupa, depicting royal figures in scenes of reverence and devotion, symbolized the rulers' connection to Buddhist ideals. These artistic and architectural motifs traveled along with traders and monks to Southeast Asia, where they inspired similar artistic traditions, such as the iconography found in Srivijaya's Buddhist monuments.

The structured organization of Buddhist monastic communities in Andhra Pradesh also influenced Southeast Asia's religious and political landscape. Inscriptions at Amaravati and other Andhra sites refer to specific monastic lineages responsible for maintaining and expanding Buddhist complexes, suggesting an organized system that enabled sustained religious activity and the creation of lasting spiritual centers. Moreover, as Gregory Schopen, has shown, the Buddhist monastic community or Saṅgha in early and medieval India, was also "an institution with economic interests and concerns. At least one of its monastic Codes or vinayas presents it as a legal or juristic personality that owned property—both real and movable—and as a corporation that was intended to generate wealth." These codes authorized monks to participate in a diverse range of activities, with established regulations overseeing ventures such as selling rice below market price, trading in luxury fabrics, offering hospice care, and more. These practices, which would travel through and across Asia, essentially "embedded the monastery in the local economies—both agricultural and commercial—so that the monastery had vested interests in the

[64] Zin, Monika. "Nagarjunakonda: Monasteries and Their School Affiliations." *Acta Asiatica Varsoviensia*, vol. 35, 2022, pp. 315-55. https://doi.org/10.4000/annuaire-cdf.429.

local economy, and the economy had the same interests in the monastery."[65] Southeast Asian kingdoms, including Pyu and Srivijaya, adopted similar organizational models, where monastic communities played dual roles as spiritual, administrative, and economic centers. The adaptation of these Indian practices in Southeast Asia allowed Buddhism to integrate seamlessly with the emerging political structures, creating a model where religious institutions supported both spiritual pursuits and governance, much as they had in Andhra Pradesh. This demonstrates how rulers utilized Buddhism as a religious practice to legitimize and reinforce their political authority. By embedding Buddhist monasteries in the administrative framework, rulers could establish a system where religious figures were not only spiritual leaders but also administrative figures who helped to maintain order and stability. This integration of religion and governance allowed Buddhism to function as a tool for political control, much like it did in Andhra Pradesh, strengthening the kingdom's authority both spiritually and politically.

Conclusion: The Legacy of Early Southeast Asian Kingdoms in the Modern Era

The developments of Southeast Asia's early kingdoms, grounded in trans-oceanic commerce and Buddhist influence, continue to shape the region's cultural and political landscape today. Just as Srivijaya once controlled critical trade routes, Southeast Asia remains strategically vital in global trade, particularly through maritime routes like the Strait of Malacca. Today, the Association of Southeast Asian Nations (ASEAN) continues the region's historical emphasis on cooperation, non-interference, and strategic partnerships. ASEAN's diplomatic approach and its balancing of global powers echo the adaptability that early Southeast Asian politics demonstrated by integrating external influences to strengthen their own identities.[66]

Buddhism, which once served as a unifying force and symbol of state legitimacy, remains deeply embedded in Southeast Asian cultural identity, especially in countries like Thailand, Myanmar, and Cambodia, while in Indonesia and Malaysia, Islam's influence prevails in structuring social relations. It is beyond the scope of this paper to consider this historical development in the spread of Islam, although many scholars have analyzed it. The concept of the righteous ruler—a legacy of the Buddhist idea of the chakravartin or "universal monarch"—continues to influence political rhetoric and national identity in these nations. This historical relationship between political power and religious patronage can be observed in the continued reverence for Buddhist leaders and the role of religious institutions in public life, where they serve as stabilizing forces much as they did in the early centuries. In essence, the legacy of Southeast Asia's first kingdoms is visible in the region's enduring cultural syncretism, its strategic economic importance, and its unique model of regional diplomacy. Understanding the foundations of these kingdoms provides insights not only into Southeast Asia's rich past but also into the values and structures that shape its role in the contemporary world.

---

[65] Schopen, Gregory. "The Business Model of a Buddhist Monasticism: Acquiring Productive Assets." *Hualin International Journal of Buddhist Studies*, vol. 2, no. 2, 2019, pp. 217-49. https://doi.org/10.15239/hijbs.02.02.08.

[66] Chalermpalanupap, Termsak. "ASEAN: Managing External Political and Security Relations." *Southeast Asian Affairs*, 2014, pp. 53–75. *JSTOR*, http://www.jstor.org/stable/44112065. Accessed 9 Jan. 2025.

Works Cited

Alamgir, Jalal. “Myanmar's Foreign Trade and Its Political Consequences.” Asian Survey, vol. 48, no. 6, 2008, pp. 977-96. JSTOR, https://doi.org/10.1525/as.2008.48.6.977.

Aung-Thwin, Michael A. “The Pyū Millennium.” The Mists of Ramanna: The Legend That Was Lower Burma, University of Hawai’i Press, 2005, pp. 13–42. JSTOR, http://www.jstor.org/stable/j.ctt1wn0qs1.5. Accessed 9 Jan. 2025.

Beaujard, Philippe, and S. Fee. "The Indian Ocean in Eurasian and African World-Systems before the Sixteenth Century." Journal of World History 16, no. 4 (2005): 411-65.

Bopearachchi, Osmund. “Indian Ocean Trade through Buddhist Iconographies.” The Maritime Silk Road: Global Connectivities, Regional Nodes, Localities, edited by Franck Billé et al., Amsterdam University Press, 2022, pp. 243–66. 244. JSTOR, https://doi.org/10.2307/j.ctv2x00w7b.13. Accessed 10 Dec. 2024

Borell, Brigitte, and Laure Dussubieux. “Exceptional Potash Glass Artifacts Excavated at Tissamaharama (Sri Lanka).” Journal of Glass Studies, vol. 64, 2022, pp. 33–57.

Buckley, B. M., et al. “Climate as a Contributing Factor in the Demise of Angkor, Cambodia.” Proceedings of the National Academy of Sciences (PNAS), 2010.

Chalermpalanupap, Termsak. “ASEAN: Managing External Political and Security Relations.” Southeast Asian Affairs, 2014, pp. 53–75. JSTOR, http://www.jstor.org/stable/44112065. Accessed 9 Jan. 2025.

Christie, Jan Wisseman. “Javanese markets and the Asian sea trade boom of the tenth to thirteenth centuries A.D.,” Journal of the Social and Economic History of the Orient, 41, 3 (1998): 344-81.

Cœdès, G., ed. Walter F. Vella, trans. Susan Brown Cowing. The Indianized States of Southeast Asia. University of Hawaii Press, 1970.

Connan, Jacques, Seth Priestman, Tom Vosmer, Abhirada Komoot, Hossein Tofighian, Buyuk Ghorbani, Michael H. Engel, Alex Zumberge, Thomas van de Velde, “Geochemical analysis of bitumen from West Asian torpedo jars from the c. 8th century Phanom-Surin shipwreck in Thailand,” Journal of Archaeological Science, Volume 117, 2020, 105111, ISSN 0305-4403, https://doi.org/10.1016/j.jas.2020.105111.

Department of Ancient Near Eastern Art. “Trade Routes between Europe and Asia during Antiquity.” Heilbrunn Timeline of Art History, Metropolitan Museum of Art, Oct. 2000, www.metmuseum.org/toah/hd/trade/hd_trade.htm.

Fletcher, R. (2010). "From Angkor to the Future: Heritage and Archaeology"

Fogelin, Lars. “Sacred Architecture, Sacred Landscape: ”Ray, Himanshu Prabha, and Carla M. Sinopoli. Archaeology as History in Early South Asia. New Delhi: Aryan Books International, 2005.

Guy, John, et al. Lost Kingdoms : Hindu-Buddhist Sculpture of Early Southeast Asia. Metropolitan Museum of Art, 2014.

Hall, Kenneth R. Maritime Trade and State Development In Early Southeast Asia. E-book, Honolulu: University of Hawai'i Press, 1985, https://hdl.handle.net/2027/heb31047.0001.001.

Hall, Kenneth R. “The ‘Indianization’ of Funan: An Economic History of Southeast Asia’s First State.” Journal of Southeast Asian Studies, vol. 13, no. 1, 1982, pp. 81–106. JSTOR, http://www.jstor.org/stable/20070472. Accessed 9 Jan. 2025.

Kimura, Jun. “Archaeological Evidence of Shipping and Shipbuilding Along The Maritime Silk Road.” The Maritime Silk Road: Global Connectivities, Regional Nodes, Localities, edited by Franck Billé et al., Amsterdam University Press, 2022, pp. 97–127. JSTOR, https://doi.org/10.2307/j.ctv2x00w7b.8. Accessed 10 Dec. 2024.

Kossak, Steven., et al. The Arts of South and Southeast Asia. Metropolitan Museum of Art, 1994.

Laliberté, André. "Religious Traditions in Politics: Buddhism." Oxford Research Encyclopedia of Politics. February 25, 2019. Oxford University Press.

Liebner, Horst Hubertus. "South-East Asian Maritime Spaces and Networks: Spatial History and Social Dynamics in the Maritime Regions of Southeast Asia." Scholars' Bank, University of Oregon, scholarsbank.uoregon.edu. Accessed 5 Jan. 2025.

Manguin, Pierre-Yves. “The Archaeology of the Early Maritime Trading Polities of Southeast Asia,” in Southeast Asia: From Prehistory to History. Eds. Peter Bellwood, Dr Ian Glover. Routledge, 2004.

Ray, Himanshu Prabha, and Carla M. Sinopoli. Archaeology as History in Early South Asia. New Delhi: Aryan Books International, 2005.

Ray, Himanshu Prabha. Coastal Shrines and Transnational Maritime Networks across India and Southeast Asia. Abingdon, Oxon: Routledge, 2021.

Schopen, Gregory. “The Business Model of a Buddhist Monasticism: Acquiring Productive Assets.” Hualin International Journal of Buddhist Studies 2, no. 2 (2019): 217-49. https://dx.doi.org/10.15239/hijbs.02.02.08.

Sen, Tansen. “Maritime Southeast Asia Between South Asia and China to the Sixteenth Century.” TRaNS: Trans-Regional and National Studies of Southeast Asia, vol. 2, no. 1, 2014, pp. 31-59.

Stark, M. T. “Early Mainland Southeast Asian Landscapes in the First Millennium CE.” 2006.

Thapar, Romila. The Penguin History of Early India. London, England: Penguin Books, 2003.

Vickery Michael. Funan Reviewed: Deconstructing the Ancients. In: Bulletin de l'Ecole française d'Extrême-Orient. Tome 90-91, 2003.

Wade, Geoff. “An Early Age of Commerce in Southeast Asia, 900–1300 CE.” Journal of Southeast Asian Studies, vol. 40, no. 2, 2009, pp. 221–265, 222. https://doi.org/10.1017/S0022463409000149.

Weisshaar, Hans-Joachim and S. Dissanayake. "The Citadel of Tissamaharama and the Torrents of Spring," in *South Asian Archaeology 2007: Proceedings of the 19th Meeting of the European Association of South Asian Archaeology in Ravenna, Italy, July 2007*, ed. P. Callieri and L. Colliva (Oxford: BAR International Series 2133, 2010), 361-368.

Weisshaar, Hans-Joachim. “Ancient Tissamaharama: The formation of urban structures and growing commerce,” in Sila Tripati, ed. Maritime Contacts of the Past: Deciphering Connections amongst Communities. Delta Book World, 2015. 208.

Zin, Monika. “Nagarjunakonda: Monasteries and Their School Affiliations.” Acta Asiatica Varsoviensia, vol. 35, 2022, pp. 315-55. https://doi.org/10.4000/annuaire-cdf.429.

Long-Term Exposure to Trauma: ACES By Saxon Mercado

Introduction

The purpose of this paper is to elaborate on how trauma impacts people, especially how adverse childhood experiences (ACEs) affect someone throughout adulthood. I would like to answer the question: "How does trauma affect someone in the long term?" My main goal with this paper is to spread knowledge and awareness of the impacts of trauma. And to have a better insight on it. More specifically, I would like to express to people that even if ACEs have long-lasting impacts throughout their lives, there are ways to cope with it and build resilience and strength. And it's not always going to feel like a tremendous weight on your back forever. Researching trauma, childhood adversity, is important due to its long-lasting effect on mental health and overall development. Childhood adversity refers to traumatic experiences, which could be a wide range of situations, including forms of abuse: emotional, sexual, and neglect (emotional and physical). (Eeruyar et al. 2023). It stunts emotional growth, leading to abnormal behaviors that persist through adulthood. For example, one study (Scully et al., 2020) discussed that 20-25% of adolescents face mental health challenges that stem from adverse childhood experiences, correlating with anxiety, depression, and more. This project specifically looks into how early life obstacles influence youth throughout their adulthood.

Understanding Trauma and ACEs

Adverse childhood experiences (ACEs) are traumatic events that happen in childhood, including violence, negligence, and abuse. Overall it is a different experience for everyone and could impact them emotionally and physically long-term, by altering brain development or affecting how the body reacts to stress (National Child Traumatic Stress Network [NCTSN], n.d.). Though ACEs have their cons, they also have their pros. For instance, those who have experienced ACEs, those who have experienced ACES may develop healthier coping mechanisms at a young age and spread better awareness and understanding to individuals who might not get the concept.

It is important to understand ACEs because trauma does not just affect a person's mental health; it impacts how they relate to others and navigate the world. For example, children who have gone through trauma may have trouble forming secure attachments with caregivers, and without these vital relationships, they may struggle to trust others, regulate their emotions, and develop a sense of safety (NCTSN, n.d.). When caregivers are abusive or neglectful, children may experience upsetting attachment bonds, which may lead them to think the world around them is unsafe. Lack of secure attachment causes more disruptions such as emotional deadening and heightened awareness, as pubescents would probably struggle to understand the unpredictable behavior within an unstable caregiver. Over time these responses will stick and will stick and manifest themselves in intimate partner relationships.

Other Emotional and Behavioral Consequences due to ACES

Children who have experienced a complex trauma (cumulative, long-term exposure to interpersonal trauma) often have difficulty identifying, expressing, and managing emotions

(NCTSN, n.d.). They may be struggling to understand or communicate how they feel, leading to internal or external stress reactions. These emotional challenges can result in major depression, anxiety, or anger (NCTSN, n.d.) Their emotional responses may be unpredictable or feel explosive, and even small reminders of past traumatic events can be triggering and may produce an unhealthy response afterward.

Scully et al. (2020) expressed that family functionality has a major impact on someone's overall well-being. They found that poor family dynamics was one of the factors that led to ACEs, which increased issues with depression and anxiety. Parents who also struggle with mental health problems often struggle to express healthy behavior towards their children. So family relationships have a crucial impact on children's mental states and can contribute to mental health problems depending on the functionality of the family dynamic. Limitations of this research are that the samples were not as diverse as they could have been, limiting the ability to generalize the study in different contexts. To gain more of a generalisability of this, I would suggest adding more diverse samples, including different racial or cultural backgrounds, to deeply understand how family functioning ACEs impact mental health worldwide instead of a specific type of group.

How attachment styles affect adulthood

Attachment styles reflect how you behave in romantic relationships, and they usually develop early on in childhood, depending on family dynamics and the type of relationship you have with your caregivers. These attachment styles are characterized by five different categories: secure attachment, avoidant, anxious, dismissive-avoidant, disorganized, and fearful-avoidant attachment (The Attachment Project, 2020).

Intimate partner violence (IPV) is the abuse or aggression that occurs in romantic relationships. This specifically occurs within current spouses or former partners, or even dating relationships. IPV can vary depending on the severity of the situation, ranging from chronic or severe episodes over multiple years. IPV can involve behaviors such as physical violence, sexual violence, stalking, or psychological aggression (Centers for Disease Control and Prevention, 2024.).

An article by Speranza et al. (2022) focuses on attachment disorganization and IPV, more specifically a sample of women who experienced IPV. The study was done to show how early childhood trauma and attachment disorganization contribute to IPV in adulthood. Researchers wanted to understand the concept of whether attachment issues, especially disorganized attachment styles resulting from traumatic early experiences, could help explain why some women fall into IPV. They found that women with a history of IPV had higher rates of attachment disorganization. Compared to the other control group, they also reported more ACEs (abuse, neglect, etc.). Therefore, concluded that early trauma and attachment disorders contribute to partner violence in some women. One limitation of this study was the use of small, homogeneous samples, which may limit the ability to generalize the findings to broader populations. They only picked a certain range of women to do the study, so it may not have presented all the diverse experiences of women at different life stages. A future study could

include a bigger, more diverse sample with varying ages in women (younger, older, etc.), to see if it would result in the same results as the original study

How it possibly all links together

Adverse Childhood Experiences (ACEs) play a crucial role in shaping an individual's emotional and psychological framework. These experiences, which include abuse, neglect, and exposure to violence, can have long-lasting effects on a person's ability to form healthy interpersonal relationships. When children are subjected to such conditions, their basic sense of trust and security is often compromised. This disruption in their foundational experiences can lead to the development of insecure attachment styles, such as anxious, avoidant, or disorganized attachment.

Anxious attachment is characterized by a preoccupation with relationships and a heightened sensitivity to relational cues, often leading to clinginess and fear of abandonment. Avoidant attachment, on the other hand, may manifest as emotional distance and difficulty in expressing emotions. Disorganized attachment, often resulting from severe trauma, can lead to unpredictable behavior and confusion in relational patterns. These attachment styles are not just limited to childhood but persist into adulthood, influencing how individuals perceive and react within intimate relationships.

When individuals with insecure attachment styles enter adulthood, they might unknowingly seek partners who reflect early familial experiences, even if such relationships are unhealthy. This can predispose them to Intimate Partner Violence (IPV), as they might either tolerate abusive behavior or become involved in a cyclical dynamic of control and fear. Additionally, those with disorganized attachment might find it challenging to establish stability within relationships, increasing the likelihood of chaotic or abusive interactions.

Moreover, the trauma from ACEs can manifest as emotional dysregulation, poor impulse control, or a heightened stress response. These psychological patterns further complicate relationships, as they can create environments ripe for conflict and misunderstanding. The perpetuation of these patterns can result in a cycle where those affected by ACEs may either fall victim to or perpetuate IPV, influenced by their persistent insecure attachment styles.

Ultimately, the interconnections between ACEs, IPV, and attachment styles highlight the importance of early intervention and therapeutic strategies. By addressing these issues in childhood or at the onset of problematic behaviors, it may be possible to alter the trajectory towards more secure attachment patterns and healthier adult relationships. Recognizing these patterns can be the first step toward breaking the cycle and fostering environments that support emotional resilience and relational stability.

How to cope and resilience within ACE

What is resilience? It refers to the ability to persevere through any adversity that a person might have to overcome in their life and use this experience for personal growth. When talking about it within ACEs, resilience helps individuals manage emotions more sustainably and healthily. Coping mechanisms you can incorporate in your personal life are hobbies such as art, music, and writing or finding a social group, practicing mindfulness breathing, and exercising.

All of these things I mention can help someone regulate their emotions and build more resilience over a long-term period.

Thomas & Jacque (2017) explored ACE, and how it influences creativity within adults that do performing arts. They elaborate on how artists with ACEs experienced higher levels of emotional abuse and neglect compared to a generalized population. However, many of the artists experience resilience, using this experience to fuel creativity and generate professional success. Protective factors that help involve resilience are finding a safe outlet, more specifically in this study. Performing artists found expressing themselves creatively helped them process and heal from any past childhood trauma. Art had served as a safe place to manifest complex emotions and experiences, which offered mechanisms to help process their trauma. Another protective factor was community support and being able to connect with other individuals who may have experienced similar things, which brings a sense of belonging and understanding within a community of people.

Conclusion

In reviewing the connections between Adverse Childhood Experiences (ACEs), Intimate Partner Violence (IPV), and attachment styles, it becomes evident how these elements significantly influence one another. Individuals exposed to ACEs often develop attachment styles that may predispose them to vulnerability in relationships, potentially increasing the risk of IPV. Recognizing these connections is crucial for developing effective coping strategies and fostering resilience.

Resilience and perseverance are key in overcoming the challenges posed by these adverse experiences. Emphasizing these traits can empower individuals to break the cycle of negativity and to seek healthier relationships. It's important to understand that there are effective ways to cope with such adversities, whether through therapy, support groups, or personal development strategies.

Above all, it is crucial to remember that no one is alone in facing such challenges. Support networks, mental health professionals, and community resources are available to help individuals navigate and heal from the impacts of ACEs and IPV. Understanding and addressing these issues holistically can lead to healthier attachment styles and improved interpersonal relationships, ultimately leading to a more fulfilling life. This approach not only supports individuals in their personal journeys but also contributes to a broader societal understanding and reduction of these complex issues.

Works Cited

Centers for Disease Control and Prevention. (2024, May 16). About Intimate Partner Violence. Intimate Partner Violence Prevention. https://www.cdc.gov/intimate-partner-violence/about/index.html

Eruyar, (2023). Understanding adverse childhood experiences (ACEs) and subsequent adult psychopathologies through the lens of attachment [Review of Understanding adverse childhood experiences (ACEs) and subsequent adult psychopathologies through the lens of attachment]. https://dpl6hyzg28thp.cloudfront.net/media/Eruyar_et_al_2023.pdf

National Child Traumatic Stress Network. (2019). What is child trauma? The National Child Traumatic Stress Network. https://www.nctsn.org/what-is-child-trauma

Scully, C., McLaughlin, J., & Fitzgerald, A. (2020). The relationship between adverse childhood experiences, family functioning, and mental health problems among children and adolescents: A systematic review. Child Abuse & Neglect, 106, Article 104555. https://doi.org/10.1016/j.chiabu.2020.104555.

The Attachment Project: Learn Attachment Theory from Experts. (2020). Attachment Project. https://www.attachmentproject.com/

Thomson, P., & Jaque, S. V. (2018). Childhood adversity and the creative experience in adult professional performing artists. Frontiers in Psychology, 9. https://doi.org/10.3389/fpsyg.2018.00111.

Thulin, M., Wåhlstedt, C., & Wretstrand, A. (2020). The role of childhood adversity in the development of creativity: A study of performing artists. Psychology of Aesthetics, Creativity, and the Arts. https://dpl6hyzg28thp.cloudfront.net/media/Thulin_et_al_2020.pdf.

The Epigenetic, Environmental, and Microbial Intersection of Crohn's Disease
By Sammy Naser

Abstract

Crohn's Disease (CD), a debilitating subtype of Inflammatory Bowel Disease (IBD), affects over half a million Americans and is characterized by chronic inflammation of the digestive tract. Living with CD can profoundly impact an individual's quality of life, often requiring lifestyle modifications, dietary changes, and ongoing medical management. Both genetic and environmental factors contribute to the development and progression of CD. Genes such as *NOD2*, ATG16L1, and IL23R play vital roles in immune system function, influencing the body's ability to sense and respond to bacteria in the digestive tract. (Torres et al). Here, we explore the interactions between these genes and the gut microbiome to understand their collective impact on disease severity, treatment response, and clinical symptoms.

This study also investigates the dynamic relationship between genetic predisposition and environmental triggers in activating CD. Environmental factors, including viral or bacterial infections, can induce an autoimmune response, leading the immune system to attack the body's own cells and trigger inflammation. Despite being carriers of the gene and possessing genetic markers, not all individuals develop the disease, underscoring the complex interplay between genetics and the environment. The ultimate goal of the research is to develop a more comprehensive understanding of the complex relationship between genetics and environmental triggers on the onset of CD, which could help disease prevention, pave the way for effective treatments, and potentially lead to a cure, because the exact cause of the disease remains unknown to this day.

Introduction

CD, a chronic inflammatory bowel disease named after Dr. Burrill B. Crohn, was first described in 1932. It can affect any part of the gastrointestinal (GI) tract and is characterized by symptoms that include abdominal pain, diarrhea, weight loss, fatigue, and rectal bleeding. The symptoms may vary in severity and may come and go. The inflammation associated with CD can lead to scar tissue formation and ulcers, primarily affecting any part of the GI tract, but it most commonly occurs in the ileum (the end of the small intestine) and the colon. The inflammation can be patchy, with normal areas of bowel in between affected areas. While the exact cause is unknown, a combination of genetic, environmental, and immune system factors is believed to contribute to its development, classifying it as an autoimmune condition. (Torres et al). This autoimmune response occurs when the body's immune system mistakenly attacks healthy tissues in the GI tract, leading to inflammation and various symptoms. (*Cleveland Clinic).*

Diagnosing CD involves a comprehensive approach that includes medical history, physical examination, blood tests, imaging studies, colonoscopies, and biopsies of affected tissue. Although there is no cure, various treatments, such as anti-inflammatory drugs, immunosuppressants, and biologics, aim to manage symptoms and control inflammation. In severe cases, surgery may be required to remove damaged portions of the intestine. *(Mayo Clinic).*

Complications of CD include strictures, fistulas, and nutritional deficiencies, all of which can have significant consequences. Strictures result in the narrowing of the intestine, fistulas create abnormal connections within the intestine or with other organs, and nutritional deficiencies can lead to various health issues, especially chronic diseases such as cancer, osteoporosis, and many other complications. *(Mayo Clinic)*.

Chronic inflammation, an indicator of CD, is the response to sudden damage to the body. In order to heal, the body mounts an immune response by releasing inflammatory cells to the site of injury. However, long-term release of inflammatory (T-Cells) cells in the absence of infection can cause cells to attack body tissues, leading to harmful inflammation. This pattern of chronic inflammation can lead to adverse symptoms such as abdominal pain, chest pain, fatigue, fever, joint pain or stiffness, mouth sores, and skin rashes. (*Cleveland Clinic)*.

In recent years, advancements in the field of CD have included the development of biological therapies targeting specific components of the immune system, particularly in the GI tract. Personalized medicine, based on genetic studies and a better understanding of the immune system, has gained emphasis in treatment approaches. Some studies have explored the role of exclusive nutrition and specific diets, while imaging technologies like colonoscopy and capsule endoscopy have improved diagnosis and monitoring.

Ongoing research continues to investigate the underlying causes of CD, focusing on genetic, environmental, and gut microbial factors. The gut microbiome consists of many diverse communities of microorganisms and microbiota in the digestive tract, playing an essential role in digestion, immune function, and metabolism. In CD, imbalances in gut microbiome diversity can result in inflammation and contribute to disease progression. The gut microbiome's role in the disease is a significant area of exploration. (Núñez-Sánchez et al). Understanding the complex interactions between the microbiome and the immune system may provide important insights into disease development and potential treatment strategies. Clinical trials are also testing new drugs and treatment approaches to enhance therapeutic options.

While much progress has been made, there are many uncertainties in the field of CD, including the exact cause of CD and the absence of a cure. While genetic, environmental, and immune system factors are implicated, the specific triggers and mechanisms leading to the development of the disease are not fully understood. Today, there is still no cure for CD. Treatment focuses on managing symptoms, inducing and maintaining remission, and preventing complications. *(Mayo Clinic)*. Achieving a complete cure remains an outstanding goal. However, tangible challenges include the need for more precise and effective individualized treatment strategies and understanding the long-term outcomes and potential complications associated with current treatments. Balancing the benefits and risks of medications and treatments, especially over extended periods, remains a critical consideration in managing this complex and challenging disease.

Crohn’s Disease and Genetics

The genetics of CD play a significant role in its development, and understanding the specific genes involved may shed light on the complexities of the condition. (Cirino). One key factor in the genetic foundation of CD is the *NOD2* gene, which is located in a crucial position

concerning Crohn's pathology. *NOD2* encodes a protein, also named NOD2 (Nucleotide-binding oligomerization domain-containing protein 2), that plays a pivotal role in the immune system's response to bacterial invasion in the gastrointestinal tract. *(MedlinePlus)*. In CD, variations in the *NOD2* gene can lead to altered protein expression, affecting the immune system's ability to sense and respond appropriately to harmful bacteria. The effects and impacts of these genetic variations are important. Studies have demonstrated distinct patterns of NOD2 protein expression in CD patients compared to those without the condition. (Gao et al). The altered expression of NOD2 in Crohn's patients contributes to an overactive immune response, leading to chronic inflammation in the digestive tract. This increased immune activity can result in the characteristic symptoms of CD, such as abdominal pain, diarrhea, and rectal bleeding. The effects of *NOD2* gene variations also extend beyond the gastrointestinal tract. The NOD2 protein, when misregulated, has been associated with systemic inflammatory responses, impacting various tissues and organs. (Gao et al). This large-scale impact highlights the interconnectedness of genetic factors in CD and the potential for effects beyond the digestive system.

Overall, the many different genes, particularly the *NOD2* gene, and their encoded proteins play a crucial role in the development and progression of CD. Understanding the effects of genetic variations on protein expression provides insights into the dysregulation of the immune system in Crohn's patients, contributing to chronic inflammation and associated symptoms.

The intricate interplay between genetics and microbial factors in CD is further exemplified by the pattern-recognition receptor NOD2, which also plays a crucial role in regulating host immunity and maintaining gut homeostasis. Loss of function mutations in *NOD2* are strongly associated with CD, highlighting its significance in the disease pathology. (Gao et al). A notable aspect of NOD2's function is its ability to sense bacterial muropeptides, and understanding how microbial variations influence NOD2 signaling and host pathology is essential.

Recent research has delved into the impact of microbial factors, particularly the Firmicutes peptidoglycan remodeling enzyme, DL-endopeptidase, on NOD2 ligand levels in the gut and its repercussions on colitis outcomes. (Gao et al). Gao et al. 2022 conducted a study on a global cohort of 857 individuals, revealing that CD patients exhibited a decrease in DL-endopeptidase gene abundance, which negatively correlated with colitis severity. When fecal microbiota from CD patients with low DL-endopeptidase was introduced into mice, they became more likely to develop colitis, highlighting the role of microbial dysbiosis in disease susceptibility. The therapeutic implications of these findings are highly crucial. Managing DL-endopeptidase, particularly in its active form, alleviated colitis through the NOD2 pathway. Additionally, therapeutic interventions aimed at restoring NOD2 ligands, such as using a DL-endopeptidase-producing Lactobacillus salivarius strain or mifamurtide, a clinical similarity of muramyl dipeptide, demonstrated strong anti-colitis effects. This comprehensive study suggests that the depletion of DL-endopeptidase contributes to CD pathogenesis by affecting NOD2 signaling, providing a therapeutically modifiable target. The intricate details of how microbial factors influence NOD2 function shed light on potential paths for therapeutic

interventions in CD, emphasizing the need for a comprehensive understanding of both genetic and environmental components in disease management.

Environmental Factors

Environmental factors play a significant role in the complex fundamentals of CD, contributing to the variability in symptom severity and disease progression. An individual may carry the gene but not experience any symptoms of it. However, certain triggers from the environment may lead to the onset of CD. It is important to draw attention to triggers such as the flu, shedding light on the broader spectrum of environmental influences that individuals with Crohn's need to navigate. Beyond viral infections, other environmental factors, such as dietary choices, stress levels, and exposure to certain medications, can also impact the course of CD. (Chen et al).

The flu, as a specific example, introduces an additional layer of concern for individuals with Crohn's. Viral infections not only pose a direct threat to the immune system but can also trigger inflammatory responses in the gut, potentially leading to a flare-up of Crohn's symptoms. (Tresca; Tinsley et al). This highlights the interconnectedness between the immune system, environmental exposures, and the intricate balance required for maintaining gut health. (Cirino).

Another example of environmental exposure is DNA Methylation, which controls the way DNA is subsequently made into protein and how DNA is accessed and transcribed. This process results in individuals with the same genotype expressing different phenotypes based on variations in DNA methylation. Oxidative stress gene expression, DNA methylation, and gut microbiota interaction may trigger CD. Another study investigates the intricate relationship between oxidative stress (OS) and CD, recognizing OS as a key pathophysiological mechanism in the disease. (Xu et al). This research explores the impact of various factors, including environmental influences, intestinal inflammation, gut microbiota, and epigenetic changes, on OS-related genes in CD. Using a multi-omics summary data-based Mendelian randomization (SMR) approach, the study identifies putative causal effects and underlying mechanisms of OS genes in CD.

Through comprehensive methods, OS-related genes were extracted, and intestinal transcriptome datasets were analyzed to identify differentially expressed genes (DEGs) related to OS in CD. The integration of CD genome-wide association study (GWAS) summaries with expression quantitative trait loci (eQTLs) and DNA methylation QTLs (mQTLs) from blood samples was performed using SMR methods. This approach prioritized blood OS genes associated with CD risk, including *BAD*, *SHC1*, *STAT3*, *MUC1*, and *GPX3*. Additionally, putative intestinal genes were identified, with three involved in gene-microbiota interactions: *MUC1*, *CD40*, and *PRKAB1*. The identification of genes like *BAD*, *SHC1*, *STAT3*, *MUC1*, and *GPX3* associated with CD highlights their roles in apoptosis, immune response dysregulation, inflammation, mucosal barrier function, and oxidative stress. (Xu et al). These findings suggest potential targets for therapeutic intervention, such as drugs targeting inflammatory pathways, mucosal healing agents, antioxidants, and microbiota-modulating therapies. While directly targeting genetic differences may be challenging, understanding these pathways can guide the development of personalized treatments to manage CD symptoms more effectively.

The study highlights the importance of OS genes in the regulation of CD, influenced by DNA methylation and host-microbiota interactions. This multi-omics integration study provides valuable insights for future targeted functional research, suggesting potential therapeutic interventions and pathways for disease prevention in CD.

Understanding and managing these environmental influences become crucial factors of a holistic approach to CD. Incorporating preventive measures during flu seasons, adopting dietary modifications that align with individual sensitivities, and implementing stress-reducing strategies are integral components of comprehensive disease management. The recognition of environmental triggers highlights the importance of personalized care plans that consider both genetic predispositions and external factors that can impact the complex nature of CD. (Li and Shi). By addressing environmental influences, individuals with Crohn's can take proactive steps toward minimizing symptom flare-ups and enhancing their overall quality of life. To minimize symptom flare-ups and improve quality of life in CD, individuals can take proactive steps to reduce exposure to oxidative stressors. These steps include adopting a balanced diet rich in antioxidants, such as fruits, vegetables, and whole grains, which can help neutralize and reduce oxidative damage in the body. Additionally, managing stress levels through relaxation techniques like meditation or yoga can help mitigate the inflammatory response triggered by stress, thereby reducing OS. Regular exercise is also beneficial, as it promotes overall health and can help combat OS by increasing antioxidant enzyme activity. Lastly, avoiding smoking and limiting alcohol consumption can further minimize exposure to oxidative stressors, contributing to better symptom management and overall well-being in CD. (Scarallo and Lionetti).

Microbiome

The human gut microbiome, consisting of a vast community of bacteria, viruses, and yeast, has emerged as a pivotal factor in the pathogenesis and severity of CD. (Khanna and Raffals). The gut microbiota, comprising over 100 trillion microorganisms, establishes a symbiotic and mutually beneficial relationship with the host, influencing crucial physiological processes, including immune system development, intestinal homeostasis, behavior, and host metabolism. From birth, the gut microbiota plays a critical role in shaping these processes, and any disturbance in its balance, known as dysbiosis, is associated with various metabolic and gastrointestinal conditions, including CD and ulcerative colitis. (Khanna and Raffals).

Diet is recognized as a key influencer of the gut microbiome, and its role in maintaining a well-balanced and healthy gut microbial microenvironment is key. The interplay between the host, gut microbiota, and environmental factors contributes to the multifactorial pathogenesis of CD. The imbalance in the gut microbiome is implicated in the abnormal response within the gastrointestinal tract, leading to chronic inflammation of CD. Understanding these host-microbe interactions is vital for developing effective therapeutic interventions for CD.

In recent years, the exploration of nutritional interventions aimed at regulating the composition of the gut microbiota has gained some popularity. Exclusive enteral nutrition (EEN) has emerged as a successful non-pharmacological strategy for CD management. It is entirely formula-based, meaning no solid foods for around four to twelve weeks for most people, and is designed to induce remission in individuals with CD. EEN has been recommended as the

first-line induction therapy for children with CD, achieving outcomes comparable to traditional corticosteroid treatments while avoiding growth restriction and ensuring complete nutritional coverage. (Scarallo and Lionetti). Despite the clear efficiency of EEN in clinical management, the precise mechanisms underlying its positive outcomes remain uncertain. This suggests the need for further research into the intricate relationship between dietary interventions, the gut microbiome, and CD pathogenesis.

The quest for more cost-effective and sustainable mitigation strategies for CD has led researchers to investigate the potential of dietary modifications. The development of specific diets tailored to modulate the gut microbiota has shown promise as a complementary approach to conventional pharmacological treatments. A recent example is the Crohn's Disease Exclusion Diet (CDED), which has emerged as a promising therapeutic approach for managing CD, as demonstrated by a retrospective cohort study conducted at the Tel Aviv Medical Center. (Fliss-Isakov et al). The study analyzed real-world data from CD patients treated with the CDED between 2018 and 2021. Results showed that among patients with clinically active disease, a large proportion reached the clinical remission stage at both week 6 and week 12, with a positive association observed between high adherence to the CDED and clinical remission. Additionally, for patients undergoing remission maintenance, a high remission rate was maintained at week 12. (Fliss-Isakov et al). These findings highlight the potential of the CDED as a promising intervention for various CD indications, highlighting the need for further validation through larger, prospective, and controlled studies. The CDED, which is designed to mitigate factors harmful to the gut microbiome and intestinal immunity while maintaining nutritional balance, offers a new dietary therapy with the potential to address the increasing rate of CD worldwide.

As pharmaceutical interventions, including anti-inflammatory therapies and biologics, often exhibit limitations in the long-term maintenance of intestinal integrity and are associated with significant costs and side effects, the exploration of nutritional interventions becomes vital for improving the quality of life of CD patients. Another interesting study on the microbiomes, published in Cell, illuminates a connection between oral health and CD, offering insights into the intricate interplay between the oral and gut microbiomes. (Abelar). The researchers initially focused on the gut microbiome and identified an overgrowth of foreign bacteria in individuals with CD, typically found in the oral cavity rather than the gut. This discovery prompted a shift in the study's focus to investigate the relationship between the gut and oral microbiomes. Of particular interest to the study was the role of periodontitis, an inflammatory condition caused by harmful bacteria in the gums. The study proposes that these bacteria, once inflamed, can enter the bloodstream, eventually reaching the gut. In the gut, these bacteria activate the inflammasome in colonic mononuclear phagocytes, triggering inflammation associated with CD. (Abelar). The study suggests that, under normal circumstances, the good bacteria in the gut would counteract these harmful bacteria. However, in patients with IBD, the good gut bacteria are weakened, making them less effective in combating the invading bacteria from the mouth. This imbalance leads to inflammation, contributing to the symptoms of IBD.

Furthermore, the study highlights another mechanism by which bacteria from the gums affect the bowel system by activating the body's T-cells, which can also develop inflammation in

the intestines. The activation of T-cells adds another layer to the complex relationship between oral health and the development or exacerbation of CD. (Abelar).

Dr. Martin Abelar, a dentist in San Diego, California, emphasizes the critical link between oral health and the overall well-being of the body. (Abelar). He underscores the mouth as the gateway to the rest of the body, indicating that the treatment of the mouth can significantly impact the performance of other systems in the human body. While maintaining good oral hygiene may not serve as a cure for CD or IBD, it may play a role in lessening the severity of the disease and easing some of the painful symptoms by controlling the oral bacteria that can contribute to gut inflammation. (Abelar).

This study contributes to a broader understanding of the systemic nature of health and emphasizes the importance of considering oral health as a factor in the complex etiology of CD. Addressing oral health through regular brushing, flossing, and dental check-ups may have a positive impact on the regulation of oral bacteria, potentially offering a complementary approach to managing CD symptoms. However, further research may be needed to explore specific interventions and long-term effects.

The intricate interplay between the host, gut microbiota, and dietary factors is central to the pathogenesis and severity of CD. Non-pharmacological interventions, especially nutritional approaches targeting the gut microbiota, hold much promise as cost-effective strategies to complement existing treatments. These approaches have gained increasing popularity and importance. (Núñez-Sánchez et al). Further research is essential to analyze the specific mechanisms by which these interventions influence CD outcomes, paving the way for more effective and sustainable management of this complex and debilitating genetic disorder.

Future Directions

There has been significant progress made in the field of CD over the last several years, but it still lacks definitive treatments and a cure. As this field of research progresses and evolves the understanding of CD and how to treat it effectively, and until a cure is discovered, the following are strategies for expanding current ways in which the field can go about effectively treating CD:

Precision Medicine Approaches: As our understanding of the genetic and environmental factors underlying CD continues to evolve, there is a growing opportunity for the development of personalized medicine approaches. Future research should focus on integrating genetic profiling, microbiome analysis, and environmental exposure data to customize treatment strategies to individual patients. This personalized approach has the potential to optimize therapeutic outcomes and minimize adverse effects.

Microbiome Modulation: The gut microbiome plays a crucial role in the pathogenesis of CD, influencing immune system function and disease progression. (Khanna and Raffals). Future research should delve deeper into understanding the complex interactions between the microbiome and host physiology, as well as how microbial dysbiosis contributes to disease development. Therapeutic approaches aimed at modulating the microbiome, such as probiotics, prebiotics, fecal microbiota transplantation, and microbial-targeted therapies, hold significant

promise and warrant further investigation. These methods could inspire the development of new therapeutic targets.

Pathways to Targets: While current treatment options for CD primarily target inflammation and immune system modulation, there is a pressing need to identify novel therapeutic targets. Future research should explore emerging routes involved in disease pathogenesis, such as the role of the environment, immune system, and microbial dysbiosis. (Khanna and Raffals). Investigating these pathways may uncover new approaches for drug development and therapeutic intervention.

Longitudinal Studies and Biomarker Discovery: Longitudinal studies tracking disease progression over time are essential for identifying predictive biomarkers of disease activity, treatment response, and long-term outcomes in CD. Future research should prioritize the establishment of large-scale cohorts with comprehensive clinical, genetic, environmental, and microbiome data to facilitate biomarker discovery efforts. These biomarkers could help in early diagnosis, prediction, and treatment selection, ultimately improving patient care and management.

Integration of Omics Technologies: Integrating multi-omics technologies, including genomics, transcriptomics, proteomics, and metabolomics, holds high potential for unraveling the complex molecular mechanisms underlying CD. Future research should leverage these high throughput approaches to gain insights into the heterogeneity of disease onset and severity, identify molecular subtypes, and illuminate key pathways driving disease pathogenesis. (Chen et al). By integrating data across multiple omics layers, researchers can gain a more comprehensive understanding of disease biology and identify novel therapeutic targets.

Patient-Centered Outcomes Research: Beyond traditional clinical treatments and outcomes/endpoints, future research in CD should prioritize patient-centered outcomes, such as quality of life, symptom severity, and functional status. Engaging patients in research design and decision-making processes is essential to ensure that study outcomes are meaningful and relevant to those living with the disease. By incorporating patient perspectives, researchers can better understand the holistic impact of CD and develop interventions that address the needs and priorities of individuals affected by the condition.

The Interesting concept of chaos theory in relation to Crohn's

In his book *Chaos: Making a New Science*, James Gleick explores the concept of chaos theory, which demonstrates that systems can be highly sensitive to initial conditions, where small changes in inputs can lead to drastically different outcomes. Chaos theory reveals that in certain complex systems, it is not possible to predict everything, and only the order in which events happen can be adjusted; they cannot be prevented or massively altered. This idea ties into CD in several profound ways.

Like the chaotic systems described in Gleick's work, CD is unpredictable. We still do not fully understand how CD develops, which genes are definitely involved, or why it affects some people more severely than others. Even though we have identified potential genetic markers, like NOD2, and environmental triggers, the exact mechanisms that lead to the onset and progression of CD remain elusive. The disease behaves unpredictably, much like a chaotic system where only

the sequence of events might be controlled or adjusted, but predicting the specific outcome is incredibly difficult.

In the context of CD, chaos theory can help explain why certain treatments work well for some patients but fail in others. The intricate interplay of genetic, environmental, and microbial factors creates an unpredictable system where targeted therapies may not always produce the same results across individuals. The personalized medicine approaches mentioned in the Future Directions section, such as precision medicine, microbiome modulation, and the integration of omics technologies, are attempts to bring some order to this chaos by tailoring treatments to the specific conditions of individual patients. However, even with advanced tools and personalized strategies, the treatment of CD is not universally successful because, like a chaotic system, the disease can respond differently in each case. Even in remission, the disease may come back, depending on specific behavioral and environmental conditions. Until we fully understand the underlying mechanisms of CD, its treatment will continue to be an exercise in managing chaos. This is why there is still no cure and why current treatments are limited in their effectiveness.

In summary, CD mirrors chaos theory in its unpredictability and complexity. While we can attempt to modify the order of treatment interventions, the disease's inherent variability often makes it difficult to predict outcomes. This recognition underscores the need for continued research and innovation as we strive to bring greater clarity and precision to the treatment of this complex and chaotic disease.

Conclusion

In conclusion, this research sheds light on the intricate interplay between genetic predisposition, environmental triggers, and immune system dysregulation in the development and progression of CD. Through an exploration of specific genes such as NOD2, and their interactions with the gut microbiome, this research deepens our understanding of the underlying mechanisms influencing disease severity, different treatment responses, and clinical symptoms.

By illuminating the roles of these genes in immune system function and bacterial recognition in the microbiome within the digestive tract, this research provides valuable insights into potential targets for therapeutic intervention. Understanding the dynamic relationship between genetic factors and environmental influences is crucial for developing more personalized and effective treatment strategies, with the ultimate goal of alleviating symptoms, achieving remission, and potentially finding a cure or more efficient treatments for CD. (Chen et al).

Despite significant advancements in the field, many challenges continue, including the complexity of disease etiology and causation, the lack of a universal cure, and the need for more precise and customized approaches to treatment. Continued research efforts are essential to unraveling the mysteries and unknowns of CD, improving diagnostic methods, and expanding therapeutic options. By addressing these challenges directly, we can strive towards better outcomes, improved quality of life, and ultimately, a brighter future for individuals living with this debilitating condition.

Works Cited

Abelar, Martin. "Oral Bacteria May Worsen Crohn's Disease Symptoms." *Aesthetic Dentistry of San Diego*, 20 Oct. 2020, www.keepsmilingsandiego.com/blog/oral-bacteria-may-worsen-crohns-disease-symptoms/. Accessed 15 Mar. 2024.

Cirino, Erica. "Crohn's Disease: Is It in Your Genes?" *Healthline*, 2 Oct. 2020, www.healthline.com/health/crohns-disease/genetic. Accessed 15 Mar. 2024.

Chen, Yueying et al. "Role of Environmental Factors in the Pathogenesis of Crohn's Disease: A Critical Review." *Springer*, 16 Nov. 2019, https://link.springer.com/article/10.1007/s00384-019-03441-9 Accessed 15 Mar. 2024.

Fliss-Isakov, Naomi et al. "Crohn's Disease Exclusion Diet for the Treatment of Crohn's Disease: Real-World Experience from a Tertiary Center." *Journal of Clinical Medicine*, vol. 12, no. 16, 21 Aug. 2023, p. 5428, doi:10.3390/jcm12165428.

Gao, Jie, et al. "Gut Microbial DL-Endopeptidase Alleviates Crohn's Disease via the NOD2 Pathway." *Cell Host & Microbe*, vol. 30, no. 10, 2022, pp. 1435-1449.e9, doi:10.1016/j.chom.2022.08.002. www.cell.com/cell-host-microbe/fulltext/S1931-3128(22)00395-X?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS193131282200395X%3Fshowall%3Dtrue. Accessed 31 Jan. 2024.

"Inflammation." *Cleveland Clinic*, https://my.clevelandclinic.org/health/symptoms/21660-inflammation. Accessed 15 Mar. 2024.

Khanna, Sahil, and Laura Raffals. "The Microbiome in Crohn's Disease: Role in Pathogenesis and Role of Microbiome Replacement Therapies." *Gastroenterology Clinics of North America*, vol. 46, no. 3, Sept. 2017, pp. 481-492, doi:10.1016/j.gtc.2017.05.004. https://www.sciencedirect.com/science/article/abs/pii/S0889855317300584?via%3Dihub

Li, Na, and Rui-Hua Shi. "Updated Review on Immune Factors in Pathogenesis of Crohn's Disease." *World Journal of Gastroenterology*, vol. 24, no. 1, 7 Jan. 2018, pp. 15-22, doi:10.3748/wjg.v24.i1.15.

Mayo Clinic Staff. "Crohn's Disease." *Mayo Clinic.*, 6 Aug. 2022, www.mayoclinic.org/diseases-conditions/crohns-disease/symptoms-causes/syc-20353304. Accessed 15 Mar. 2024.

"NOD2 Gene." *MedlinePlus*, www.medlineplus.gov/genetics/gene/nod2/. Accessed 15 Mar. 2024.

Núñez-Sánchez, María A., et al. "Crohn's Disease, Host–Microbiota Interactions, and Immunonutrition: Dietary Strategies Targeting Gut Microbiome as Novel Therapeutic Approaches." *International Journal of Molecular Sciences*, vol. 23, no. 15, 2022, p. 8361, doi:10.3390/ijms23158361.

Scarallo, Luca, and Paolo Lionetti. "Dietary Management in Pediatric Patients with Crohn's Disease." *Nutrients*, vol. 13, no. 5, 11 May 2021, p. 1611, doi:10.3390/nu13051611.

Tinsley, Andrew et al. "Increased Risk of Influenza and Influenza-Related Complications Among 140,480 Patients With Inflammatory Bowel Disease." *Inflammatory Bowel Diseases*, vol. 25, no. 2, Feb. 2019, pp. 369–376, https://doi.org/10.1093/ibd/izy243.

Torres, Joana et al. "Crohn's Disease." *The Lancet*, vol. 389, no. 10080, 29 Apr. 2017, pp. 1741-1755, doi:10.1016/S0140-6736(16)31711-1.

Tresca, Amber J. "IBD and the Flu." *Verywell Health*, 1 Aug. 2023, www.verywell health.com/ibd-and-the-flu-5086526. Accessed 31 Jan. 2024.

Xu, Shu, et al. "Oxidative Stress Gene Expression, DNA Methylation, and Gut Microbiota Interaction Trigger Crohn's Disease: A Multi-Omics Mendelian Randomization Study." *BMC Medicine*, vol. 21, no. 1, 11 May 2023, p. 179, doi:10.1186/s12916-023-02878-8.

Global Income Inequality and Upward Mobility By Emre Elgormus

1. Introduction

Income inequality is a global issue that impacts economies and societies worldwide. It is the uneven distribution of income among individuals, leading to significant disparities in wealth and opportunities. This research paper explores the multifaceted nature of income inequality, its historical evolution, and its current trends. By analyzing data from sources such as the World Bank, OECD, and IMF, the paper aims to uncover the underlying causes of income inequality and propose effective policies to address it.

The significance of this topic lies in its far-reaching implications for social cohesion, economic stability, and individual opportunities. Income inequality can undermine social unity, limit economic mobility, and contribute to broader societal problems. To understand these dynamics, the paper will first review the historical background of income inequality, examining key periods from the Pre-Industrial Era to the present day. Following this, the current picture of income inequality will be assessed, focusing on global trends, regional disparities, and the social impact of income inequality.

The research will then delve into the methods used to collect and analyze data, including a review of existing literature and statistical analysis. This will be followed by an examination of global and regional trends in income distribution, a comprehensive analysis of how varying levels of access to education affect economic opportunities and lifetime earnings, and an exploration of the structural and barriers that hinder upward mobility across different socioeconomic groups. The paper will conclude with a discussion of findings and policy recommendations aimed at reducing income inequality and fostering a more equitable distribution of resources.

This research reveals that income inequality is deeply rooted in historical and structural factors, exacerbated by unequal access to education, healthcare, and economic opportunities. Our analysis shows that targeted policy interventions, such as improving educational access, progressive taxation, and social safety nets, can significantly reduce income disparities. By addressing the systemic barriers to upward mobility and focusing on equitable resource distribution, this research concludes that sustainable progress toward a fairer society is achievable. These findings aim to guide policymakers in creating strategies that not only reduce income inequality but also enhance social cohesion and economic stability.

2. The History of Income Inequality

Income inequality has been a significant issue throughout history, shaping societies and economies. Understanding its historical background is essential for addressing the problem today, as persistent income disparities contribute to social unrest, limit economic mobility, and hinder overall economic growth. This section highlights two key periods: the Pre-Industrial Era and the Industrial Revolution Era.

Pre-Industrial Era

Before the advent of industrialization, most economies were primarily agricultural. This period, often referred to as the Pre-Industrial Era, was characterized by a society where wealth and land were concentrated in the hands of the nobility and the church. The majority of the population consisted of peasants who worked on land owned by the nobility. These peasants engaged in subsistence farming, producing just enough food to feed their families with little to no surplus. During this period, around 80-90% of the population in European countries were involved in agriculture, with land ownership heavily skewed towards the top 5-10% of society.(1)

Economic mobility was extremely limited during this time. Studies show that in England during the 14th to 17th centuries, social mobility was almost non-existent, with only about 1-2% of the peasant population able to move up to a higher economic status within their lifetime. The rigid class system made it difficult for individuals to move out of their assigned social status. Education and opportunities for economic advancement were scarce, particularly for the lower classes. In the 17th century, literacy rates in Europe were very low, with estimates suggesting that only about 30% of men and 10% of women were literate, and among the lower classes, these rates were even lower (OECD). The nobles and the church not only controlled the land but also wielded significant political and social power, reinforcing the economic divide while peasants faced high rents, taxes, and various feudal obligations, which further restricted their ability to accumulate wealth or improve their living conditions. This created a significant economic disparity between the wealthy landowners and the common people. The feudal system entrenched this inequality, as peasants were often bound to the land and lacked the freedom to seek better opportunities elsewhere.(1)

Industrial Revolution

The Industrial Revolution, beginning in the late 18th century, marked a major turning point in history. It brought rapid industrial growth and significant social changes. Factories and new machinery transformed production processes, leading to the mass production of goods. This era saw a shift from agrarian economies to industrial ones, with many people moving from rural areas to cities in search of work. However, this period also highlighted significant economic disparities. Factory owners and industrialists sustained substantial wealth, while the working class often faced harsh conditions. Workers in factories endured long hours, low wages, and unsafe environments. Child labor was also common, with children working in dangerous and unhealthy conditions for minimal pay. For this reason, during the Industrial Revolution, school enrollment rates were significantly low, with only about 20% of children attending school regularly by the late 19th century, reflecting the economic necessity of child labor over education.(2)

The significant contrast between the factory owners and the struggling workers led to the rise of labor movements. Workers began to organize into unions, demanding better wages, shorter working hours, and safer conditions. These movements were crucial in pushing for social reforms. Despite some reforms taken by governments, income inequality remained a significant issue. The Industrial Revolution created modern economic systems but also revealed the large

gap between social classes, showing the need to continually address economic inequality and protect workers' rights.

3. Current Picture of Income Inequality

Global Trends

After World War II, many countries experienced a period of significant economic growth and stability. For instance, in the United States, real GDP grew at an average annual rate of about 4.2% from 1945 to 1973, reflecting strong economic expansion and recovery from the war (World Bank). This era, often called the "Golden Age of Capitalism," saw the introduction of various policies aimed at reducing income inequality. Governments implemented progressive taxation, where higher incomes were taxed at higher rates, to ensure a more equitable distribution of wealth. For example, in the 1950s, the top marginal tax rate, the tax rate on additional income earned, in the United States was as high as 91% on incomes over $400,000 (equivalent to about $4 million today), which was aimed at reducing income inequality (OECD). Moreover, social security systems and welfare programs were also established to provide support to the unemployed, elderly, and vulnerable populations. During this time, many developed countries saw a decrease in income inequality, like in Sweden, the Gini coefficient, statistical measure of income inequality within a population, fell from 0.35 in 1960 to 0.27 in 1980, demonstrating a significant decrease in income inequality due to progressive policies and social welfare programs (OECD). Strong labor unions played a crucial role in negotiating better wages and working conditions for workers. Education became more accessible through expanded public schooling and higher education funding, allowing individuals from diverse backgrounds to acquire skills and improve their economic prospects. For example, the GI Bill in the United States provided veterans with tuition assistance, significantly increasing college enrollment and upward mobility. Investments in public infrastructure, such as the construction of highways and affordable housing projects, along with improvements in healthcare services, further contributed to reducing economic disparities and fostering a more equitable society.(3)

However, from the late 20th century onwards, even though overall income inequality levels have gotten better throughout the world regions, income inequality has worsened in some countries, mostly due to globalization and technological advancements. Globalization has led to significant profits for multinational companies and their executives, but the benefits have not been evenly distributed. The Gini coefficient, which measures income inequality, increased significantly in many developed countries. In the United States, the Gini coefficient rose from 0.36 in 1980 to 0.48 in 2020, reflecting a substantial increase in income inequality. This increase was primarily driven by a growth in the incomes of the wealthiest households, with the top 1% of earners seeing their real incomes grow by over 200% during this period, while the bottom 50% experienced no change or even decline in real incomes. This growing disparity highlights how the benefits of economic growth have been disproportionately concentrated at the top of the income distribution.(9/ World Bank). Similarly, in the United Kingdom, the Gini coefficient increased from 0.29 in 1980 to 0.35 in 2020 (OECD). Technological advancements contributed to this trend, as automation and digitalization created high-paying jobs for skilled workers, such

as software developers, data scientists, and IT managers, with median annual salaries ranging from $100,000 to $150,000. However, these advancements also reduced opportunities for less-skilled workers, such as factory line operators and clerical staff, whose jobs were either automated or outsourced, with wages typically stagnating around $30,000 to $40,000 per year. Economic policies in many countries began to favor deregulation, lower taxes for the wealthy, and reduced social spending, exacerbating income inequality. In the United States, the top marginal tax rate dropped from 70% in 1980 to 37% by 2020, primarily benefiting the wealthy. The financialization of the economy, where financial markets and institutions gained greater influence, further concentrated wealth among the top earners, with the top 1% of households in the U.S. holding approximately 38.6% of total wealth by 2021, up from 22% in 1980.(OECD-Tax Policy Reforms 2021/ World Inequality Database) As a result, the gap between the rich and the poor widened, leading to renewed debates on how to create a fairer economic system.(4)

In less developed countries, income inequality in urban areas has become increasingly pronounced. In India, about 90% of urban workers are employed informally, which often results in lower wages and job insecurity. Similarly, in Nigeria, around 60% of urban workers face precarious employment conditions, worsening economic disparities in these regions(World Bank- Informality in Indian Economy- Nigeria Urbanization Review). Additionally, limited access to quality education and healthcare in urban areas also exacerbates income inequality. In Indian cities, approximately 15% of children from low-income families are out of school, limiting their future economic opportunities (UNESCO). In Nigerian urban centers, nearly 40% of people lack access to basic healthcare services (World Bank). These barriers make it increasingly difficult for individuals in urban areas to improve their economic situation and their living conditions.(5)

These increasing disparities in income levels worldwide are a significant concern. Understanding the factors driving global income inequality is crucial for developing effective policies to address it. Later in this paper, data analysis will compare income inequality in different regions, highlighting the extent of the problem and potential solutions.

Regional Disparities

Income inequality varies widely across regions and countries, influenced by historical, political, and economic factors. In regions like Latin America and Sub-Saharan Africa, high income gaps often come from long-standing historical factors, weak governance, and limited economic opportunities. For instance, Brazil's Gini coefficient, a measure of income inequality, was 0.53 in 2021, indicating high inequality (World Bank). Brazil's colonial and slavery past has created deep economic divides, with land and wealth concentrated among a few. For example, the large landholdings, or latifundios, established during colonial times have persisted, limiting access to land for small farmers. Ineffective governance, such as the failure to implement land reforms, intensifies these issues, allowing wealth and land to remain in the hands of a small elite.(6)

Sub-Saharan Africa faces severe high income inequality rates, driven by political instability, conflict, and weak institutions. In countries like the Democratic Republic of Congo, ongoing conflict and corruption have stunted economic growth and perpetuated extreme poverty.

In Nigeria, the economic divide is stark, with the richest 10% of the population earning 40% of the national income, while the poorest 10% earn only 1.4% (World Bank). Similarly, South Africa's high Gini coefficient of 0.63 in 2021 highlights the deep economic divide between the affluent urban elite and impoverished rural areas. This inequality is compounded by limited access to quality education and healthcare. In Nigeria, about 10.5 million children are out of school, and around 70% of the population lacks access to adequate healthcare services (UNESCO, World Bank). These barriers prevent many individuals from acquiring the skills and resources necessary to improve their economic status, reinforcing the cycle of poverty and inequality.(5)

In contrast, Northern European countries like Sweden, Denmark, and Norway have implemented effective policies to reduce inequality and promote social mobility. Sweden's Gini coefficient was 0.28 in 2021, reflecting lower income inequality compared to other regions (World Bank). These countries have strong social welfare systems, progressive taxation, and policies ensuring equal access to education and healthcare. For example, Sweden's comprehensive welfare system supports families, the unemployed, and the elderly, helping to mitigate income inequality.(7)

Social Impact

High income inequality harms society by making it difficult for people to move up economically, creating a cycle of poverty. Children from low-income families often have less access to quality education, limiting their future job opportunities and perpetuating economic disparities. High income inequality creates tensions between the rich and the poor, leading to social unrest and resentment.(9) For instance, in countries with high income inequality, such as South Africa, there are higher rates of crime and social unrest. According to the OECD, countries with greater income inequality often experience lower levels of trust in government institutions. In the U.S., income inequality has been linked to increased political polarization, with trust in government declining from 40% in 2001 to around 20% for the whole population in 2021 (Pew Research Center database). These factors can contribute to political instability and social discontent.(8)

Solving income inequality is crucial for creating fairer societies and stronger economies. Countries with lower income inequality tend to have higher social cohesion, better health outcomes, and more stable economies. Policies promoting equal access to education and healthcare, progressive taxation, and social welfare programs will help reduce income inequality and its negative impacts.

4. Methods

This research paper employs a combination of data collection, literature review, and quantitative analysis to investigate the impacts of income inequality. The goal is to provide a comprehensive understanding of income inequality and offer evidence-based policy recommendations. The methods outlined below detail the approach taken to gather, analyze, and interpret the relevant data.

Data Collection

Data for this research paper is gathered from various sources to provide a comprehensive analysis of income inequality. Key datasets include wage gap data, which examines disparities in earnings between different demographic groups; Gini coefficient data, which measures income distribution within countries; and government education spending data, which analyzes investments in education as a factor in income inequality. These datasets are sourced from reliable organizations such as the World Bank, OECD, and IMF to ensure reliability and accuracy in the analysis.

Literature Review

A thorough review of existing research articles and books are conducted to understand current perspectives on income inequality. Keywords such as wealth gap, poverty, educational inequality, and minimum wage guide the search for relevant literature. Notable works from Thomas Piketty, as well as articles by Ellora Derenoncourt, and more, provides foundational insights and support for the study. These sources help frame the research within the broader academic context and offer critical analyses relevant to the research questions.

Data Analysis

Quantitative analysis is employed to interpret the collected data. Statistical methods are used to compare trends in income distribution over time and across regions, examine relationships between education levels, government policies, and income inequality, and visualize findings using charts, graphs, and tables to present them effectively and enhance understanding for the reader. This approach ensures that the research findings are based on robust and transparent data analysis, allowing for clear and actionable insights.

5. **Analysis**

Trends In Income Distribution

Income distribution globally has shown increasing inequality over recent decades, influenced by economic growth, globalization, and government policies. For example, the Gini coefficient in the United States rose from 0.36 in 1980 to 0.48 in 2021, reflecting a significant increase in income inequality. In contrast, countries like Denmark have maintained a relatively stable Gini coefficient around 0.28, due to strong social welfare systems and progressive taxation(OECD)(World Bank). Moving on to Latin America, Brazil's Gini coefficient decreased from 0.60 in the early 2000s to 0.53 in 2019, thanks to policies like Bolsa Família, a conditional cash transfer program that has helped reduce poverty and inequality.(6) On the other hand, Sub-Saharan Africa continues to face high levels of inequality, with South Africa's Gini coefficient remaining one of the highest in the world at 0.63 in 2019, driven by historical disparities and unequal access to resources.(10) Globalization has exacerbated income inequality within many developing countries. For instance, in India, the top 1% held 21.7% of the national income in 2021, up from 6% in 1982, indicating a growing concentration of wealth among the elite. Similarly, China's rapid economic growth has led to significant income disparities between

urban and rural areas, with the urban Gini coefficient at 0.41 compared to the rural 0.30 in 2020. (3)(5)

Government policies play a crucial role in shaping income distribution. Countries like Norway or Denmark, with policies emphasizing equitable education, healthcare, and social protection, maintain lower levels of inequality. In contrast, the United States, with less progressive tax policies and weaker social safety nets, continues to see rising inequality. Overall, these trends underscore the complex relationship between economic factors and government policies in shaping income distribution, with significant variations observed across different regions and countries.

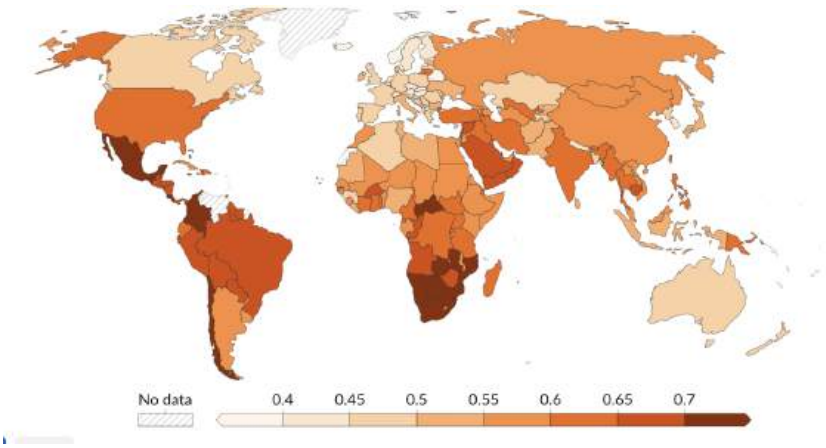


Figure 1

Figure 1 above illustrates Gini coefficient levels in different countries around the world

Impact of Education on Economic Opportunities

Education significantly influences job prospects and earning potential. Individuals who attend well-funded schools or higher education institutions often have better opportunities in the labor market. In the United States, those with a bachelor's degree or higher earned a median weekly wage of $1,305 in 2022, compared to $781 for high school graduates without a college degree. This disparity underscores how educational attainment directly affects income levels and economic opportunities.(11)

The disparities in educational opportunities between different socio-economic groups contribute to income inequality. Students from lower-income families often face challenges such as underfunded schools, lack of access to advanced coursework, and limited resources for college preparation. This disadvantage restricts their future earning potential and perpetuates the

cycle of poverty. In Brazil, a country with relatively high income inequality, individuals from wealthier families are more likely to complete higher education, leading to greater economic opportunities. Meanwhile, those from poorer backgrounds often struggle to achieve the same level of education, which limits their access to well-paying jobs. This cycle is called the poverty trap. It can be said that, because of these given reasons, wealth is sticky in countries like Brazil or Turkey with lower-income families having almost no opportunities for good education.

Globally, the relationship between education and income inequality is also evident. Countries with more equitable access to education tend to have lower levels of income inequality. For example, in Scandinavian countries like Norway, where 100% of children attend primary school and around 93% complete secondary education (UNESCO), the Gini coefficient—measuring income inequality—is among the lowest in the world, at approximately 0.27 in 2021 (World Bank). This is in sharp contrast to countries with significant disparities in educational access, such as South Africa. In South Africa, where only about 60% of students complete secondary education and approximately 40% of the population lacks access to quality education beyond the primary level (UNESCO), the Gini coefficient remains one of the highest globally, at 0.63 in 2019 (World Bank). These disparities underscore how systemic barriers to education contribute to high levels of income inequality in such regions.(OECD-Education at a Glance 2022)

Ellora Derenoncourt's work on economic mobility further illustrates the impact of education on economic opportunities. In her research, Derenoncourt emphasizes that geographical mobility—moving to areas with better educational resources and job opportunities—can significantly improve economic outcomes, particularly for individuals from disadvantaged backgrounds. This reinforces the importance of education as a pathway to economic mobility and reducing income inequality.(12)

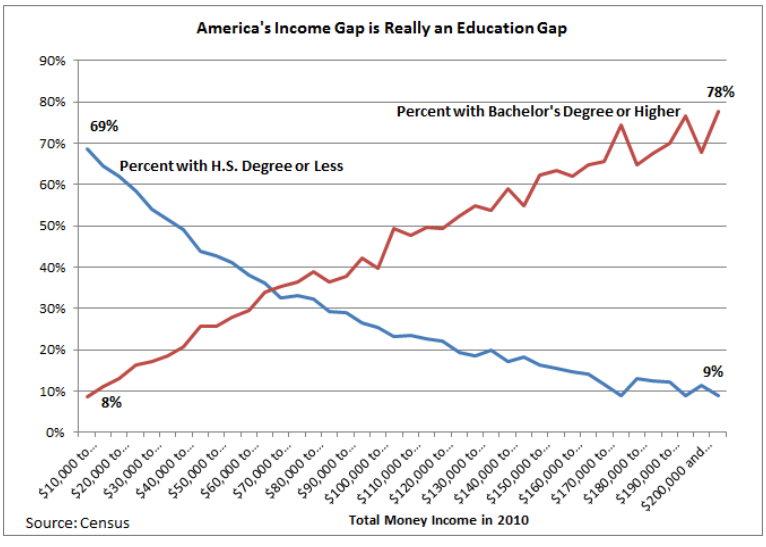


Figure 2

Figure 2 above shows that as people's income rises, they are more likely to have a college degree or higher, in the US. By contrast, those with the lowest incomes are most likely to have a high school education or less
In the US, just 8 percent of those at the lowest income level have a college degree while 78 percent of those earning $250,000 or more have a college degree or advanced degree. At the other end of the income scale, 69 percent of low-income people have a high school degree or less, while just 9 percent of those earning over $250,000 have just a high school degree.

Opportunities for Moving Up

Economic mobility, or the ability to improve one's economic status, varies widely across different regions and socio-economic groups. In many advanced economies, such as those in Northern Europe, systems are designed to facilitate upward mobility through progressive taxation, robust social safety nets, and universal access to high-quality education and healthcare. In Denmark, a country renowned for its strong social policies and significant investments in education, the intergenerational elasticity of income—indicating the extent to which parents' income influences their children's income—is relatively low at approximately 0.22. This suggests higher economic mobility compared to countries with higher elasticities. In contrast, the U.S., with less comprehensive social policies, has an intergenerational elasticity of around 0.47 (OECD), indicating lower economic mobility and a stronger correlation between parents' and children's incomes. This data highlights how well-designed social policies and educational investments can significantly enhance economic mobility, allowing individuals from lower-income backgrounds to improve their economic status more effectively.



Figure 3
The map shown in Figure 3 illustrates upward mobility rates in different countries. It can be seen that with many social policies, Northern European countries have the highest upward mobility rates

Moreover, in countries with relatively higher income inequality, such as Turkey, barriers to upward mobility can be considerable. Structural issues, including disparities in educational quality, unequal access to healthcare, and varying minimum wage laws, create obstacles for individuals from disadvantaged backgrounds. For example, a report from the Brookings Institution highlights that children born into low-income families in the U.S. are less likely to move up the income ladder compared to their peers in countries with more equitable systems. This is due to factors such as underfunded schools and limited access to affordable higher

education. The poverty trap is especially pronounced in these settings, where insufficient opportunities for skill development and economic advancement keep generations locked in a cycle of poverty.(13)

The poverty trap, also known as the cycle of poverty, refers to the different conditions that keep individuals and families in poverty across generations. This cycle is driven by a lack of access to critical resources such as quality education, healthcare, and stable employment opportunities. This means children born into low-income households are more likely to attend underfunded schools, limiting their skill development and future earning potential. Without targeted interventions, such as investments in education and healthcare or policies aimed at reducing wealth disparities, this cycle becomes increasingly difficult to break. This qualification makes it a crucial topic for income inequality and upward mobility. As discussed earlier, in countries like Brazil, Turkey and even the US, this cycle is already almost impossible to break, increasing income inequality levels for these countries each day.(16)

To address these barriers and enhance opportunities for upward mobility, several comprehensive policy interventions can be considered. Investing in education, particularly for marginalized communities, can provide individuals with the skills and qualifications needed to access better job opportunities. Implementing progressive tax policies and increasing investments in social safety nets can help reduce economic disparities and increase upward mobility. Additionally, promoting equal access to healthcare and addressing discrimination in hiring and promotions can ensure that all individuals have a fair chance to improve their economic status no matter what race or where you belong to.(12)(14)

Regional economic disparities can also impact economic mobility. In the U.S., there are significant differences in economic opportunities based on geographic location. Areas with high paucity levels often lack the necessary resources and infrastructure to support upward mobility. Research by the Equality of Opportunity Project indicates that children growing up in poorer regions of the U.S. face significantly lower chances of achieving higher income levels compared to those in more wealthy areas. Regional differences in economic opportunities highlight the need for targeted efforts to address local barriers and support people's ability to improve their economic situation.(15)

Addressing income inequality and improving economic mobility requires a multi-faceted approach that includes both immediate and long-term strategies, however. Short-term measures might include expanding access to vocational training programs and increasing minimum wage levels to ensure that even low-income workers can achieve a decent standard of living. Long-term strategies should focus on systemic changes such as reforming education systems to ensure equal quality across all regions and investing in community development projects to improve economically distressed areas. By adopting a comprehensive approach that tackles both immediate needs and underlying systemic issues, policymakers can create a more equitable environment where economic mobility is accessible to all individuals, regardless of their background or location.(World Bank-World Development Report 2020)

6. Discussion and Policy Recommendations

Income inequality remains a critical issue impacting global economic stability, upward mobilty, and social cohesion. This research highlighted the significant role that disparities in education and job opportunities play in perpetuating income inequality. Individuals from lower socio-economic backgrounds face considerable barriers to upward mobility throughout the globe, particularly in regions where access to quality education is limited. The contrast between countries with equitable educational systems, like Norway, and those with pronounced disparities, such as the United States or countries with even worse income distribution like South Africa, underscores the importance of equal access to education in mitigating income inequality.

The findings indicate that economic policies greatly influence income distribution. Nations with progressive taxation, well-funded public education, and strong social welfare programs, such as Finland, tend to experience lower income inequality and greater social mobility. Conversely, countries with regressive tax policies and inadequate social safety nets often see widening income gaps. Thomas Piketty's work on the importance of progressive taxation and social policies in reducing inequality further supports these conclusions.(9)

To address income inequality effectively, governments must focus on policies that enhance educational access and economic opportunities. Investing in high-quality education across all levels is crucial for equipping individuals with the skills needed for better job prospects and higher wages. For instance, the OECD's 2022 report highlights that each additional year of schooling increases an individual's earnings by an average of 10% (OECD). Furthermore, implementing progressive tax policies and expanding social safety nets are essential for redistributing wealth and supporting those most affected by economic disparities. Data from the World Bank shows that progressive tax systems, which impose higher tax rates on higher income brackets, can reduce income inequality by up to 7% (World Bank). Not only only reforming the taxing system, but also changing the marginal tax rates is an important component in creating a better income distribution. In the US, the marginal tax rate dropped from 70% in 1980 to 37% in 2020 — as can be seen from figure 4 right below—, which was discussed earlier in the paper, detrimentally affecting the Income Inequality levels in the country during this period of time.



Figure 4

Additionally, social safety nets such as unemployment benefits and social transfers have been shown to significantly reduce poverty rates and income inequality. For example, social protection programs in Finland have been effective in lowering the Gini coefficient— a measure of income inequality—by approximately 10% (World Bank). These measures collectively highlight the importance of targeted policies in mitigating income inequality and fostering a more equitable society.

Addressing income inequality also necessitates a focus on labor market reforms that ensure fair wages and job security. Many countries have seen stagnation in real wages for low-income workers, which worsens income inequality over time. Raising the minimum wage to a living wage can significantly improve income distribution. Evidence from the International Labour Organization (ILO) indicates that increasing the minimum wage in several countries led to an average reduction in poverty rates by 6%. Furthermore, promoting job training and vocational programs can help workers transition into higher-paying jobs and adapt to changing job market demands. By investing in job training and education programs, governments can enhance workers' skills and increase their earning potential.

Solving systemic issues such as gender and racial inequality is crucial for reducing further income disparities. Women and minority groups often face additional barriers in the labor market, including wage gaps and limited career advancement opportunities. Policies aimed at promoting equal pay for equal work and supporting diversity and inclusion in the workplace are essential for achieving broader economic equity. For example, gender pay gap initiatives and affirmative action policies have been shown to improve economic outcomes for underrepresented groups. Countries that have implemented gender parity policies experienced a 12% increase in female labor force participation and a reduction in gender income disparity (World Economic Forum). Comprehensive approaches that address these systemic issues will further contribute to reducing overall income inequality.

Reducing income inequality requires a comprehensive approach that tackles the root causes of disparities in education and economic opportunities. Adopting inclusive policies and investing in human capital will help governments create more equitable societies and promote both social justice and economic growth.

7. Conclusion

Income inequality is a critical global issue, impacting economic stability, social cohesion, and individual opportunities. This research traced its evolution from the Pre-Industrial Era to the present, revealing how technological advancements and economic shifts have shaped the current landscape. Today, globalization and technological change have deepened disparities, especially between wealthier and less developed nations.

Education emerged as a key factor in reducing inequality, offering a path to better economic outcomes and social mobility. However, unequal access perpetuates disparities, particularly in regions with weak governance. Racial inequality, particularly in the U.S., further exacerbates income gaps, underscoring the need for policies addressing both economic and social dimensions.

The research also highlighted the social impact of inequality, including diminished trust in institutions and social fragmentation. Case studies of government policies revealed both successes and challenges, with Northern Europe demonstrating effective strategies for reducing disparities, while regions like Latin America and Sub-Saharan Africa struggle with systemic barriers.

Addressing income inequality requires region-specific strategies that account for local conditions and challenges is a crucial part. In the Middle East, diversifying economies beyond oil dependence is important. Investing in sectors such as technology and renewable energy can create new job opportunities and reduce income gaps. Additionally, implementing progressive tax policies and enhancing social safety nets can contribute to a fairer distribution of wealth. Promoting gender equality by increasing women's participation in the workforce and leadership roles will further support inclusive economic growth. These region-specific approaches, tailored to local needs and conditions, can significantly help in reducing income inequality and fostering more equitable economic development.

This paper calls for a broad approach to reducing income inequality, focusing on better access to education, healthcare, and social support, while also addressing racial and gender inequalities. Overall, the research highlights the need for fair and inclusive policies to create a more just and equal society with better welfare.

Works Cited

1. Allen, Robert C. "Economic structure and agricultural productivity in Europe, 1300–1800." European Review of Economic History 4.1 (2000)
2. Lucas, Robert E. "The industrial revolution: Past and future." Lectures on economic growth 109 (2002)
3. Piketty, Thomas. "Capital in the twenty-first century." Trans. Arthur Goldhammer/Belknap (2014)
4. Brada, Josef C., and E. M. Bah. "Growing income inequality as a challenge to 21st century capitalism." Italian Association for the Study of Economic Asymmetries, Working Paper 1402 (2014).
5. Milanovic, Branko. Global inequality: A new approach for the age of globalization. Harvard University Press (2016)
6. Gasparini, Leonardo, and Nora Lustig. "The rise and fall of income inequality in Latin America." (2011)
7. Income Inequality Data and Trends, OECD
8. Stiglitz, Joseph E. The price of inequality: How today's divided society endangers our future. WW Norton & Company, (2012)
9. Piketty, Thomas. The Economics of Inequality: Harvard University Press. Harvard University Press, (2015)
10. Income Inequality Trends in Sub-Saharan Africa, UNDP(2017)
11. United States Bureau of Labor Statistics, (2022)
12. Derenoncourt, Ellora. "Can you move to opportunity? Evidence from the Great Migration." American Economic Review 112.2 (2022): 369-408
13. Isaacs, Julia B. The Economic Mobility of Black and White Families, Brookings Institution (2007)
14. Derenoncourt, Ellora, et al. "Racial inequality, minimum wage spillovers, and the informal sector." Unpublished manuscript (2021)
15. Chetty, Raj, et al. "Where is the land of opportunity? The geography of intergenerational mobility in the United States." The quarterly journal of economics 129.4 (2014)
16. Duflo, Esther, and Abhijit Banerjee. Poor economics. Vol. 619. New York: PublicAffairs, (2011)

—

Figure 1: https://ourworldindata.org/economic-inequality

Figure 2: https://taxfoundation.org/data/all/federal/census-data-shows-inequality-linked-education-not-taxes/

Figure 3: https://en.wikipedia.org/wiki/Global_Social_Mobility_Index

Figure 4: https://taxpolicycenter.org/briefing-book/how-do-federal-income-tax-rates-work

Treatment for Veterans with PTSD By Josephine connell

Abstract

Researching the best way to treat people with Post traumatic stress disorder (PTSD) becomes more pressing everyday. The veteran community has had a particularly high rate of veterans who are diagnosed every year. PTSD affects their functions and quality of life throughout the day and night. The new forms of treatments have been less studied due to the uncertainty about their effectiveness. This paper uses a mixture of literature reviews and interviews with various people at the VA who specialize in treating PTSD to explain how new treatment works and why they are less studied. New treatments for PTSD work in several ways and often apply techniques that are used in traditional treatment plans. EDMR and KAP are used to desensitize a person to the trauma that they faced while pharmacological treatments are used to help minimize the symptoms and make it easier to get through a typical day. However, many new treatment routes haven't been proven as effective as typical treatment routes and therefore are less researched. The goal of doctors who treat PTSD is to find what will help the most and hurt the least and therefore traditional methods are used the most often. Key words: PTSD, Psychological Treatment, Pharmacological Treatment, and Alternative Treatment Pathways, Eye movement Desensitization and Reprocessing (EMDR), Ketamine Assisted Psychotherapy (KAP).

Introduction

Across the United States 6-8% of the population is impacted by Post Traumatic Stress Disorder, or PTSD, which is a debilitating mental disorder (Ressler). PTSD is caused by a traumatic event in a person's life which varies from person to person and 30-40% of PTSD risk is hereditary (Ressler). The risk can also be increased based on a persons, race, socioeconomic status, lower rank in military, low level of education, amount of deployments, previous mental health struggles, lack of social support, gender, and military sexual trauma (Reisman). Some symptoms include intrusive memories, negative mood changes, avoidance, depersonalization, or feeling detached from their own body, and derealization, or feeling as though their surroundings are not real (Harnett). PTSD effect's multiple locations in the brain including the prefrontal cortex, hippocampus, and amygdala which are necessary portions of the brain in helping a person process fear and responding with emotions (Harnett). PTSD is a critical field of medical studies because many people with this disorder struggle to get through their everyday lives; from nightmares, ignoring people in their lives, and being unable to work (VA). This effects the veteran community at an even greater rate with approximately 6% of male veterans and 13% of female veterans having PTSD (National department of Veterans Affairs). Starring after the Vietnam war there was a large amount of stigma around PTSD and other mental health conditions because they were thought to not exist (VA).The goal is to determine what forms of treatment are the most effective and if new treatment could be beneficial to some people. There is a large amount of research that has been done on cognitive processing therapy and pronged exposer because they have been proven extremely effective. However, there is little research

about EMDR and KAP because they have yet to be proven entirely effective. The information in this paper comes from literature reviews and interviews with people at the VA.

Methods

In this paper google scholar and pub med were used to find reliable resources. Key words searched were PTSD, brain impact, Alternative treatment pathways, Psychological treatment, pharmacological treatment, Cognitive Processing Therapy, Ketamine Assisted Psychotherapy, Eye Movement Desensitization and Reprocessing, and Prolonged Exposure. Exclusion criteria was set for dated information and opinion based papers. Information was gathered after an outline was made based off of what would best fit into the order of the paper. Requirements for inclusion were that the studies were after 2010 and had a clear study objectives and findings, or that the information in the paper was synthesized from reliable sources. Researching authors and researchers and making sure they had the credentials for research and degrees was an important part making sure the authors were reliable.

Historical treatment methods

There are three forms of treatments, Psychological Treatment, Pharmacological Treatment, and Alternative Treatment Pathways (Reisman). Doctors look at what treatments will help the most and do the least harm when deciding which route would be the most effective (VA).

Psychological treatments such as Eye movement Desensitization and Reprocessing (EMDR) help to reprocess traumatic information. This in-turn decreases the impact on the patient's psyche. EMDR requires the patient to recall the traumatic event while focusing on external stimulation such as eye movements and hand taps (Reisman). They need to recall multiple parts of a traumatic memory and reassociate those events with a different and positive representation(Chemtob). This form of treatment is used less frequently because it has yet to be proven entirely effective (VA).

Pharmacological treatments include first line medications and second line therapies. First line medications include Sertraline and Paroxetine (Reisman). These drugs are prescribed as an antidepressant to increase serotonin levels in the brain. In a treatment studying the acute treatment of panic disorders it was found that Sertraline and Paroxetine are of equal effectiveness in terms of treating a panic disorder (Bandelow). Second line therapies include Mirtazapine, Tricyclic Antidepressants, and Monoamine Oxidase Inhibitors (Reisman). These drugs work to prevent enzymes and receptors from blocking the production of neurotransmitters such as serotonin, norepinephrine, and dopamine. Mirtazapine blocks the alpha-2 adrenergic receptors in both the cell bodies and nerve terminals (Sheffler). This promotes the release of norepinephrine which is a neurotransmitter that works to regulate arousal, attention, stress reactions, and cognitive function (Hussain). Tricyclic antidepressants work to increase serotonin and norepinephrine levels by preventing reabsorption in the brain and also binds to the M1 and H1 receptors (Sheffler). Monoamine Oxidase Inhibitors block the monoamine oxidase enzymes that break down neurotransmitters such as norepinephrine, dopamine, and serotonin(Sheffler).

Ketamine Assisted Psychotherapy

Ketamine Assisted Psychotherapy, or KAP, is a newer treatment for PTSD as the FDA approved it in 2019. Ketamine is a dissociative anesthetic that acts as an antidepressant and has psychoactive properties (Debora, Vivian, Julia). The ketamine is infused 24 hours prior to each therapy session with an IV and following the infusion the PTSD patient does multiple sessions of prolonged exposure therapy (Drozdz). Ketamine mainly effects the glutamate system in the brain by blocking N- Methyl - D- Asparate , NMDA receptors which activates the Alpha – Amino – Hydroxy – 5 – Methyl – 4 – Isoxazolepropionic Acid Receptor, AMPA receptors (Debora, Vivian, Julia). Through the antidepressant benefits KAP can give rapid relief of anxiety and depression and help to form a more neutral connection with a traumatic experience (Debora, Vivian, Julia). The psychoactive properties come from higher dosages in which patients can increase emotional openness, reduce defensiveness, have a transpersonal experience and gain insights into the traumatic experience. This allows for a therapist to be able to help a patient change their outlook on the experience while remaining in a comfortable head space, or in come cases go beyond their sense of self (Debora, Vivian, Julia). KAP has been proven to cause large reductions in anxiety and depression (Drozdz). This can allow for two different benefits. First it can remove fear by decreasing a conditioned fear response overtime by repeated exposure to the fear whether it is an object or an event. Second is the restructuring of the traumatic memories by making them more manageable (Debora, Vivian, Julia). It has been shown to have a decrease in how severe the PTSD symptoms are — this was measured by the PLC-5 (Drozdz).

Discussion

In summary, medicine is changing every day and one treatment route isn't right for all patients. However, many new forms of treatment for PTSD have shown a lot of promise, but not enough to be determined to be the most helpful and the least harmful. This advances the knowledge of all different types of medicine and what each treatment does and how/ where it creates impact in both the mind and the body. Limitations within determining effective's of treatment pathways are high because of how long clinical studies take. Doctors put a lot of effort into determining new and more effective treatments but it takes a long time and new ones will come about very frequently and more ideas come about every day

Work cited

Harnett NG, Goodman AM, Knight DC. PTSD-related neuroimaging abnormalities in brain function, structure, and biochemistry. Exp Neurol. 2020 Aug;330:113331. doi: 10.1016/j.expneurol.2020.113331. Epub 2020 Apr 25. PMID: 32343956.

Ressler KJ, Berretta S, Bolshakov VY, Rosso IM, Meloni EG, Rauch SL, Carlezon WA Jr. Post-traumatic stress disorder: clinical and translational neuroscience from cells to circuits. Nat Rev Neurol. 2022 May;18(5):273-288. doi: 10.1038/s41582-022-00635-8. Epub 2022 Mar 29. PMID: 35352034; PMCID: PMC9682920.

Vivian W. L. Tsang, Dragos C. Ragazan, Pamela Kryskow, Zach Walsh & Shannon Dames. (2023) A Pilot Study Comparing a Community of Practice Program with and without Concurrent Ketamine-Assisted Therapy. Journal of Psychoactive Drugs 0:0, pages 1-10.

Julia M. K. Freind, Fernando R. Beserra, Bruno S. Menezes & Daniel C. Mograbi. (2023) Therapeutic Protocols Using Ketamine and Esketamine for Depressive Disorders: A Systematic Review. Journal of Psychoactive Drugs 0:0, pages 1-17.

Debora Oliveira, Rodrigo Fontenele, Jeremy Weleff, Mehmet Sofuoglu & Joao P. De Aquino. (2023) Developing non-opioid therapeutics to alleviate pain among persons with opioid use disorder: a review of the human evidence. International Review of Psychiatry 0:0, pages 1-20.

Reisman M. PTSD Treatment for Veterans: What's Working, What's New, and What's Next. P T. 2016 Oct;41(10):623-634. PMID: 27757001; PMCID: PMC5047000.

Bandelow B, Behnke K, Lenoir S, Hendriks GJ, Alkin T, Goebel C, Clary CM. Sertraline versus paroxetine in the treatment of panic disorder: an acute, double-blind noninferiority comparison. J Clin Psychiatry. 2004 Mar;65(3):405-13. doi: 10.4088/jcp.v65n0317. PMID: 15096081.

Sheffler ZM, Patel P, Abdijadid S. Antidepressants. [Updated 2023 May 26]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK538182/

Drozdz, S. J., Goel, A., McGarr, M. W., Katz, J., Ritvo, P., Mattina, G. F., … Ladha, K. S. (2022). Ketamine Assisted Psychotherapy: A Systematic Narrative Review of the Literature. Journal of Pain Research, 15, 1691–1706. https://doi.org/10.2147/JPR.S360733

Interview with woman at VA not to be named — stays anonymous\

Chemtob, C. M., Tolin, D. F., van der Kolk, B. A., & Pitman, R. K. (2000). Eye movement desensitization and reprocessing. In E. B. Foa, T. M. Keane, & M. J. Friedman (Eds.), Effective treatments for PTSD: Practice guidelines from the International Society for Traumatic Stress Studies (pp. 139–154). The Guilford Press.

Exploring Sustainability in Cosmetics for Policy Making and Marketing Communication: A Case Study on Sephora's approach and practices By Ege Yazlar

Abstract

This study aims to explore sustainability practices in the cosmetics industry in an attempt to further inform policy making and marketing communications. The objectives of the research are multifaceted, aiming to investigate Sephora as one of the biggest cosmetic giants, its practices and communication initiatives, in order to examine the current progress in the cosmetic industry in terms of sustainable production, sustainable consumption and circular economy. By delving into Sephora's case, Sephora's public Chemical Policy Reports and scholarly articles, this study is aiming to shed light into the cosmetic industry's progress towards sustainability. Eventually, this study aims to showcase the ongoing development on sustainability and shift needed in the cosmetic industry to a more environmentally conscious future. This aligns with the industry's current objectives to address the changing consumer attitudes in the sector.

Introduction

Sustainability, and sustainability agendas, in the cosmetics industry, as well as in many other industries have been growing from the business, as well as the consumer perspective. We are witnessing increasing concerns about product safety, environmental impact and social impact, while further concerns are being raised from unfair trading (Bom et al, 2019). Whilst these concerns keep growing, the cosmetics companies are pursuing sustainable initiatives on all phases of their product life cycle, including packaging. However research has identified that this shift has taken place for the purpose of cost saving and marketing communication objectives, aside from other factors (Civancik-Uslu et al., 2019, L'Haridon et al., 2018). It is evident that there are plenty of "clean'', "green'', "environmentally friendly'' descriptions on cosmetics products in the market while we see that the definitions are not standardized, neither understood precisely by consumers (Boivin, 2010). This research is exploring how the cosmetics industry, through utilising the case of Sephora, as a key retailer worldwide, is incorporating sustainability principles and practices, with particular focus on the role sustainability plays in their marketing communications in the sector. The study is particularly interested in how businesses in the sector communicate and incorporate in their communication tactics and strategies the importance of sustainability in line with sustainable production, consumption and circular economy. The study has identified that there is an overwhelming lack of studies on the actual and specific environmental activities of cosmetics and personal care companies, so further research is needed to map the landscape of sustainability in the cosmetics sector. In addition, most consumer environmental activities have been studied from a branding and marketing perspective, however there is a need for research that drives consumer environmental behavior from a different perspective**.** Lastly, this research also aims to provide insights for policy making in the sector.

Sustainable production and the case of Sephora

In order to understand the practical relevance of the current situation in the industry, we are looking into Sephora's case and their recent developments in sustainability matters and

practices. Sephora is the world's leading global prestige beauty retailer of personal care and beauty products, offering close to 500 brands alongside its own private label called the Sephora Collection. Its product range includes cosmetics, skincare, fragrance, nail color, beauty tools, body products, and hair care items. (Linkedin, 2024; Shastri, 2024) It operates in 34 markets with more than 3000 points of sales which include flagship stores and over 30 e-commerce sites with its 52,000 employees globally (LVMH Company website, 2024). Sephora claims that as a company its ambition is to connect customers and beauty brands through an immersive personalized seamless experience across every touchpoint in its store and create a welcoming beauty shopping experience for all (Patov, 2024). Unlike the typical small perfume shop, the concept of the store layout is to provide a large and open sales area and an environment where customers can more easily access and try the products. This concept was unique at the time of the origin of Sephora and revolutionized the market (Putnal, 2009). By the end of 2000, Sephora had become a major international brand, with more than 460 stores. (Marka Fikirleri Editor, N.D.) The company has been recognized by Forbes as one of America's Best Employers four years in a row, from 2018-2021. For three years in a row, from 2019-2021, Sephora scored 100% on the Human Rights Campaign's Corporate Quality Index, the national benchmarking tool on corporate policies and practices pertaining to LGBTQ+ employees. (Sephora Company Website, 2024).

Sephora has a commitment to contribute to sustainable development through 4 pillars which are: caring for the climate, engaging in responsible resourcing, eco design and product transparency, reducing, reusing and recycling. They have set a target to reduce the carbon footprint (Scope 1 & 2 CO2 emissions) of its stores, headquarters, and distribution centers by 50% by 2026. They implemented energy-saving solutions, such as more efficient lighting in stores and other facilities across the globe in order to achieve 100% renewable energy in various markets (mainly developed markets like France and the United States), with more markets to come. They are also claiming that they apply best practice tools and third party assessors to help track reductions in their indirect emissions across the whole value chain, and hold themselves accountable to the progress. (Sephora Company Website, 2024)

Furthermore, they have joined an effort with Pact Collective called Beauty (Re)Purposed Program which cooperates with cosmetic industry-related brands in creating a take-back initiative to address the problem of hard-to-recycle packaging in the beauty industry. Sephora is running this campaign at the moment across 600 stores in the US and Canada. The partnership with Pact Collective is in accordance with the mission of Sephora (which is to achieve recycling not only with the clients but also the broader beauty industry) and helps to bring the recycling process to a whole new level, garnering clients' awareness of their responsibility. Participation in the Beauty (Re)Purposed Program for consumers begins with separating the packaging, as a part of the transforming process, at home and making sure there is no liquid or product inside. Similarly, customers can simply just leave their old and empty beauty products containers in any store of Sephora without any hassle, thus, supporting the company's overall target of lowering waste management and sustainability (Zwieglinska, 2023).



Fig. 1 Sephora's seals related to sustainability - Used across offline stores and online stores

Sephora recently launched 2 new sustainability seals for its products across stores in April 2024 (Smith, 2024). To earn the 'Clean at Sephora' seal (Fig. 1), products must be free from over 50 different ingredients, including sulfates (SLS and SLES), parabens, phthalates, and more. The objective is to provide healthier beauty choices to the customers without compromising on the effectiveness by banning several potentially harmful ingredients in their products. 'Planet Aware at Sephora', (Fig. 1) meanwhile, is a seal that highlights "brands that are making commitments across four sustainability topics from ingredients to corporate commitments to packaging" (Sephora Company Website, 2024).

In 2020, Sephora also launched a number of labels (Beauty Packaging, 2020) , including "Good for you" for products whose formulas contain at least 90% ingredients of natural origin, and "Good for a better planet" for products whose packaging is eco-designed or made from sustainable natural materials (Fig. 2). In addition, Sephora store salesperson in the Dubai Mall of Emirates branch claimed that Sephora Collection product range is made from recyclable packaging partially, like some plastics are degradable. At the same time, she claimed that their famous Lip Stain liquid lipsticks packaging has been recently redesigned to have lesser plastic material while keeping the same product volume. (Fig. 3) She also mentioned that some products in Sephora Collection have a Vegan label (Fig. 2) for products with plant-based ingredients, which are known for their gentle and nourishing properties. (Sephora Company Website, 2024)



Fig 2. Sephora's sustainability seals- Used across stores and online stores



Fig. 3 Point of Sales Communication for Sephora Collection Lip Stain product

In addition to the above information about Sephora's actions, initiatives and position, the company has launched in 2019 a pblic Chemical Policy report. They first reported data in 2020. In order to understand the information chases by the company, an analysis of their Chemical Policy reports has been conducted. The reports are dated from 2020 to 2023, and are organized based on the goals set at the original Policy document on 2019, as follows:

High-Priority Chemicals: Sephora has identified a list of high-priority chemicals for reduction and elimination from their assortment in 2019. They had set a target to reduce the number of products that contain these chemicals in the nine chemical categories by 50% in the next 3 following years. Not only that, they had also announced that their private label products satisfy the requirements on quality, traceability, and product safety, including adhering to a unique internal RSL (restricted substance list) that goes beyond the 1,300+ ingredients restricted by the European Union (Sephora public Chemicals Policy, 2019)

Safer Alternatives: Sephora's second goal in this report was encouraging the replacement of high-priority chemicals with safer alternatives and working toward avoiding regrettable substitutes. They had announced that they would support industry partners on innovation and green chemistry practices to help find safer alternatives, plus provide resources for brands to evaluate ingredients and help find safer alternatives to chemicals of concern. (Sephora public Chemicals Policy, 2019)

Transparency and Consumer Knowledge: Sephora also committed to work together with its brands and suppliers to disclose ingredient information for products they sell online and in store. By 2020, they targeted 100% of formulated beauty and personal care products to have all intentionally added ingredients listed on our US Sephora.com website. (Sephora public Chemicals Policy, 2019)
In addition, in 2023 Sephora has updated the Policy and added the followings under each goal, as follows:

High-Priority Chemicals-This list of Sephora is updated as per the latest developments in product safety generally, but also considering environmental justice and the types of ingredients used in products marketed to people of color. Sephora has also announced that they will continue to include these considerations in their policy implementation and their future updates.

The high-priority list is expanded with additional groups of chemicals which are Antimicrobials, Antioxidants, Bisphenols, Borates, Ethanolamine, Fragrances of concern, Heavy metals, Isothiazolinones, PFASs.(Sephora public Chemicals Policy. 2023)

Safer Alternatives- Sephora committed to continue this goal further by providing training to their brand partners, where they have partnered with industry leaders to provide tools and resources to suppliers designing products, and by offering products that meet third-party product standards. They aimed to facilitate greater access to resources that support safer product design and report on the number of products that meet alternative criteria, such as those that meet applicable third-party certification programs (e.g., Cradle to Cradle, etc.) (Sephora public Chemicals Policy, 2023).

Transparency and Consumer Knowledge- They committed to advance this goal further by having all Clean at Sephora products to list intentionally added ingredients, plus 24 fragrance allergens defined in EU regulations by 2024. For the rest of the products, Sephora agreed to disclose these fragrance allergens and those to be defined by the US FDA on their US Sephora.com website by 2024. To increase transparency of fragrance ingredients, they would also work to educate and explain the function and origin of ingredients by 2030 or earlier. Lastly, they would encourage all brands with formulated products to provide full disclosure of product ingredients down to 100 ppm, including fragrance ingredients, on Sephora's website by 2030 (Sephora public Chemicals Policy, 2023).

The data from the reports has further been organised based on the goals set, and the following results have been obtained (Table 1). Sephora has reduced the number of products with one or more of the chemicals on the initial list of high-priority chemicals in their assortment by 39.5% between 2019 and 2022. This led to 95.5% of their assortment not containing high-priority chemicals from the initial list.

Table 1.Sephora's High Priority Chemicals Ingredient Target Overview 2019-023

High Priority Chemicals	2019	2020	2021	2022	2023
Formaldehyde & formaldehyde donors Parabens Phthalates Miscellaneous preservatives Ethanolamine compounds	Reduce # of products with these categories by **50%** in 3 years	**14.3%** reduction in SKUs with 1 or more HPC from Dec 2019 to July 2020	**28%** reduction in SKUs with 1 or more HPC from Dec 2019 to July 2021	**39.5%** reduction in SKUs with 1 or more HPC from Dec 2019 to July 2022	**95.5%** of assortment has no HPC from the initial list

Coal tar compounds Aluminum salts Chemical sunscreen Miscellaneous solvent		**93.8%** of SKUs sold on Sephora.com have no HPC as of July 2021	**94.8%** of SKUs sold on Sephora.com have no HPC as of July 2021	**95.5%** of SKUs sold on Sephora.com have no HPC as of July 2022	HPC list is expanded with; Antimicrobials Antioxidants Bisphenols Borates Ethanolamine Fragrances of concern Heavy metals Isothiazolinones PFASs Phthalates Quaternary ammonium com. Siloxanes UV filters Packaging specific

Since 2019, Sephora has contracted external experts to enable suppliers to identify alternatives that may be safer. They had continued year by year to fund and collaborate with these and additional industry partners in innovation and green chemistry practices such as Change Chemistry (formerly the Green Chemistry & Commerce Council), Novi Connect, ChemFORWARD, and the Environmental Defense Fund (EDF). This includes amplifying verified safer alternatives on cloud-based platforms including ChemFORWARD and Novi Connect, listed as “ChemFORWARD SAFER”. They further supported the development of 26 chemical hazard assessments in 2021 that provide the information needed to understand if an ingredient may be a safer alternative (Table 2).

Table 2. Sephora’s Safer Alternatives Ingredient Target Overview 2019-2023

Safer Alternatives	2019	2020	2021	2022	2023
	Encouraging the replacement of HPC with safer alternatives and avoid regrettably substitute	50 brands utilizing tools to identify safer alternatives through pilot with Novi Connect by the end of 2020 4 safer alternative assessment portfolio analyses developed in partnership with ChemForward 73% of total HPC for which safer alternatives have been	26 chemical hazard assessments (CHAs) funded by Sephora These CHAs will be utilized in the Novi Connect tool, making them more accessible for brands with high-priority chemicals to find fully assessed safer alternatives 45 brands onboarded (all Sephora brands have	Sephora funded CHAs are used to qualify ingredients by trade name and if they satisfy the rigorous program rules, earn the “ChemFORWARD SAFER” designation. SAFER trade name ingredients are now fully	Since 2019, Sephora have supported external experts that enable suppliers to identify alternatives that may be safer. We will continue to fund and collaborate with these and additional industry partners in innovation and green chemistry practices such as Change Chemistry (formerly the

		assessed in one of ChemForward's alternative portfolio analysis 11 ChemForward alternative portfolio analyses exist for Beauty and Personal Care even though not all funded by Sephora (and 15 total for all industries) 271 alternatives have been assessed by credible 3rd parties that already exist (for e.g. from the Safer Chemical Ingredient List or SCIL list) and have been incorporated into ChemForward for easy access by the supply chain $200K USD Sephora has invested in identifying Safer Alternatives	Access beginning August 2021) All brands at Sephora, including Accelerate brands, are able to utilize tools to vet, source, and increase transparency by identifying safer alternatives with Novi Connect	integrated into the Novi tool and available to all brands. 121 brands onboarded	Green Chemistry & Commerce Council), Novi Connect, ChemFORWARD, and the Environmental Defense Fund (EDF). This includes amplifying verified safer alternatives on cloud-based platforms including ChemFORWARD and Novi Connect, listed as "ChemFORWARD SAFER". 26 chemical hazard assessments is supported to provide the information needed to understand if an ingredient may be a safer alternative.

Sephora has achieved their goals of listing intentionally added ingredients on their US website by 2022, even though they targeted to reach 100% by 2020. The delay in achieving this goal mainly came from fragrance products delayed development on this goal (Table 3).

Table 3. Sephora's Transparency & Consumer Knowledge Target Overview 2019-2022

Transparency & Consumer Knowledge	2019	2020	2021	2022
	By 2020, Sephora aims for 100% of formulated beauty and personal care products to have all intentionally added ingredients. listed on US Sephora.com website.	**98.0 %** of SKUs on Sephora.com that have ingredient info (All SKUs excluding fragrance)	**99.8 %** of SKUs on Sephora.com that have ingredient info (All SKUs excluding fragrance)	**99.6 %** of SKUs on Sephora.com that have ingredient info (All SKUs excluding fragrance
		95 % of SKUs on Sephora.com that have ingredient info (All SKUs)	**98 %** of SKUs on Sephora.com that have ingredient info (All SKUs)	**99.5 %** of SKUs on Sephora.com that have ingredient info (All SKUs)

In Sephora's public Chemicals Policy report 2022, they have also included progress on their transparency programs even though they have not committed any goals in advance. As per this progress table (seen below in Table 4) Clean at Sephora labeled brands increased by 47.7% from 2019 and Clean+ Planet Positive labeled brands remarkably increased by 55% from 2021 to 2022.

Table 4. Sephora's Transparency Programs Target 2022

Transparency Programs: Clean and Sustainable Products			
Program	Brand Count	Growth	Time
Clean at Sephora	130 41.8% of total	47.7 % increase from 2019	July 2022
Clean + Planet Positive	40 12.9% of total	55.0% increase from 2021	

Sephora has incorporated a sustainable approach in its products manufacturing, the chosen and used ingredients, as well as conveying sustainability promises to its consumers through different labels launched and progression plans published throughout the years. As a leading cosmetics retailer, Sephora is making a substantial contribution to the industry and can lead the way forward.

In conclusion, Sephora, similar to other cosmetic companies in industry, as identified in the study done by Kolling, Ribeiro and Medeiros in 2022, has given greater emphasis in the design and sourcing stages on sustainability efforts, while less effort has been directed to social aspects and other phases of the life cycle. They have initiated some efforts to induce sustainable consumption as well. Consequently, we will delve into the industry developments on sustainable consumption, production and circular economy in the next section to further investigate the progress.

Sustainable consumption, production and circular economy in the cosmetics industry

Sustainable consumption (SC) in cosmetics refers to the use of goods/ingredients that respond to basic needs and brings a better quality without compromising on the outcome, while minimizing the use of natural resources, toxic materials and minimizing excessive use of water in production. The main principle behind this is not to jeopardize the future generations' resources. Although the decision to adopt a sustainable commitment relies on those principles, it is still usually driven predominantly by the expectation of obtaining economic benefits. This is the case in terms of developing higher-value products or gaining more customers (Civancik-Uslu et al., 2019, L'Haridon et al., 2018). Meanwhile, sustainable products have gained great strength within the beauty industry lately. Consumers look for cosmetics with natural elements, socially responsible approaches, and packaging that is less aggressive to the environment (Lin et al., 2018). Sustainable products for personal care are connected to a healthy and eco-friendly lifestyle, and it has been found that these products are increasingly popular with female consumers (Pudaruth, Juwaheer, & Seewoo, 2015).

Tamashiro et al. (2014) explain that consumers are not fully aware of ecologic cosmetics information and benefits, but behave with a high level of ecological affection and present ecological concerns. Regular consumers make decisions based on personal benefits, while environmentally conscious consumers tend to assess the positive environmental and social impacts (Muralidharan & Sheehan, 2016; Xu, Prybutok, & Blankson, 2019). Consumers who purchase sustainable products are more likely to repeat such shopping behavior (Askadilla & Krisjanti, 2017). The greater the individual's attitude toward the environment, the more likely they are to consume sustainable market offerings (Radons, Battistella, & Grohmann, 2016; Silva, Lima Filho, & Freire, 2015). In addition, the level of environmental awareness and engagement is directly connected to the information that the consumer has on the issue (Silva, Lima Filho, & Freire, 2015). Hence, the more aware and informed consumers are, the more they become engaged with sustainable market options. Considering these, the shift to sustainability sensitivity or ecological concerns in the new generations is evident, and therefore the importance of SC is becoming higher than ever.

A recent study by Dubey and Agarwal (2024) has explored the revolutionary shift occurring in the cosmetic industry towards sustainability and circular economy initiatives. The objectives of the research are multifaceted, aiming to investigate the current landscape of sustainability practices within the industry, identifying key drivers behind this transformation, analysing the effectiveness of existing initiatives, and proposing strategies for further advancement. The methodological approach of their research is mainly qualitative through secondary sources with data from online sources, including examination of companies' websites, annual reports, CSR reports, sustainability reports, circularity reports, and any other accessible information related to the companies' initiatives in circular economy practices, zero wastage, and product recycling. Six cosmetic companies were mainly observed that are leading in sustainability initiatives; and Sephora was one of them. The study identifies that the cosmetic industry's primary challenge in adopting a circular economy is the complex nature of its supply chain. Secondly, due to the nature of the products, many formulations incorporate diverse materials, some of which may be difficult to recycle or repurpose, therefore the need for innovative packaging solutions that prioritize recyclability and biodegradability becomes paramount. Consumer behaviour and awareness also play a pivotal role in incorporating more sustainable production methods and transition to a circular economy in the sector. One key conclusion of their research is that sustainability is a long-term movement where consumers endeavour to engage, and their participation in this movement, are key for the effectiveness of the process. The reasons for companies to favour a circular economy approach varies from keeping an eye on environmentally friendly consumers, having a competitive edge, and avoiding certain legal charges. However, the study highlights obstacles in the transition, for example the need for heavy investment. On the other hand, it illustrates that the long-term advantages, in terms of brand reputation and ecological impact, surpass the drawbacks at the onset. Despite all the aforementioned important contributions and insights, the data in the study is limited, therefore primary data or a wider variety of data can be added to give more precise and context-specific conclusions and recommendations.

Viljoen (2024) states that all the stages in the life cycle of a cosmetic product, from the initial design and raw material sourcing, through manufacturing, packaging, distribution, consumer use, and the post-consumer use (disposal) phases are essential to be considered in the circularity of it. Life Cycle Thinking (LCT) (Fig. 4) demonstrates that all phases in a product life cycle are interdependent, therefore any sustainability initiatives implemented in one phase should not adversely affect another phase (Bom et al, 2019). The approach demonstrated in the study shows that an efficient way to ensure sustainability is through a closed system to be adopted, where as little as possible is brought in from the outside. This approach can be adapted in the cosmetic industry, especially when considering the sourcing of raw materials (Doolan, 2024).

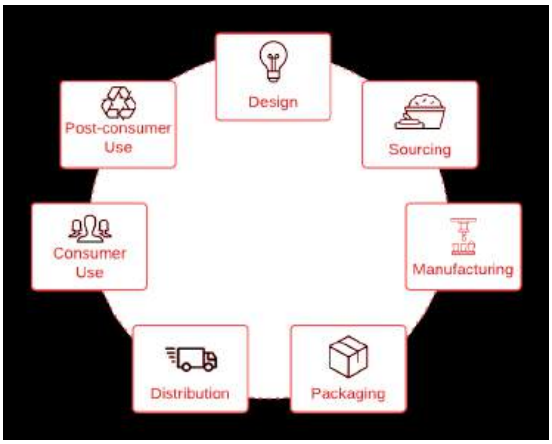


Fig. 4. Cosmetic product life cycle - sustainability at the forefront. Life Cycle Thinking. Adapted from Bom et al.3

A study done by Reid (2022) tells that the shift to sustainability in the cosmetics industry has been driven by increasing demand from consumers for products that have a low environmental impact (in relation to both ingredients and product packaging), but also are safe and tolerable for the user (Fortunati, Martiniello, & Morea, 2020). Many large cosmetics manufacturers have therefore elected to use life cycle assessments to analyze the full environmental impacts of their products and to identify how their sustainability position can be improved (Vital, 2013). This involves assessing the impacts of raw materials extraction, materials processing, transportation emissions, and disposal methods, while then implementing changes to minimize the impacts of these processes (Vital, 2013). An example is the Lush group that seeks to implement principles of eco-design during the ‘design’ and ‘sourcing’ phases of cosmetics manufacturing, by working with small local producers and selecting plant-derived ingredients with a view to preserve biodiversity (Fortunati, Martiniello, & Morea, 2020). The design phase of cosmetics development has a large impact on the sustainability of the final product, and the selection of ingredients for the formulation of cosmetics is considered one of the most crucial steps that determines the sustainability of the finished products (Bom et al.,2019). On the other hand, the performance of ingredients is also important, and natural or sustainable ingredients may not be as effective as their synthetic or unsustainable alternatives (Bom et al.,2019). For example, plastic microbeads were preferred over natural abrasives such as ground fruit pips, because their smoothness is less damaging to the skin and they improve the visual appeal of a product (Chang, 2015).

While Reid (2022) states that the benefits of the CE are well established, there are still many obstacles associated with putting the theory into practice. One of these challenges is ensuring that the widespread adoption of sustainable practices occurs at a scale that would allow the true benefits of the CE to be realized. Attempting to maintain present norms and mechanisms of consumerism within the CE is likely to be unsuccessful (Hobson, 2020). Rather, profound transitions and transformations in consumers' thinking and lifestyles are required. This will involve disrupting present modes of consumption and fundamentally changing the way in which consumers perceive their relationship to goods and services (Kirchherr, 2017; Hobson, 2020; EEA, 2018). Therefore, a transition to a CE will only be successful if a substantial fraction of consumers are prepared to adopt and participate in circular consumption practices.

Another essential component to foster sustainable consumption and promote circularity in the sector, is communication. How key information is communicated and shared with the consumer is paramount in creating awareness, educating and correcting any misconceptions. A study done by Fischer and Reinermann (2021) has explored different concepts of communication and sustainable consumption, while it conceptualized and explained how they are related to each other. Their study is based on a mixed methodology by triangulating quantitative data analyzed via descriptive statistics and qualitative data collected via open-ended questionnaires. These were then coded to be analyzed via thematic text analysis in order to observe conceptualizations of sustainable consumption and communication (SCC). This led to the development of a typology for SCC. Key conclusions shown that sustainable consumption communication is a young, growing, disciplinarily diverse, highly fragmented and sectorized field of research. Four types of SCC research were described, conceptualizing and examining communication as an approach to either (1) behavior change, (2) self-empowerment, (3) systems change, or (4) sustainable consumption discourses. Research findings show that sustainable consumption is largely conceptualized as individual behavior change, especially in the purchase of products, and communication is mainly studied with regards to how message design and transmission promote behavioral change.

Agrawal (2023) has pointed out that it is necessary to provide consumers with more information to raise awareness of eco-friendly products; however, consumers may prioritize affordability, accessibility and availability over environmental friendliness. Moreover, the low efficiency of products can be a barrier to sustainable purchasing intentions among consumers. Therefore, it is crucial to implement measures to lower the hurdles for eco-friendly products such as insufficient product information, lack of variety, low quality, and cost. As a result of a Bain & Company survey, one in five US consumers expected retailers to help them with sustainable shopping by lowering costs, raising awareness, and providing sufficient information about the products in stores (Batista, 2023). Sephora's implementation of Good for, Clean and Planet Aware communications in stores among other products are a good example of this as these products are perceived of similar pricing and quality to the other products for their shoppers. Sephora's commitment to grow their transparency programs is also created to promote changes in consumers behaviours over the years. Having discussed the importance of providing more information and education on sustainable products, Sephora has employed advertising and marketing communications to provide information on their sustainably sourced, designed and

manufactured products. Sephora is also incorporating affordability, alongside the quality, to the consumer in its sustainability strategy (Fig. 5).



Fig 5. Sephora Advertisements from USA

To conclude, with sustainability in the cosmetics sector being new and evolving, it is crucial to set norms, standards and certain measures to formulate a reliable framework for the industry. Consumers look for cosmetics with natural elements, socially responsible approaches, and packaging that is less aggressive to the environment. Consumers now are more aware and better informed, and hence they become more engaged with sustainable market options. However, this shift in consumer behaviour needs to be addressed by the organisations through long-term strategies to ensure that their participation contributes to the effectiveness of the sustainable development and growth. This can be established via encouraging and incorporating circularity in the sector, which can offer long-term advantages, in terms of brand reputation and ecological impact, and surpass the drawbacks at the onset (i.e. heavy investment). While there are long-term advantages, profound transitions and transformations in consumers' thinking and lifestyles are required to be made. This will involve disrupting present modes of consumption and fundamentally changing the way in which consumers perceive their relationship with the organisations in their market transactions. As a concluding remark, it is crucial to implement measures to lower the hurdles for eco-friendly products towards the consumer and allow room for education, transparency, awareness and co-creation. Hence it is imperative to understand the implications for policy making in the sector to ensure good practice and conduct, and pave the way forward. The next section elaborates on this.

Implications for policy making and marketing communications in the sector

By exploring the case of Sephora, as well as key areas in the sector, the information gathered can help to identify potential implications for policy making while providing some further context for marketing communication purposes. To start with, a study by Lim and Kwon (2023) explores in detail how the cosmetics and personal care companies are being proactive in

addressing environmental issues despite the complexity of industry. The study does not only observe them, but also looks into the OEM (Original Equipment Manufacturer), the ODM (Original Design Manufacturer) and the retail companies (Lim and Kwon, 2023). This approach is particularly important when considering policy making in the sector as it takes into account the role and the impact of the entire supply chain, and not only elements of it. Their study also found that the cosmetics and personal care companies (brands, OEM/ODMs, retailers) are more encouraged to take responsibility and hold key roles in environmental issues, and they identified three main levels of responsibility: a) when they have different key roles towards the environment and should be proactive in order to secure future-oriented goals, b) when they ensure transparency of progress, and c) when they communicate and ensure objectivity of results. While the study conceptualizes the levels of accountability in the sector, it still falls short to address the followings:

(1) The absence of published goals, targets, or commitments, or the lack of sufficient information being disclosed, may not mean that the company is not doing enough towards the environment. Each company has a different time frame for report disclosure, comparisons at the same point in time are limited.

(2) The need to study consumer attitudes and behaviors towards the environment.

A research done by Goncalves & Santos in 2022 has analyzed the consumers' behaviors associated with purchasing sustainable cosmetics. Three specific objectives guided this study; the main beliefs involved in sustainable cosmetics purchase, showing that behavioral and control beliefs stood out. The second objective, analyzing the influence of these beliefs on the behavior of purchasing sustainable cosmetics, presented that behavior has more influence than control beliefs on the purchase intention. Finally, the third objective examined the influence of environmental engagement on purchase behavior, and showed that an individual's engagement in environmental issues has a medium effect on the behavior. In contrast, normative beliefs have a weak effect, and control beliefs have no effect. These findings can be considered allude with a study done by Kolling, Ribeiro and Medeiros in 2022, in which it has been shown that cosmetic companies demonstrate advances concerning Corporate Social responsibility, but have not yet reached the necessary maturity to consider in an integrated way the dimensions of sustainability or all phases of the life cycle in product design, as advocated by Design for Sustainability. Their study essentially shows that change in this sector will be possible based on the influence of some policies that aim, in addition to regulating, to encourage the growth of the industry and the businesses in it. In addition, promoting educational campaigns for the consumer is needed. The more educated the consumer is the more confident they are, hence the more control over their decisions they have. As a result they will demonstrate more engagement with environmental issues and this eventually pulls their consumer behaviour towards more environmentally sustainable products.

While Bozza (2022) states that green cosmetics do not have their own definitions, internationally harmonized standards, and minimum requirements in the ingredient percentage as defined by legislation. The current best practice as per Bozza (2022) would be specific

guidelines for cosmetic ingredients of natural origin in Europe, and the clarifications of cosmetic terms (FDA) and green guides of environmentally friendly products (Federal Trade Commission) in the United States which facilitate companies on how to use environmental marketing claims such as ‘organic’, and ‘green’, or ‘carbon offset’. They also provide general principles to facilitate consumer interpretation to avoid greenwashing (Federal Trade Commission, n.d.; U.S. FDA, 2020).

In conclusion, the cosmetics sector underscores the need for policies and marketing strategies that address environmental issues across the entire supply chain. The importance of transparency, consumer education, and regulatory clarity to promote sustainability are critical for the sector. While companies are increasingly responsible, challenges remain in goal-setting and consumer engagement. Standardized definitions for green cosmetics are also essential to prevent greenwashing and ensure authentic progress in the industry.

Conclusions

Sephora’s ongoing commitment to sustainability illustrates a significant effort to not only reduce its environmental impact but also to lead the cosmetic industry toward more sustainable practices. The company has integrated sustainability across multiple aspects of its operations, from carbon footprint reduction and waste management to responsible sourcing and product transparency. Through partnerships, innovative programs, and the introduction of sustainability-focused product labels, Sephora is actively fostering both environmental responsibility and consumer awareness.

On the other side, the cosmetics industry is increasingly adopting sustainable consumption and circular economy practices driven by growing consumer demand for eco-friendly products. While challenges such as complex supply chains and the need for recyclable packaging persist, consumer awareness and behavior are key to this transformation. This also states that effective communication, transparency, and consumer education are crucial in promoting sustainable choices. Ultimately, the industry's shift toward sustainability offers long-term benefits for both the environment and brand reputation, requiring continued innovation and collaboration across the entire product life cycle.

Moreover the cosmetics industry clearly requires comprehensive policies and marketing approaches that address environmental issues throughout the entire supply chain. Many companies are becoming more proactive in their environmental efforts, while establishing measurable goals, ensuring consistent progress, and engaging consumers effectively are still challenges. The lack of standardized definitions and guidelines for "green" cosmetics further complicates the situation, allowing for the risk of greenwashing, where misleading environmental claims could undermine genuine efforts. To address these issues, the industry must prioritize clear regulations, provide educational campaigns for consumers to better understand sustainability claims, and ensure that environmental responsibility is embedded throughout the product life cycle. Only through coordinated industry efforts, improved consumer awareness, and stronger regulations can the cosmetics industry make meaningful strides toward true environmental sustainability.

Works Cited

D. Civancik-Uslu, R. Puig, S. Voigt, D. Walter, P. Fullana-i-Palmer ‘’Improving the production chain with LCA and eco-design: application to cosmetic packaging’’ Resource Conserv. Recycl. (2019) Last modified July 3, 2019. Accessed October 19, 2024. https://doi.org/10.1016/j.resconrec.2019.104475

Boivin, Caroline ‘’In Search of a Green Product Definition’’ Innovative Marketing (2010) Accessed October 15, 2024. https://www.academia.edu/download/88839763/journals.pdf

Lin, Y., Yang, S., Hanifah, H., & Iqbal, Q. ‘’An Exploratory Study of Consumer Attitudes toward Green Cosmetics in the UK Market’’ Administrative Sciences, 8(4), 1-14 (2018) Last modified November 6, 2018. Accessed October 12, 2024. https://doi.org/10.3390/admsci8040071

Pudaruth, S., Juwaheer, T. D., & Seewoo, Y. D. ‘’Gender-based differences in understanding the purchasing patterns of eco-friendly cosmetics and beauty care products in Mauritius’’ Social Resp. Journal 1(1), 179-198 (2015) Accessed October 12, 2024. https://doi.org/10.1108/SRJ-04-2013-0049

Tamashiro, H. R. S., Silveira, J.A.G., Mantovani, D. M. N., & Campanário, C. R. A. A. ‘’Aspectos determinantes do consumo de produtos cosméticos verdes. Revista de Administração e Inovação’’ INMR Inno. & Man. Review 11(1), 238-262 (2014) Accessed October 15, 2024. https://doi.org/10.5773/rai.v11i1.1206

Muralidharan, S., & Sheehan, K. ‘’“Tax” and “Fee” Message Frames as Inhibitors of Plastic Bag Usage Among Shoppers: A Social Marketing Application of the Theory of Planned Behavior’’ Social Marketing Quarterly 22(3), 200-217 (2016). Accessed October 7, 2024. https://doi.org/10.1177/1524500416631522

Askadilla, W. L., & Krisjanti, M. N. ‘’Understanding Indonesian green consumer behavior on cosmetic products: Theory of Planned Behavior Model’’ Journal of Microencapsulation, 15(2), 7-15 (2017) Accessed October 24, 2024. https://doi.org/10.17512/pjms.2017.15.2.01

Radons, D. L., Battistella, L. F., & Grohmann, M. Z. ‘’Generation and gender as moderators on pro-environmental purchase behavior’’ Pensamiento y Gestión, 41, 148-173 (2016). Accessed October 20, 2024. http://www.scielo.org.co/scielo.php?pid=S1657-62762016000200007&script=sci_abstract&tlng=en

Silva, F. Q., Lima Filho, D. D., & Freire, O. ‘’The influence of environmental concern and attitudes towards sustainable consumption on beef meat purchase intention’’ Revista de Administração da UFSM, 8(3), p. 463-481(2015) Accessed November 2, 2024. http://doi.org.br/10.5902/1983465911110900

Linkedin “About company/ Sephora” Accessed December 30, 2024. www.linkedin.com/company/sephora/about/

Shastri, Aditya ‘’A detailed SWOT Analysis of Sephora- Unveiling their secrets to success’’ IIDE (2024) Accessed November 2, 2024. https://iide.co/case-studies/swot-analysis-of-sephora/

LVMH Company Website. “Selective Retailing’’ Accessed October 3, 2024. www.lvmh.com/our-maisons/selective-retailing/sephora

Patov, Aslan “How Sephora Enhances Customer Experience (CX) Through Personalization and Digital Innovation” Renascene (2024) Accessed November 20, 2024. www.renascence.io/journal/how-sephora-enhances-customer-experience-cx-through-personalization-and-digital-innovation

Putnal, O. “15 Brand Names Decoded’’ Woman’s Day Lifestyle (2009)Accessed October 10, 2024. https://www.womansday.com/life/a1139/15-brand-names-decoded-100977/

Marka Fikirleri Editor “Sephora nasil kuruldu?’’ Accessed November 7, 2024. https://www.markafikirleri.com/sephora-nasil-kuruldu/

Sephora Company Website. “About us”. Accessed October 20, 2024. https://www.sephora.com/beauty/about-us

Sephora Company Website. “Sustainability’’. Accessed October 20th, 2024. https://www.inside-sephora.com/en/commitments/sustainability

Zwieglinska, Zohia “Sephora launches national take-back program Beauty (Re)purposed’’ Glossy (2023) Accessed October 27, 2024 https://www.glossy.co/beauty/sephora-launches-national-take-back-program-beauty-repurposed/

Smith, Sophie “How Sephora is championing transparency with its new global sustainability seals’’ The Industry.Beauty (2024) Accessed November 1, 2024. https://theindustry.beauty/how-sephora-is-championing-transparency-with-its-new-global-sustainability-seals/

Sephora Company Website “Planet Aware” Accessed October 20, 2024. https://www.sephora.com/beauty/planet-aware

Sephora Public Chemicals Policy 2019. Accessed December 1, 2024. Sustainable Packaging Guide Sephora Public Chemicals Policy

Sephora Public Chemicals Policy Progress Update 2020 Report. Accessed December 1, 2024. Sephora Public Chemicals Policy Progress Update

Sephora Public Chemicals Policy 2nd Progress Update 2021. Accessed December 2, 2024. Sephora Public Chemicals Policy 2 n d Progress Update

Sephora Public Chemicals Policy 2022 Progress Update Report. Accessed December 2, 2024. Sephora Public Chemicals Policy 2022 Progress Update

Sephora Public Chemicals Policy 2023 Report. Accessed December 3, 2024. Sustainable Packaging Guide Sephora Public Chemicals Policy 2023

Beauty Packaging Website “Sephora unveils new green labels’’ Beauty Packaging (2020) Accessed October 20, 2024 Sephora Unveils New Green Labels | Beauty Packaging

Dubey & Agarwal ‘’Revolutionizing the Cosmetic Industry: Pioneering Sustainability and Circular Economy Initiatives’’ResearchGate (2024) Accessed October 15, 2024 Revolutionizing the Cosmetic Industry: Pioneering Sustainability and Circular Economy Initiative

Fischer & Reinermann ‘’Sustainable consumption communication: A review of an emerging field of research’’ Journal of Cleaner Production (2021) Accessed October 3, 2024. https://www.sciencedirect.com/science/article/pii/S0959652621010994

Lim & Kwon ‘’Sustainable Assessment of the Environmental Activities of Major Cosmetics and Personal Care Companies’’ Last modified July 13, 2023. Accessed October 10, 2024. https://www.mdpi.com/2071-1050/15/18/13286

Kerry-Anne Viljoen (nee White) ‘’Sustainability and Circular Economy in Cosmetic Formulations: Life Cycle Thinking’’ IFSCC (2024) Accessed October 17, 2024. https://ifscc.org/wp-content/uploads/2024/09/Kerry-Anne-Viljoen-Essay-IFSCC.pdf

Bom, S., Jorge, J., Ribeiro, H. M. & Marto, J. A step forward on sustainability in the cosmetics industry: A review. Journal of Cleaner Production. 225, 270–290 (2019) Accessed December 12, 2024. https://www.sciencedirect.com/science/article/abs/pii/S0959652619309655

Doolan, K. ‘’Circularity in the beauty industry: Q&A with MD at Weleda UK Cosmeticsdesign-europe.com (2024) Accessed October 20, 2024 https://www.cosmeticsdesign-europe.com/Article/2024/03/04/Circular-beauty-Q-A-with-MD-at-Weleda-UK/

Reid, Jane ‘’What a waste! Exploring the circular economy for cosmetics products’’ The University of Auckland (2022) Accessed DEc 1, 2024.
https://researchspace.auckland.ac.nz/bitstream/handle/2292/61269/Reid-2022-thesis.pdf?sequence=1&isAllowed=y

Hobson, K. (2020). ‘Small stories of closing loops’: social circularity and the everyday circular economy. Climatic Change, 163(1), 99-116 (2019) Accessed Dec 4, 2024. ‘Small stories of closing loops’: social circularity and the everyday circular economy | Climatic Change

Kirchherr, J., Piscicelli, L., Bour, R., Kostense-Smit, E., Muller, J., Huibrechtse-Truijens, A.,& Hekkert, M. ‘’Barriers to the circular economy: Evidence from the European Union (EU)’’ Ecological economics, 150, 264-272 (2018) Accessed November 20, 2024. https://www.sciencedirect.com/science/article/pii/S0921800917317573

Fortunati, S., Martiniello, L., & Morea, D. ‘’The strategic role of the corporate social responsibility and circular economy in the cosmetic industry’’. Sustainability, 12(12)(2020) Last modified June 17, 2020. Accessed November 17, 2024. https://www.mdpi.com/2071-1050/12/12/5120

Vital, X. ‘’Sustainability: How the Cosmetics Industry is Greening Up’’ Wiley Online Library (2013) Accessed October 2, 2024. Sustainability: How the Cosmetics Industry is Greening Up | Wiley Online Books

Chang, N. J., & Fong, C. M. ‘’Green product quality, green corporate image, green customer satisfaction, and green customer loyalty’’ African journal of business management (2010)

Accessed December 10, 2024.
https://academicjournals.org/journal/AJBM/article-full-text-pdf/DD16E2424448

Kolling, C., Ribeiro J., Medeiros J. ‘’Performance of the cosmetics industry from the perspective of Corporate Social Responsibility and Design for Sustainability’’ Sustainable Production Consumption (2022) Last modified December 15, 2021. Accessed December 2, 2024.
https://www.sciencedirect.com/science/article/abs/pii/S2352550921003493

Bozza, A., Campi, C., Garelli, S., Ugazio, E., Battaglia, L., ‘’Current regulatory and market frameworks in green cosmetics: the role of certification. Sustain. Chem.Pharm. 30.(2022) Last modified Sept. 29. 2022. Accessed December 2, 2024.
Current regulatory and market frameworks in green cosmetics: The role of certification - ScienceDirect

Federal Trade Commission Website (2024). “Green guides’’. Accessed December 10th, 2024
https://www.ftc.gov/news-events/media-resources/truth-advertising/green-guides

Goncalves J., Santos A., Kieling A., Tezza R ‘’The Influence of environmental engagement in the decision to purchase sustainable cosmetics: An analysis using the theory of planned behaviour’’ Scielo Brasil (2022) Accessed November 16, 2024.
https://www.scielo.br/j/reaufsm/a/ndNdhp8WVXqN8wnPh8LNcvn/

Nath, V.; Agrawal, R. 2023. ‘’Barriers to consumer adoption of sustainable products—An empirical analysis’’ Soc. Res, Journal 19(3) (2022) Accessed Dec 1, 2024.
Barriers to consumer adoption of sustainable products – an empirical analysis

Batista, L.; Larsen, A.T.; Davis-Peccoud, J.; Häuptli, M. ‘’Sustainability in Retail: PracticalWays to Make Progress’’ Bain & Company (2022) Accessed October 2024.
https://www.bain.com/insights/sustainability-in-retail/

The Politicization of Ultras in Soccer By Dylan Shah

Introduction

During the late 1900s, as soccer gained popularity worldwide, soccer fans in countries such as Turkey, Argentina, and Italy increasingly became more involved in the politics of those nations. The game quickly became more than just a mere pastime for the bourgeoisie and took a greater role in society. What factors contributed to the politicization of the ultras in soccer? Among the factors that contributed to the politicization of the ultras were the globalization of the game, violence between supporters and the police, and the threat of the government to lifestyles. In certain countries, the ultras ultimately became politically organized in opposition to the state and state policies, which caused a heightened battle for power between the state and the ultras.

Context

While soccer and politics have been connected since the early years of the sport, the globalization and commodification of soccer in the late 20th century both popularized the game and politicized it. For many, soccer was a *weapon of emancipation*, a way to express themselves, and also acted as a release valve for social and political tensions in their community.[67] In Argentina, the Argentine Football Association (AFA) , established in the late 19th century, had a presence and influence on the development of soccer, as relationships were strong between AFA leadership and the government. Around the same time, there was an expansion of urban populations and, therefore, the working class, against which the government used soccer to control this part of the population.[68] Violence was often an issue between ultras and the police, but also between the fans themselves, often due to political concerns. During 1995, a Genoa fan in Italy was stabbed to death by a Milan Ultra, likely because the Genoa fan was a communist militant.[69] These acts of brutality often led to ultras being used synonymously with hooligans, though hooliganism often comes with a much higher level of violence and scuffling, while ultras received their name for their passionate fandom.

During the 1990s, with the introduction of satellite televisions, soccer fans worldwide could view games from almost any location, and watch leagues they never could before. The globalization of soccer was also driven by increased media and press coverage of the sport. While this period helped further the development of the game, it also led to a greater distinction between the higher and middle-class fans, which marked the start of global politicization.[70] For example, in Turkey, after the AKP party won control of the country in the early 2000s and pushed their conservative values, ultras especially felt a threat to their freedom and began to use soccer as a way to express their concerns.

---

[67] Mickaël Correia, *A People's History of Soccer* (Las Vegas, NV: Pluto Press, 2023), 5.
[68] Dağhan Irak, "Istanbul United: Football Fans Entering the 'Political Field'," in *Everywhere Taksim: Sowing the Seeds for a New Turkey at Gezi*, by Isabel David and Kumru F. Toktamış (Amsterdam University Press, 2015), 137, JSTOR.

[69] Vic Duke and Liz Crolley, Football, Nationality and the State (London: Routledge, 2014), 117.
[70] Irak, "Istanbul United," 149.

Analysis

Globalization

The globalization of soccer gave ultras a voice that extended beyond their home country, giving attention to their opposition to the state. After Erdogan became Prime Minister of Turkey in 2003, he began to implement an authoritarian ideology, causing the Turkish ultras to protest. The Human Rights Watch noted that during Erdogan's rule, "the AKP has shown increasing intolerance of political opposition, public protest, and critical media."[71] The politicization of ultras commonly happens in places with authoritarian governments, such as the AKP, where no dissenting voices are allowed. In Turkey, the ultras have been part of the opposition to the government, and they are especially difficult to silence. As globalization had led to the spread of the game through critical media, more attention has been paid to leagues such as the Turkish Super League, bringing with it negative coverage of the AKP. The ultras contributed to this focus by protesting against the authoritarian state, which has, in turn, led the state to take violent measures to regain control.

Police Violence

In Italy, the 2001 protests confronting the G8 summit in Genoa were led by preexisting anti-globalization movements, and among the protesters were Italian ultras. James Montague's book *1312: Among the Ultras*, chronicles the experiences of ultras around the world, highlighting how a mix of globalization and violent response by police to fans also contributed to the politicization of ultras in soccer. These anti-globalization protests soon turned violent between protesters and police. A journalist for *La Gazzetta dello Sport* and former ultra, Pierluigi Spagnolo, said "You didn't [used to] go to the stadium to fight against police…It largely begins at Genoa with the G8 and you see the police show this ugly side of itself."[72] Spagnolo highlighted how the Italian ultras initially just went to the stadiums to support their team and enjoy the match. However, the police violence targeting the anti-globalization protests led ultras to turn political in stadiums to express their opposition to the unnecessary bloody violence that occurred.

These clashes led to larger struggles for power between the ultras and the police. Montague also highlights Rafa di Zeo, one of the leaders of La Doce, a group of Boca Juniors ultras, who was banned by the government for 6 years. Di Zeo mentioned how the government and police had attempted to control him and other members of the group, saying that the police "are afraid of our power and of our violence…The times when we had a fight and they won, it was because they fired at us you know? In this country to kill a cop is 'valuable.'"[73] The value in killing a cop for La Doce was not only so they could rid the cops of their weapons, but more importantly to show the cops employed by the state that the ultras would not back down, even if

---

[71] Zachary Andrews, "Inception, Legitimization, and Preservation of Competitive Authoritarian Regimes in Turkey," *The Commons: Puget Sound Journal of Politics* 2, no. 1 (2021): 15, JSTOR.

[72] James Montague, *1312 : among the Ultras : a Journey with the World's Most Extreme Fans* (London: Ebury Press, 2020), 69.

[73] Montague, *1312 : among*, 40.

they have less firearms and equipment. Because it was considered a prize to kill a cop, it highlighted the immense amount of hatred the two groups had for each other, since police were the visible enactment of the government's controlling policies.

Government Threat

The Gezi protests in Turkey in 2013, during which 48 people were killed, started as a protest against turning a park into an industrial shopping mall. However, it soon spread to a country-wide protest more broadly against the AKP, including concerns about freedom of press rights, as well as opposition to political hegemony. According to Selim, a Turkish ultra who was present at the protests, "The government saw there was a threat in the football stadium…They thought: We have to do something about this or we will face some open protests."[74] Especially after the Gezi protests led to the unification of rival fans, ultras became politicized and used chants and banners in soccer stadiums to send anti-AKP messages. Since the ultras were some of the most powerful voices in the Turkish community, the response from the AKP was to put restrictions on who could enter the stadiums by implementing policies such as an online ticketing system to crack down on fan behavior. However, the expression of political views by ultras in soccer stadiums during matches was powerful because that is when their political messages received the most attention from the state.

Conclusion

Three reasons that contributed to the politicization of ultras in soccer were the spread of the sport through globalization, violence between ultras and the police increasing the hatred between both parties, and the ultras' concern about governments attempting to use the sport to control lifestyles. Historically, ultras received their name from the Latin word ultrā, meaning "beyond". While ultras take their fandom "beyond" that of regular supporters, the influence of the ultras has clearly extended "beyond" the realm of soccer. Even in 2024, there are instances of ultras using their platform to promote political messages, such as Black Clad ultras at the 2024 Euros, who used their presence to promote their nationalistic ideology.

[74] Montague, *1312 : among*, 290.

Works Cited

Andrews, Zachary. "Inception, Legitimization, and Preservation of Competitive Authoritarian Regimes in Turkey." The Commons: Puget Sound Journal of Politics 2, no. 1 (2021). JSTOR.

Correia, Mickaël. A People's History of Soccer. Las Vegas, NV: Pluto Press, 2023.

Duke, Vic, and Liz Crolley. Football, Nationality, and the State. London: Routledge, 2014.

Irak, Dağhan. "Istanbul United: Football Fans Entering the ‘Political Field‘." In Everywhere Taksim: Sowing the Seeds for a New Turkey at Gezi, by Isabel David and Kumru F. Toktamış, 137-52. Amsterdam University Press, 2015. JSTOR.

Montague, James. 1312 : among the Ultras : a Journey with the World's Most Extreme Fans. London: Ebury Press, 2020.

Peer Influence on Spending Habits: The Role of Group Chats on Social Media Among Bangladeshi Teenagers By Zahia Zakaria

Background

Social media have rapidly reshaped the way adolescents interact with peers; this phenomenon influences nearly every aspect of teenagers' lives, including financial behavior. In Bangladesh, places such as Instagram and WhatsApp are important sites of interaction among teenagers engaging in peer conversations, especially in group chats through which trends, opinions, and recommendations are flowing nonstop. At this stage of development, it is very impressionable because of social acceptance and the need for teenagers to belong to a circle of friends. These subjects in discussions among peers will more often result in social pressure when talking about lifestyle, brands, and purchasing decisions that could lead to a change in spending habits for individuals. Having an easy avenue to share links of products, promotions, and new trends within the conversations of group chats amplifies these influences.

Although research into the issues of peer influence and social media has been conducted all over the world, very few portions of evidence point out how this dynamic plays within the Bangladeshi context. Cultural factors, socioeconomic differences, and the particular way Bangladeshi teens use social media set this apart from other regions. The nuance of peer influence via group chats is important to understand in countries like Bangladesh, where smartphone and social media use among teenagers surged in recent years.

The study shall investigate how spending habits are formed in the case of group interactions through Instagram and WhatsApp by detailing peer pressure, social comparison, and group dynamics. The study shall offer insight into Bangladeshi teenagers and their contribution to local and global discussions about the intersection of social media, peer influence, and adolescent financial behavior.

Objectives

The general objective is to determine or establish the influences on spending due to peers among teens. The aims are:

- The types of peer influences in social media group chats that might bring changes in spending habits among Bangladeshi adolescents.
- To establish the level at which peer pressure drives the teenagers' expenditure
- To identify The categories of products or services most influenced by peer recommendations within these chats.
- Assessing what contribution the parameter of social status, and especially that of group dynamic mechanisms, does towards shaping spending habits in a Bangladeshi teenager:.
- To assess the effect of peer influence spending on expenditure of other items/ family spending To find out its ultimate effect on society.

Literature Review

Peer influence plays a critical role in shaping teenagers' spending habits, with this being particularly exacerbated in the modern digital age. Developmentally, adolescents are at such a stage of life where they need to be socially accepted and peer-validated; therefore, they tend to easily succumb to the influence from their peers (Brown & Larson, 2009).

The same studies showed that adolescents often imitate the purchasing behavior of their peers, especially recommendations from close friends or members of social groups they seek to belong to (Valkenburg, Peter, & Schouten, 2006).

The scope and depth of peer influence have been amplified by social media. Today, social interaction centers on WhatsApp, Instagram, and small group chats, where teenagers/young consumers share product suggestions, discuss what is trending, or share their experiences of making purchases. Social comparison: Social media also tends to increase the amount of social comparison experienced by teenagers; teens have a growing sense of pressure to keep pace with the spending of others to maintain their social status.

Moreover, research has shown that peer influence within online groups not only dictates immediate purchase decisions but also alters general spending habits, often in ways that would not have been considered by teenagers themselves. These trends, becoming increasingly pronounced in developing economies like Bangladesh with increasing smartphone and internet penetration, have limited research regarding this particular context.

The contributions of this work complement the existing literature in view, for example, of the role specific to group chats via WhatsApp and Instagram applications among Bangladeshi adolescents within the setting of shaping their habits of expenditure onto which a number of locally contextualizing socio-cultural factors of peer influencing are likely to be overlaid.

Methodology

Based on both primary and secondary data sources, Approach: Mixed method approach qualitative and quantitative. Qualitative: KII, key informant interview: Interview with two parents of teenagers alongside the owner of one of the most popular online stores in Bangladesh. Quantitative: based on the survey of an individual. Sample size- 60 google survey

Survey period

Analysis of data, coding of data using excel. A sample of 60 English medium students all of ages between 13-18 were collected. About 40% of them were within the 13-14 age group.

A majority of the respondents fall within the age bracket of 15-17 years, followed by 40% falling within the 13-14-year bracket. This age group is more prone to peer influence because of the nature of their growth stage, in which social acceptance becomes a central concern in their behavior and decision-making. The gender distribution indicates that 72.5% of respondents are male and 27.5% are female. This might indicate that male consumers would be more interested in gadgets and entertainment, while females would be more vulnerable to fashion and beauty. With respect to family income, it was found that 48% of the surveyed teenagers belonged to a family income group of less than 1 lakh, indicating that even in lower-income groups with budget constraint issues, peer pressure seems to act as an influencing force in expenditure.

Peer Pressure and Spending Habits : In the questions of discussions about trends, asking for recommendations, and purchasing similar products, the data indicates a highly positive relationship exists between peer pressure and spending: Discussion of Trends: 62.8% discuss trends either frequently or sometimes, meaning social interaction is a huge role in how they form an opinion about the product and the trend. This suggests that peer influence is an intrinsic part of the decision-making initiatives, as many as 57.1% of the respondents seek recommendations from friends before making any purchase. Buying Similar Products: A full 66.7% of respondents said they purchase the same products as their friends, a surefire sign that peer pressure is at play in purchasing decisions. These figures show that peer pressure acts very much through either social interaction or recommendations in relation to how much teenagers spend on something, many preferring to keep up with the group even though they probably can't afford it.

Apart from the survey, this section shows how peer pressure factors into a teenager's spending habit. More context is provided by insights from the owner of a trendy jewelry shop known as Allthatsheknowsbd with ver 36K followers targeted at teenagers; the owner cited that teenagers like to fashion or wear trending jewelry or popular branded clothes, especially if the jewelry is perceived to be highly fashionably accepted in society among teenagers..

Most teenagers are usually found buying jewels that match with those of their friends, a result driven by the need for acceptance and social conformity. "Marketing campaigns influence such decisions since teenagers are easily influenced to buy well-marketed products through social media and influencers, thereby further fuelling the consumption choice because of peers, the owner added. This finding confirms the survey results that showed teenagers are very much influenced by their friends' views when it comes to buying preferences, especially for fashion and accessory categories. To this respect, two parents of 16 year old girls have been interviewed in the frame of above mentioned research, with both parents' being neighbors, both parents emphasized that peer influence considerably affects the spending habits of their daughters and most times unconsciously too. They related how their daughters influence one another through their exposure to new brands and trends, through everyday conversations like about makeup. This then intensifies through social activities where they said there was a recent wedding in their family for which the girls all went out to go jewelry shopping with one another and the end selection was made with the girl friends' likings in mind. Further, the parents described that the girls have the habit of going window shopping together to amuse themselves, supplementing peer influence on consumer trends. This points out how social interaction among friends influences both choices and spending behaviors.

Conclusion

This study has identified and determined the level of the role that peer influence plays in shaping spending habits among Bangladeshi teenagers, particularly through social media group chats. These results point to the fact that with proper peer pressure, teenagers are motivated to buy things which perhaps otherwise would not have been purchased by them according to the group trend. These behaviors are further compounded by social status considerations, in which teenagers believe that purchasing items that are trending or popular helps them to retain their social standing.

The findings postulate that the Bangladeshi context, with rapid technological adoption and cultural importance of brand recognition, provides fertile ground for peer influence to thrive. With social media continuing to evolve, grasping these dynamics will be essential for parents, educators, and policymakers in their quest for ways through which the effects of peer-driven consumerism among teenagers are mitigated.

Understanding the Neuroscience of Play - the Impact of Play on Brain Development
By Anahita Gupta

Abstract

Play is a fundamental component of children's development, serving as the foundation for emotional, social, and cognitive growth. Beyond its recreational value, neuroscience research underscores play's critical role in brain development. This study investigates the brain regions associated with social cognition, emotional regulation, and executive function, emphasizing how play shapes neural circuits and enhances neuroplasticity. It examines the influence of physical, imaginative, and social play on the developing brain, highlighting their contributions to essential skills like learning, creativity, and problem-solving. By exploring the neuroscience of play, this research provides insights for policymakers, educators, and caregivers, advocating for the creation of environments that foster meaningful play experiences to support optimal brain development in children.

Keywords

Neuroscience, Play, Brain development, Neuroplasticity, Cognitive development

Introduction

As an activity that promotes happiness, exploration, and socialization, play has long been seen as fundamental to children. New insights from neuroscience, however, have shed light on play's more fundamental role in brain development. The ability to freely explore one's environment, solve problems, and create emotional resilience are all skills that children naturally acquire via play (Pellis, & Pellis, 2009). Not only can play improve one's mood and sense of well-being in the short term, but it also helps build neural networks that are essential for later success in school and in life (Panksepp, 2007). By evaluating the effects of various types of play on neuroplasticity and important brain areas, this research hopes to close the gap between conventional wisdom and the scientific evidence supporting play's significance in brain development. The significance of play in infancy and beyond may be better grasped with this knowledge.

The science of early childhood development

Investments in early childhood may be informed by the science of early brain development. Decades of neuroscience and behavioral study have established these fundamental ideas, which assist to show how the development of children, especially from birth to the age of five, lays the groundwork for a sustainable and affluent society (Grindal, et al., 2012).
The human brain develops gradually, piece by piece. The process of building the brain's fundamental architecture starts long before birth and continues into maturity. The strength of that framework is determined by the quality of the early experiences that provide the groundwork for subsequent learning, health, and behavior. According to Phillips and Shonkoff (2000), there is a formation of over 1 million new synaptic connections each second throughout the first few years of life. In order to make brain circuits more efficient, connections are decreased via pruning after

this phase of fast proliferation. Development of primary sensory pathways, such as those for vision and hearing, occurs before the emergence of language and later higher-level cognitive abilities. Brain circuits grow and shrink in a specific sequence, with more sophisticated networks forming from more basic ones (Black, et al., 2017).

Figure 1: Human brain development

The brain develops because of the combined effects of heredity and environmental factors. Researchers have discovered that the "serve and return" bond between children and their parents or other community carers plays a significant role in this developmental process. Using chattering, facial expressions, and gestures, young toddlers instinctively seek out engagement, and adults mimic these behaviors in response. Inconsistent or incorrect answers, or a lack of them altogether, might cause the brain to develop differently than anticipated, which in turn can cause differences in how people learn and act (Black, et al., 2017).

As we get older, our brain's plasticity diminishes. Early in infancy, the brain is more malleable, or "plastic," and can adapt to a variety of situations and interactions; but, as it develops and takes on more complicated tasks, the brain becomes less malleable and adaptable. For instance, during the first year of life, a baby's brain begins to develop a preference for the sounds of the language he or she has been exposed to, while simultaneously beginning to lose the capacity to distinguish between sounds in other languages. These neural circuits are harder and harder to change as we age, even as the "windows" for learning a new language and other abilities stay open. Influencing a baby's growing brain architecture is simpler and more successful than rewiring sections of its circuitry in adulthood due to early plasticity (Shonkoff, et al., 2012).

Over the course of a person's lifetime, their cognitive, emotional, and social skills are intricately linked. The brain is an intricate network of interconnected parts that work together in perfect harmony to carry out all its many tasks. The building blocks of human growth are emotional health and social competence, which, when combined, provide a solid base for developing cognitive capacities. Prerequisites for academic and occupational success as well as community involvement and leadership in later life include mental and physical well-being, social competence, and cognitive-linguistic abilities that develop throughout childhood (Campbell, & Howitt, 2024).

Damage to the developing brain's architecture from toxic stress may cause behavioral and cognitive issues as well as physical and mental health issues later in life. New research shows that the growing brain is particularly vulnerable to the harmful effects of chronic, unwavering stress in early infancy, which may result from factors like as severe mother depression, prolonged maltreatment, or extreme poverty. An integral part of normal growth is the body's ability to deal with mild, transient physiological reactions to stressful situations; toxic stress, on the other hand, is the strong, unrelieved activation of this system. Processes that form the structure of the growing brain incorporate toxic stress into the body when the buffering protection of adult support is lacking (Britto, et al., 2017).

Neuroscience and learning through play

The neuroscience literature in brief

Research on humans with normal and atypical development, as well as animal models, mostly contribute our present knowledge of how each aspect of playful encounters might facilitate learning processes. Although animal models may provide light on potential human brain systems, it's important to remember that there are significant differences between the two. Research on human cortical networks in adults also sheds light on this fascinating area of the brain, however in a more developed and less easily affected form than in young children. In light of this, we do a literature review while, for the most part, allowing room to speculate about how each trait can influence children's learning. Furthermore, whereas these studies demonstrate the potential contributions of the five traits to learning, very few examine the causal link between play and education. There is still need for more investigation in this area as well.

Interconnected and holistic learning

Before delving into the five traits of enjoyable experiences, it's helpful to put them in the broader perspective of how the brain develops. What we do know is that the brain's regions do not develop and govern learning alone, but rather that they are strongly interconnected and rely on continuous and specialized input from outside sources. The results show how the five features of play-based learning help children's developing brains activate and build interrelated systems that support their learning.

Our perspective on learning via experiences is comprehensive, which means that it pertains to the growth of several areas instead of a predetermined set of academic metrics. When we talk about learning in the brain, we're referring to the neuronal ability to receive and process information from a variety of sources, or multimodal inputs. Proper development of learning behaviors in children typically requires inputs across various modalities. For instance, a baby may improve her visual acuity, facial recognition, phoneme recognition, and secure attachment skills via face-to-face contact with a caregiver, which also gives her auditory and social-emotional inputs (Fox, et al., 2010). As a result, these outcomes will help her develop abilities in language, cognitive control, and emotion management as she grows older. Multimodal inputs that promote linked learning networks may be provided by engaging in playful learning experiences that are marked by pleasure, meaning, active participation, iteration, and social interaction. Thus, our experiences shape our development from a young age depending on their quality. We use this context to examine the connections between the various traits and these mental operations in our review (Duckworth, 2016).

Figure 2: Medial view of the brain and the areas related to the five characteristics
Effects of play on the learning brain

Possibilities for play promote brain development and enhancement of memory.

Marion Diamond and colleagues released a groundbreaking study on rat brain development in 1964. In a groundbreaking experiment, the neuroscientists housed some rats in dull, solitary captivity while others were raised in lively, toy-filled colonies. Brain scans revealed that compared to their "impoverished" counterparts, the "enriched" rats had much thicker

cerebral cortices. Rats reared in stimulating circumstances have larger brains, as later studies corroborated.

A chemical necessary for the proliferation of brain cells, BDNF is secreted in response to play and exploration.

Once again, because no one has yet found out how to test this on people in an ethical manner, the proof is derived from rats: Rats' brain levels of brain-derived neurotrophic factor (BDNF) rise during episodes of rough-and-tumble play (Gordon et al., 2003). When it comes to the development and upkeep of brain cells, BDNF is king. According to Huber et al. (2007), rats who are given the freedom to explore also show an increase in BDNF levels.

Children are better able to focus on schoolwork when they are given frequent, short breaks to play.

Children benefit academically from playtime, an unstructured interval when they are allowed to play without adult supervision, according to numerous experimental research. The following evidence is also circumstantial: Short breaks are provided every half an hour in the classrooms of the world's top students, the Chinese and the Japanese (Stevenson and Lee 1990). Keep in mind that PE lessons can't hold a candle to unstructured playing (Bjorkland and Pellegrini 2000). In and of itself, physical exercise has substantial positive effects on cognitive function. Contrary to popular belief, playtime is much more beneficial than physical education sessions. That, according to the researchers, is due to the fact that physical education courses depend too much on regulations set by adults. A play break can only be beneficial if it is really fun. How many minutes should playtime last? While the exact duration of the most beneficial recesses is unknown, it seems that 10-to 30-minute breaks work best. Ten-to-twenty-minute breaks improved classroom attention in one little study of children aged four to five. According to Pellizgrini and Holmes (2006), the inverse was true during recesses that were up to 30 minutes long.

Communication and the positive effects of play

Research has shown that play, and more specifically symbolic, make-believe play, may aid in the acquisition of linguistic abilities. As an illustration: Edward Fisher, a psychologist, reviewed 46 research on the positive effects of play on cognition (Fisher 1999). "Results in improved performances in both cognitive-linguistic and social affective domains" was his conclusion after seeing "sociodramatic play," or when children act out imaginary scenarios in a group. The ability of British children to engage in symbolic play was assessed in research including children aged 1 to 6 years old (Lewis et al 2000). Symbolic activities, such putting a teddy bear in place of something that was missing, were given to the children. Children with higher symbolic play scores also exhibited greater receptive and expressive language abilities, according to the study's authors. These findings persisted even after accounting for the child's age. Additionally, new studies show that building with blocks helps kids learn new words. This article discusses the advantages of play and building toys in more detail.

Creative problem-solving is fostered through play

Convergent and divergent problems are defined differently by psychologists. One and only one answer will do for a convergent issue. There are many ways to fix a diverging situation. The way children play may help them learn to solve issues with different solutions, according to several studies. As an example, Pepler and Ross (1981) conducted an experiment in which they gave toddlers two different kinds of playthings. A few children were provided with tools for convergent play, such as pieces of a puzzle. We provided other children with blocks and other tools for creative play. The children were allowed free time to play before being evaluated on their problem-solving skills. When tested on divergent tasks, children whose play items encouraged diverse thinking did better. In addition, they were more imaginative in their approaches to finding solutions (Pepler and Ross 1981). Another research found that children who played make-believe were better able to solve issues in a variety of ways (Wyver and Spence 1999). The capacity to address divergent situations was enhanced in children who received instruction in role-playing. Additionally, children who were taught to handle divergent issues also exhibited higher rates of make-believe play.

Improving mathematical abilities via play

This is a fascinating tale: Wolfgang, Stannard, and Jones (2001) conducted a longitudinal research that followed students from the time they were four years old all the way through high school, measuring the intricacy of their block play and their academic achievement. High school arithmetic performance was shown to be predictive of how complicated a child's block play was. Those whose preschool block play was more complex performed better academically and enrolled in more math courses (including honours) when they were in high school. I know what you're thinking: these findings only show that students who do well in preschool also do well in high school. Oh, it's far from easy. After researchers accounted for a child's IQ, the correlation between block play and arithmetic proficiency persisted. Thus, it is reasonable to assume that the act of playing with blocks had an effect on the children's cognitive growth.

IN WHAT WAYS CAN WE MAP IMPORTANT BRAIN NETWORKS FOR PLAY?

There are a number of ways to go about trying to design the brain circuits that regulate the rat's play behaviour. What does play look like when a certain neural system or brain region isn't there? That's one example. This method has served as the foundation for behavioral neuroscience studies for a long time. Depending on whether it's aiming at neuroanatomical or neurochemical substrates, a system may be up- or down-regulated. By injecting substances that temporarily suppress activity in that section of the brain, it is possible to permanently damage or incapacitate that area. Looking into a rat's brain while it's playing is another way to address the problem. Is it true, therefore, that different brain pathways are activated while we play? By combining metabolic markers like c-fos with well-defined and isolated behaviors, an incredibly precise metabolic map of the brain during that activity may be generated (Dragunow & Faull, 1989). Data from each of these methods should, ideally, corroborate one another's findings about the function of the studied region or system.

Cortical modulation of play

There has been a conceptual and historical bias towards first isolating cortical and subcortical processes as potential brain mechanisms involved for regulating play in rats. Paul MacLean's theory of the triune brain, which places a premium on subcortical limbic regions as essential for a wide variety of behaviors shared by mammals, has had a major impact on this. It seems to reason that the neurological bases of play would mostly be located in rather ancient subcortical circuits that all mammals share, given that play emerged as an adaptive behavioral approach relatively early in mammalian development. When neuroscientific investigations on playfulness started in the late 1970s and early 1980s, this was indeed one of the first ideas that was explored.

Decorticate rats showed a higher propensity to react to nape touches with partial rotations when tested both before and after puberty. This finding suggests that juvenile decorticate play was more similar to adult rat play. Furthermore, these rats did not alter their behaviour in response to the partner's status. Subsequent research by Pellis's group indicated that several regions of the brain seem to regulate these distinct features of play. As an example, according to Kamitakahara et al. (2007), male rats that have motor cortex lesions do not exhibit age-related changes in play strategies, such as responding with partial rotations before and after puberty. In contrast, Pellis et al. (2006) found that rats that had injury to the orbitofrontal cortex were unable to adapt their behaviour depending on their partner's condition. Damage to the medial prefrontal cortex in rats makes them respond more naively when provoked, for as by running away instead of engaging in complicated play (Bell et al., 2009). Lesions produced after puberty had similar effects on adult play, making it more probable that the removal of these cortical regions, rather than having a negative impact on neurodevelopmental processes, is responsible for these results.

Brain development as 21st-century skills

21st Century Skills, Play and Brain Development

Nowadays, it's quite simple and almost limitless to get information. In our modern day, having knowledge is not sufficient on its own. Knowledge should be taught last in schools, claims Harari (2018). This is affecting people's expectations of schooling. Just imparting knowledge to students is no longer sufficient for schools. Children need to be taught to think critically, solve problems, communicate effectively, work together, be creative, and be adaptable in order to succeed in the modern world (National Research Council, 2012). Additionally, kids need to know how to put knowledge to use, sort it into relevant categories, and draw connections to real-life situations (Harari, 2018). People in the modern day should strive to improve their skill sets and knowledge bases. Brain growth certainly plays a role in the acquisition of these abilities. One of the greatest methods to encourage brain growth and abilities appropriate to the modern world is to start early with education. During this time, a person's brain develops, which is fundamental for learning. The brain is responsible for a wide variety of functions, including interpreting sensory data, controlling and regulating mental operations like thinking, remembering, and decision-making, and processing information from many sources. Kids need their brains to grow and develop properly so they can learn everything (Schlegel, 2011).

Brain and Motor Development

In addition to developing motor skills and physical exercise, children's unstructured play includes pretend play, building with blocks, imitating grown-ups, and finding creative uses for everyday things. According to Burdette and Whitaker (2005), unstructured play may have a greater impact on brain development than structured physical exercise. Burdette and Whitaker (2005) discovered that free play improved social abilities including attentiveness, closeness, and emotional affect, in addition to the cognitive advantages previously discovered. Motor skill development paves the way for youngsters to engage in physical interactions with others. Playing games with friends is a great way for kids to practise teamwork and communication. Team sports promote teamwork, communication, and collaboration.

Brain and Language Development

Language develops in young children via their play (Hirsh-Pasek et al., 2009). New research shows that methods that focus on the whole kid—teaching the child how to learn rather than just what they need to know—are more effective in fostering language development. Young children's language development is aided by play, particularly when it is free and unstructured. A baby's IQ, academic success, and social skills are most strongly predicted by the amount of words they hear daily. That is why plays are beneficial for the development of both language and the brain. Modern competencies, including the ability to communicate and work together, need proficiency in more than one language. The ability to articulate one's thoughts and feelings clearly and communicate with others is a talent that is essential in modern society. Children who are too young to play outside safely and who are glued to their screens all day long as a result of this century's fast technological advancements may hone their motor skills—and one of the 21st-century abilities is social development—through play.

Brain and Social Development

A youngster needs to engage in social interactions and play with others for optimal development. The significance of interacting with older children and adults was highlighted by Vygotsky (1966/1976) prior to the availability of high-tech neuroimaging investigations. Their interactions, he argues, give rise to play and, by extension, higher cognitive capabilities. Adult adjustment is better predicted by a child's social abilities rather than by intelligence, academic performance, or classroom conduct (Hartup, 1992). Children learn to interact with others, understand their environment (see to social development for more on this), and adjust to new circumstances all via play. The youngster gains competence in sharing, working together, and solving problems via play. The child's play aids in this process since it helps them navigate their social and environmental environments, which are becoming more complicated. Helping children develop their social skills and brains, play is an important part of childhood. Both the growth of one's intellect and one's social abilities have an effect on one another.

Brain and Emotional Development

For healthy growth and development, kids need to hang out with their friends and play with other kids. Until sophisticated neuroimaging techniques were available, Vygotsky

(1966/1976) emphasised the importance of children's and adults' interactions. He contends that as a result of these interactions, play and, subsequently, enhanced cognitive capacities, emerge. Rather than IQ, grades, or behaviour in class, adult adjustment is best predicted by a child's social skills (Hartup, 1992). Through play, kids learn to communicate with others, explore their surroundings (for more on this, see social development), and adapt to different situations. As they play, kids learn important life skills like sharing, cooperating, and problem-solving. Because it helps them make sense of their increasingly complex social and environmental surroundings, children's play is an important part of this process. Play is a crucial component of childhood since it aids in the development of children's social skills and brains. Improving one's social skills and one's intelligence both impact the other.

Brain Development and Video Games

The widespread use of video games has sparked a renewed focus on the potential negative effects on children's growth and development, on human cognition and behaviour, and on society at large. Concerned and interested parties include those in the fields of psychology, education, politics, parenting, medicine, and everyone else who want to understand its impact. Despite video games' increasing popularity, very little research has looked at how they affect brain function biologically.

Haier et al. (1992) used positron emission tomography (PET) to monitor changes in regional glucose metabolism in individuals who had just mastered a comprehensive visuospatial/motor task. He reasoned that while playing video games decreased glucose metabolism in the brain generally, it was likely because the participants' brain circuits, or neurons per circuit, were not used as much when the work was over. Koepp et al. (1998) carried out an extra PET investigation. Both studies found that better performance on the pitch was associated with either reduced glucose metabolism or greater dopamine release. Using conventional cerebral functional imaging has its advantages for evaluating the spatial resolution of deep brain activity and cognitive processes. Patients undergoing paediatric investigations must be highly restrained throughout measurements. On the other hand, Feng et al. (2007) found that kids who play video games could have better brain rotation skills. Regular gamers of action video games had a cognitive advantage of around half a standard deviation, according to a meta-analysis of data spanning the years 2000–2015. Action games, on the other hand, seem to improve top-down attention and spatial cognition more than other genres of video games (Bediou et al., 2018). Video game play impacts brain networks while engaged in the 'Flow' experience, as indicated in the systematic study by Palaus et al. (2017). This experience is crucial for skill-difficulty balance and includes the entire activation of somatosensory networks via the reward system. But there are certain negative impacts of playing video games, including on brain growth, attention, and cognitive abilities. The research highlights the need to explore the neurological bases of video game impacts, however it doesn't do so for the neural correlates of these effects but rather for their effects on cognition and behaviour. The primary areas that benefit most from playing video games are motor skills and concentration.

Parents and teachers often express concern that kids' excessive screen time poses health risks to their brain development, especially when it comes to video games. But the data to back

up or disprove this worry is lacking. Many worry that children's brain development could be severely impacted by excessive screen time, a lack of physical activity, and a preference for gaming over other activities like playing with friends or participating in sports. As we've shown, video games can help with brain growth, but they only do so in little ways. Other factors, including physical exercise, social connections, and interactions with peers, are far more important. A child's brain growth will be impeded if they are lacking in these areas.

Conclusion

Play is more than just a hobby; research in the neuroscience of play shows that it's crucial for a child's healthy brain development. Executive function, emotional control, and social skills are foundational for future success in life, and children build these abilities via play. Because of the strong correlation between play experiences and improved neuroplasticity and cognitive flexibility, this study highlights the significance of designing spaces that encourage diverse types of play. The significance of making play a top priority in educational and developmental programs is growing as our knowledge of the link between play and brain development expands. People who are encouraged to play while studying are more likely to be creative, resilient, and adaptable, which is good for society overall.

Works Cited

1. Bediou, B., Adams, D. M., Mayer, R. E., Tipton, E., Green, C. S., & Bavelier, D. (2018). Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills. Psychological Bulletin, 144(1), 77–110.
2. Bell, H. C., McCaffrey, D. R., Forgie, M. L., Kolb, B., & Pellis, S. M. (2009). The role of the medial prefrontal cortex in the play fighting of rats. Behavioral neuroscience, 123(6), 1158.
3. Bjorklund, D. F., & Pellegrini, A. D. (2000). Child development and evolutionary psychology. Child development, 71(6), 1687-1708.
4. Black, M. M., Walker, S. P., Fernald, L. C., Andersen, C. T., DiGirolamo, A. M., Lu, C., ... & Grantham-McGregor, S. (2017). Early childhood development coming of age: science through the life course. The Lancet, 389(10064), 77-90.
5. Black, M. M., Walker, S. P., Fernald, L. C., Andersen, C. T., DiGirolamo, A. M., Lu, C., ... & Lancet Early Childhood Development Series Steering Committee. (2017). Advancing Early Childhood Development: From Science to Scale 1: Early childhood development coming of age: Science through the life course. Lancet (London, England), 389(10064), 77.
6. Britto, P. R., Lye, S. J., Proulx, K., Yousafzai, A. K., Matthews, S. G., Vaivada, T., ... & Bhutta, Z. A. (2017). Nurturing care: promoting early childhood development. The lancet, 389(10064), 91-102.
7. Burdette, H. L., & Whitaker, R. C. (2005). Resurrecting free play in young children: looking beyond fitness and fatness to attention, affiliation, and affect. Archives of pediatrics & adolescent medicine, 159(1), 46-50.
8. Campbell, C., & Howitt, C. (Eds.). (2024). Science in early childhood. Cambridge University Press.
9. Diamond, A. (2013). Executive functions. Annual review of psychology, 64(1), 135-168.
10. Dragunow, M., & Faull, R. (1989). The use of c-fos as a metabolic marker in neuronal pathway tracing. Journal of neuroscience methods, 29(3), 261-265.
11. Duckworth, A. L. (2016). Grit: The power of passion and perseverance. Simon &Schuster.
12. Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. Psychological Science, 18, 850–855.
13. Fisher, E. P. (1992). The impact of play on development: a meta-analysis. Play & Culture.
14. Fox, S. E., Levitt, P., & Nelson III, C. A. (2010). How the timing and quality of early experiences influence the development of brain architecture. Child development, 81(1), 28-40.
15. Gordon, N. S., Burke, S., Akil, H., Watson, S. J., & Panksepp, J. (2003). Socially-induced brain 'fertilization': play promotes brain derived neurotrophic factor transcription in the amygdala and dorsolateral frontal cortex in juvenile rats. Neuroscience letters, 341(1), 17-20.
16. Grindal, T. A., Hinton, C., & Shonkoff, J. P. (2012). The science of early childhood development. Defending childhood: Keeping the promise of early education, 13.

17. Haier R. J., Siegel, B. V., MacLachlan, A., Soderling, E., Lottenberg, S., & Buchsbaum, M. S. (1992). Regional glucose metabolic changes after learning a complex visuospatial/motor task: a positron emission tomographic study. Brain Research, 570, 134–43.
18. Harari, Y. N. (2018). 21 Lessons for the 21st Century:'Truly mind-expanding... Ultra-topical'Guardian. Random House.
19. Hartup, W. W. (1992). Having friends, making friends, and keeping friends: relationships as educational contexts (ED345854). ERIC.
20. Hirsh-Pasek, K., & Golinkoff, R. M. (2003). Einstein never used flash cards. Rodale.
21. Huber, R., Tononi, G., & Cirelli, C. (2007). Exploratory behavior, cortical BDNF expression, and sleep homeostasis. Sleep, 30(2), 129-139.
22. Kamitakahara, H., Monfils, M. H., Forgie, M. L., Kolb, B., & Pellis, S. M. (2007). The modulation of play fighting in rats: Role of the motor cortex. Behavioral Neuroscience, 121(1), 164.
23. Koepp, M. J., Gunn, R. N., Lawrence, A. D., Cunningham, V. J., Dagher, A., Jones, T., Brooks, D. J., Bench, C. J., & Grasby, P. M. (1998). Evidence for striatal dopamine release during a video game. Nature, 393, 266– 268.
24. Lewis, Jill Boucher, Laura Lupton, Samantha Watson, V. (2000). Relationships between symbolic play, functional play, verbal and non-verbal ability in young children. International Journal of Language & Communication Disorders, 35(1), 117-127.
25. National Research Council. (2012). Education for life and work: Developing transferable knowledge and skills in the 21st century. National Academies Press.
26. Palaus, M., Marron, E. M., Viejo-Sobera, R., & Redolar-Ripoll, D. (2017). Neural basis of video gaming: A systematic review. Frontiers in Human Neuroscience, 11, 248.
27. Panksepp, J. (2007). **Play, ADHD, and the construction of the social brain: Should the first class of psychomotor stimulants be entraining endogenous play systems?** Journal of Neuropsychoanalysis, 9(2), 73-83.
28. Pellis, S. M., Hastings, E., Shimizu, T., Kamitakahara, H., Komorowska, J., Forgie, M. L., & Kolb, B. (2006). The effects of orbital frontal cortex damage on the modulation of defensive responses by rats in playful and nonplayful social contexts. Behavioral neuroscience, 120(1), 72.
29. Pellis, S., & Pellis, V. (2013). The playful brain: Venturing to the limits of neuroscience. Simon and Schuster.
30. Pepler, D. J., & Ross, H. S. (1981). The effects of play on convergent and divergent problem solving. Child Development, 1202-1210.
31. Phillips, D. A., & Shonkoff, J. P. (Eds.). (2000). From neurons to neighborhoods: The science of early childhood development.
32. Pinker, S. (2007). The mystery of consciousness. Time, 169(5), 58-70
33. Schlegel, S. (2011). Mind in the making: the seven essential life skills every child needs. Journal of Developmental & Behavioral Pediatrics, 32(5), 401.

34. Shonkoff, J. P., Richter, L., Van Der Gaag, J., & Bhutta, Z. A. (2012). An integrated scientific framework for child survival and early childhood development. Pediatrics, 129(2), e460-e472.
35. Singer, D. G., Golinkoff, R. M., & Hirsh-Pasek, K. (Eds.). (2006). Play= Learning: How play motivates and enhances children's cognitive and social-emotional growth. Oxford University Press.
36. Singer, D. G., Golinkoff, R. M., & Hirsh-Pasek, K. (Eds.). (2006). Play= Learning: How play motivates and enhances children's cognitive and social-emotional growth. Oxford University Press.
37. Wolfgang, C. H., Stannard, L. L., & Jones, I. (2001). Block play performance among preschoolers as a predictor of later school achievement in mathematics. Journal of Research in Childhood Education, 15(2), 173-180.
38. Wyver, S. R., & Spence, S. H. (1999). Play and divergent problem solving: Evidence supporting a reciprocal relationship. Early Education and Development, 10(4), 419-444.

Neurodiverse Activity Recognition Using Wearable Technology and Machine Learning
By Aashna Parsa

Abstract

Human Activity Recognition (HAR) significantly enhances quality of life by facilitating the monitoring and understanding of diverse activities, particularly for neurodiverse individuals. This study focuses on utilizing machine learning algorithms to predict learning activities through smartwatch sensor data collected from eight participants. The collected data was preprocessed to remove noise or irrelevant information, enabling effective model training using algorithms such as K-Nearest Neighbors, Random Forest, and Multilayer Perceptron. Random Forest was identified as the most accurate classifier based on cross-validation. The model's performance was validated using precision, recall, F1 score, confusion matrix and accuracy. The results reveal strong predictive capability, highlighting the feasibility of leveraging wearable technology for unobtrusive activity monitoring. These findings provide insights for improving personalized interventions for neurodiverse populations. Future research includes expanding dataset diversity, exploring real-time implementation, and addressing ethical considerations in behavior monitoring.

Key terms

Human Activity Recognition, K-Nearest Neighbor, Random Forest, Multilayer Perceptron, cross-validation, F1 scare, neurodiverse

Introduction

Human Activity Recognition (HAR) is rapidly becoming a vital area of research, offering significant benefits to various populations. HAR is utilized in fields such as healthcare for patient monitoring, sports for evaluating athletes, and senior care for detecting health anomalies. In recent years, the intersection of wearable technology and machine learning has emerged as a fertile ground for advancing personalized healthcare and activity recognition systems. This paper explores the application of these cutting-edge technologies in the context of neurodiversity—a domain often characterized by varying behavioral patterns and unique needs. Individuals on the neurodiversity spectrum, including those with autism, ADHD, and other cognitive differences, often experience a diverse range of daily activities and interactions that challenge traditional activity recognition frameworks. The objective of this study is to develop a robust approach that can accurately recognize and adapt to the unique activity patterns of neurodiverse individuals, thereby enhancing quality of life and fostering greater autonomy.

Through the deployment of state-of-the-art wearable devices capable of continuous physiological and behavioral data collection, combined with machine learning algorithms, this study presents a framework that addresses the complex dynamics of neurodiverse activity recognition.This project aims to identify the learning activities of neurodiverse individuals using sensor data from eight participants, ultimately assisting those with Attention Deficit Disorders (ADDs) in maintaining focus.

The project is driven by the need to understand the critical tools and techniques involved in HAR, including data collection, activity recognition, and the analysis of complex, real-world activities. Historically, HAR relied on specific sensor data metrics, like accelerometer readings, and simple rules-based methods, which limited the ability to identify complex movements and patterns. Recently, machine learning has provided a new avenue for overcoming these limitations by leveraging a broader range of sensor data.

In our study, we leveraged machine learning solutions to analyze and interpret these intricate patterns. The data was collected from smartwatch built-in sensors, chosen for their ability to unobtrusively and accurately capture participants' movements and physiological changes.

Materials and methods

This section outlines the processes used for data collection, preprocessing, and machine learning model evaluation. This includes a description of the wearable sensor setup, data cleaning steps, and the implementation of cross-validation to ensure robust model training and testing. The section also highlights the machine learning algorithms employed and the rationale for selecting the most effective model.

Dataset Description

The WLA4ND dataset comprises data gathered using wearable sensor technology, capturing learning activities from a cohort of eight neurodiverse participants. This group included three females and five males, all of whom have intellectual and developmental disabilities. Participants offered self-reports on their types of disabilities, revealing a demographic composition of seven White-Caucasian individuals and one African-American. Notably, all participants indicated right-hand dominance, with an average age of 28.5 years, spanning an age range from 23 to 31 years. However, patient nine had only three activities in their data set, so keep that in mind when looking at the patient 9 results. Additionally, patient 4 was omitted from the original data set “due to an incomplete reading task” (Zheng et al. 4)

Participant	Age	Gender	Neurodiversity
1	24	M	Autism Spectrum Disorder
2	27	M	Down Syndrome
3	23	M	ADHD
5	29	M	Intellectual and Developmental Disability
6	31	F	ADHD
7	32	F	Down Syndrome
8	31	M	Down Syndrome
9	31	F	Intellectual and Developmental Disability

Table 1: Demographic profile and Neurodiversity of participants (from Zheng et al. 7)

Data collection focused on common learning activities performed by these individuals, utilizing smartwatch sensors to capture tasks such as reading, writing, typing, answering follow-up questions, and engaging in off-task behaviors. The dataset was enriched by five built-in sensors of the smartwatches, encompassing 13 distinct feature dimensions: 12 derived from four motion sensors and one from the heart rate monitor. Specifically, motion-based data included three-axis vectors (x, y, z) from the gyroscope (measuring raw rotations), accelerometer (capturing raw accelerations), linear acceleration (excluding gravitational forces), and gravity sensors (assessing acceleration vectors). Additionally, heart rate data was recorded in terms of beats per minute (bpm) using a singular vector. The dataset contains approximately 500 raw data points per second across the five sensor origins, offering a rich resource for detailed activity recognition analysis. The dataset is labeled with six activity categories: 'read' for reading, 'writeQ&A' for writing questions and answers, 'write' for general writing, 'type' for typing, 'rest' for taking a break, and 'off' for being off-task. These labels help in training models to recognize and analyze the different learning and non-learning activities exhibited by neurodiverse participants.

Methodology

The initial stage of dataset preprocessing involved addressing the presence of null values, which can impede the performance of machine learning models. These null values were systematically identified and removed. Once clean, the dataset was organized for model training and evaluation by splitting it into separate subsets: training and testing datasets. To evaluate the machine learning models robustly, cross-validation was employed as the method of choice. This approach involves dividing the dataset into several smaller subsets or folds. The model is iteratively trained on a combination of these subsets while validated on the remaining portions. Cross-validation offers a comprehensive evaluation by ensuring that every data point gets a chance to be in both training and validation sets, thereby providing a more accurate and reliable assessment of the model's performance. By implementing cross-validation, we aim to provide a robust metric of model performance.

The model was initially trained with data from the first patient using several algorithms. For K-Nearest Neighbors (KNN), the focus was on tuning the number of neighbors, which determines how many nearby data points contribute to the classification of each sample, with typical values such as 3, 5, and 7 being tested to find the best balance between model simplicity and accuracy. For Random Forest, the maximum depth of each decision tree was adjusted to control the complexity of the model, with depth values ranging from 1 to 7, and the number of estimators, or trees in the forest, explored with counts like 60, 80, or 100. These hyperparameters directly affect the model's ability to capture intricate patterns in the data while managing overfitting. In the case of Multilayer Perceptron (MLP), the learning rate, dictating the step size at each iteration toward a minimum of the loss function, was varied between 0.000001 and 0.05, and the number of epochs, representing how many times the learning algorithm processes the entire dataset, was explored with values ranging from 10 to 100 to ensure effective convergence without overfitting. Cross-validation was conducted during the training phase to determine which algorithm with its respective hyperparameters exhibited the highest accuracy.

The final accuracy of the model was assessed on each patient's test data, and these results were utilized to generate a comprehensive classification report and confusion matrix for each test sample. These evaluation metrics offered insights into the model's performance, guiding further refinement.

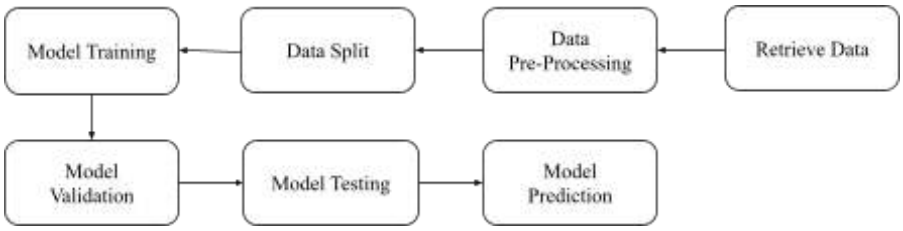


Figure 1. Flowchart of the proposed methodology

Results

Validation Results

1.K-Nearest Neighbors

Patient ID	Best K Value	Accuracy	Precision	Recall
1	**2**	**0.961**	**0.966**	**0.966**
2	**2 & 3**	**0.965**	**0.965**	**0.965**
3	**3**	**0.972**	**0.972**	**0.972**
5	**2 & 3**	**0.971**	**0.971**	**0.971**
6	**3**	**0.971**	**0.971**	**0.971**
7	**3**	**0.95**	**0.95**	**0.95**
8	**2**	**0.94**	**0.94**	**0.94**
9	**2**	**0.994**	**0.994**	**0.994**

Table 2: Validation Results for KNN

2.Random Forest

Patient ID	Best Depth Value	Best number of estimators	Accuracy	Precision	Recall
1	**7**	**20**	**0.964**	**0.961**	**0.961**
2	**7**	**50**	**0.981**	**0.981**	**0.981**
3	**7**	**20**	**0.971**	**0.976**	**0.976**
5	**7**	**20**	**0.977**	**0.974**	**0.974**
6	**7**	**70**	**0.985**	**0.985**	**0.985**

7	**6**	**70**	**0.929**	**0.929**	**0.929**
8	**7**	**60**	**0.947**	**0.947**	**0.947**
9	**6**	**70**	**0.998**	**0.998**	**0.998**

Table 3: Validation Results for Random Forest

3.Multi-Layer Perceptron

Patient ID	Best number of epochs	Best Learning Rate	Accuracy	Precision	Recall
1	**80**	**0.05**	**0.938**	**0.89**	**0.89**
2	**80**	**0.05**	**0.925**	**0.906**	**0.906**
3	**70**	**0.00001**	**0.9**	**0.896**	**0.896**
5	**30**	**0.000001**	**0.948**	**0.941**	**0.945**
6	**50**	**0.01**	**0.95**	**0.947**	**0.947**
7	**20**	**0.05**	**0.892**	**0.873**	**0.873**
8	**70**	**0.001**	**0.851**	**0.834**	**0.834**
9	**30**	**0.0001**	**0.974**	**0.974**	**0.974**

Table 4: Validation Results for MLP

After evaluating the validation results of the model on all the patient data, Random Forest emerged as the most accurate model, and thus, it was employed for further testing:

Test Results

Patient ID	Test Accuracy Score	Precision	Recall	F1 Score
1	**0.969**	**0.97**	**0.97**	**0.97**
2	**0.981**	**0.97**	**0.98**	**0.98**
3	**0.954**	**0.95**	**0.95**	**0.95**
5	**0.977**	**0.8**	**0.78**	**0.78**
6	**0.985**	**0.98**	**0.98**	**0.98**
7	**0.938**	**0.92**	**0.94**	**0.93**
8	**0.947**	**0.96**	**0.97**	**0.96**
9	**1.0**	**1.0**	**1.0**	**1.0**

Table 5: Classification Report for Random Forest Applied to Individual Patient Data

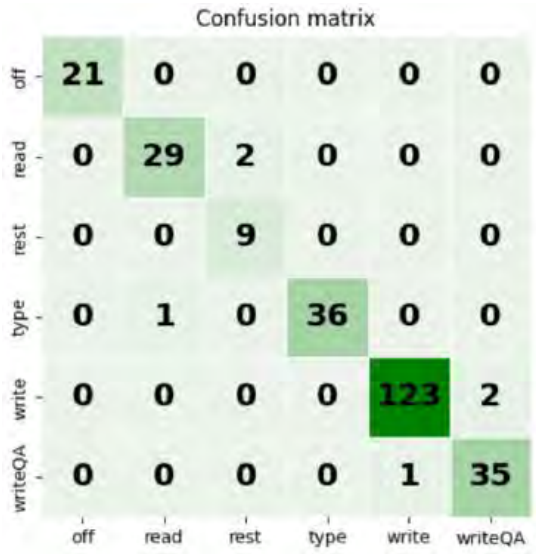


Figure 2: Confusion matrix for Patient 1

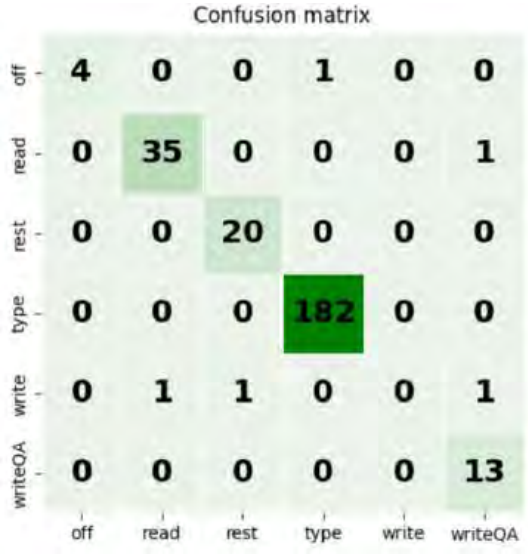


Figure 3: Confusion matrix for Patient 2

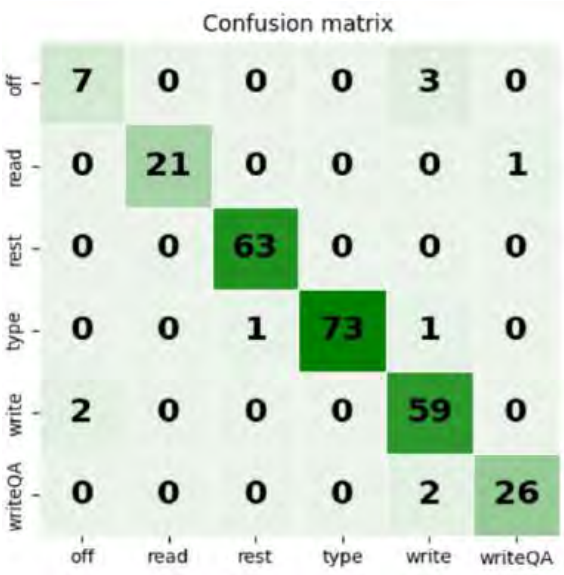


Figure 4: Confusion matrix for Patient 3

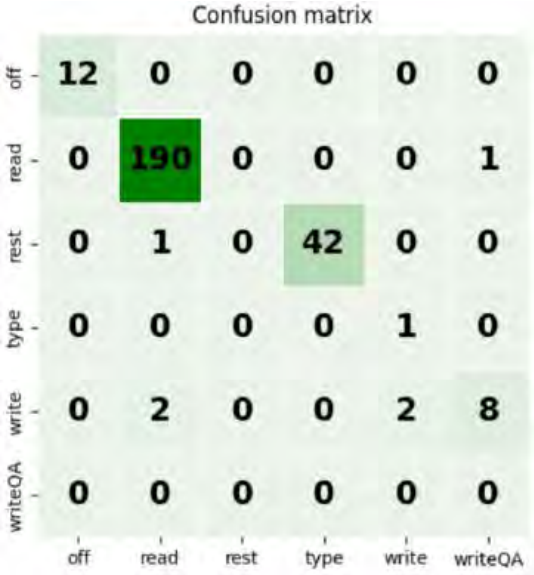


Figure 5: Confusion matrix for Patient 5

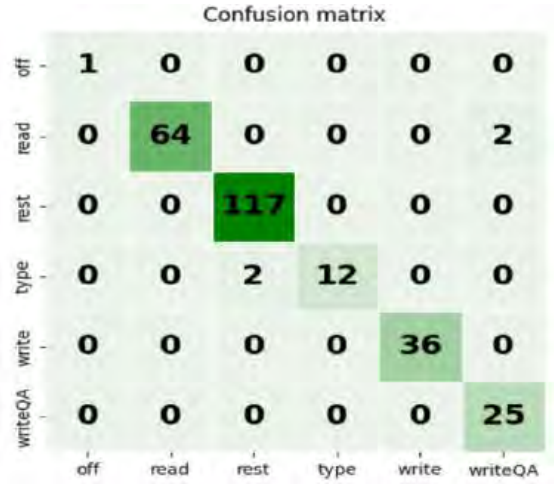


Figure 6: Confusion matrix for Patient 6

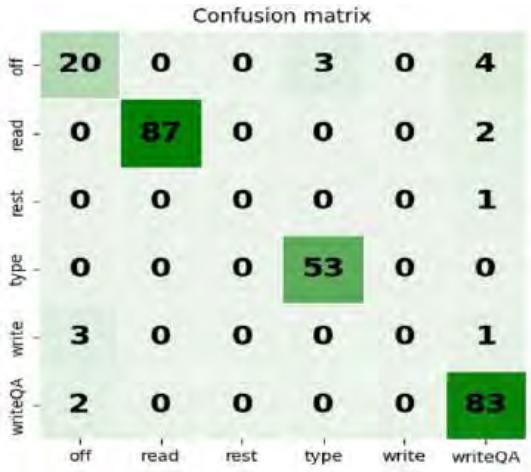


Figure 7: Confusion matrix for Patient 7

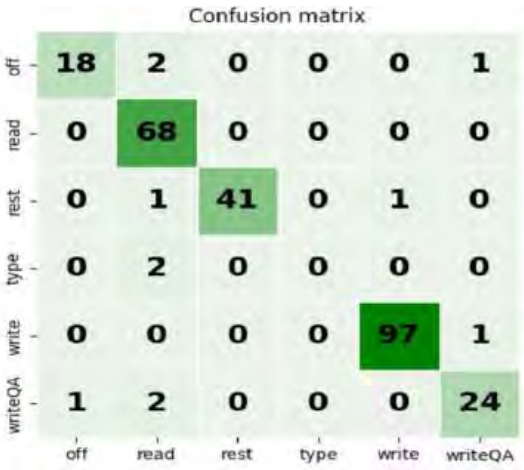


Figure 8: Confusion matrix for Patient 8

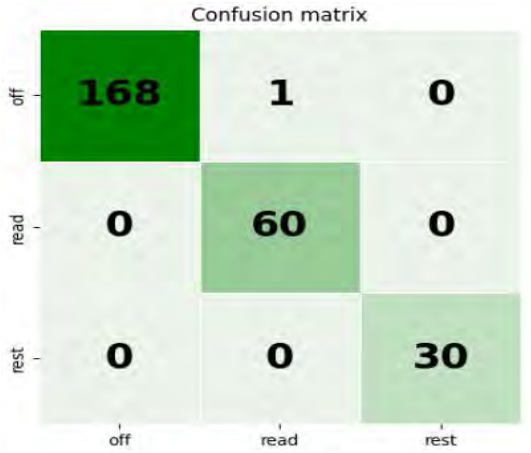


Figure 9: Confusion matrix for Patient 9

Discussion

The metrics in Table 1 offer a comprehensive evaluation of the model's performance in predicting activities among neurodivergent individuals. Using Patient 1's data as an illustrative example, we examine the classification process and the significance of confusion matrices, accuracy, precision, recall, and F1 scores.

Confusion matrices reveal specific activities where prediction errors occur, highlighting potential areas for model improvement. In a multi-class setup, each class is considered separately, with the matrix showing predictions versus actual activities. For instance, examining the "read" column of Patient 1's confusion matrix, the model predicted reading 30 times; in actuality, 29 were correct, while one was typing. Such insights underscore the model's robustness and direct model refinements.

Accuracy provides an overall performance snapshot, calculated as the ratio of total correct predictions to total predictions across all classes:

$$Accuracy = (Total\ Correct\ Predictions)/(Total\ Number\ of\ Predictions)$$

While informative, it should be viewed alongside precision and recall, especially in class-imbalanced datasets.

Precision, for multi-class problems, can be calculated using macro-averaging (treating all classes equally) or micro-averaging (accounting for the size of each class). Precision for a specific class is given by:

$$Precision = (True\ Positives)/(True\ Positives + False\ Positives)$$

For Patient 1, precision in the reading class can be directly interpreted using this formula.

Recall measures the model's capacity to identify true positives and can also be calculated using macro or micro-averaging in multi-class contexts. Recall is defined as:

$$Recall = (True\ Positives)/(True\ Positives + False\ Negatives)$$

For Patient 1, recall highlights the model's effectiveness in detecting true activities across all classes.

The F1 Score offers a balance between precision and recall, crucial for understanding model performance, especially with potential class imbalances. In a multi-class setting, it can also be calculated using macro or micro-averaging. The F1 score is given by the harmonic mean:

$$F1\ Score = 2 \times ((Precision \times Recall)/(Precision + Recall))$$

For Patient 1, an F1 score greater than 0.9 indicates excellent model performance.

These metrics altogether provide a balanced assessment of the model's performance, guiding further refinements and ensuring effectiveness across all patients and activities.

The evaluation of the three algorithms—K-Nearest Neighbors (KNN), Random Forest, and Multilayer Perceptron (MLP)—revealed distinct trends in their performance across the validation and test phases. Several key observations emerged from the analysis, highlighting both commonalities and unique behaviors of these models when subjected to different hyperparameters.

K-Nearest Neighbors (KNN) consistently showed high performance with optimal k values ranging from 2 to 3. The accuracy and precision metrics were particularly consistent across different patients, indicating the model's robustness. For example, Patient 3 achieved an exceptionally high accuracy of 0.972 with k=3. This suggests that KNN was well-suited to the underlying data structure, effectively leveraging the proximity of neighboring data points to make accurate predictions.

Random Forest emerged as the most accurate model during validation, with accuracy scores consistently above 0.96. The model demonstrated stability across various depths and numbers of estimators, with depth values typically around 6 to 7 and estimator counts between 20 and 70. This indicates the algorithm's strength in managing complex data patterns and avoiding overfitting, courtesy of its ensemble approach.

Multilayer Perceptron (MLP)'s performance was more variable, largely influenced by the chosen learning rate and number of epochs. While it achieved decent results, such as 0.95 accuracy for Patient 6 with a learning rate of 0.01, it generally lagged behind the other models. The tuning of hyperparameters like learning rate, which ranged from 0.000001 to 0.05, greatly impacted its ability to converge effectively.

When comparing models, Random Forest consistently outperformed the others in terms of higher accuracy and more reliable precision and recall. Its resilience across different patients and adaptability to various data complexities made it the model of choice for further testing. KNN, while effective, showed sensitivity to the optimal selection of k values. In contrast, MLP required careful adjustment of hyperparameters and still exhibited more fluctuation in performance compared to the other two models.

These validation trends translated effectively into test metrics, where Random Forest delivered high test accuracy, often exceeding 0.95, even achieving a perfect score of 1.0 for Patient 9. The precision and recall scores reflected the model's ability to generalize well from training and validation situations to new, unseen data.

An interesting insight was Random Forest's consistently high performance irrespective of the patient data, pointing towards its robustness and generalization capabilities. In contrast, MLP's reliance on optimal hyperparameter tuning became evident, suggesting potential benefits from more advanced methods like hyperparameter search or tuning to enhance its reliability. The absence of a one-size-fits-all k value in KNN highlights the importance of patient-specific tuning to maximize performance. Random Forest's success in both validation and testing phases makes it a strong candidate for practical applications in this context.

Conclusion

This study underscores the potential of leveraging smartwatch sensor data to accurately predict the learning activities of neurodiverse individuals through machine learning models. Among the models evaluated, the Random Forest algorithm proved to be the most effective, outperforming K-Nearest Neighbors and Multilayer Perceptron models in terms of accuracy across various patient datasets.

We used machine learning tools like accuracy, precision, recall, and F1 scores to assess the effectiveness of the model. The high precision, recall, and F1 scores observed demonstrates the model's ability to reliably identify activities while minimizing errors, thereby contributing to a deeper understanding of neurodiverse behaviors. These results affirm the viability of unobtrusive monitoring tools in supporting personalized interventions that aim to enhance the quality of life for neurodiverse populations.

Further research

While the current study provides significant insights into the potential of using smartwatch sensor data for human activity recognition in neurodiverse individuals, further research is essential to address existing challenges and explore further opportunities. Our future research directions include:

To enhance the model's generalizability, future research should be dedicated to gathering data from a larger and more diverse cohort. This would involve recruiting participants from varied age groups, ethnic backgrounds, and neurodiverse profiles. By doing so, the robustness and applicability of the model across different demographics will be significantly improved, ensuring that the system can cater to a broader range of user needs.

Investigating more sophisticated feature extraction and selection techniques can capture subtle patterns in the data that might be missed through conventional methods. This could involve the use of deep learning approaches to automatically learn representations from raw sensor data or identifying domain-specific features that are particularly relevant to neurodiverse activities. Enhanced feature engineering can lead to improved prediction accuracy and a deeper understanding of the nuanced behaviors exhibited by neurodiverse individuals.

Initial explorations into the real-time application of the HAR model could yield immediate feedback mechanisms that are responsive to the needs of the user. This involves optimizing algorithms for deployment on wearable devices, ensuring low processing latency, and creating adaptive interfaces that provide actionable insights to the user or caregiver, facilitating timely interventions supportive of the individual's focus and behavioral needs.

Extending the model's capabilities to recognize a wider range of activities and environmental contexts will better accommodate the complexity and variability inherent in neurodiverse behaviors. This may involve categorizing activities across different domains of daily life, including social interactions and sensory engagements, which are critical to the comprehensive understanding of neurodiverse experiences.

Enriching the HAR dataset by incorporating information from supplementary data streams—such as environmental sensors, digital feedback from educational platforms, or input from caregivers—can enhance the model's contextual awareness and predictive performance. This multidisciplinary data integration could facilitate a more holistic view of the individual's activity and well-being.

Ensuring ethical standards are meticulously upheld in monitoring neurodiverse individuals is paramount. This includes prioritizing data privacy, securing informed consent explicitly aligned with the neurodiversity of participants, and maintaining transparency about data usage. Developing guidelines and frameworks to address these issues will help build trust and acceptance among users and stakeholders.

By addressing these areas, future research can significantly advance the field of human activity recognition. This will not only lead to improved technological solutions but also offer tangible benefits for neurodiverse individuals, enhancing personalized care and supporting their autonomy.

Works Cited

Angerbauer, Simon, et al. “Comparing Human Activity Recognition Models Based on Complexity and Resource Usage.” MDPI, Multidisciplinary Digital Publishing Institute, 13 Sept. 2021, www.mdpi.com/2076-3417/11/18/8473.

Ha, Nam, et al. “Machine Learning‐enabled Smart Sensor Systems - Ha - 2020 - Advanced Intelligent Systems - Wiley Online Library.” Wiley Online Library, 14 July 2020, onlinelibrary.wiley.com/doi/full/10.1002/aisy.202000063.

IBM. “What Is Random Forest?” IBM, 19 Dec. 2024, www.ibm.com/think/topics/random-forest.

IBM. “What Is the K-Nearest Neighbors Algorithm?” IBM, 19 Dec. 2024, www.ibm.com/think/topics/knn#:~:text=The%20k%2Dnearest%20neighbors%20(KNN)%20algorithm%20is%20a%20non,used%20in%20machine%20learning%20today.

Ibrahim, Mostafa. “An Introduction to the F1 Score in Machine Learning.” W&B, 28 Feb. 2024, wandb.ai/mostafaibrahim17/ml-articles/reports/An-Introduction-to-the-F1-Score-in-Machine-Learning--Vmlldzo2OTY0Mzg1#:~:text=For%20example%2C%20if%20a%20model,performance%20between%20precision%20and%20recall.

Jaiswal, Sejal. “Multilayer Perceptrons in Machine Learning: A Comprehensive Guide.” DataCamp, DataCamp, 28 July 2024, www.datacamp.com/tutorial/multilayer-perceptrons-in-machine-learning.

Keen, Martin. “What Is the K-Nearest Neighbor (KNN) Algorithm?” YouTube, IBM Technology, 2 Sept. 2024, www.youtube.com/watch?v=b6uHw7QW_n4.

Shah, Deval. “Human Activity Recognition (HAR): Fundamentals, Models, Datasets.” V7, 27 Mar. 2023, www.v7labs.com/blog/human-activity-recognition.

Zheng, Hui, et al. “WLA4ND: A wearable dataset of learning activities for young adults with neurodiversity to provide support in education.” Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility, 17 Oct. 2021, pp. 1–15, https://doi.org/10.1145/3441852.3471220.

Impact of Dietary Management in Estrogen Regulation on the Prevention of Relative Energy Deficiency in Sport (REDs) By Katherine Mullane

Abstract

Relative Energy Deficiency in sport (REDs) is a syndrome individuals develop from lacking sufficient amounts of energy compared to the energy demanded during exercise. Similarly, significant symptoms of REDs is Low Energy Availability (LEA): a condition found more commonly in females. Sparse supplies of energy result in many harmful impacts on athletes' physiological health. Low levels of estrogen can lead to complications including low bone mineral density (BMD) and disruptions of the menstrual cycle. To reverse or prevent the effects of energy deficiency, individuals need to consume specific micro and macronutrients. These nutrients can be ingested to supply the body with more energy, raising levels of estrogen and repairing menstrual and bone function. The purpose of this paper is to review studies on consumption of essential macro and micronutrients to help athletes manage and mitigate the physiological impact of REDs.

Introduction

By definition, LEA, or Low Energy Availability, is insufficient energy intake relative to expended energy through exercise[1]. REDs, Relative Energy Deficiency in sport, is a syndrome of depleting health and performance when athletes are exposed to LEA[1]. Despite its commonality in female athletes, REDs is rarely discussed and not well known to a greater population of athletes and coaches. Athletes who are prone to developing REDs are likely to participate in high intensity sports such as running. Both REDs and LEA can leave long-lasting impacts on athletes which can stay with individuals longer than their athletic careers[2]. These impacts include lowered bone density from disrupted hormonal levels, disrupting performance and overall health. Without sufficient energy, the hypothalamic-pituitary-ovarian axis is suppressed, resulting in altered production of estrogen[3]. Menstrual and ovarian functions become dysregulated. Thus, bone mineral density is lowered, which impacts bone regeneration and increases the risk of injury[4]. Athletes are frequently told to eat "healthier" when faced with health issues; however, most athletes do not know what to eat and what best enhances their performance to keep them healthy and avoid injuries. Micronutrients such as vitamin D, calcium, and iron contribute to healthy bone cells and energy levels[5]. Furthermore, lower levels of macronutrients, including carbohydrates and protein, are associated with REDs[6]. Therefore, this literature review is to help athletes find a balanced diet with sufficient macro and micronutrients to avoid or mitigate the effects of REDs. Specifically, REDs should not occur when athletes get enough energy from certain nutrients to match the amount of energy they expend[2].

Literature Review

REDs and LEA

REDs, as stated previously, is a syndrome of declining health due to individuals being exposed to Low Energy Availability (LEA)[1]. LEA is a state in which one doesn't get a sufficient amount of energy compared to the energy being expended. The decline of health can be defined by multiple injuries mainly due to stress[4] and by the disruption of important body functions such as the menstrual cycle and the regrowth of bone cells. To avoid overuse, overtraining, burnout, and consequently REDs, individuals can look towards nonsurgical rehabilitation or management[7]. To minimize risk, athletes can optimize energy availability and treat body system dysfunctions by moderating types and amounts of food intake[1].

Macronutrients

Dietary management plays a crucial role in preventing REDs by restoring lost energy and allowing the body to function healthily. Carbohydrate, protein, and fat consumption is vital to avoid LEA and its negative health consequences (Figure 1).

Carbohydrates are sugar molecules that are broken down into simpler compounds known as glucose. Glucose is the body's main source of energy. LCHO, or low carbohydrate intake (<3.0g/kg), is positively correlated with low energy availability. Thus, the lower the carbohydrate intake, the lower energy availability[6]. Furthermore, lower carbohydrate levels are associated with negative endocrine and metabolic responses from bodies. A study conducted by doctors of sport & health sciences in Iceland tried to find a correlation and causation between low carbohydrate intake and LEA. Individuals in this study who fell under the group LEA+LCHO (had both low energy availability and low carbohydrate intake) were deficient in other vital micro and macronutrients[6]. All aspects of a diet are important to reduce the chances of REDs. With lower levels of carbohydrates, individuals can lack essential nutrients to promote a healthy body. Women experiencing LCHO reported disruptions in their periods, conveying that menstrual cycles and endocrine systems are impacted by carbohydrate intake (Figure 2).

Proteins also impact systemic functions in the body due to the macronutrient's role in hormonal regulation, enzyme functions, and being the main building blocks of cells. As a result of the importance of protein for the menstrual cycle, a higher intake of protein should be prioritized[2]. Consumption of proteins allows muscle mass to grow and repair. Recommended protein levels are 1.2-2.0 g/kg BW[6]. The LEA+LCHO group displayed <1.0 g/kg BW of protein, expressing that protein deficiency is also related to low energy availability.

Similarly, fats sustain metabolic and hormonal systems. Fats should be about 20 to 35% of an individual's total energy consumption or intake[2]. Women burn more macronutrients when exercising than men, as found in a study conducted by Hausswirth et al, emphasizing the significance of LEA in active females who cannot get an adequate amount of macronutrients. In addition, the connection between LEA and the lack of macronutrients expresses a correlation between the two factors. Although there are many aspects that play into the lack of energy, consuming vital macronutrients – such as carbohydrates, proteins, and fats – ensures an increase in energy availability and will decrease the chances of having LEA. Some foods high in protein

are red meat, fish, and eggs. Carbohydrates consist of wheat or grain products such as pasta noodles, bread, and rice. Fats are found in avocados, nuts, and oils.

Micronutrients

Similar to macronutrients, micronutrient consumption is vital to ensure energy availability is optimal (Figure 1). Despite the variety of micronutrients, vitamin D, calcium, iron and ferritin, and magnesium are essential due to their positive impact on bone mineral density (BMD). Having a healthy BMD ensures that future breakage or fracturing of bones – which is a symptom of REDs – is less likely to occur.

Vitamin D, a crucial micronutrient, helps improve bone health and reduce the risk of bone stress injuries. Vitamin D also helps maintain normal neuromuscular functions in individuals[2]. According to the IOC, sufficient Vitamin D intake is >30ng/mL ~ 80 nmol/L. Vitamin D deficiency is defined as <50 nmol/L[6]. Furthermore, athletes with LEA and LCHO had lower levels of vitamin D. In other words, low vitamin D intake is associated with low carbohydrate intake and low energy availability. To raise vitamin D levels, foods, and supplements should be consumed. Eggs, dairy products (milk and cheese), salmon, herring, and mackerel have abundant amounts of vitamin D[2].

Calcium, more commonly associated with bone growth, is crucial for maintaining optimal BMD. In support of this statement, a study shows that 85% of female runners with low bone density didn't meet the recommended intake for calcium[2]. Athletes with nutritional disorders, or menstrual disruptions, are recommended to take about 1500 to 2000 mg/day of calcium. Higher dietary calcium intakes protect athletes from stress fractures. Two studies concluded that low calcium intake increased the risk of stress fractures[5]. Furthermore, the main dietary sources of calcium include dairy and milk products. Spinach, kale, and beans are plant products high in calcium. Both lactose and vitamin D augment calcium absorption. Only 500mg of calcium can be ingested at one time, so it is crucial to spread calcium intake throughout the day.

Ferritin, a protein that stores iron (Fe) is important for oxygen transport as hemoglobins in the bloodstream. If iron levels are suppressed, LEA can develop due to the diminishment of energy efficiency, lowering of maximal oxygen consumption, and the athlete's lack of ability to work at an optimal level[2]. The recommended Fe consumption per day is about 18 mg/day. In addition, it is recommended that females increase the estimated average requirement by 70% due to the menstrual cycle[2]. Iron is lost through menstruation, sweat, and the gastrointestinal tract. Therefore, females in demanding and high-impact sports (such as distance running) need to consume iron to avoid fatigue and decreased performances[4]. Meat products have high iron content and can be consumed to raise iron levels. Supplementation, particularly in pill, liquid, drops, or capsule form, is another way to increase iron consumption. Intestinal absorption of ferrous sulfate, citrate, and gluconate can substitute for consuming foods such as meat. By supplementing or consuming more foods rich in iron, athletes can improve their athletic capabilities and fight against LEA.

Magnesium, another vital micronutrient, helps ensure electrolyte balance and keeps the endocrine system in check. This micronutrient also ensures the body takes in oxygen at a proper level. Women in sports with high LEA rates express that they have low magnesium intake[2]. Moreover, athletes need about 20% more magnesium than standard individuals. Magnesium aids other micronutrients to support bone health as well[9]. According to a book on dietary reference intakes, magnesium's Recommended Dietary Allowances (RDAs) are based on personal size and weight rather than a universal amount of magnesium. For instance, females in the 14 to 18-year-old range should take 300 mg (12.5 mmol)/day of magnesium with a reference weight of 57 kg[10]. Foods high in magnesium include bananas, potatoes, nuts, legumes, dark chocolate, and cereal. Supplementation of magnesium is important if individuals can't reach recommended magnesium levels.

A balanced diet comprising of macronutrients and food rich in micronutrients, or vitamins, is essential to maintain a sufficient EA and reduce the development of LEA or further REDs.

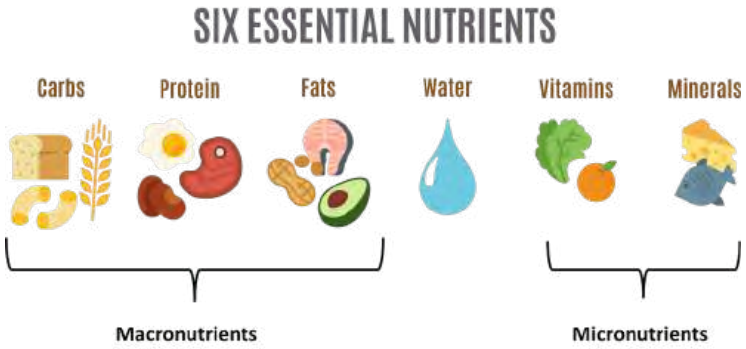


Figure 1 — Essential nutrients comprise macro and micronutrients[11]. Carbohydrates, proteins, and fats are macronutrients while micronutrients consist of vitamins and minerals. Vitamin D is a well known micronutrient which promotes bone health. All of these nutrients are proven to be helpful when raising Energy levels if they are consumed at a recommended level.

REDs Impact on Estrogen and the HPG Axis

As seen in the macronutrient section, women and men are different when it comes to how each gender expends energy. Women secrete sex-specific hormones from glands such as ovaries. Controlling the ovaries, the hypothalamic-pituitary-gonadal axis (HPG) is a system of endocrine glands controlling reproductive processes. The ovaries produce estrogen which is a hormone vital for reproductive and bone health (Figure 2). As found in female athletes, long-term LEA can lead to low bone mineral density. In addition, these athletes with low BMD have menstrual dysfunction[8]. Osteoblasts and osteoclasts exist in balance in bones. Both of these cells are important to maintain bone health. Osteoblasts build bones while osteoclasts break down bones. Estrogen reduces the number of osteoclasts. When estrogen is low, osteoclasts increase in numbers, offsetting the balance of bone remodeling towards more bone destruction (Figure 2). About 28% of females in a low BMD study reported having no menstrual bleeding despite

having gone through puberty and having a cycle in the past[8]. Therefore, there is a clear connection between LEA and menstrual cycles. The more LEA is present, the more strain on the menstrual cycle and overall bodily health.

Due to the HPG axis and estrogen's regulation of the menstrual cycle, these systems are also impaired by low energy availability and REDs[4]. The macro and micronutrient segment highlights that without proper nutrition, individuals will most likely develop LEA. Females with low levels of essential nutrients have an inhibited HPG axis. Due to this impairment, there is an inhibition of the release of gonadotropin-releasing hormone (GnRH), and consequently, low production of luteinizing hormone[9], follicle-stimulating hormone, and estrogen[3] (Figure 2). Inhibition of these systems causes females to have a higher risk of anovulation. Estrogen deficiency is linked to menstrual irregularities and the development of stress fractures[5]. Disruption of the menstrual cycle can lead to severe health complications and impede athletic abilities. Therefore, by avoiding dietary deficiencies, female athletes can avoid LEA and disruptions of menstrual processes[3] and reduce bone breakdown.

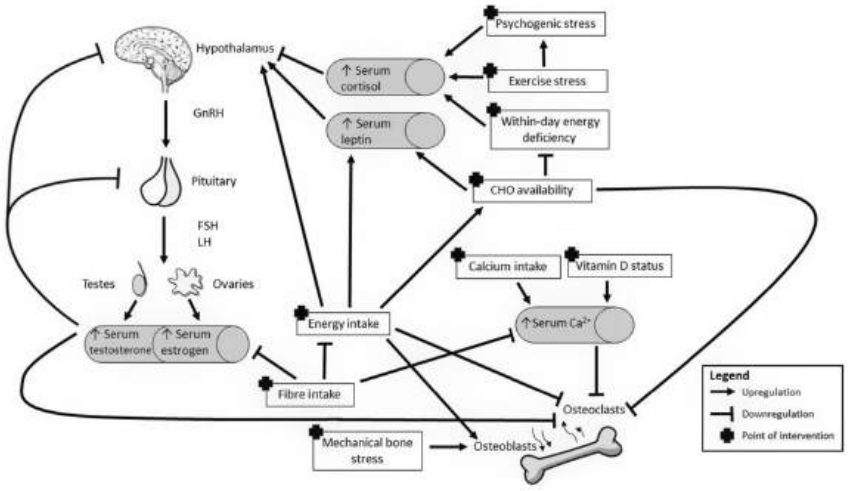


Figure 2 — This image depicts the function of the HPG axis and how hormones affect bodily processes[9]. The HPG axis consists of the hypothalamus, the pituitary gland, and the ovaries/testes, which produces estrogen. Furthermore, this image expresses how calcium and Vitamin D (micronutrients) intake can affect osteoclasts and osteoblasts, affecting BMD and the HPG axis as a result.

Alternative Perspective

Dietary management has a minimal effect on the prevention of REDs. Some factors other than lack of sufficient nutrition can cause REDs. There are many different approaches to mitigating the effects of REDs, causing an emphasis on nutritional balance to be diminished.

Conclusion

Biases

As a review paper, this text contains and utilizes information from published medical articles and studies conducted by credible sources. These studies note that all athletes and individuals, used as study subjects, are different and have differing nutritional necessities. For instance, everyone has varying EA levels which affects the amount of nutritional intake (of macro/micronutrients) to ensure that LEA isn't developed. Despite the recommended amounts of intake of macro and micronutrients, the amount of energy expended by each person differs. Furthermore, many studies, such as the LCHO study, used self-reporting surveys to get an accurate idea of how much food participants consume each day. Surveys can lead to small forms of bias as the participants can fill out surveys knowing how to answer certain questions to appear different than their true selves. Accordingly, nutritional intake was subject to error in memory when asked to be recorded. Measuring EA was another process that could involve some bias as there are portion estimates needed to figure out how much energy is being consumed and how much is being expended. Finally, despite REDs and LEA being the main causes of stress fractures and disruptions of the menstrual cycle, individuals struggling with these issues may have other health problems that can't be solved with just nutritional changes.

Summary

As individuals are exposed to LEA, REDs can develop, causing many vital bodily functions to be disrupted. In order to minimize the risk of REDs, individuals, specifically female athletes, can manage the types or amount of food intake to optimize energy availability. Macronutrients with the ability to reduce the chances of developing LEA include carbohydrates, proteins, and fats. All of these macronutrients play a key role in raising energy availability and keeping systems and glands, such as the endocrine system, in check. Furthermore, micronutrients are also crucial for keeping energy availability at a healthy level. Vitamin D and calcium ensure healthy bones and a healthy BMD, while iron sustains a high aerobic capacity. Magnesium supports energy uptake and metabolic processes. These micronutrients keeps the bodily processes in good condition and reduce risk of LEA. Supplementation of micronutrients is important if the consumption of foods rich in these nutrients isn't meeting the daily consumption requirements. Furthermore, without sufficient amounts of macro and micronutrients, suppression of the hypothalamic-pituitary-ovarian axis occurs, causing menstrual dysfunction. If consumption of the previously listed macro and micronutrients is increased to reach optimal levels, LEA and REDs will occur less in female athletes.

Further Implication

Athletes can be more aware of what needs to be consumed for their bodies to stay healthy. Furthermore, athletes should know the dangers of REDs and LEA and how sufficient nutrition is vital to maintain healthy bodily functions. Parents and coaches can be more vigilant of REDs in female athletes, helping them develop good nutrition to avoid menstrual disturbances and bone fractures. Overall, this literature review ensures that REDs is acknowledged in athletic communities to help athletes mitigate the impacts and development of LEA.

Works Cited

Mountjoy, Margo, et al. “2023 International Olympic Committee’s (IOC) Consensus Statement on Relative Energy Deficiency in Sport (REDs).” British Journal of Sports Medicine, vol. 57, no. 17, 1 Sept. 2023, pp. 1073–1097, https://doi.org/10.1136/bjsports-2023-106994.

Grabia, Monika, et al. “Female Athlete Triad and Relative Energy Deficiency in Sport (REDs): Nutritional Management.” Nutrients, vol. 16, no. 3, 1 Jan. 2024, p. 359, www.mdpi.com/2072-6643/16/3/359, https://doi.org/10.3390/nu16030359.

Allaway, Heather C.M., et al. “The Physiology of Functional Hypothalamic Amenorrhea Associated with Energy Deficiency in Exercising Women and in Women with Anorexia Nervosa.” Hormone Molecular Biology and Clinical Investigation, vol. 25, no. 2, 1 Jan. 2016, https://doi.org/10.1515/hmbci-2015-0053.

“UpToDate.” Uptodate.com, 2024, www.uptodate.com/contents/running-injuries-of-the-lower-extremities-in-adults-risk-factors-and-prevention/print. Accessed 25 Dec. 2024.

Wentz, Laurel, et al. “Dietary and Training Predictors of Stress Fractures in Female Runners.” International Journal of Sport Nutrition and Exercise Metabolism, vol. 22, no. 5, Oct. 2012, pp. 374–382, https://doi.org/10.1123/ijsnem.22.5.374.

Birna Vardardottir, et al. “Patterns of Energy Availability and Carbohydrate Intake Differentiate between Adaptable and Problematic Low Energy Availability in Female Athletes.” Frontiers in Sports and Active Living, vol. 6, 9 May 2024, https://doi.org/10.3389/fspor.2024.1390558. Accessed 14 May 2024.

“UpToDate.” Www.uptodate.com, www.uptodate.com/contents/whats-new-in-sports-medicine-primary-care.

Mursu, Jaakko, et al. “Dietary Intake, Serum Hormone Concentrations, Amenorrhea and Bone Mineral Density of Physique Athletes and Active Gym Enthusiasts.” Nutrients, vol. 15, no. 2, 12 Jan. 2023, p. 382, https://doi.org/10.3390/nu15020382. Accessed 6 May 2023.

Kuikman, Megan A., et al. “A Review of Nonpharmacological Strategies in the Treatment of Relative Energy Deficiency in Sport.” International Journal of Sport Nutrition and Exercise Metabolism, vol. 31, no. 3, 2020, pp. 1–8, https://doi.org/10.1123/ijsnem.2020-0211.

National Academies of Sciences, Engineering, and Medicine. 1997. Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride. Washington, DC: The National Academies Press. https://doi.org/10.17226/5776.

Burke, Deanna. “Macronutrients and Food Groups: Part 1 of 7.” ClubWorx, 6 Oct. 2021, clubworx.net/macronutrients-and-food-groups-part-1-of-8/.

Unraveling Nemaline Myopathy:
Genetic Mutations, Pathophysiology, and Emerging Therapeutic Approaches
By Ishaan Murth

Abstract

Nemaline myopathy is a rare muscle disorder affecting 1 in 50,000 people, with a wide range of severity. This disorder is characterized by muscle weakness and issues with skeletal mobility, and is caused by mutations in genes like NEB or ACTA1. These mutations disrupt the proteins encoded by the genes which ultimately leads to less effective muscle function. These mutations in the genes lead to the formation of nemaline bodies in muscle fibers which disrupts the muscles ability to contract and movement of the muscle becomes more limited. This paper dives into the pathophysiology of nemaline myopathy along with the genetics involved with it, especially looking into the impacts mutations have on the proteins produced by their respective gene. This paper will also talk about current treatments for nemaline myopathy which are very limited, as treatments only include physical therapy and there is no cure currently for the disease. Researchers are working on cures including gene therapy and ASOs. Current treatments are just meant to help stabilize the muscles and give patients a stronger foundation to live upon, increasing quality of life at the same time.

Introduction

Nemaline myopathy is a group of muscle disorders that primarily affect the body's mobility,that vary in severity, and are diagnosed in early childhood or infancy. (1) Nemaline myopathy (NM) is marked by nemaline bodies which are rod or thread like structures that appear in skeletal muscle cells of people with NM, and are visible by a muscle biopsy. (2) Nemaline myopathy is prevalent in approximately 1/50,000 live births worldwide, although some isolated communities have much higher incidence like the Amish where nemaline myopathy is detected in an estimated 1/500 deaths. (3) NM is an inherited genetic disorder where there can be mutations in one of several genes, such as NEB which accounts for 50% of all NM cases, ACTA1 which accounts for 15-25% of cases, and TMP3. (4) These genes play roles in the function and structure of skeletal muscles, while the mutations lead to the distinct muscle weakness in NM. NM is usually inherited in an autosomal recessive pattern which means that both copies of the gene in each cell have mutations. (4) There are many forms of NM which can affect the patients lifespan due to the severity of the disease. The neonatal form of nemaline myopathy is the most severe form because infants who are born with this will have severe respiratory issues along with muscle weakness. Among individuals born with this form of NM, lifespan ranges from a few months to a couple of years, with most passing away during their early lives due to major respiratory complications or failure. (5) With other milder forms of nemaline myopathy people can live through adulthood or even experience a typical life span although still experiencing health problems and muscle weakness. (6) Regardless of the severity or form of NM most individuals with the disease experience some form of chronic muscle

weakness which weakens the muscle overtime and can eventually lead to a disability which may require different forms of motor assistance. Progressive weakness of respiratory muscles or chronic respiratory failure can dramatically change an individual with NM's quality of life as they would need to be placed on ventilatory support. (5) If not treated correctly, these respiratory issues can lead to complications such as pneumonia. (7) Individuals with NM can also develop hypotonia and skeletal deformities, such as scoliosis which is a sideways curve of the spine. These deformities may require surgery and will further limit mobility. (4) Overall nemaline myopathy significantly can affect an individual's quality and length of life.This review paper will cover the genetics and mutations that cause NM, the pathophysiology of NM, and finally an in depth summary of treatments and diagnosis and how it has changed and will change overtime.

Genetics

Nemaline Myopathies are caused by mutations in a variety of genes that play a role in the function of skeletal muscle where the issue is at its worst. The two most common genes are NEB, which accounts for 50% of all nemaline myopathy cases, and ACTA1, which accounts for 15-25% of all cases. Additional genes like TPM2, TPM3, TNNT1, and CFL2 also contribute to the disease, but are connected to less common forms of the already rare nemaline myopathy disease (4). Although these genes can cause nemaline myopathy variants they are all very unique. One of these genes is the NEB gene which provides instructions for making a protein called nebulin. Within skeletal muscles, nebulin is found inside sarcomeres which are necessary for the muscle to contract, which means that they tense up and become smaller and tighter (8). Muscle contraction is necessary for muscle growth, and when the muscle does not contract the body starts to think that it doesn't need to use the muscle anymore causing it to decrease in size and strength (9). In patients with nemaline myopathy, the muscles do contract but are often weak or inefficient. Over 60 mutations in the NEB gene have been found that cause nemaline myopathy (10). Some mutations insert or delete genetic material within the NEB gene, whilst others change a single amino acid resulting in the lack of ability to produce the necessary nebulin (11).

The ACTA1 gene, like the NEB gene, provides a set of instructions to create alpha-actin (a-actin) which is an important protein for cell movement. (12) A-actin is part of the actin family where acta1 plays a role of maintaining the structure of actin filaments (13) which are important for cell division, cytokinesis, muscle contractions and more, meaning that a mutation in the ACTA1 gene can lead to severe forms of nemaline myopathy (14). The TPM2 and TPM3 genes encode tropomyosin isoforms that are also meant to stabilize actin filaments (15).

The primary inheritance pattern of nemaline myopathy is autosomal recessive, and that means that both copies of the affected gene will have to carry mutations for the disease to present (16). The autosomal recessive inheritance is more typical for mutations in the NEB, TPM2, and TPM3 genes (Figure 2). This doesn't mean that autosomal recessive is the only possible inheritance for this disease. In fact, ACTA1 cases of NM can exhibit autosomal dominant inheritance, but this is rare and can result in more severe cases of the disease. One interesting

thing regarding the genetics of nemaline myopathy is the “super repeat” which is a structural unit found in some genes but commonly NEB (17). In different species, nebulin contains super-repeats, each with a similar domain composition that consists of smaller repeats called simple repeats, each containing an actin-binding site (Figure 1) The number of super repeats was shown to correlate with thin filament lengths (18). Super repeats are essential for regulating the length of actin filaments in muscle fibers, but mutations within this region can disrupt the proper structure of nebulin causing defective muscle contraction ability (19). Malfunctions in these regions can also cause errors during DNA replication (18). This matters as it will cause heavily disrupted muscle development and function.

Pathophysiology

Nemaline myopathies are defined by nemaline bodies which are rod-like structures that are found in the muscle fibers of affected individuals with nemaline myopathy (3). Nemaline bodies consist of sarcomeric proteins and actin, and the accumulation of these disrupts normal muscle fiber function (20). Nemaline bodies are formed by miscomposed or aggregated folded proteins, and this is the result of mutations in the ACTA1 or NEB genes (21). The presence of abnormal proteins interrupts proper sarcomere positioning and organization (Figure 3). This matters because sarcomeres are responsible for muscle contraction which is disrupted in nemaline myopathy (21). The difference in a healthy muscle cell is that sarcomeres are arranged optimally which results in efficient muscle contraction which is necessary for mobility, but in nemaline myopathy the organization of sarcomeres is disrupted by the nemaline bodies that weaken the muscle fibers (19).

Researchers have conducted various tests and experiments on different animals to understand nemaline myopathy better, most notably being tests on mice. Studies have found that in mouse models there are shorter thin filament lengths and decreased force generation by muscle contraction. (22) There are two main types of NM that have been studied in mouse models: ACTA1 and NEB. In NEB mouse models research has shown that there aren't major reductions in mouse lifespan, and in nebulin deficient mice nemaline rods are not consistently shown. NEB mutations for mice can be characterized by both missense and nonsense mutations (23). There have been other studies and tests for nemaline myopathy in other animals like zebrafish. In an ACTA1 zebrafish model, impaired swimming ability is present along with disorganized muscle structure that is proven by the nemaline bodies present, and in zebrafish there are 3 different types of nemaline bodies (24). One reason why studying zebrafish for nemaline myopathy is beneficial for research is because genetic manipulation comes easier than in mammals. Some of the most studied models in mice are nebulin knockout mice and mice with specific nebulin mutations. Nebulin knockout mice completely lack nebulin expression and this causes severe muscle weakness (23). Unfortunately although these models have been crucial in studying nebulin in nemaline myopathy their phenotype limits the usefulness for studying long term disease progression. Recently a compound heterozygous mouse model(Neb Y2303H,Y935X) has been developed, and this more closely mimics the genetics of most patients

with nemaline myopathy (23). This is because this model resembles nemaline bodies and muscle weakness of those similar to a human without early lethality (23).

Treatment

The treatment options for nemaline myopathies are limited because of how rare the disease is and the complexity of the genetics that come into play for it. As of today there are no cures for nemaline myopathy, but extensive research is being conducted into myopathies, especially rod body myopathy (1). Currently treatments are used to help improve patients' quality of life (25) and also to help with managing respiratory complications and skeletal muscle weakness (26). Most of these treatments involve physical therapy and if needed mechanical ventilation to support severe respiratory issues (26). Some patients with NM undergo surgery to fix skeletal deformities which is similar to surgery for scoliosis (27).

Recent studies in nemaline myopathy have been focusing on the molecular mechanisms of the disease, with the goal of addressing the harmful genetic mutations and proteins that are causing nemaline myopathy in individuals in the first place (27). Due to the recent boom in popularity of CRISPRCas9, gene therapy has gained more attention than ever before. Gene therapy, a technique that involves modifying an individual's genes, or introducing repaired genes into a muscle cell affected with a certain disease, is a promising route to go for future treatment regarding Nemaline myopathy (28). Scientists have been conducting different research and exploring possibilities to deliver functional copies of genes like NEB or ACTA1 into muscle tissues, which could improve and possibly restore proper muscle function (29). Unfortunately approaches like this are still in the early experimental stage and could take years to develop.

Another area of treatment that is being experimented on regards the use of ASOs, which are short for antisense oligonucleotides. ASOs are short RNA or DNA molecules that are used to change the expression of certain genes (30). ASOs haven't been directly linked to nemaline myopathies specifically but more around the area of congenital myopathy. ASOs could theoretically be used to fix splicing defects in NEB or ACTA1 and also moderate the production of proteins like nebulin or actin. ASOs have been experimented on in humanized versions of mice models (31).

Although gene based therapies look to be the most promising as of now, a lot of other research is going on for pharmacological based treatments. A promising approach for nemaline myopathy is targeting different pathways that impact muscle growth and repair. Myostatin is a protein that inhibits and restricts muscle growth and development, and researchers hope to be able to create a myostatin inhibitor that can block the activity of myostatin to promote muscle regeneration (32). Overall nemaline myopathy has several promising therapies, but due to the rarity of the disease and cost of research there is still no cure.

Conclusion

To conclude, nemaline myopathy is a rare genetic disorder which affects different muscles, causing defects in mobility and the ability to contract muscles. NM is caused by

mutations in genes like NEB or ACTA1, which are the 2 most common mutations and also the most significant mutations related to the disorder (3). These proteins are crucial for muscle function but the mutation leads to the disorganization of sarcomeres and the formation of nemaline bodies which causes troubles with muscle contraction, overall worsening muscle weakness. The nemaline bodies restrict muscles' ability to contract. Symptoms of nemaline myopathy can vary, ranging from slight muscle weakness to life- threatening respiratory issues that can eventually lead to death (3). There is still no cure for nemaline myopathy yet due to the rarity of the disease, but researchers have future plans with a lot of theoretical upside to them. Current options for nemaline myopathy primarily focus on improving the individual with the disease's quality of life, and examples of this are physical therapy and ventilatory support (26). Future treatments for nemaline myopathy are so far aimed to correct the genetic defects regarding the disease itself. This can be approached through gene therapy including CRISPR-Cas9, which can be helpful to replace the mutated genes responsible for causing nemaline myopathy. Unfortunately for NEB which also happens to be the main causative gene, gene therapy will be very troublesome. The nebulin protein is one of the biggest proteins in the human body and because of this when gene therapy is in play and the corrected version of a gene is getting delivered to a cell, due to the size of the protein it's impossible to deliver the whole gene. Researchers are trying to get around this by replacing the protein with short miniature proteins that they hope will make a small difference.

To conclude, nemaline myopathy is a very harmful and rare muscular disorder that affects 1:50,000 individuals from birth. Although there are currently no cures for it, scientists and researchers are working hard to find a solution using the new technologies they have in place today whilst minimizing cost-spending on a rare disorder like nemaline myopathy.

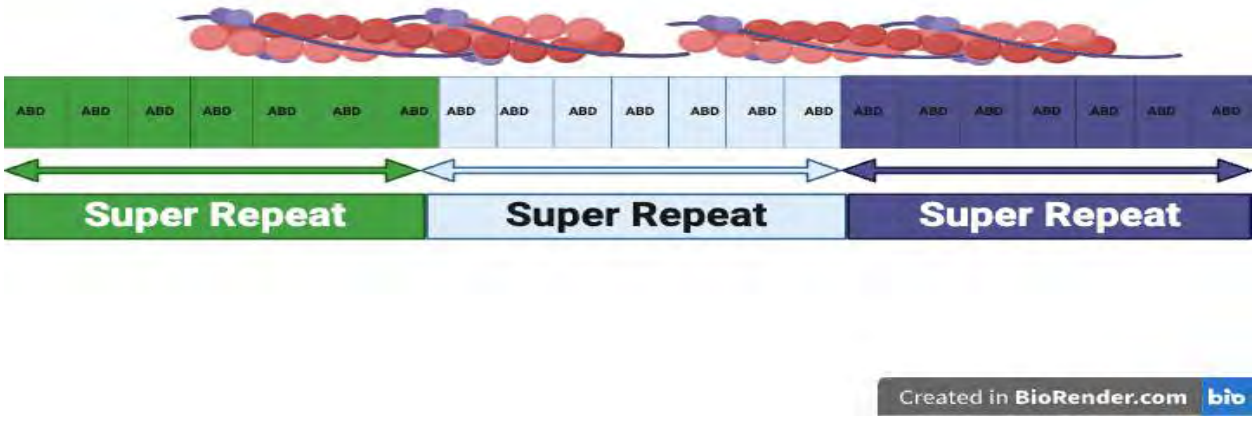


Figure 1. Super repeats in the NEB gene of nemaline myopathy. This figure highlights the structural pattern of nebulin, showing 3 super repeats with 7 simple repeats per super repeat. Mutations affecting the super repeats disrupt its ability to bind to actin (protein showed in red), leading to structural issues within the muscle.

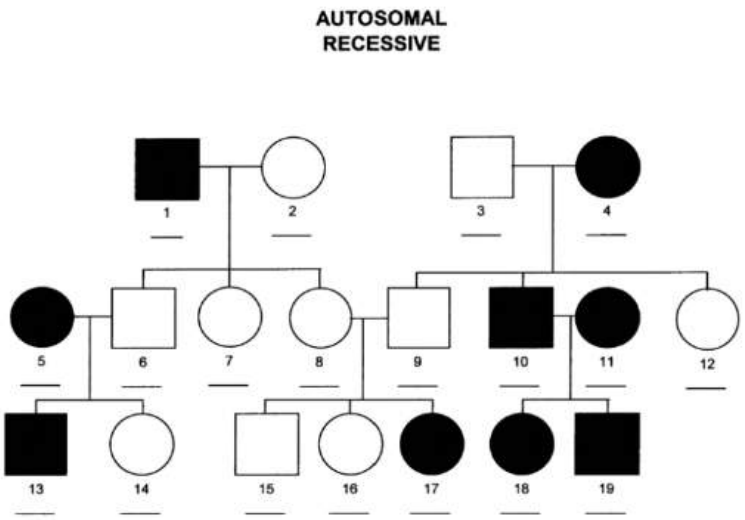


Figure 2. Pedigree showing autosomal recessive inheritance in nemaline myopathy. This pedigree shows autosomal inheritance within nemaline myopathy displaying that affected individuals carry 2 copies of the mutated gene, one from each parent (carrier). Image taken from Michigan genetics research center.

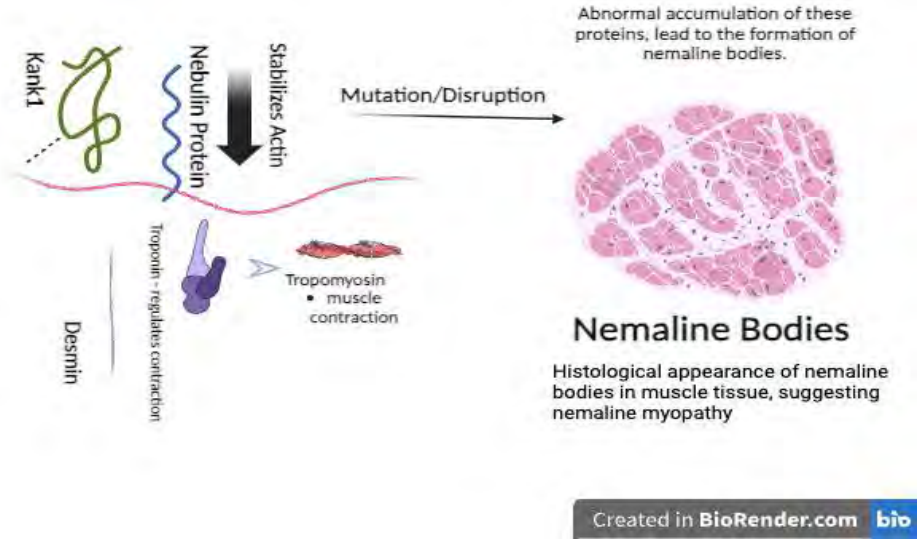


Figure 3. Diagram showing how proteins are affected regarding the formation of nemaline bodies in nemaline myopathy. This diagram shows the impact of mutations on the proteins Kank1, Nebulin, desmin, and troponin. These mutations lead to the accumulation of nemaline bodies (shown in the histology image) which disorder the regular structure of muscle fibers. These interferences lead to problems with muscle contraction and an overall weakness of the muscle.

Works Cited

Muscular Dystrophy UK [Internet]. [cited 2024 Aug 17]. Nemaline myopathy. Available from: https://www.musculardystrophyuk.org/conditions/a-z/nemaline-myopathy/

Sewry CA, Laitila JM, Wallgren-Pettersson C. Nemaline myopathies: a current view. J Muscle Res Cell Motil. 2019;40(2):111–26.

Nemaline Myopathy - Symptoms, Causes, Treatment | NORD [Internet]. [cited 2024 Jul 30]. Available from: https://rarediseases.org/rare-diseases/nemaline-myopathy/

Nemaline myopathy: MedlinePlus Genetics [Internet]. [cited 2024 Aug 17]. Available from: https://medlineplus.gov/genetics/condition/nemaline-myopathy/

Akuamoah-Boateng G, Stetson RC, Kaemingk BD, Bieber DA, Brumbaugh JE. Profound Hypotonia and Respiratory Failure due to Suspected Nemaline Myopathy in a Preterm Infant. AJP Rep. 2021 Apr;11(2):e91–4.

Cleveland Clinic [Internet]. [cited 2024 Aug 18]. Nemaline Myopathy, Nemaline Body Disease: Symptoms, Causes & Treatment. Available from: https://my.clevelandclinic.org/health/diseases/24180-nemaline-myopathy

Yeşilbaş O, Şevketoğlu E, Kıhtır HS, Ersoy M, Petmezci MT, Akkuş CH, et al. A rare structural myopathy: Nemaline myopathy. Turk Arch Pediatr Pediatri Arş. 2019 Mar 1;54(1):49–52.

NEB nebulin [Homo sapiens (human)] - Gene - NCBI [Internet]. [cited 2024 Sep 7]. Available from: https://www.ncbi.nlm.nih.gov/gene/4703

Cleveland Clinic [Internet]. [cited 2024 Sep 7]. Muscle Atrophy: Causes, Symptoms & Treatment. Available from: https://my.clevelandclinic.org/health/diseases/22310-muscle-atrophy

NEB gene: MedlinePlus Genetics [Internet]. [cited 2024 Sep 7]. Available from: https://medlineplus.gov/genetics/gene/neb/

Li F, Buck D, De Winter J, Kolb J, Meng H, Birch C, et al. Nebulin deficiency in adult muscle causes sarcomere defects and muscle-type-dependent changes in trophicity: novel insights in nemaline myopathy. Hum Mol Genet. 2015 Sep 15;24(18):5219–33.

ACTA1 gene: MedlinePlus Genetics [Internet]. [cited 2024 Sep 7]. Available from: https://medlineplus.gov/genetics/gene/acta1/

Alpha Actin - an overview | ScienceDirect Topics [Internet]. [cited 2024 Sep 7]. Available from: https://www.sciencedirect.com/topics/medicine-and-dentistry/alpha-actin

PubChem. ACTA1 - actin alpha 1, skeletal muscle (human) [Internet]. [cited 2024 Sep 7]. Available from: https://pubchem.ncbi.nlm.nih.gov/gene/ACTA1/human

Lambert MR, Gussoni E. Tropomyosin 3 (TPM3) function in skeletal muscle and in myopathy. Skelet Muscle. 2023 Nov 7;13:18.

Arts WF, Bethlem J, Dingemans KP, Eriksson AW. Investigations on the inheritance of nemaline myopathy. Arch Neurol. 1978 Feb;35(2):72–7.

Laitila J, Lehtonen J, Lehtokari VL, Sagath L, Wallgren-Pettersson C, Grönholm M, et al. A nebulin super-repeat panel reveals stronger actin binding toward the ends of the super-repeat region. Muscle Nerve. 2019 Jan;59(1):116–21

Gohlke J, Tonino P, Lindqvist J, Smith JE, Granzier H. The number of Z-repeats and super-repeats in nebulin greatly varies across vertebrates and scales with animal size. J Gen Physiol. 2020 Dec 18;153(3):e202012783.

Szikora S, Görög P, Mihály J. The Mechanisms of Thin Filament Assembly and Length Regulation in Muscles. Int J Mol Sci. 2022 May 10;23(10):5306.

Ilkovski B, Cooper ST, Nowak K, Ryan MM, Yang N, Schnell C, et al. Nemaline Myopathy Caused by Mutations in the Muscle α-Skeletal-Actin Gene. Am J Hum Genet. 2001 Jun;68(6):1333–43.

Labasse C, Brochier G, Taratuto AL, Cadot B, Rendu J, Monges S, et al. Severe ACTA1-related nemaline myopathy: intranuclear rods, cytoplasmic bodies, and enlarged perinuclear space as characteristic pathological features on muscle biopsies. Acta Neuropathol Commun. 2022 Jul 9;10:101.

Tinklenberg JA, Slick RA, Sutton J, Zhang L, Meng H, Beatka MJ, et al. Different Mouse Models of Nemaline Myopathy Harboring Acta1 Mutations Display Differing Abnormalities Related to Mitochondrial Biology. Am J Pathol. 2023 Oct;193(10):1548–67

Laitila JM, McNamara EL, Wingate CD, Goullee H, Ross JA, Taylor RL, et al. Nebulin nemaline myopathy recapitulated in a compound heterozygous mouse model with both a missense and a nonsense mutation in Neb. Acta Neuropathol Commun. 2020 Feb 17;8(1):18.

Sztal TE, Zhao M, Williams C, Oorschot V, Parslow AC, Giousoh A, et al. Zebrafish models for nemaline myopathy reveal a spectrum of nemaline bodies contributing to reduced muscle function. Acta Neuropathol (Berl). 2015 Sep;130(3):389–406

Laitila J, Wallgren-Pettersson C. Recent advances in nemaline myopathy. Neuromuscul Disord. 2021 Oct 1;31(10):955–67.

van Kleef ESB, Langer D, van Engelen BGM, Ottenheijm CAC, Voermans NC, Doorduin J. Inspiratory Muscle Training in Nemaline Myopathy. J Neuromuscul Dis. 10(5):825–34.

Moreno CAM, Artilheiro MC, Fonseca ATQSM, Camelo CG, de Medeiros GC, Sassi FC, et al. Clinical Manifestation of Nebulin-Associated Nemaline Myopathy. Neurol Genet. 2023 Jan 25;9(1):e200056.

kateemmert. A Foundation Building Strength. 2024 [cited 2024 Sep 26]. Building Strength: Promising Developments in Gene Therapy for Nemaline Myopathy. Available from: https://buildingstrength.org/blog/building-strength-promising-developments-in-gene-therapy-for-nemaline-myopathy/

Fisher G, Mackels L, Markati T, Sarkozy A, Ochala J, Jungbluth H, et al. Early clinical and pre-clinical therapy development in Nemaline myopathy. Expert Opin Ther Targets. 2022 Oct 3;26(10):853–67.

Nandy A, Tankisi H, Krøigård AB, Dalager MG, Hvidbjerg MS, Schrøder HD, et al. Sporadic late onset nemaline myopathy with concurrent dermatological symptoms responding to immunosuppressive treatment. BMC Neurol. 2023 Jun 16;23:233.

Schoch KM, Miller TM. Antisense oligonucleotides: Translation from mouse models to human neurodegenerative diseases. Neuron. 2017 Jun 21;94(6):1056–70.

Lindqvist J, Granzier H. Pharmacological Inhibition of Myostatin in a Mouse Model of Typical Nemaline Myopathy Increases Muscle Size and Force. Int J Mol Sci. 2023 Oct 12;24(20):15124.

Music Therapy: An Alternative Treatment for Tourette's Syndrome By Ellie Chen

Abstract

Around 1.4 million people in the U.S. are diagnosed with Tourette's Syndrome (TS), also referred to as Tourette Disorder (TD), which is a condition that affects the nervous system. It is typically presented with sudden, repeated motor and vocal movements known as tics. Its symptoms in addition to the tics are diverse, with people often retaining other conditions such as attention deficits, obsessive-compulsive disorder, anxiety and depression, learning disabilities, etc. The symptoms of TS and its associated conditions often lower one's quality of life. Unfortunately, the exact pathophysiology processes behind TS are largely unknown, although it is likely a cause of various complex genetic factors. As a result, a curative measure for this condition has yet to be found. Due to the lack of definitive treatment, part of the effort of studying TS is to find means to alleviate its symptoms. Medications and behavioral therapy have been applied in the past decades. Music therapy, while not thoroughly explored as an alternative to improving symptoms of TS, has shown to be a promising alternative treatment. It has been used to treat similar conditions to TS before, such as Parkinson's Disease, Huntington's Disease, etc. During music therapy, the patient listens or plays various types of instruments and genres, or discusses the meanings of such with a music therapist. Scientists have been able to link the decrease of tics with activities incorporating a substantial amount of focus, such as with athletics or music. There is evidence showing that the areas of the brain affected by TS also had some changes associated with listening to music. These studies have shown that music therapy is a viable option for lessening TS symptoms, and potentially an alternative to medications with some unique superiorities as compared to the medications presently in use: no side-effects, non-invasiveness, be enjoyable, and improvement of overall health.

Intro/Analysis

This paper aims to determine the effectiveness and significance of music therapy as an alternative, noninvasive treatment to Tourette's Syndrome with minimal side effects compared to other forms of treatment and the plausibility of music therapy being implemented in the future as a more widespread, standard form of care. Tourette's Syndrome (TS) is a neurological disorder that is characterized by, and usually causes, rapid and/or repeated involuntary sound and movements – called tics. There are two major groupings, vocal (sounds) and motor (movements). The severity of these tics fluctuate. The symptoms (tics) can be exaggerated by many factors – environmental, external and internal, etc. Studies have shown that anything from fatigue and anxiety to distraction, and even being observed could affect tic rate in those afflicted (NHS, 2017; Royal Children's Hospital Melbourne, n.d.; Hasan, 2020; Cleveland Clinic, n.d.). While TS can be chronic, it usually isn't – it onsets during childhood and tends to recede with age, especially during late teens and early adulthood. TS also occurs more often in males as opposed to females. TS was discovered in the 19th century around 1884-1885, by a French doctor named Gilles de la Tourette, who studied under Jean-Martin Charcot, who is largely

believed to have been the “father of modern neurology” (Lajonchere et al., 1996). While there are no cures, there are a few different treatment options that aim /to reduce the general symptoms and aid in the controlling of tics, such as dopamine-reducing medication, injections, antidepressants, and various types of therapy (Mayo Clinic, n.d.).

The cause of TS is largely unknown. It is believed that the deviation of a few neurotransmitters may contribute to it. In TS patients, the amount of dopamine, a neurotransmitter that aids motor and reward systems in the brain, was found to produce in excess in the striatum (Leisman and Sheldon, 2022). The striatum is one of the places where dopamine is located. It is involved in reward processing and function. It also has other important functions, such as motor (which correlates to TS, i.e. body movement spasms) and executive functions (Bamford and Bamford, 2019). Executive functions refers to everyday, basic tasks, such as planning, following distractions, and self-control (Harvard University, 2019). This signifies importance in this area of the brain and its correlation to TS.

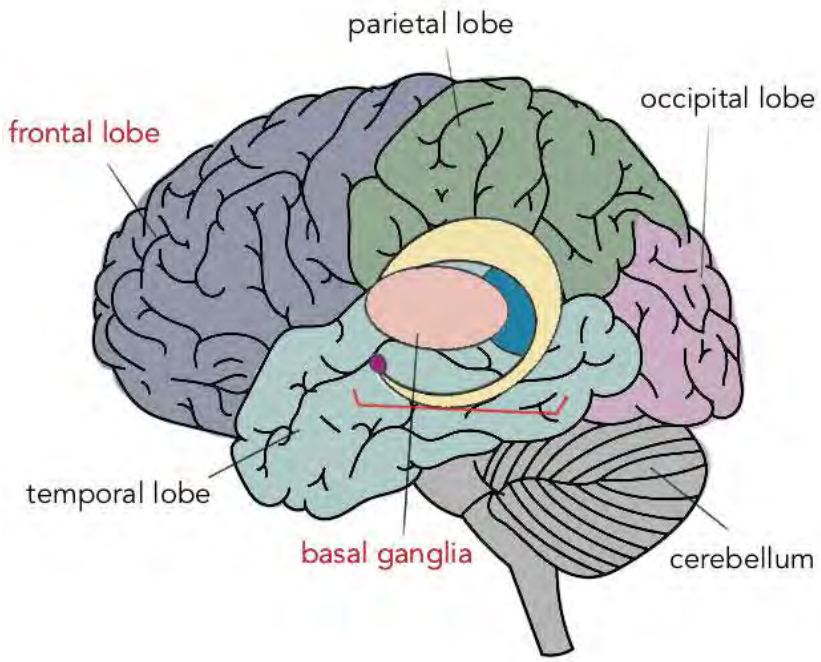


Fig. 1. Color-coded regions of the brain.

An overproduction of dopamine has often been linked to amplification in normal neurological behavior, such as increased aggression or lessened self-control. An overproduction of dopamine has also been associated with neurological disorders, such as schizophrenia and ADHD. The higher dopamine levels appear to trigger the shift of balance between the excitatory and inhibitory neural pathways more toward the excitatory side, producing more tics (Cleveland

Clinic, 2022). Excitation and inhibition refers to the communication of neurons through electrical currents; excitatory pathways help to pass along information between neurons, and inhibitory pathways regulate this passage (Ippolito, 2020).

TS also appears to be associated with inhibitory nerve pathways, such as the levels of gamma-aminobutyric acid – GABA for short. These are the neurotransmitters that calm and lessen nerve cells' activities (Cleveland Clinic, 2022).

GABA is a key neurotransmitter for maintaining balance between excitation and inhibition in the nervous system in order for neural circuits to function properly. GABA specifically is an inhibitory neurotransmitter that is responsible for keeping excitation in check in neurons (Allen et al, 2024). A study has shown a decrease in GABA levels, specifically within the striatum, in those afflicted with TS (Draper et al., 2014). Additionally, those with other neurological disorders similar to TS, such as ADHD or even general anxiety, are found to have a similar reduction in GABA levels (Draper et al., 2014). In TS, both excitatory and inhibitory pathways function abnormally. As the affected child matures, changes in the brain and the rest of the body help to increase GABA levels, which then balance out the neural pathways (Schür et al., 2016). This re-balancing of neurotransmitters is most likely the reason for the lessened symptoms and TS rates in the adult.

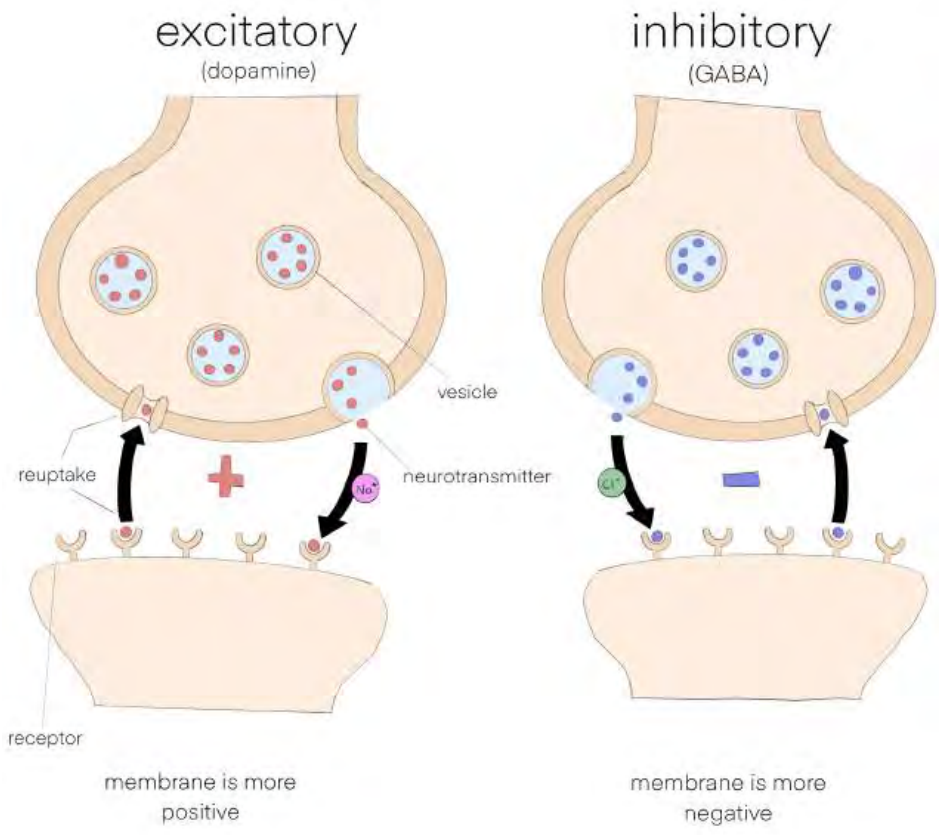


Fig. 2. Function of two amino acids according to the main (predominant) neurotransmitters.

In terms of medication treatment for TS, there are a few types. The first type is the medications helping reduce dopamine levels. Those medications include fluphenazine, haloperidol, risperidone, pimozide, and tetrabenazine (Mayo Clinic, 2018). Unfortunately, almost all of those medications carry a few significant side effects from weight gain to severe depression.

An additional treatment would be botulinum – mainly known as Botox – injections. Although most widely known for its cosmetic use, these injections can actually help to lessen tics as well (Burns, 2016). Botox functions by blocking certain chemical signals connecting nerves to muscle, which effectively weakens/paralyzes muscle tissue and stops it from contracting. Because of this, the muscle is unable to normally respond to the signals from the brain that would usually cause tics, and thus lessens the amount. There are a few different types of botulinum injections, including Botox, which is the most common, as well as Xeomin, Dysport, and Myobloc. However, although usually safe to use when a licensed health provider administers it, this treatment is relatively invasive compared to other common remedies. In addition, there are a few side effects to face, such as infection at the location of the injection. There is also a possibility that the treatment may spread to other parts of the body, which would result in unwanted muscle weakness, vision and/or breathing problems, allergic reactions, etc. (Mayo Clinic, 2018).

Medications used for ADHD are also a possible choice for treatment to help the decreased amount of focus, attention, and concentration as the major symptoms for both conditions. These medications usually involve the ingredients methylphenidate or dextroamphetamine. Besides having a possible allergic reaction, ADHD medications may actually increase the amount of tics for certain people with TS, although there is also contradictory evidence, making it a trade-off between different symptoms (Law and Schachar, 1999). It is well known that all of these treatments have side effects. Music therapy is an alternative treatment that may be able to alleviate the symptoms of TS, yet carries essentially no side-effects. It is non-invasive, enjoyable, and beneficial for overall health.

The main impacted parts of the brain in TS patients are the basal ganglia – of which the striatum is a part of it, and the frontal lobes. These areas are presumably responsible for the symptoms or tics (National Institute of Health, n.d.). As music has the ability to enable people to concentrate, it can be seen as a promising treatment. Music therapy may help to restore disrupted neural networks. In addition, the frontal lobe of the brain is also one of the key brain structures in decision-making and goal-directed behavior (Leckman et al., 2010). As music itself requires a significant amount of concentration and attention, as a means of training, listening to music might help pivot back to the neural balance of excitatory and inhibitory circuits, and accordingly alleviate the disrupting symptoms resulting from TS.

Music Therapy as Treatment

Music in general has a variety of benefits, including improving efficiency of general tasks, enhancing memory, improving anxiety symptoms, and even acting as a pain reliever

(McCrary et al., 2022). Music is also something most people experience, due to it being easily accessible on a daily basis. Given all of the listed benefits of music, music therapy has naturally become a tool to help with a multitude of conditions, from bettering overall lifestyle to being the focus of research that may make it a feasible and effective alternative to treatment of neurological conditions such as TS.

Music therapy encompasses a number of musically-oriented activities, also known as rhythmic auditory stimulation, such as dancing to tunes, singing, instrumental performance, or even writing down one's particular feelings about a piece (American Music Therapy Association, 2005). Although not investigated as widely on TS, there have been studies done on similar movement-related conditions. In particular, a study on the effect of music therapy on Parkinson's Disease, which similarly affects the central nervous system and hinders intentional movement, found that functions in the frontal lobe, which is the driving force of the executive control network, greatly improved (Devlin et al., 2019). The study also examined the effects of music therapy on other movement disorders, including TS. In particular, the study focuses on rhythmic auditory stimulation, also known as RAS. This particular type of music therapy makes use of a metronome or a music cue, adjusted to the patient's movements, typically gait or other 'biologically rhythmic' movements.

Parkinson's Disease is a movement disorder that is characterized by uncontrollable movements, namely tremors; it is linked to neuron damage in the brain, which causes dopamine levels to be lowered (National Institute on Aging, 2022). Other symptoms include mental and behavioral changes, which tend to progress with time, speech impairments, etc. Similar TS, there are no curative treatment options available at present. The study found that there were significant improvements in the patients' balance and gaits, and noted that even within groups with music therapy (Devlin et al., 2019). Additionally, therapeutic singing was another tested intervention in the study. The study found that singing improved both cognitive and motor functions in patients with Parkinson's Disease, as well as general quality of life improvements. Although instrumental studies were conducted as well, they were often sparse and not well-organized; however, many showed some degree of improvements, suggesting that further, more organized studies may be needed.

Another tested disorder from the study was progressive supranuclear palsy (PSP) without definitive treatment methods. It occurs when cells in the brain, specifically cells that control body movements and coordination are damaged or destroyed (Mayo Clinic, 2024). Music therapy is one of the very few treatments that has been proven to relieve symptoms of PSP, and were able to have greater amounts of movement.

Though the effects of music on TS have not been widely studied, there are still a few studies showing the promising benefits of music on TS, which have found even greater success than those in other movement disorders. One study in particular looked into various types of music therapy on a few different movement disorders, such as Parkinson's Disease and Huntington's Disease, as well as Tourette's Syndrome. The types of therapy included rhythmic auditory stimulation, rhythmic speech cueing, auditory perception training, etc. The study found

that TS symptoms were successfully reduced in all tested interventions (Devlin et al., 2019). Because of this, music therapy could be considered a novel treatment without the side effects or invasiveness of other forms.

Conclusion

Music therapy in most forms, especially during musical performance, is able to positively impact the symptoms of Tourette's Syndrome. While TS is often associated with decreased gamma-aminobutyric acid levels, music therapy may help increase GABA levels in key places in the brain, which subsequently restores the balance of neurotransmitters in the brain. TS also decreases with goal-directed behavior or increased focus and concentration levels, all of which is in the nature of musical therapy and training. Comorbidities of TS have also seen positive results with exposure to musical therapy. Thus, musical therapy might be a viable form of treatment for those diagnosed with TS. However, while the vast majority of studies point to improvements, there is still some conflicting evidence and basis behind the results of musical intervention and its connection to brain function. Much of this controversy could be alleviated with technological advancements allowing for better understanding of the brain's anatomy, functions, and in the case of TS, interaction between dopamine and GABA within the brain. There are also various types of music and ways to express it, and because of this, certain types of music therapy and intervention may not work for some TS patients. This form of treatment may need to be individualized. Due to the vast amount of music genres and subgenres to choose from, one is able to fine-tune their form of musical intervention to their preference. Additionally, musical intervention appears to be slightly less effective for TS patients in certain situations, namely, that symptoms tend to return after therapy. Given the side effects in other treatments and the general benefits of music in life, it is an acceptable compromise much lower than that of other remedies. The fluidity and the plethora of musical forms in the world allows for a research window to have a better understanding of the disorder while looking for innovative cures and treatments that are sure to be studied further in the future.

Works Cited

Allen, Mary J., et al. "GABA Receptor." *StatPearls*, StatPearls Publishing, 2024. *PubMed*, http://www.ncbi.nlm.nih.gov/books/NBK526124/.

Balancing Act in the Brain: Excitatory and Inhibitory Activity – Max Planck Florida Institute for Neuroscience. https://mpfi.org/balancing-act-in-the-brain-excitatory-and-inhibitory-activity/. Accessed 17 Nov. 2024.

Bamford, Ian J., and Nigel S. Bamford. "The Striatum's Role in Executing Rational and Irrational Economic Behaviors." *The Neuroscientist : A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, vol. 25, no. 5, Oct. 2019, pp. 475–90. *PubMed Central*, https://doi.org/10.1177/1073858418824256.

Burns, Meghan. "Botox - It's Not Just for Wrinkles Anymore." *Tourette Association of America*, 15 Aug. 2016, https://tourette.org/botox/.

Devlin, Kerry, et al. "Music Therapy and Music-Based Interventions for Movement Disorders." *Current Neurology and Neuroscience Reports*, vol. 19, no. 11, 11, Nov. 2019, p. 83. *DOI.org (Crossref)*, https://doi.org/10.1007/s11910-019-1005-0.

"Dopamine: What It Is, Function & Symptoms." *Cleveland Clinic*, https://my.clevelandclinic.org/health/articles/22581-dopamine. Accessed 17 Nov. 2024.

Draper, Amelia, Mary C. Stephenson, et al. "Increased GABA Contributes to Enhanced Control over Motor Excitability in Tourette Syndrome." *Current Biology*, vol. 24, no. 19, 19, Oct. 2014, pp. 2343–47. *PubMed Central*, https://doi.org/10.1016/j.cub.2014.08.038.

Draper, Amelia, Mary C. Stephenson, et al. "Increased GABA Contributes to Enhanced Control over Motor Excitability in Tourette Syndrome." *Current Biology*, vol. 24, no. 19, Oct. 2014, p. 2343. *pmc.ncbi.nlm.nih.gov*, https://doi.org/10.1016/j.cub.2014.08.038.

"Gamma-Aminobutyric Acid (GABA): What It Is, Function & Benefits." *Cleveland Clinic*, https://my.clevelandclinic.org/health/articles/22857-gamma-aminobutyric-acid-gaba. Accessed 27 Mar. 2024.

Kids Health Info : Tics. https://www.rch.org.au/kidsinfo/fact_sheets/Tics/. Accessed 13 Nov. 2024.

Lajonchere, C., et al. "Gilles de La Tourette and the Discovery of Tourette Syndrome. Includes a Translation of His 1884 Article." *Archives of Neurology*, vol. 53, no. 6, June 1996, pp. 567–74. *PubMed*, https://doi.org/10.1001/archneur.1996.00550060111024.

Law, Samuel F., and Russell J. Schachar. "Do Typical Clinical Doses of Methylphenidate Cause Tics in Children Treated for Attention-Deficit Hyperactivity Disorder?" *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 38, no. 8, Aug. 1999, pp. 944–51. *DOI.org (Crossref)*, https://doi.org/10.1097/00004583-199908000-00009.

Leckman, James F., et al. "Neurobiological Substrates of Tourette's Disorder." *Journal of Child and Adolescent Psychopharmacology*, vol. 20, no. 4, Aug. 2010, p. 237. *pmc.ncbi.nlm.nih.gov*, https://doi.org/10.1089/cap.2009.0118.

Leisman, Gerry, and Dana Sheldon. "Tics and Emotions." *Brain Sciences*, vol. 12, no. 2, Feb.

2022, p. 242. *pmc.ncbi.nlm.nih.gov*, https://doi.org/10.3390/brainsci12020242.
"Parkinson's Disease: Causes, Symptoms, and Treatments." *National Institute on Aging*, 14 Apr. 2022, https://www.nia.nih.gov/health/parkinsons-disease/parkinsons-disease-causes-symptoms-and-treatments.
"Progressive Supranuclear Palsy - Symptoms and Causes." *Mayo Clinic*, https://www.mayoclinic.org/diseases-conditions/progressive-supranuclear-palsy/symptoms-causes/syc-20355659. Accessed 17 Nov. 2024.
Schür, Remmelt R., et al. "Brain GABA Levels across Psychiatric Disorders: A Systematic Literature Review and Meta-Analysis of (1) H-MRS Studies." *Human Brain Mapping*, vol. 37, no. 9, Sept. 2016, pp. 3337–52. *PubMed*, https://doi.org/10.1002/hbm.23244.
"Tics." *Nhs.Uk*, 23 Oct. 2017, https://www.nhs.uk/conditions/tics/.
Tics (for Teens). https://kidshealth.org/en/teens/tics.html. Accessed 13 Nov. 2024.
Tourette Syndrome | National Institute of Neurological Disorders and Stroke. https://www.ninds.nih.gov/health-information/disorders/tourette-syndrome. Accessed 17 Nov. 2024.
Tourette Syndrome - Diagnosis and Treatment - Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/tourette-syndrome/diagnosis-treatment/drc-20350470. Accessed 1 Sept. 2024.
"What Are Tics and Tic Disorder?" *Cleveland Clinic*, https://my.clevelandclinic.org/health/diseases/tics-and-tic-disorders. Accessed 13 Nov. 2024.
"What Is Executive Function? How Executive Functioning Skills Affect Early Development." *Center on the Developing Child at Harvard University*, https://developingchild.harvard.edu/resources/what-is-executive-function-and-how-does-it-relate-to-child-development/. Accessed 17 Sept. 2024.
What Is Music Therapy? | What Is Music Therapy? | American Music Therapy Association (AMTA). https://www.musictherapy.org/. Accessed 17 Nov. 2024.

Black Holes in Our Universe By Arjun Konijeti

Abstract

A black hole is a singularity with such immense gravity that it can suck in objects, and bring them to the event horizon where they will be frozen in time. This paper explains several variants of black holes, including stellar mass black holes, intermediate mass black holes, supermassive black holes, and primordial black holes. When an object approaches a black hole, it is sucked into the accretion disk, where it spins and emits waves ranging from radio waves to x-rays, and afterwards falls into the event horizon.To observe black holes, we can use particle jets of gamma rays that come from the accretion disk of a black hole in order to locate the black hole. We are also able to find unusual readings of time dilations or movement of stars to trace a black hole's location. We can also detect when two black holes are colliding, or when a black hole is sucking in another object.

Introduction

As an object approaches a black hole, an observer will see the object begin to slow down in time. With the use of new, advanced technology, black holes from far away, big or small, can be seen and researched by astrophysicists around the world. This paper describes the unusual behavior of an object approaching a black hole, and how it can behave differently depending on the type of black hole. These types range from simple, stellar mass black holes to the supermassive black holes found at the center of galaxies. This paper also describes ways in which we can observe this behavior with the use of high-power technology.

Different Types of Black Holes

In our universe, there are many different types of black holes. Black holes can range from only tens of times the mass of our sun to thousands upon thousands times that mass (NASA, 2024). A black hole's gravity depends on its mass. A larger black hole will be much stronger, and can even hold galaxies. A smaller black hole, on the other hand, will only be able to hold relatively small amounts of matter. The different types of black holes consist of stellar mass black holes, supermassive black holes, intermediate black holes, and primordial black holes.

Stellar mass black holes are the most commonly found in our universe. They are the most commonly known, being formed when a star of approximately twenty times the mass of the sun, at the very least, collapses onto itself. When a star collapses this way, it explodes into a supernova. A supernova occurs when a star loses its energy, and then condenses and breaks down with such a high density that it forms a singularity. A singularity is a point in space that is infinitely dense, causing an object to have such a strong gravity that it forms a black hole. There are millions of these throughout the universe, and possibly in the Milky Way alone.

When a star of this size dies, it may also turn into a neutron star. Neutron stars are approximately the size of a small city on Earth, but are so dense that even a spoonful of the matter in one could weigh a billion tons. Neutron stars spin incredibly fast, and create pulsars,

which are beams of electromagnetic radiation that form on both sides of the star. Neutron stars are so dense that they have mountains only centimeters tall, at the most. When a neutron star dies, it can then become a stellar mass black hole.

Supermassive black holes are located at the center of most galaxies. They are the largest black holes that have been observed in our universe, ranging from tens of thousands to even billions of times the mass of our sun. The Milky Way's supermassive black hole is called Sagittarius A, which was discovered in the mid-1990's. Today, it is unclear how these black holes came to be. Understanding black holes play a crucial role in understanding galaxies. Galaxies revolve around supermassive black holes, and by understanding the supermassive black hole itself, we can form theories of galaxy formation in addition to helping understand how galaxies truly function.

There is a massive gap in the size range between stellar mass black holes and supermassive black holes. The type of black hole that fills this gap is called an intermediate black hole. Intermediate black holes, however, may not even exist. There is no concrete evidence that they even exist, or have ever existed. Possible candidates are scattered throughout the universe, but none of them are confirmed.

The idea of an intermediate mass black hole was created due to this immense gap in size. People believed that there must be some size of a black hole that could exist within this gap, leading to the creation of this new category. It would be unclear as to how an intermediate mass black hole would have formed, as most stars aren't large enough to condense into a black hole large enough to be considered an intermediate mass black hole. Because the formation of supermassive black holes is unknown, intermediate mass black holes, if they exist, could have possibly been formed the same way.

The last type of black hole is called a primordial black hole. These black holes were most likely formed in the first few moments after the birth of the universe. These black holes were formed by incredibly dense material, but as the universe expanded, there were no longer conditions for making primordial black holes. Like the intermediate mass black hole, it is unclear if these ever existed, or if they still exist today. This is because most of the primordial black holes with less mass would have easily evaporated due to the extreme temperatures of the universe in its first stages. This means that only the largest would have survived, although it is highly unlikely due to the larger ones also evaporating or losing energy later on.

How Do Objects Behave Around Black Holes?

The effects of a black hole on other objects depend primarily on its mass. A larger black hole will have a stronger gravitational pull, and a smaller black hole will be weaker. Larger black holes, such as a supermassive black hole, could hold entire galaxies. Stellar mass black holes, which are far smaller, may only be able to suck in smaller stars and planets. The larger the black hole, the stronger the gravity.

Black holes are extremely dense. This density is the reason that a black hole's gravity is so strong, and can attract even light. The density of a black hole is the result of a singularity. A

singularity is a single point in space that is infinitely small and infinitely dense. As the size of the point approaches the limit of zero and the density approaches the limit of infinity, the singularity grows in power.

When an object gets too close to a black hole, it will begin to undergo changes in shape and time. First, the object will experience tidal forces (NASA, 2023). Tidal forces are similar to high and low tides, because of their difference in gravitational pull from one side to the other.

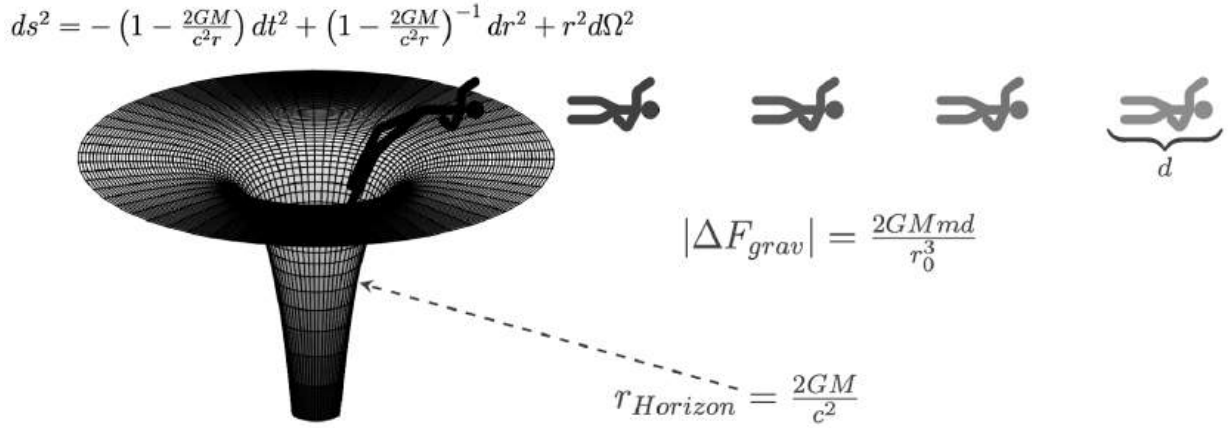


Figure 1: This diagram depicts the physical change of a human body as it approaches the event horizon of a black hole, otherwise known as spaghettification. It also includes the Schwarzschild radius, which is the radius of the event horizon ($\frac{2GM}{c^2}$) and uses it along with other equations to mathematically explain the change in gravitational force on the human (SciTechDaily, 2021).

Because the side of the object closer to the black hole will experience a much stronger gravitational pull than the other side that is further from the black hole, it will be stretched. This difference increases as the object moves closer to the black hole, and results in the object being pulled apart into a string. This process is called spaghettification (see Figure 1).

Tidal forces may also result in “pancake detonation”. This is where the object is flattened rather than pulled apart, and it mostly occurs in supermassive black holes. The object is compressed by tidal forces, and then this is followed by an explosive release of thermonuclear energy.

As mentioned earlier, the object will undergo changes in time. Time will begin to slow down, as the object is approaching the speed of light. When an object approaches the speed of light, time slows down asymptotically until the object is infinitely close to the limit. Time dilation occurs because massive objects, such as black holes, stretch the fabric of spacetime. This can be noticed here on Earth, as well. For example, people on the surface of Earth are affected by gravity more than people on the International Space Station (ISS), meaning that clocks will tick more slowly down on Earth than in space. When an object is infinitely close to the speed of light, it will slow down until completely frozen.

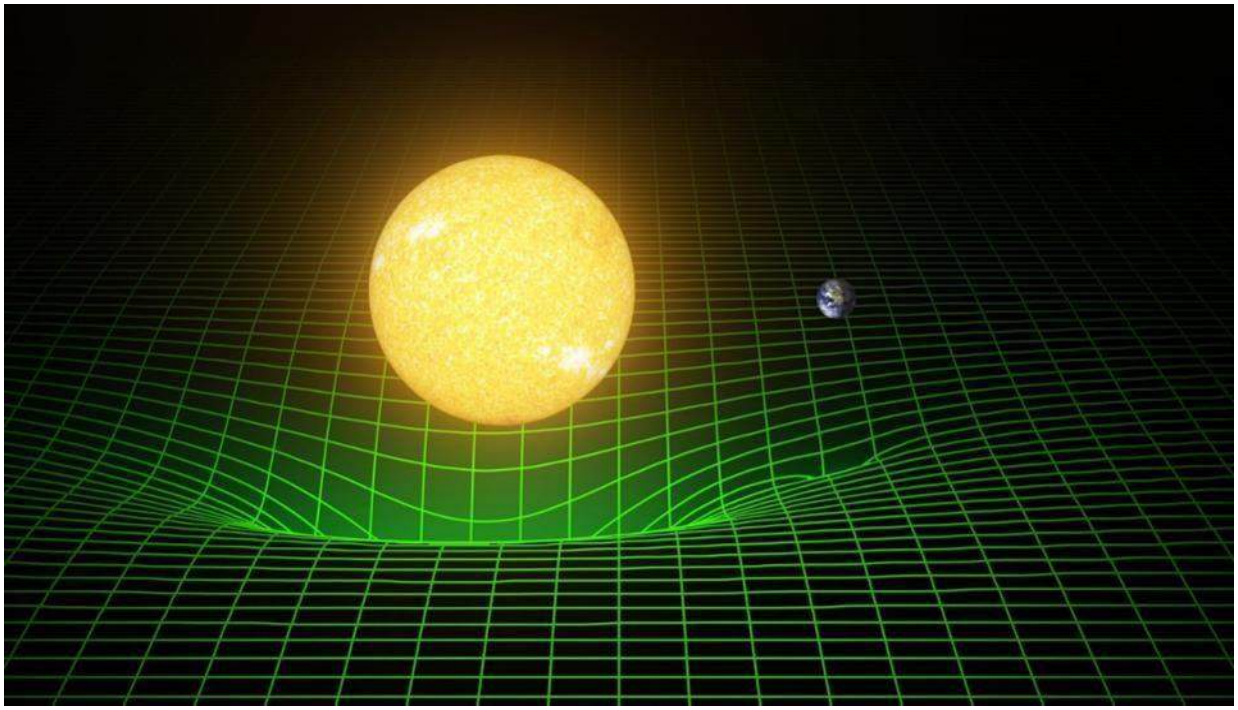


Figure 2: This image represents the curvature in a two-dimensional visualization of space time, being affected by the gravity of the sun and the gravity of Earth. It compares the two gravitational forces, as well, showing that the gravitational force of the sun is much greater than that of Earth (LIGO, 2016).

The fabric of spacetime refers to the influence of gravitational pull of large masses on time. The more gravity a celestial object has, the more it affects time. This can be visually represented by placing a sphere directly above an infinitely large plane, and then curving the plane spherically to bend it along the curve of the celestial object. This can be seen in Figure 2.

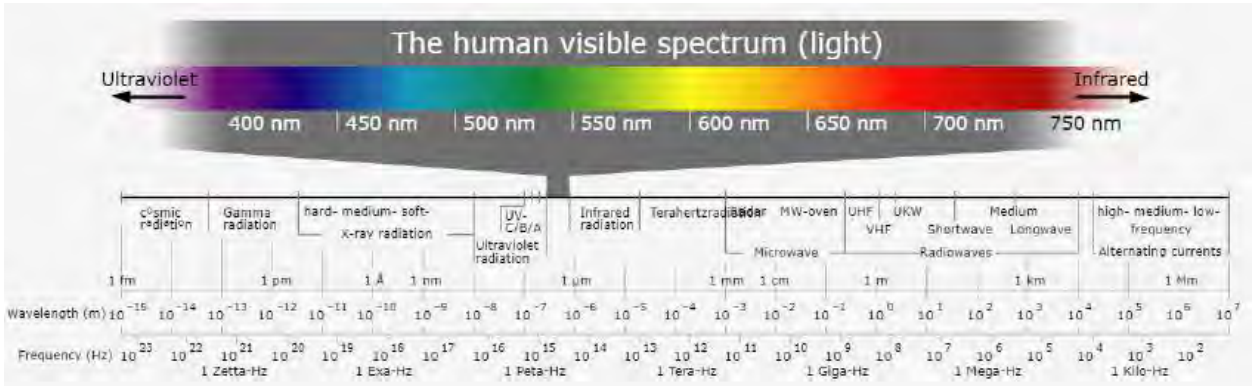


Figure 3: This diagram represents the wavelengths of different forms of energy in the form of radiation. It ranges from radio waves, which have the longest wavelength, to gamma rays, which have the shortest wavelength (PennState, 2012).

As an object moves closer to a black hole, it will fall into the accretion disk (see Figure 4). The accretion disk is a spherical disk that moves around a black hole, and is completely made

up of hot, swirling matter. The matter in the accretion disk is flattened, and will continue to fall until it reaches the event horizon. The event horizon is the “point of no return”, where an object is lost forever into the black hole. Upon entry into the accretion disk, the matter will emit radio waves, and then accelerate to x-rays as it gets closer to the event horizon. Figure 3 shows the frequencies and different types of radiation. Radio waves have a lower frequency of 10^2 Hz to 10^4 Hz, whereas x-rays and gamma rays have a much higher frequency of 10^{18} Hz to 10^{24} Hz. The matter forms x-rays due to friction. This energy can also be redirected into a particle jet, which is emitted in the form of gamma rays.

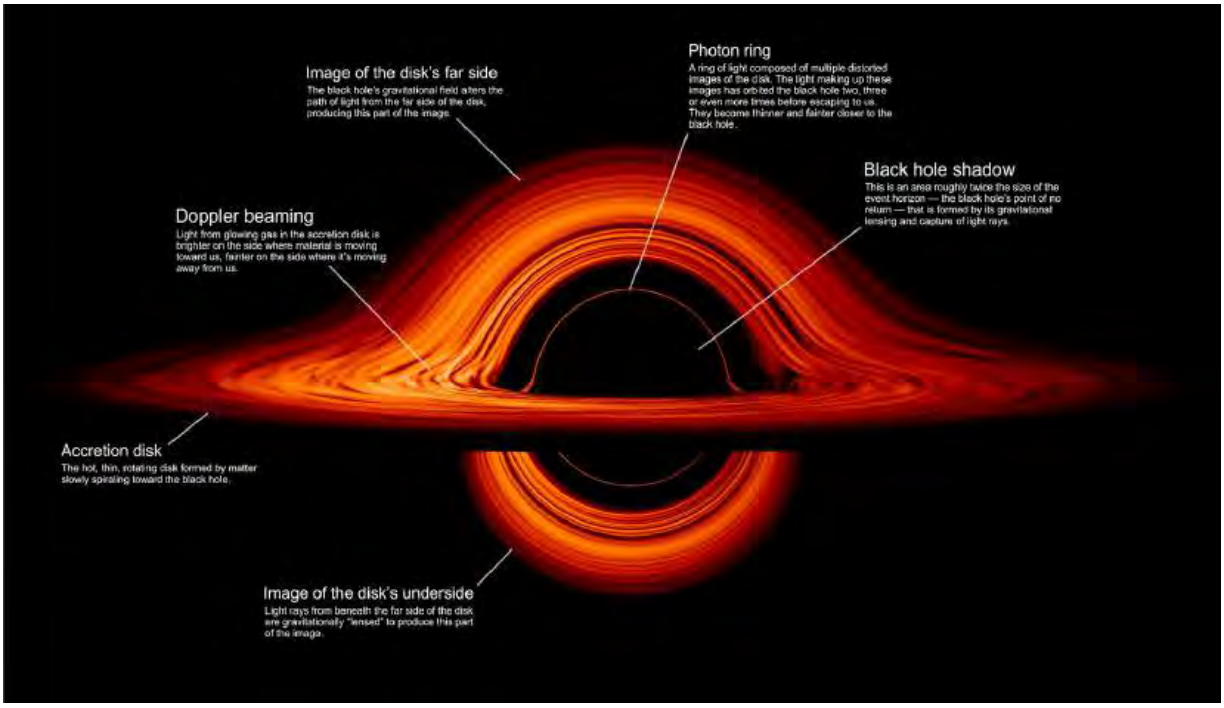


Figure 4: This diagram displays the anatomy of a black hole, including parts such as the accretion disk, photon ring, shadow, and more. It gives a brief and general explanation of each part (SciTechDaily, 2024)

Figure 4 shows the different parts of a black hole. Although the accretion disk is spherical, this image is how our technology on Earth captures the light emitted from the black hole. Doppler beaming is where the light rotating in our direction appears brighter, whereas the light rotating away from us appears to be more dim. The photon ring is the ring of light on the inside of a black hole, and this light has orbited the black hole several times before escaping and shooting off in different directions.

Objects fall into black holes when they can no longer maintain a stable orbit. A black hole can have orbits surrounding it, as most other large masses do. A supermassive black hole, for example, has an entire galaxy orbiting around it. Any object, including light, can be sucked into a black hole. When an object reaches the event horizon, it is unclear as to what exactly happens.

Because even light is unable to escape the gravitational force of a black hole, it is pulled into the event horizon. This means that, because the light is trapped, it cannot reach Earth. This is why humans are unable to observe the center of a black hole. We are, however, able to observe the accretion disk, where some light escapes the black hole and shoots off in all directions.

The object orbiting the black hole can be of nearly any size. It can be a simple, small molecule or polymer to even a large star or planet. A smaller object will simply fall into the accretion disk along with the rest of the matter. A star, on the other hand, will break down into bits and make a noticeable impact on the accretion disk. Eventually, when a black hole consumes more and more matter, it will grow, and therefore have a stronger gravitational pull.

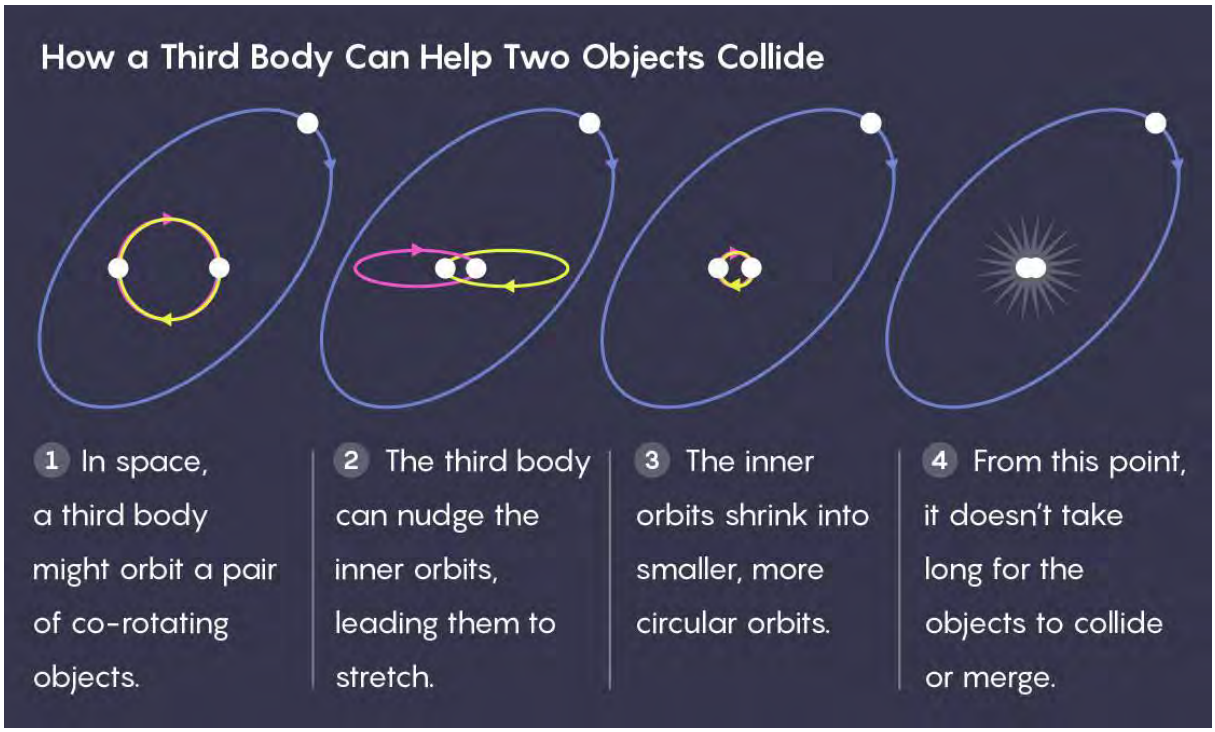


Figure 5: This is a step-by-step demonstration of two celestial objects rotating around each other and inevitably colliding with the additional force of a third celestial object (QuantaMagazine, 2019).

When two masses are orbiting around each other, they begin to attract. They orbit each other, but it is highly unlikely that they collide. For them to collide, there would need to be a third object to throw off the stable gravity between the first two objects. This can occur when the third object is orbiting the first two objects at an angle. If the angle is large enough, then that could set up just the right orientation so that the two objects crash (see Figure 5).

The two masses would need to be of similar size in order to maintain a stable orbit around each other. If one of the black holes is significantly larger than the other, then they would easily collide due to the unequal force of gravity and the immensely strong gravitational force.

This also applies to black holes. Two black holes are able to orbit around each other when spaced apart appropriately. This would mainly occur between stellar mass black holes, but it may

also occur when two galaxies are orbiting around each other. If the two black holes are accelerating toward one another, or simply unstable in orbit, they will collide.

We can even notice this within our own galaxy, the Milky Way. The Milky Way galaxy will collide with the Andromeda galaxy in approximately 4.5 billion years, and the galaxies will collide to form an even larger one.

When two black holes maintain a stable orbit around each other, they will not collide until the influence of a third object throws them off course. The third black hole will orbit the first two at an angle, forcing the two black holes out of the stable orbit (QuantaMagazine, 2019).

Observation of Black Holes and Dwarf Galaxies

There are several ways in which humans are able to detect black holes from millions of light years away. These methods use very high power technology that has only existed since the mid-1900's, and is still developing today. We are able to detect black holes through the use of particle jets, ripples in space time, the accretion disk, and movement of stars (NSF, 2019).

Particle jets are emitted from the matter in the accretion disk. When the matter in an accretion disk is in an unstable orbit around the black hole, it can be shot out in the form of gamma radiation. These particle jets are extremely powerful—so much so that if they manage to come near Earth, we can detect them using new, advanced technology. When we find the location of the jets, we can use the strength of the radiation to determine the location of the black hole.

We can also trace the location of a black hole by finding ripples in space time. When two black holes collide, their immense gravitational force can cause a ripple in the fabric of space-time. These ripples can make their way to Earth, but by the time they do, they are too weak to be noticed by a human without any other tools or devices. Using modern-day technology, we can detect the ripples in space time and then use that information to find the black holes.

Another, perhaps more obvious way of finding black holes is by seeing them with a telescope. We can find the accretion disk of a black hole, and use telescopes, such as the James Webb, to find the black hole. After using other methods to find the exact location of a black hole, we can take images of the black hole using telescopes.

Finally, we can observe a black hole by noticing a change in the movement of stars. Usually, stars will move in a relatively uniform way when orbiting around the supermassive black hole at the center of the galaxy. However, this is not always the case. When there is a unique change in the movement of a star, then there must be another celestial object to act upon it. The object would have to be big enough, or have enough gravitational pull, in order to alter the movement of the stars.

The most likely explanation to this would be a black hole, and using math and physics, we can trace the location of the black hole. The black hole would also have to be acting on multiple celestial objects, so we can check other stars in order to confirm the existence of a larger object.

We can also use the observations to discover more about black holes. The observations can be used in order to make inferences about a black hole's creation, or possibly the creation of the universe. We can make conclusions and explanations of unusual findings, and then connect those findings to other observations, which will provide further insight to the laws that govern our universe.

One example of this is when astronomers found supermassive black holes in relatively small galaxies. Originally, it was thought that supermassive black holes could not exist in these small galaxies, however, it was later discovered that they could be after further investigation. Supermassive black holes are more likely to hold larger, more complex galaxies than simpler ones such as these, being the reason it was not believed that they could exist in dwarf galaxies.

Dwarf galaxies are relatively uncharted territory. This is because astronomers do more research and investigation in larger galaxies, which have much more to offer than small galaxies. Upon finding that there could be supermassive black holes in smaller galaxies, astronomers could use the information to make more conclusions about the first, earliest black holes in the universe. It could also help to give an explanation about the distribution of matter in the universe, or even within a single galaxy.

In dwarf galaxies, there is minimal gravitational force. This lack of gravity is the reason that dwarf galaxies are so small, and why they cannot support a large number of celestial objects. In addition to this, dwarf galaxies are not as round and circular as other, larger galaxies. Their orbits do not align as neatly, thus making it more difficult to do research on them.

In general, larger galaxies are the product of much smaller galaxies merging together to form an even bigger one. This means that galaxies such as our own Milky Way or the Andromeda galaxy were formed from dwarf galaxies. Dwarf galaxies, therefore, were likely the earlier galaxies of the universe that did not collide with other dwarf galaxies.

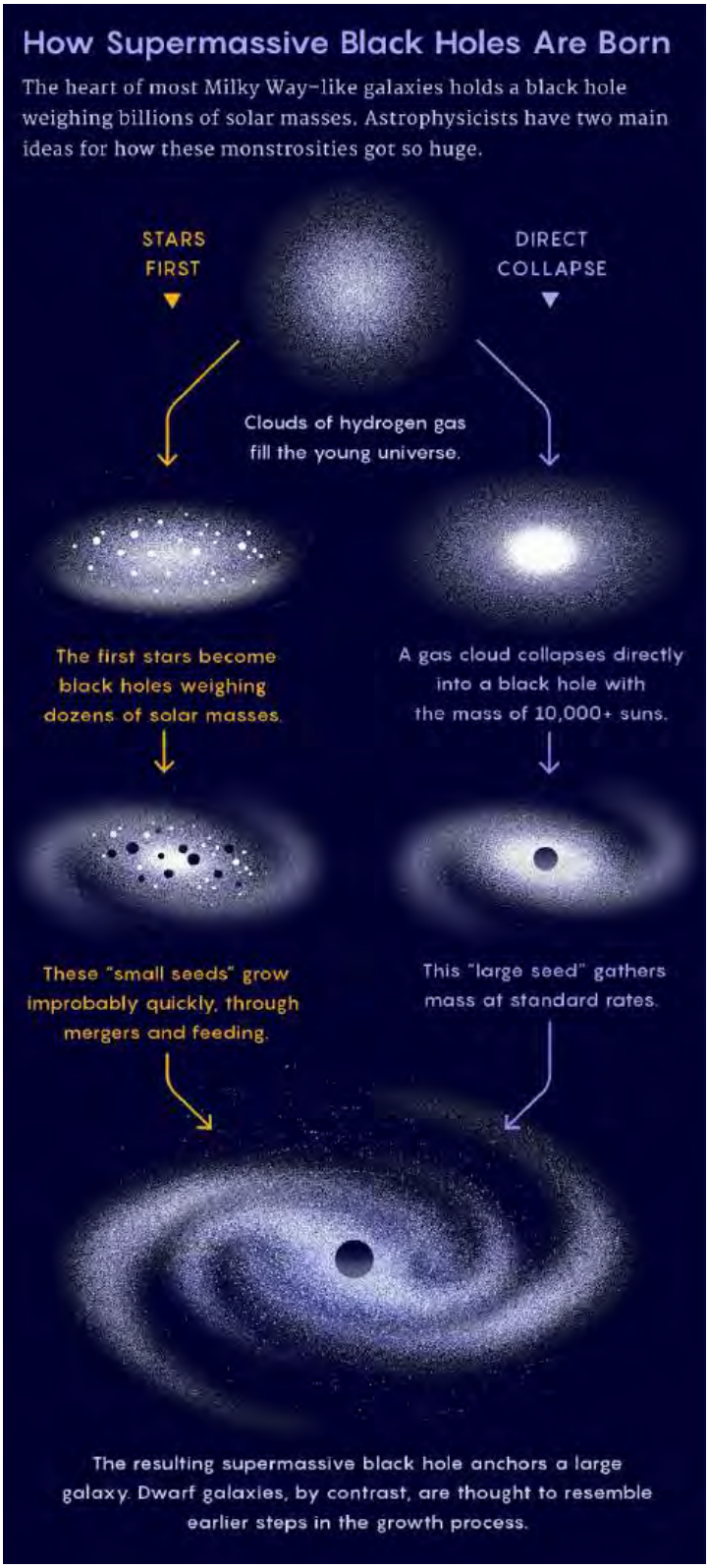


Figure 6: This shows how hydrogen gas evolves into a supermassive black hole through two different processes. One of them shows the stars forming before the black holes, while the other shows the hydrogen gas directly collapsing into a black hole (QuantaMagazine, 2022).

The same concept can be used with black holes. It is theorized that the creation of a supermassive black hole is from smaller black holes, as shown in Figure 6. It begins with clouds of hydrogen gas (H_2) which gather to form small stars. Eventually, these stars will die and form small black holes, which will then combine to form a supermassive black hole.

However, there is another way that supermassive black holes may have formed. This is from a direct collapse from the hydrogen gas into a black hole. After this collapse, the black hole will grow rapidly and ultimately create the supermassive black hole.

While there are no confirmed guesses as to how supermassive black holes are actually formed, these are the most likely answers. Astrophysicists, along with other physicists or scientists use their observations of celestial objects in addition to theoretical mathematics and physics in order to find a solution to the mysteries of the universe.

Dwarf black holes may not always collide to form a larger black hole. The dwarf black hole will remain as a seedlike, smaller black hole and could form a dwarf galaxy. The dwarf galaxy will be quite difficult to find using telescopes, largely due to the lack of matter surrounding the black hole in the accretion disk.

Another reason that it might be difficult to observe supermassive black holes in dwarf galaxies is due to the fact that they are unable to outshine the stars around them. Many will go undetected for this reason and will not be registered by telescopes.

One way that we can find these supermassive black holes in dwarf galaxies is from the x-rays that are able to permeate through dense gas clouds (QuantaMagazine, 2022). If we are unable to use a telescope to find the dwarf galaxies, then we can resort to other methods of finding any radiation from them. These can allow us to pinpoint the location of the dwarf galaxies.

However, not all dwarf galaxies even have a supermassive black hole. In fact, most of them don't. A supermassive black hole is not required to form dwarf galaxies. This is because the celestial objects in a dwarf galaxy will simply orbit around each other in a circular motion. There are few celestial objects in a dwarf galaxy, meaning that it would not be difficult to sustain this orbit.

Astronomers have found iron atoms in the accretion disk of supermassive black holes in dwarf galaxies that have lost nine electrons due to severe damage. The reason that the astronomers were able to detect this is due to the change in color of the accretion disk surrounding the black hole.

They then argued that the reason the iron atoms had undergone such damage was due to immense heat. The gas surrounding the atoms could have been heated up enough to take this many electrons due to the incredible amount of energy.

A Mathematical Perspective

We have seen many black holes in our universe, from stellar mass black holes within our own Milky Way to supermassive black holes in dwarf galaxies or larger galaxies. We have also

seen black holes that can range from tens of times the mass of the sun to millions of times the mass of the sun.

However, there exist black holes beyond just the ones that we can see with modern technology. These are called mathematical black holes. Previously, we have been using astrophysical black holes, which have a variety of physical factors that make them imperfect.

Mathematical black holes, on the other hand, are very different from astrophysical ones. Mathematical black holes don't have any limiting factors, and are essentially perfect, symmetrical black holes that exist in an infinite plane (QuantaMagazine, 2019). They can be infinitely large, as well. These are completely theoretical, meaning that we cannot find them in our universe.

There are many ways that mathematical black holes are able to help us in the real world. We are able to apply theories onto them, and use the mathematical black holes to test those theories through computer simulations or through mathematical equations.

One example of this is when Stephen Hawking theorized that black holes can evaporate. Because a black hole's surface corresponds to its entropy, Hawking was able to form a connection between thermodynamics and gravity, thus leading him to find that black holes can, in fact, have a temperature and evaporate.

Obviously, this led to several problems. First, a black hole's radiation contains no information, so how could a black hole evaporate? Second, If a black hole collapses and evaporates, what happens to the information inside? Does it disappear, or does it come out of the black hole somehow? In addition, this would be incompatible with quantum theory.

However, it is a purely mathematical situation. The temperature of a black hole is far too small to be observable. Also, black holes won't actually evaporate for hundreds of billions of years, so we can only make predictions for now. Although, solving this would certainly help with our understanding of quantum mechanics along with space and time, which could ultimately help us in finding a theory of quantum gravity.

While we cannot do physical experiments, we can do theoretical ones. Theoreticians have done experiments with a single quantum of information in order to find an explanation, and they found that black holes are able to juggle information very efficiently.

Another way that we can experiment with mathematical black holes is through the use of different dimensions of space and several shapes of black holes. Theoretically, if we expand the dimensions beyond the third or fourth dimension, we can find other answers in other dimensions. This could allow us to find many other explanations and pathways to answers when encountered with complex problems.

Theorists can also study black holes in a universe with a negative cosmological constant, where black holes can be infinitely extended planes. This theoretical situation could also help in other fields, such as describing collections of particles that interact strongly with one another or in quantum physics.

Mathematical black holes can be used to study complex quantum systems, and they allow theoretical physicists to find links between thermodynamics, gravity, quantum information, and

condensed matter physics. They can help to develop quantum gravity theories or could even help to predict behavior of certain fluids. They have many uses and can be applied in a variety of theoretical scenarios, helping to prove theories or solve problems.

Conclusion

Black holes are complex celestial objects that form from a star collapsing at the end of its life. It consists of a singularity, event horizon, accretion disk, and more. It has such immense gravity that it can pull anything, even light, into its accretion disk and is able to warp space and time.

The four types of black holes are stellar mass black holes, intermediate mass black holes, supermassive black holes, and primordial black holes. Stellar mass black holes are simple, and form at the end of a star's life. They exist throughout galaxies and the universe. Intermediate mass black holes are unconfirmed to even exist, and are larger than stellar mass black holes but smaller than supermassive black holes and were initially theorized in order to fill the size gap between the two. Supermassive black holes are found at the center of galaxies, and it is unclear as to how they formed. Primordial black holes are also unconfirmed, and are thought to have existed at the very beginning of the universe.

A black hole affects objects around it depending on its mass. The larger a black hole is, the more force it has on an object. Black holes are incredibly dense, being the reason that they can attract objects with such a strong force. When an object comes close to a black hole, it will start to experience tidal forces and spaghettification. Tidal forces are where one part of the object is experiencing a stronger gravitational force than the other side due to it being closer, causing spaghettification. Spaghettification is where an object is stretched from tidal forces. However, the object can also become compressed by tidal forces and result in "pancake detonation," where the object will explode and release thermonuclear energy. The object will also begin to undergo changes in time because of the theory of general relativity. As an object goes faster, time slows down. At the event horizon, the object will become completely frozen. The larger an object is, the more it warps space time and the more it will slow down time. When an object moves closer to a black hole, it will fall into the accretion disk, and will eventually turn into radiation that is bright enough to be seen through a telescope. When an object is unable to maintain a stable orbit around the black hole, it will fall into the event horizon. If two objects are orbiting each other, they are stable and will not collide. However, with the help of a third celestial object, the two objects can collide if the third object is orbiting the first two at a large enough angle. This will cause the two objects to collide.

We can observe black holes through particle jets, ripples in space time, the accretion disk, and movement of stars. Particle jets can shoot off of the accretion disk in the form of gamma radiation, which we can detect using technology and find the location of the black hole. We can also find ripples in space time that are often formed from two black holes colliding. The accretion disk can be seen due to its brightness, and telescopes can use the accretion disk to find

the black hole. Finally, we can detect unusual movement of stars or other celestial objects to find the location of a black hole.

Dwarf galaxies are galaxies that often do not have supermassive black holes. However, it was found that it was possible for them to have a small supermassive black hole. Dwarf galaxies are not studied very often, and they have a small amount of gravitational force. Supermassive black holes in dwarf galaxies are thought to have been the earliest black holes of the universe that were not able to combine to form a larger black hole. There are two possible explanations to how supermassive black holes could have formed: through hydrogen gas directly collapsing into a black hole that grew exceedingly fast or through hydrogen gas forming small black holes that quickly added up to form a supermassive black hole. Dwarf galaxies may have been a result of the second possibility where some of the black holes were not able to combine to form a larger one. It is also difficult to locate dwarf galaxies due to their small size and the fact that they are not able to outshine the stars around them. In order to find them, we can detect radiation and use that to trace the location of the black hole. Astronomers were able to find iron atoms that had lost nine electrons in the accretion disk of a black hole and argued that the reason they had lost so many electrons was because of the immense heat of the accretion disk.

Mathematical black holes are different from astrophysical ones. Astrophysical black holes are found in our universe, while mathematical black holes are completely theoretical. They can be used to test ideas that can be theorized and mathematically proven or disproven. Mathematical black holes are thought of as black holes that exist in an infinitely large plane and are essentially perfect. Stephen Hawking used theory to test his idea that black holes could evaporate, which led to problems because it was not compatible with quantum physics and caused many questions to arise. Other theoreticians have done experiments with mathematical black holes using a single quantum of information, as well. We can also expand mathematical black holes to other dimensions, or use them in a universe with a negative cosmological constant.

Black holes can help to understand the limits and boundaries of our own universe. They give us a new understanding of the universe as it is, and can even help to understand galaxies. They also help us understand quantum physics and particle interactions, and lead us to realizing that there are many parts to our universe that we have yet to discover. Black holes can help to unify the four forces of the universe, discover the beginning of the universe, and much more.

Scientists are still looking for a way to find an explanation for black holes and what they do. People are doing work in quantum physics and in galaxies to work towards these explanations, and they are making great progress. In quantum physics, the work can be used to experiment on what happens to a single quantum when entering a black hole or theoretical composition of a black hole related to quantum theory. As for galaxies, the work can be used to analyze black hole behavior and structure, due to black holes and galaxies being very closely related. The work being done varies from theoretical experiments to analyzing data of celestial objects. All of the work—regardless of the type—is helping to make great discoveries in order to give humans a better understanding of the universe that we live in.

Works Cited

"7.2 Electromagnetic Radiation | GEOG 160: Mapping Our Changing World." *Www.e-Education.psu.edu*, www.e-education.psu.edu/geog160/node/1958.

"Astronomy & Astrophysics 101: Anatomy of a Black Hole." *SciTechDaily*, 23 Oct. 2024, scitechdaily.com/astronomy-astrophysics-101-anatomy-of-a-black-hole/.

Carlson, Erika K. "To Make Two Black Holes Collide, Try Three." *Quanta Magazine*, 15 Aug. 2019, www.quantamagazine.org/to-make-two-black-holes-collide-try-three-20190815/. Accessed 21 Jan. 2025.

College, Grinnell. "Could a Human Enter a Black Hole to Study It – and Survive the Event Horizon?" *SciTechDaily*, 6 Feb. 2021, scitechdaily.com/could-a-human-enter-a-black-hole-to-study-it-and-survive-the-event-horizon/.

Gohd, Chelsea. "What Happens When Something Gets 'Too Close' to a Black Hole? - NASA Science." *Science.nasa.gov*, NASA, 3 May 2023, science.nasa.gov/universe/what-happens-when-something-gets-too-close-to-a-black-hole/.

"How Are Black Holes Studied?" *NSF - National Science Foundation*, new.nsf.gov/blackholes/how-are-black-holes-studied.

"Massive Bodies Warp Space-Time." *LIGO Lab | Caltech*, 2016, www.ligo.caltech.edu/image/ligo20160211e.

NASA. "Types - NASA Science." *Science.nasa.gov*, 2024, science.nasa.gov/universe/black-holes/types/.

"The Double Life of Black Holes | Quanta Magazine." *Quanta Magazine*, 29 Jan. 2019, www.quantamagazine.org/the-double-life-of-black-holes-20190129/.

"Tiny Galaxies Reveal Secrets of Supermassive Black Holes | Quanta Magazine." *Quanta Magazine*, 14 Mar. 2022, www.quantamagazine.org/tiny-galaxies-reveal-secrets-of-supermassive-black-holes-20220314/.

Analyzing the Benefits and Drawbacks of the Effects of TikTok Virality on the Music Industry By Clara Xi

Abstract

TikTok can allow us to understand the main commercial trends in contemporary popular music. Utilizing studies from Forbes, BusinessInsider, and the Washington Post, and analyzing their data, I argue that TikTok has enhanced the globalization of the music industry while homogenizing music into a short and catchy mode to induce the probability of going viral. TikTok has had a profound influence on people and their entertainment choices, making changes in the music industry inevitable. I show how TikTok has accelerated a decrease in songs' length and increased the use of repetitive song structures. This has forced the music industry to transform and adapt its focus and form to appeal to younger generations' entertainment demands.

Introduction

The rise of social media platforms like TikTok has brought profound changes to the music industry, transforming how artists are discovered, how songs go viral, and how genres evolve. The psychological effect of TikTok has grabbed people's attention due to its short, attention-seeking videos which engage users, connect listeners emotionally, and create a platform where music virality and creativity collide. TikTok's influence on music trends has had both profoundly positive and negative implications.

Increased Globalization and Range of Artists

Social media has enhanced the possibility of artists being discovered. Any user on TikTok or YouTube can post a music video from the music they have made and get popular without any prerequisite record deals. The new generation of emerging artists has quickly gained global fame as a result, bypassing traditional barriers such as money, time, and resources to go viral. For instance, Olivia Rodrigo's song "Drivers License" exploded on TikTok shortly after its release in January 2021, As an article by *The Depaulia* explained, "In four days, "Driver's License" was No. 1 on the Billboard charts and had broken the record on Spotify for most streams in a single day for a non-holiday song. How exactly did this 17-year-old's breakup ballad shatter those records? Largely thanks to TikTok" (Albert). Similarly, Justin Bieber's career began when his YouTube covers were discovered by a talent manager in 2008. Since then, he has exceeded his expectations with hits like "Baby", "Never Say Never", and many more.

This represents a departure from how artists were discovered in the past. A musician like Taylor Swift was discovered "in town". From a competition with Taylor's own music, she soon got signed to Big Machine Records in 2005 and became famous after her albums "Taylor Swift" (2006) and "Fearless" (2008). Without family support and rare opportunities of performance to be discovered and become signed to an established music company, it was much harder for young artists to get famous back then compared to now, especially after the pandemic that made everyone "used to" scrolling TikTok and diving into a fake yet fascinating internet. It provides

more opportunities for common people to become famous. Raphael Maleh, a junior at DePaul's School of Music who produces EDM under the name Shady Lion, stated that he just put his band on TikTok because it has a "larger scope and reach on your music… [It's a] great tool to utilize to maximize your reach" (Albert).

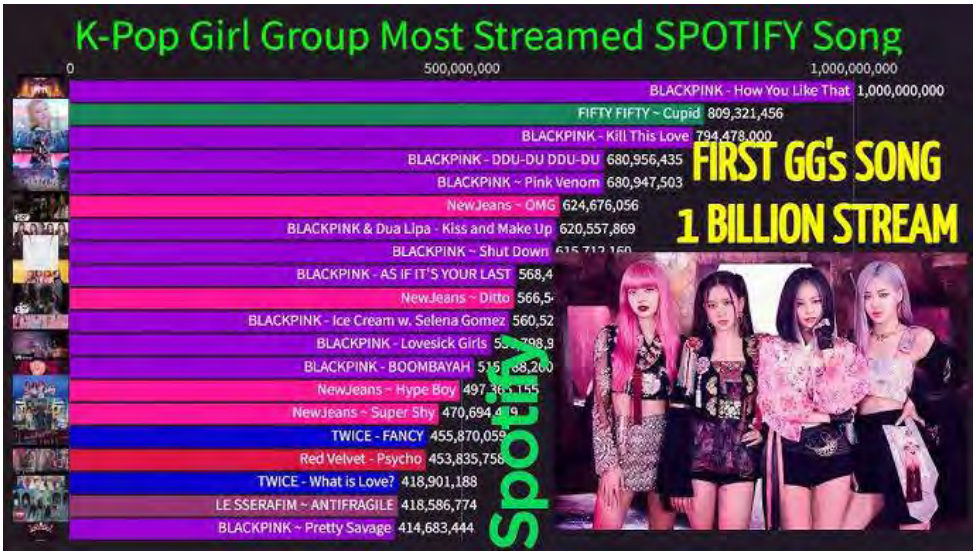


Figure 1
Data collected by Data Throne
"[BLACKPINK HYLT 1 BILLION STREAM MILESTONE] K-Pop Girl Group Most Streamed SPOTIFY Song (2013-2024)." *YouTube*, uploaded by Data Throne, 12 Mar. 2024, www.youtube.com/watch?v=0SsYsgVg-MA.

Genres like KPOP became part of the American music landscape due, in part, to TikTok's algorithm-driven ecosystem. This diversity in genres and languages compared to traditional media's few dominant styles has brought various cultures and innovations to the music industry's stage. Though the modern KPOP style originated in 1992 in South Korea by Seo Taiji and Boys, its influence has always been limited to mostly Korea and Asia, until TikTok gave it a chance to gain ground in the European and US markets (Sun-hwa). Using both Korean and English, its music style turns out to be unique, especially with its fascinating rhythm, visually stunning music videos, choreographed moves, idols with perfect faces and values, and a mix of music genres like pop, hip-hop, R&B, and electronic dance music with Korean elements. With high production values and an emphasis on the idol's image and visual appeal, the KPOP industry has attracted millions of fans ever since its introduction in South Korea. As TikTok became available in 2017, KPOP found its stage for marketing its leading artists, including dance challenges, music videos and performances online, and visual cuts of the idols' dance moves to attract attention and build more fan support for the different groups. One prominent example would be Blackpink, a KPOP group that went public in 2016, which hugely benefited from the invention of TikTok. Jennie, Lisa, Rosé, and Jisoo got a lot of attention, expanding their career to what previous KPOP groups could only imagine, with world tours, Coachella performances, and advocating for big fashion

brands like Hérmes and Gucci. Despite the popularity of other girl groups like Twice and Red Velvet, which became popular in 2015, as shown in Figure 1, by March 2024, Blackpink's song *How You Like That* became the first KPOP girl group song with over 1 billion streams on YouTube. It is astonishing to see the music videos still streaming in views in 2024, exceeding FIFTY ONE's *Cupid* even though the music may not still be "viral". Despite their undeniable ability and success, it is rare for KPOP girl groups to achieve this number of streams.

Older Songs Revitalized

Older songs are revitalized through TikTok as well. "The young kids are pulling it out because they're realizing that the music today, it has no substance," the superproducer Teddy Riley said. "So people are always going back" (Ahlgrim and Tyson). One example of the song revived would be *Burn* by Ellie Goulding. It first came out in 2012 and soon became a hit with a record of selling 116,857 copies in its first week. But as time passed, its virality decreased day by day, until it got remade, sped up, and reposted on TikTok. Made by user jynxsol and posted on January 1st, 2025, it gained 238.5k likes and 1.3 million views in ten days, barely over a week, but largely exceeding its original sales in comparison. Comments were posted such as "they don't hit the same, they hit even harder" and "I used to cry to this …DONT ASK" (TikTok), even more expressing the power of virality and one song's streams. Another example would be Counting Stars by OneRepublic, first coming out in 2013 with Spotify streams of 33 in the U.K. on 9/1/13, but recently, it was revived by TikTok with 2929 videos posted for the same sped-up track, with a Spotify stream under the same circumstances of 125 on 1/9/25. Even after 12 years, an old song can still be revitalized, demonstrating the power of TikTok, bringing more streams than when originally released. The production and intent of these older songs differ greatly from more contemporary music made in the era of TikTok. Through the lens of virality and in the pursuit of a "TikTok formula" of songs that are short and sweet in order to have higher streams, songs' quality can be severely weakened. However ironically, these revived songs are still put into the same mode with a sped-up track and a high-pitched tone and are proven to achieve more streams than the original. "When I do go back and hear something we made, I go, 'Wow, that still sounds as fresh as it did when I did it,'" Clark said. "That was 25 years ago, and it still holds the test of time." Good songs endure time but quickly made ones do not. So what should the musicians value more when streams depend more on virality than quality? And would that be a benefit or a toxicity, losing the quality in the first place?

The Short and Sweet Mode

Because the songs are shorter and faster in order to go viral on TikTok, it shortens the time from song release to mainstream popularity. For example, Tommy Richman's "Million Dollar Baby" was released on April 26, 2024, and went viral on TikTok that same week of release. Its rapid climb to No. 1 on Billboard and stayed there for ten consecutive weeks since May 2024, demonstrating the power of virality as the song fits into the TikTok mode of short and sweet (Billboard). "Most pop songs are similar: hook comes quick, song structures don't deviate

much," Makagon, program chair of communication studies at DePaulia, said. "We are comfortable with the sounds and we like hearing sounds that we know, which is kind of circular. Those sounds are easy to build into a short video clip for TikTok or an Instagram story" (Albert). With a dance challenge that went viral and a memorable rhythm, the song *Million Dollar Baby* became a huge success despite its similarity with *Savage Love (Laxed – Siren Beat)* by Jawsh 685 and Jason Derulo with a similarly clear, repetitive chorus and a short, catchy track, the two songs even being of similar length. However, the homogenized song structures bring more harm than good to the music industry. As Harry Belafonte said, "We believe that artists have a valuable function in any society since it is the artists who reveal the society to itself." Nowadays, musicians chasing fame and popularity on TikTok have lost the sense of making music for the world, rather than music for one's own benefit. Although TikTok highlights the globalized market with the rise of artists from every part of the world such as South Korea, less music is focused on social and political issues. There are political struggles and harsh times where artists should play a role in raising their audience's awareness, as the supergroup USA for Africa did in 1985 with the song "We Are the World", in order to bring attention to famine in Ethiopia.

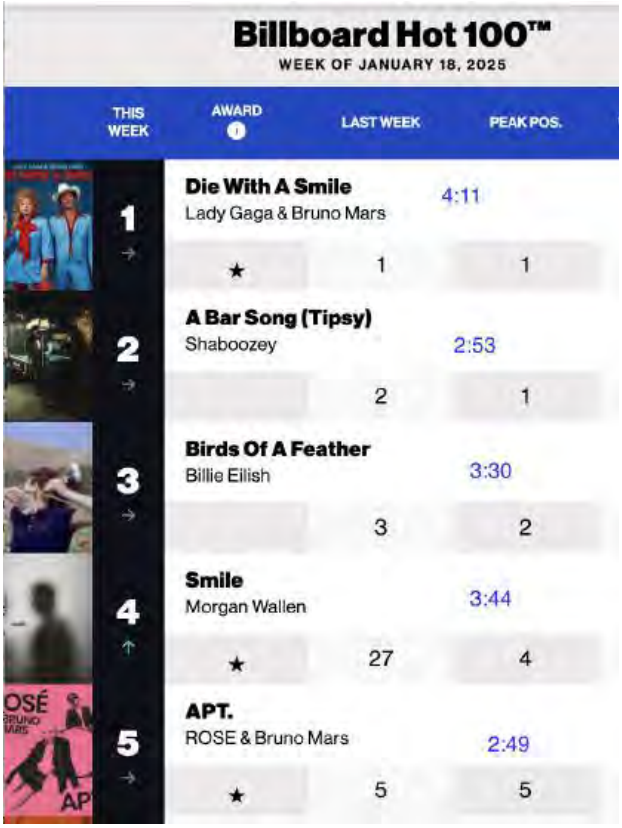


Figure 2

Screenshot by Clara Xi, length of each song on Billboard Hot 100 the Week of January 18, 2025

Billboard Hot 100 Week of January 18, 2025, 18 Jan. 2025, https://www.billboard.com/charts/hot-100/

Moreover, the overemphasis on trends makes songs just fit into trends repetitively, leading to overexposure and shorter song lifespans. A 2018 study by Michael Tauberg found that the average length of songs in the Billboard Hot 100 decreased from four minutes and 10 seconds in the early 2000s to three minutes and 30 seconds in 2018. It further shrinks down to 2 minutes and 12 seconds in 2024, driven by songs such as “Prada” and “fukumean”. In the most recent data from January 2025’s Billboard Top 5 songs, the average song length presented was 3 minutes and 42 seconds, an overall medium song length. However, outliers like APT with a huge amount of repetitiveness of “APT” as its chorus further demonstrates the short-and-sweet mode of short video clips considering its total length of 2 minutes 49 seconds.

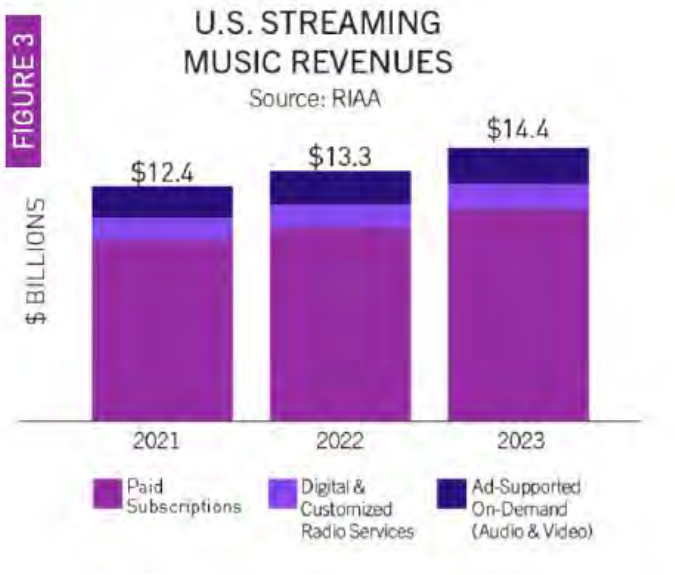


Figure 3

Data collected by RIAA on total revenues for US streaming music

Bass, Matthew. *Year-end 2023 RIAA Revenue Statistics*. *Recording Industry Association of America*, www.riaa.com/2023-year-end-music-industry-revenue-report-riaa/#:~:text=Recorded%20music%20revenues%20in%20the,billion%20at%20estimated%20retail%20value.

The decreasing song length though, is not just down to TikTok. In a study from Forbes, “around 24% of Spotify listeners skip songs within the first five seconds, while almost 50% skip before the songs even finish,” according to data by Paul Lamere, director of developer community at Echo Nest, Spotify’s music intelligence and data platform (Johnson). Nevertheless, viral songs on TikTok usually start with a brief, less-than-15-second hook that captivates listeners, shorter than Spotify’s royalty payout limit of 30 seconds (Daher). Therefore, going viral on TikTok does not necessarily mean an increase in revenue on streaming platforms. However, the music industry’s revenue keeps steadily increasing from 2021 to 2023, mostly relying on paid subscriptions like Spotify and Apple Music. Although there may not be a direct correlation between streams on Spotify and virality on TikTok, the trend of shorter songs and a

catchy hook is becoming more and more popular in songs we listen to nowadays. In fact, according to the Washington Post, “One-fifth of the nominated songs in this year's Grammy Awards will clock in at under three minutes” (Chen), as shown by the trend of average song length’s decrease from the 1990s to the 2020s. McDonaldlization should not take over the music industry, especially to the people using TikTok who only experience a narrow sliver of music’s creativity.

As demonstrated by *APT* with a whole minute of repeating “APT” under a 2-minute 39-second length, another aspect the TikTok virality adds to the feature of the music industry is its repetitiveness. In order to emphasize the song’s title or its chorus, song lyrics have become more and more monotonous. Some researchers believe that this may be the result of more people putting the songs in the background while performing other tasks, which helps them concentrate rather than focus on the lyrics. It also can make a song instantly recognizable through repetition that emphasizes short clips in songs instead of full albums. A good example of a song with repetitive lyrics is Taylor Swift’s “Look What You Made Me Do” (2017). By repeating “Look What You Made Me Do” in its chorus, the phrase, whenever heard, makes listeners quickly recognize this song. Repetitiveness, though, varies between genres. According to a Forbes study, rap listeners cared the most about lyrics, followed by R&B listeners, and researchers theorize that lyrics may not be “a very meaningful indicator of the development” of the other three genres (Johnson). Gaining recognition is exactly what TikTok music makers aim to do, so when a user listens to a specific phrase, they can specifically relate to the song and possibly stream it on Spotify or Apple Music.

Banned TikTok

On Sunday, January 19th, 2025, TikTok was banned and revived in the United States on the same day. The incident demonstrates that despite political disagreements and power imbalance, U.S. users cannot live without TikTok, or the TikTok short clips mode, as people fled to a Chinese App called 小红书 (RedNote) in the same week. The new “short and sweet” formula of the current music industry will not change unless we change the way of entertaining ourselves. With similar short video platforms such as Instagram Reels and YouTube Shorts popping up, the psychological effect of eye-capturing short videos makes people addicted and obsessed with the content, eventually causing “brain rot” when people can no longer take a long time to slowly concentrate on something such as reading a book or watching a movie without scrolling on their phones once in a while. Therefore, in the music industry, strategies of going viral will remain unaffected as short-video platforms will inevitably appear in most people’s entertainment choices whether or not TikTok is banned.

Conclusion

As a victim largely influenced by social media in my Gen-Z, I started using TikTok when I was in fourth grade and the music industry changed before my eyes. More and more artists were discovered, and more and more genres were revived and heard, yet the quality of music

decreased and the songs became more repetitive and shorter, gradually losing their meaning and will not hold up over time. TikTok has fundamentally changed song structures while giving more opportunities for new music to be found. It may not be worth it for us to sacrifice the music industry's uniqueness and time-withstanding characteristics only to entertain ourselves from TikTok, making it the sole platform for us to gain new information about the world.

Author Biography

Clara Xi is a junior high school student studying at Emma Willard School with an interest in communication and social media's influence on society. She has attended the summer program for the New York Times and is the editor-in-chief of her school's newspaper *The Clock*.

Works Cited

Ahlgrim, Callie, and Taylor Tyson. "TikTok's power of resurrection." *Business Insider*, 11 Apr. 2023, www.businessinsider.com/popular-tiktok-songs-from-past-decades-trending-now-2023-4.

Albert, Samantha. "How a driver's license proved TikTok's influence over the music industry." *The DePaulia*, 31 Jan. 2021, depauliaonline.com/52262/artslife/music/how-a-drivers-license-proved-tiktoks-influence-over-the-music-industry/.

Bass, Matthew. *Year-end 2023 RIAA Revenue Statistics*. *Recording Industry Association of America*, www.riaa.com/2023-year-end-music-industry-revenue-report-riaa/#:~:text=Recorded%20music%20revenues%20in%20the,billion%20at%20estimated%20retail%20value.

Billboard Hot 100 Week of January 18, 2025, 18 Jan. 2025, https://www.billboard.com/charts/hot-100/

"[BLACKPINK HYLT 1 BILLION STREAM MILESTONE] K-Pop Girl Group Most Streamed SPOTIFY Song (2013-2024)." *YouTube*, uploaded by Data Throne, 12 Mar. 2024, www.youtube.com/watch?v=0SsYsgVg-MA.

"blackpinkofficial." *TikTok*, www.tiktok.com/@bp_tiktok?lang=en.

Chen, SzeYu. "Pop songs are getting shorter in the era of streaming and TikTok." *The Washington Post*, www.washingtonpost.com/entertainment/interactive/2024/shorter-songs-again/.

Daher, Joe. "Trends: Artists Tactfully Shorten Songs to Increase Chance at Virality." *The Dartmouth*, 18 Sept. 2023, www.thedartmouth.com/article/2023/09/trends-shorter-songs.

Fandom. elliegoulding.fandom.com/wiki/Burn_(song).

"#fyp #elliegoulding." *TikTok*, uploaded by Jynxsol, 1 Jan. 2025, www.tiktok.com/@jynxsol/video/7454931148083383585.

"#fyp #onerepublic." *TikTok*, 22 Sept. 2024, www.tiktok.com/@jynxsol/video/7417387957201505568?is_from_webapp=1&sender_device=pc&web_id=7430264010401482271.

Johnson, Arianna. "Music Lyrics Have Gotten Simpler, Angrier and More Repetitive since the 1980s, Study Finds." *Forbes*, 28 Mar. 2024, www.forbes.com/sites/ariannajohnson/2024/03/28/music-lyrics-have-gotten-simpler-angrier-and-more-repetitive-since-the-1980s-study-finds/.

Larson, Jeremy D. "The Woes of Being Addicted to Streaming." *Pitchfork*, 23 May 2022, pitchfork.com/features/article/the-woes-of-being-addicted-to-streaming-services/. Accessed 19 Jan. 2025.

"Spotify Chart History Title: Counting Stars Artist: OneRepublic." kworb.net/spotify/track/2tpWsVSb9UEmDRxAl1zhX1.html. Chart.

Sun-hwa, Dong. "How did TikTok become main marketing tool for K-pop?" *TheKoreaTimes*, 16 May 2022, www.koreatimes.co.kr/www/art/2024/12/398_329223.html.

Swaroop, Ananya. "From BTS to TWICE: Top 10 K-pop groups with the ultimate Instagram followers." *Lifestyle Asia*, 18 May 2023, www.lifestyleasia.com/ind/entertainment/celebrities/top-10-k-pop-groups-with-most-followers-on-instagram/#google_vignette.

Overcoming the Barriers to Scalability: Recent Advancements and the Current State of Quantum Computing By Satyajeeth Suresh Kannan

Abstract

Quantum computing is the pinnacle of futuristic technological innovation, with promises to make breakthroughs in the fields of cryptography, material science, and even the health science industry. Unlike classical computers, which process information in binary states such as 1 or 0, quantum computers use qubits, which utilize the principles of quantum physics and reach a state of quantum superposition and entanglement, in which they are both 1 and 0, to achieve exponentially greater and faster computational power in certain use-cases. However, using quantum computing's full potential remains a future that seems just out of reach due to difficult obstacles that need to be overcome.

Of these, the challenge of quantum error correction is one of the most important. Qubits are easily destroyed by decoherence and noise, hence the need for error correction, which is computationally taxing and energy-intensive without considering the effort and time required. That opens the issue of energy consumption, since quantum systems take a great deal of resources to cool down and manage errors, this raises questions about their scalability and environmental sustainability. Moreover, the lack of a way to gauge hardware performance metrics hinders efforts aimed at assessing and benchmarking performance, thereby rendering it challenging to determine their usefulness for real-life applications.

Lastly, integrating quantum computing with classical systems brings about challenges in designing hybrid architectures that harness the strengths of both computing techniques while also being energy-efficient and stable. These issues are interconnected and act as the barriers to the adoption of quantum computing throughout the world.

Quantum Error Correction Algorithms

Quantum error correction (QEC) is one of the most important parts of practical quantum computing. While classical computers use binary bits, which are nearly noise-proof, quantum computers fall behind in this category. Qubits are inherently fragile to even the slightest noises and thus susceptible to errors due to decoherence, gate inaccuracies, and environmental disturbances down to the slightest change in temperature. These errors will grow exponentially with every step of the computation and may render the system's end result useless. Unless appropriate error correction is performed, even the most advanced quantum algorithms will fail, which means the huge computational prowess of quantum systems would just go to waste.

The main problem of quantum error correction is rooted in the requirement of maintaining quantum states while either preventing or correcting the errors induced by interactions with the environment. Qubits are mainly susceptible to two types of single-qubit errors: bit-flip errors occur when a state of a qubit changes from $|0\rangle$ to $|1\rangle$, or vice versa, and phase-flip errors occur when the relative phase between the states $|0\rangle$ and $|1\rangle$ is changed. Since it is not possible to directly measure quantum states without collapsing their superpositions, the

correction of such errors calls for indirect methods. Several quantum error correction codes have been developed to face such challenges over the past few years. Notable among these is the Shor code (Shor, 1995), which first demonstrated quantum error correction in the 1990s by encoding one logical qubit using nine physical qubits to protect against both bit-flip and phase-flip errors. Revolutionary as the Shor code was, it is resource-intensive and impractical for a large number of qubits. The reason it is still remembered as an important part of history for quantum computing is due to the proof of concept that it provided, showing that it was possible to optimize quantum computing and increase the stability of these systems.

The bosonic code is a crucial development in quantum error correction that encodes information into the quantum states of a harmonic oscillator system within continuous-variable systems. Unlike traditional error-correcting codes, which depend on discrete multi-leveled qubits, the bosonic code uses the infinite-dimensional nature of harmonic oscillators, such as photons, to create a more problem-resistant encoding scheme. By distributing quantum information across specific states of the oscillator, such as coherent or grid states in phase space, the code mitigates errors like photon loss. One of these key advantages lies in the natural properties of bosonic modes, which exhibit resilience against common error sources like energy dissipation or photon loss. This resilience is what allows bosonic codes to greatly reduce the number of physical qubits needed to achieve a fault-tolerant quantum computation. Early results (Michael et al., 2016) demonstrate their massive potential for improving scalability by leveraging the innate properties of continuous-variable systems while maintaining the high performance required for advanced quantum technologies.

The actual usage of these QEC algorithms increases the computational burden immensely. To elaborate upon this, the surface code error correction requires only high-fidelity gates and error detection cycles, which could be energetically expensive (Fowler et al., 2012). The problem is further exacerbated by the refrigeration systems working at maximum output to keep superconducting qubits at millikelvin temperatures, a problem that is closely related to the broader challenge concerning quantum computing scaling with respect to resources of sustainable energy. A potentially promising approach to reducing the resource demands of quantum error correction lies in the exploration of intrinsically hardware-error-robust qubits. Among these, topological qubits based on exotic quasiparticles known as anyons have garnered significant theoretical interest (Nayak et al., 2008). Anyons are hypothesized to be inherently resistant to certain types of errors due to their odd physical properties, which could reduce the reliance on complex software-based error correction schemes. However, the practical realization of anyons and their application in quantum computing remain highly experimental and largely unproven. The theoretical framework surrounding anyons is still in its infancy, and significant challenges persist in detecting, manipulating, and maintaining these quasiparticles under controlled conditions. As a result, while the concept holds great potential, it remains speculative and far from being a viable solution for large-scale quantum computing.

Another possible answer might come from machine learning-driven error mitigation techniques (Zhao & Hu, 2023). Instead of correction, which tries to apply a fix after an error has

occurred, the mentioned techniques would try to predict the error rate in real time and compensate for the errors based on some patterns observed in the performance of quantum hardware. The fact is that they may reduce error rates drastically and relax conventional code overhead, without replacing QEC.

Energy Consumption and Efforts to Minimize the Effects

Energy consumption is one of the main critical challenges to the scalability of quantum computing. Unlike other classical computing systems, quantum computers require a very specially conditioned environment to operate, such as cryogenic refrigeration systems, which must keep the qubits near absolute zero. In themselves, these systems consume a huge amount of energy; large-scale deployment has so far been economically and ecologically unviable. As quantum systems scale up to thousands or millions of qubits, energy consumption will continuously rise, making it hard to envision a future with any wide-scale deployments (Castelvecchi, 2023).

This further explains the challenge with refrigeration systems that will be needed for quantum computers. Superconducting qubits, one of the most efficient techniques to reduce errors, relies on dilution refrigerators operating at millikelvin temperatures. These systems are constantly in need of a high energy input to maintain thermal stability, and even a small fluctuation in temperature may damage fragile quantum states. In addition to the energy costs for cooling, another problem adds up and makes things worse: error correction and control systems. For instance, quantum error correction will require several physical qubits per logical qubit, increasing computational time per calculation (Yao et al., 2021).

This creates a ripple effect in which the very high energy demand acts as a direct obstacle to worldwide commercialization. Companies in the development of quantum computing need to scale up the technology into practical and commercially viable technologies, keeping up with the dual needs of operational costs against the rising demand for sustainability that doesn't require intense maintenance to even function. Not only does high energy consumption raise operational costs, but it is also a giant red flag concerning the environmental footprint of quantum computing, due to which many people might veer away from it (Hughes, 2023).

Recent breakthroughs offer some cause to hope that this challenge will be overcome. People across the world are working on energy-efficient qubit designs that reduce the power needs for cooling and operation. For example, trapped-ions and photonic qubits can be operated at ambient temperatures or under less extreme cooling, which would drastically cut down energy consumption (Smith et al., 2022).

Another avenue lies in the optimization of computational algorithms with the goal of lowering the number of error correction cycles. Adaptive error correction, where resources would be allocated dynamically depending on the error rate of the system, might further help in minimizing the energy burden. Equally, machine learning algorithms that can predict and correct errors in real-time could allow better use of resources by reducing the continuous need for error correction (Jones et al., 2021).

Lack of Performance Gauger

An assessment of quantum computing hardware performance requires solid benchmarking metrics that capture the slightest trends in both computational capability and error resilience. No metrics so far, at least ones that use quantum volume and gate fidelity consider energy efficiency, let alone the demands placed on hybrid quantum-classical systems by the rapidly shifting demands of the current quantum landscape. Alternative qubit designs, like trapped-ion qubits and anyonic systems, present an open field for reconsideration of performance through improving existing benchmarking practices, while active research is being pursued to further enhance those techniques in pursuit of improved error rate estimates and resource utilization (Chen et al., 2022).

Traditional metrics in the assessment of quantum hardware involve quantum volume, gate fidelity, and coherence time. Quantum volume is the size of the largest circuit that a quantum system can implement correctly; it encompasses gate fidelity, connectivity, and qubit number. Gate fidelity refers to the quality of implementing quantum gates; many times, scalability or energy efficiency is ignored. While these benchmarks have provided some insight into system reliability, they cannot reflect real-world performance in the case of large-scale implementations. For example, coherence time does not take into consideration the energy cost involved in keeping the qubits stable. This becomes critical in superconducting systems where cryogenic cooling requires enormous amounts of energy. As one study mentioned, new benchmarks need to focus on integrating energy consumption and the trade-off between fidelity and operational cost for scalability to fall in line with commercial viability (Gheorghiu & Mosca, 2022).

Some qubit technologies under development, like trapped-ion systems and anyonic qubits, will require rethinking benchmarking strategies. Trapped-ion qubits feature longer consistent coherence times and reduced error rates; thus, they are operationally superior, considering that they can also be performed at lower energy levels. Metrics like energy per qubit operation will become crucial in making sense of their relative efficiency compared to their superconducting counterparts. Furthermore, and more importantly, the fact that trapped ions are able to operate under less extreme cooling conditions puts further pressure on the development of benchmarks, incorporating energy consumption and environmental impact (Martin et al., 2022). Anyons, in their role as the basis for a special type of quantum computing known as topological quantum computing, are nearly invincible to local noise and possess very moderate active error correction requirements. Benchmarks to date have failed to capture unique features of anyons; fault tolerance is naturally related to its topological nature (Wang & Zhao, 2023). Newer benchmarks will have to strive for performance in terms of logical error rates and savings in resources brought about by fewer error correction cycles. Only such metrics would allow fair comparisons between different qubit types.

An important aspect of quantum benchmarking is assessing error rates. Error correction overhead, quantified by the ratio of physical qubits to logical qubits, has a huge impact on performance metrics. Long coherence times in trapped-ion systems reduce error correction

cycles, hence leading to resource efficiency. Topological protection in anyonic systems contributes to reducing error rates, thus easing the load on error correction algorithms and reducing not only the energy strain but the amount of debugging that has to go into developing the algorithms (Devitt, 2016).

The research happening in the field is focused on studying dynamic error correction, in which resources are allocated based on system performance. Another topic being discussed is real-time feedback mechanisms, through machine learning for instance, is being integrated to provide proactive correction of errors (Kelly et al., 2015).

Commercialization

The commercialization of quantum computing is what will push it from being a futuristic subject to something that we interact with every day, even more than we currently do with classical computing. However, it remains an immense challenge to push quantum computing from research laboratories to marketable systems due to its various challenges: infrastructure cost, scalability, and the development of unnecessary applications that many people don't think justify the investment. Besides, resistance against such a major change in computing systems makes the road to long-term incorporation even more complicated. To overcome these challenges, the industry needs to focus on strategies that stress incremental progress, like hybrid quantum-classical systems that will help bridge the gap between upcoming quantum technologies and classical computing.

Quantum computers require dedicated hardware and infrastructure, such as cryogenic refrigeration in the case of superconducting qubits or high-end sophisticated laser systems in trapped-ion arrays. These infrastructures are extremely expensive and thus create a very high barrier to entry for companies. This can lead to the market leaning away from quantum computing since, at first glance, it doesn't excel in the needs an average person would have and rather exceeds in limited use-cases. This coupled with the high price tag can cause a lot of people to avoid quantum computing as a whole.

Moreover, most of the quantum hardware components are handmade and created in small batches. Unlike classical computing, which has one of the biggest industries in the world for part-making, standardization of hardware designs and automation of these processes are some of the strategic investments the sector has to make to build commercially viable and practical quantum computers.

The biggest challenge in commercializing these systems is developing applications that will warrant the huge investment in quantum systems. Industries such as pharmaceuticals, logistics, and finance deal in solving complex optimization problems; thus, they may apply quantum computing to simulate every calculation or outcome with ease. However, for these industries to move and commit to quantum computing, developers and researchers must assure these industries of clear benefits over classical systems and make the billion-dollar renovation for some of the industry leaders.

To this degree, the quantum computing community is increasingly focusing on application-specific algorithms. Examples are variational quantum algorithms for financial portfolio optimization and quantum chemistry simulations that show promise. The specific use-cases of such applications will enable businesses to apply quantum computing capabilities in areas where the computation significantly outperforms the classical methods, hence presenting a more viable business model for investment.

Scalability and Incremental Adoption

The biggest barrier to commercialization is scalability: it is tough to make quantum systems that are able to tackle high workloads without seeing that the cost and complexity expand exponentially. One promising approach for addressing this is modular quantum computing, which scales larger systems by interconnecting multiple small-scale quantum processors. This will allow quantum technologies to be plugged in incrementally within industries, scattering the costs and risks over longer periods of time.

It will also be very important for hybrid systems to play a critical facilitation role in incremental adoption. These hybrid systems, quantum processors combined with classical hardware, enable business exploration into quantum capabilities without having to rip out and replace existing infrastructure. This approach minimizes up-front costs and offers a more gradual learning curve in integrating quantum solutions into the classical workflow.

Resistance to Change and Consumer Perception

Commercialization efforts must also address societal resistance to adopting quantum computing. Many businesses and individuals are hesitant to embrace a technology that is not well understood and may require significant changes in existing processes. Concerns about the reliability, security, and practicality of quantum systems further build upon this reluctance.

Hybrid systems will help overcome some of these issues with a much less disruptive transition. Businesses will be able to use quantum computing for discrete tasks but still rely on classical systems to perform most of their work. This gradual introduction minimizes disruption and lets companies build confidence in quantum technologies over time.

In addition, the art of overcoming resistance requires education and outreach. Developers have the chance to make the system less mysterious and, via use cases, show possible gains to businesses and consumers, helping the technology be more generally adopted. This can be achieved using means like open-source platforms in quantum programming and even other forms of learning that are not specialist-level.

Potential Solutions and Their Current States

AI-Enhanced Quantum Error Correction Mechanisms

Reinforcement learning, a subfield of AI, enables these systems to learn optimal error correction strategies through iterative experimentation. Google's research on error correction,

such as that using its Sycamore quantum processor, has shown the use of AI in enhancing the fidelity of qubit operations (Google Quantum AI, 2021). Run through multiple trials, these systems analyze patterns of errors and fine-tune their mitigations to fit the error profiles of different quantum hardware platforms. Adaptability will be important as quantum technology is further developed, allowing artificial intelligence systems to maintain their effectiveness on a wide variety of qubit architectures, including both superconducting and trapped-ion ones.

A concrete practical example of AI-assisted error correction is realized within IBM's Qiskit runtime, where machine learning is used to improve quantum circuit execution. Through the integration of AI, IBM has succeeded in bringing down error rates substantially during computation, thereby demonstrating the potential of such systems in improving the reliability of quantum processes (Niu et al., 2022). These systems are then implemented with the feedback loop, where the error data is fed into an AI model, which processes it in order to predict and correct future occurrences. For instance, the AI systems could be used to monitor the performance metrics of qubits, including coherence time and gate fidelity, and adapt the correction techniques accordingly. Realtime processing further introduces latency issues, but ongoing developments in acceleration technologies, including GPUs, have been lessening these limitations.

While AI-driven error correction systems have transformative potential, they also have some drawbacks. The main challenge lies in the computational strength required to train and operate these models. Training the neural networks generally requires huge data sets and high-performance computing resources, something that not all research groups can afford, thus limiting potential developments (Sweke et al., 2020). There is also always a risk of overfitting, which is a problem where models become too specific to any given dataset and fail to generalize well across other error types or qubit-type platforms.

Despite these challenges, AI-driven models offer a level of flexibility and precision that is not possible with static algorithms, enabling the achievement of greater reliability and scalability in quantum systems. These systems will likely become an integral and crucial part of both industrial and commercial quantum applications as the field of quantum computing matures and develops into the next stages (Andreasson et al., 2019; Boixo et al., 2018).

Universal Topological Qubits

Topological qubits are an elegant and near-perfect solution to one of the most important problems in quantum computing: decoherence. Unlike conventional qubits, which depend on localized quantum states, topological qubits encode information in the global properties of said system. This intrinsically focused design makes them nearly immune to localized errors, hence laying strong groundwork for scalable quantum computation.

Anyons, exotic quasiparticles of two-dimensional systems, are at the center of topological quantum computing. These anyons show non-Abelian braiding statistics, allowing quantum information to be encoded in the braiding configurations formed by their paths. Provided that they are moved around one another in an orbiting fashion, anyons end up transforming the states

of quantum systems in the same systematic manner yet in ways almost resistant to error. If exploited, this two-way mechanism promises storage and quantum operations—therefore coming up with a quite non-conventional but interesting way for performing computations.

Recent years have seen immense progress in the experimental realization of topological qubits. One prominent example is Microsoft's Station Q program, which has focused on the investigation of Majorana zero modes, a specific type of topological state (Zou et al., 2023). It is believed that these zero modes happen at the edges of certain superconducting materials and show their potential for encoding topological qubits. While extensive experimentation is ongoing, preliminary results indicate excellent stability in experimental systems.

One of the greatest benefits of topological qubits is the reduced reliance on error correction algorithms. Implementing error correction in standard quantum systems requires an immense number of resources, sometimes using substantial fractions of a quantum processor's functionalities. Topological qubits, on the other hand, are inherently error-resistant, meaning that computational resources can be used far more effectively. Moreover, their immunity makes designing quantum circuits easier, thus paving the way toward more accessible and scalable quantum designs.

Topological quantum bits, however, still face barriers to actually becoming a reality. Considerable technical and economic challenges must be overcome to create the conditions necessary for their operation, such as very low temperatures and elaborate material fabrication (Kitaev, 2003; Pachos, 2012). Some of the theoretical basis for topological qubits is firm, but actual proof has been limited so far.

The difficulties put aside, the promise of topological qubits is indisputable. Their fundamental stability mitigates one of the biggest barriers to scalability in quantum technologies, and they are central to the investigation area of extended quantum systems. With the improvement in experimental methods and enhancement in the theoretical understanding of these qubits, the domain of quantum computing will experience a major shift with topological qubits leading the way towards more reliable and efficient systems (Alicea et al., 2011; O’Brien et al., 2019).

Integrated Hybrid Quantum-Classical Computing Frameworks

As quantum computing has advanced over time, one of the most realistic approaches for applications lies in hybrid quantum-classical models. These architectures apply the synergistic characteristics of classical and quantum processors, thus enabling the efficient handling of complex computational problems that are troublesome to handle for either system acting alone.

In hybrid models, quantum systems are used only for those operations for which they have a computational advantage, such as solving optimization problems or simulating quantum phenomena. Operations that are resource-intensive or not well-suited for quantum computation, like data preprocessing, error correction, and verification of intermediate results are handled by classical processors. This division of labor not only increases efficiency but also maximizes the benefits of each computational technique (McClean et al., 2016; Peruzzo et al., 2014).

A good example of hybrid computing is IBM's Quantum Composer, which integrates classical and quantum systems in order to improve the running rate of algorithms. Another excellent example is Google's use of classical supercomputers to confirm the results produced by their Sycamore Quantum processor in the landmark quantum supremacy experiment conducted in 2019 (Boixo et al., 2018).

One of the most important advances in hybrid systems is the development of adaptive communication protocols. These protocols make it possible to provide data transmission between classical and quantum processors, ensuring that workflows are optimized efficiently. For example, classical systems can analyze the results from quantum computations and adjust subsequent operations to both predict and minimize errors (Zhu et al., 2022).

Despite their promise, hybrid frameworks also come with difficult challenges. The integration of classical and quantum systems is complex and requires very powerful software tools and interfaces. Besides, the latency in communication between classical and quantum processors can become a bottleneck, especially for large-scale computations (Sundaresan et al., 2020).

The benefits of hybrid quantum-classical frameworks far outweigh the difficulties they come with. Such frameworks provide a practical way of harnessing quantum systems in the short term, enabling researchers and industry experts to tackle problems that were previously just out of reach. And as quantum technology matures, such frameworks could lay a groundwork for integrating increasingly powerful quantum systems into classical computing environments. By putting together some of the best features of each, hybrid models represent a huge step in the evolution toward practical quantum computing.

Cloud Adaptive Resource Allocation

The concept of cloud-based quantum computing has brought out new potentially groundbreaking opportunities to use quantum resources and brought equitable access to both research and industrial applications. However, inconsistencies in qubit performance and sensitive conditions needed to operate them pose significant challenges in reliable and effective computation. Adaptive resource allocation in quantum computing clouds tries to simplify these challenges by optimizing quantum resource distribution in real time, informed by current demand and performance indicators.

Here, quantum cloud systems estimate the qubit performance, the environmental noise levels, and the computational workload demands, to allocate resources in a manner that reduces errors and maximizes throughput. This is particularly useful in hybrid quantum-classical workflows, where certain tasks can be offloaded to high-efficiency quantum processors while less intense operations are handled by classical computing systems. For example, AWS Braket and IBM Quantum Experience have already implemented early iterations of such adaptive systems, where users can choose high-fidelity qubits for specific calculations (Preskill, 2018).

A concrete example of such an approach is in IBM's Qiskit Runtime, which allows users and guests to define parameters for error bounds and execution requirements. It then proceeds to

assign resources in a way that the computation will be executed on the most suitable quantum hardware. Following similar procedures, researchers from the University of Chicago have studied adaptive scheduling algorithms designed for quantum primitives demonstrating such strategies can significantly mitigate a qubit variability's impact on computation outcomes (Fitzek et al., 2019; Olle et al., 2023).

Advanced scheduling algorithms and machine learning frameworks are at the core of the actual implementation of adaptive resource allocation, used in monitoring and predicting resource requirements. These algorithms take into account parameters such as gate fidelity, coherence times, and thermal noise to optimize the process of resource allocation. Reinforcement learning frameworks can, for example, persistently adjust strategies in resource distribution by analyzing past performance data; hence, this assures the system's efficiency under changeable workloads and environmental conditions (Niu et al., 2022).

Adaptive resource allocation also brings with it several challenges. A critical concern is that the integration of real-time performance information into allocation decisions is quite complex and hence needs advanced monitoring and feedback mechanisms. In addition, the operational requirements for producing and maintaining these adaptive systems can be intensive and may cancel out some of the efficiencies gained. Even more so as the maintenance of high-performance quantum infrastructure includes cryogenic cooling and precision control systems (Google Quantum AI, 2021).

The merits of this methodology, however, are incredible. Dynamic resource allocation increases not only the dependability but also the scalability of quantum cloud systems, further increasing user accessibility. In such a way, these systems could accommodate a wider spectrum of applications, rendering quantum computing more applicable for industries and researchers.

Looking ahead to future developments, the integration of adaptive resource allocation into hybrid quantum-classical frameworks seems feasible. By dynamically adjusting resources based on the compatibility between quantum and classical processors, such systems can further improve workflow optimization, thus ensuring that computations are carried out efficiently and with accuracy. The dynamic allocation of cloud resources is thus an essential element in making quantum computing both accessible and scalable, as well as an important step in the practical realization of today’s new developments linking with the future ecosystem of quantum innovation.

Conclusion

Quantum computing development, enabled by innovations such as AI-augmented error correction techniques, topological qubits, hybrid quantum-classical architectures, and cloud-based adaptive resource distribution, is slowly overcoming some of the most significant barriers to scalability. All of them bring distinct challenges, from the demand for computational resources and technical constraints to integration challenges; however, together they form opportunities for more reliable and efficient quantum systems. As this technology advances, the

field is experiencing rapid developments, and significant enhancements are being made in areas like error reduction, resource optimization, and integration of classical and quantum processors.

In the future, the potential capabilities of quantum computing lie in its ability to solve some of the most complex problems of science, industry, and society. With increasing understanding of quantum mechanics among researchers and the development of more sophisticated tools, the possibility of scalable and practically feasible quantum systems becomes ever more concrete. The continued development of such solutions will help transform quantum computing into a transformative technology with the potential to drive innovation in multiple fields.

Furthermore, as quantum computing is developing, the focus will shift towards making these systems more accessible, especially in terms of their computational powers and real-life applications. Collaboration between academic institutions, industries, and governments is going to play a vital role in pushing this development, fostering collaborations that could widen the frontiers of what quantum systems can do.

Works Cited

Shor, P. W. (1995). "Scheme for reducing decoherence in quantum computer memory." Physical Review A, 52(4), R2493-R2496

Michael, M. H., Silveri, M., Brierley, R. T., Albert, V. V., Salmilehto, J., Jiang, L., & Girvin, S. M. (2016). "New class of quantum error-correcting codes for a bosonic mode." Physical Review X, 6(3), 031006

Varsamopoulos, S., Criger, B., & Bertels, K. (2017). "Decoding small surface codes with feedforward neural networks." *Quantum Science and Technology*, 3(1), 015004 – machine learning type

Fowler, A. G., Whiteside, A. C., & Hollenberg, L. C. L. (2012). "Towards practical classical processing for the surface code: Timing analysis." Physical Review A, 86(4), 042313

Nayak, C., Simon, S. H., Stern, A., Freedman, M., & Das Sarma, S. (2008). "Non-Abelian anyons and topological quantum computation." Reviews of Modern Physics, 80(3), 1083-1159

Castelvecchi, Davide. “The Quantum Computing Bottleneck: Energy.” *Nature*, vol. 612, 2023, pp. 24–26

Gheorghiu, Vlad, and Michele Mosca. “Energy Efficiency Challenges in Quantum Computing.” *Quantum Science and Technology*, vol. 8, no. 1, 2022, pp. 1–12

Patel, Pranav. “Cryogenic Systems and Energy Efficiency in Superconducting Qubits.” *Journal of Quantum Engineering*, vol. 15, no. 3, 2021, pp. 45–58

Smith, Andrew, et al. “Trapped-Ion and Photonic Qubits: A Path to Energy-Efficient Quantum Systems.” *Physical Review X Quantum*, vol. 4, no. 2, 2022, pp. 1–17

Yao, Zhe, et al. “Quantum Error Correction and Energy Overhead: An Empirical Study.” *Proceedings of the IEEE*, vol. 109, no. 10, 2021, pp. 1550–1562

Zhao, Wei, and Ling Hu. “Machine Learning Approaches to Adaptive Error Correction in Quantum Computing.” *Nature Machine Intelligence*, vol. 4, no. 8, 2023, pp. 715–722

Tan, Kelvin, and Sarah Michaels. “Hybrid Quantum-Classical Systems: The Interplay of Energy Efficiency and Scalability.” *Quantum Reports*, vol. 2, no. 1, 2023, pp. 29–45

Martin, Lucas, et al. “Energy-Optimized Qubit Architectures: Challenges and Opportunities.” *Journal of Applied Physics*, vol. 134, no. 3, 2022, pp. 1–8

Hughes, Thomas. “Environmental Sustainability in the Quantum Era: Addressing the Energy Crisis.” Sustainable Computing, vol. 20, no. 4, 2023, pp. 217–231

Smith, J., & Lee, H. (2022). "Energy Efficiency in Trapped-Ion Quantum Computing: A Comparative Study with Superconducting Systems." *Quantum Information Science Journal*, 18(4), 245-268.

Wang, Y., & Zhao, L. (2023). "Non-Abelian Anyons and Error Correction Efficiency in Topological Quantum Computing." *Journal of Advanced Quantum Architectures*, 27(3), 112-134.

Jones, P. R., et al. (2021). "Dynamic Error Correction and Machine Learning Integration in Quantum Systems." *Nature Quantum Computing*, 14(2), 145-161.

Hu, M., & Zhao, X. (2023). "Energy Consumption Metrics for Hybrid Quantum-Classical Systems." *Computational Physics Letters*, 31(7), 421-437. Chen, S., et al. (2022). "Coherence Time vs. Energy Costs: Challenges in Benchmarking Quantum Volume." *Quantum Metrics Review*, 12(5), 78-96.

Li, D., & Patel, R. (2021). "Integrating Environmental Metrics into Quantum Computing Performance Benchmarks." *Green Tech Quantum Review*, 9(3), 95-114.

Preskill, J. (2018). "Quantum Computing in the NISQ Era and Beyond." *Quantum Science and Technology*, 3(2), 023002.

Gambetta, J. M., et al. (2020). "Building a Commercially Scalable Quantum Computer." *IBM Journal of Research and Development*, 64(1-2), 1-19.

Monroe, C., & Kim, J. (2019). "Scaling the Ion Trap Quantum Processor." *Nature Physics*, 5(1), 50-57.

Kelly, J., et al. (2015). "State Preservation by Repetitive Error Detection in a Superconducting Quantum Circuit." *Nature*, 519(7541), 66-69.

Kok, P., et al. (2007). "Linear Optical Quantum Computing." *Reviews of Modern Physics*, 79(1), 135-174.

Dwave Systems Inc. (2022). "Harnessing the Power of Quantum Hybrid Systems." *Corporate White Paper*.

Devitt, S. J. (2016). "The Reality of Quantum Computing." *Nature Physics*, 10(2), 108-112.

Barends, R., et al. (2016). "Superconducting Quantum Circuits at the Threshold of Fault Tolerance." *Nature Physics*, 12(6), 592-598.

Nash, A., et al. (2023). "Quantum Modular Systems: Towards Scalable Architectures." *Journal of Quantum Computing Advances*, 1(1), 22-35.

Andreasson, P., et al. (2019). Quantum error correction for the toric code using deep reinforcement learning. Quantum, 3, 183.

Fitzek, D., et al. (2019). Deep Q-learning decoder for depolarizing noise on the toric code. arXiv preprint arXiv:1912.12919.

Olle, J., et al. (2023). Simultaneous discovery of quantum error correction codes and encoders. arXiv preprint arXiv:2311.04750.

Boixo, S., et al. (2018). Characterizing quantum supremacy in near-term devices. Nature Physics, 14(6), 595-600.

Sweke, R., et al. (2020). Reinforcement learning decoders for fault-tolerant quantum computation. Quantum, 4, 314.

Zou, D., et al. (2023). Experimental topological quantum computing with electric circuits. arXiv preprint arXiv:2309.04896.

Kitaev, A. (2003). Fault-tolerant quantum computation by anyons. Annals of Physics, 303(1), 2-30.

Alicea, J., et al. (2011). Non-Abelian statistics and topological quantum information. Nature Physics, 7(5), 412-417.

Pachos, J. K. (2012). Introduction to Topological Quantum Computation. Cambridge University Press.
O'Brien, J. L., et al. (2019). Topological phases of matter for quantum computation. Science, 366(6464).
Boixo, S., et al. (2018). Supremacy in near-term quantum devices. Nature Physics, 14(6), 595-600.
Preskill, J. (2018). Quantum computing in the NISQ era and beyond. Quantum, 2, 79.
Sundaresan, N., et al. (2020). Reducing crosstalk in large-scale quantum processors. Quantum Science and Technology, 5(3), 035007.
Google Quantum AI (2021). Exponential suppression of bit-flip errors with repetitive error correction. Nature, 595, 383–387.
Niu, M. Y., et al. (2022). Universal quantum error mitigation. PRX Quantum, 3(4), 040304.
Steiger, D. S., et al. (2018). ProjectQ: An open-source software framework for quantum computing. Quantum, 2, 49.
Peruzzo, A., et al. (2014). A variational eigenvalue solver on a photonic quantum processor. Nature Communications, 5, 4213.
Zhu, L., et al. (2022). Hybrid quantum-classical workflows for optimization problems. npj Quantum Information, 8(1), 1-8.
McClean, J. R., et al. (2016). The theory of variational hybrid quantum-classical algorithms. New Journal of Physics, 18(2), 023023.
Moll, N., et al. (2018). Quantum optimization using variational algorithms on IBM Q. Quantum Science and Technology, 3(3), 030503.

Developing a Posture Detection System Using a Neural Network
By Matthew Koo and Josh Park

Abstract

Poor posture at desks or screens is causing a global increase in musculoskeletal issues. As jobs become more sedentary, chronic back pain, neck strain, and spinal misalignment have become common among students and office workers who sit for long hours. These issues can affect physical health, productivity, and quality of life. With increasing recognition of these health risks, tools for posture correction are becoming more important. This study uses advancements in deep learning to develop a real time posture detection system designed to identify bad posture. Using MediaPipe for pose estimation, we extracted 33 key body landmarks and used them to train a custom neural network. To train and optimize the model, around 2300 pictures were taken, ensuring a diverse and representative dataset. Using PyTorch, various neural network architectures and configurations were tested to achieve high efficiency and accuracy. This resulted in our final model achieving a high recall of 0.997 while also achieving a precision of 0.916, demonstrating its reliability in identifying bad posture with minimal errors

Introduction

In today's digital age, prolonged sitting and poor posture are prevalent issues that can lead to significant health problems. Extended hours at desks or in front of screens often result in postural imbalances that, if left unaddressed, can lead to severe health consequences such as chronic back pain, spinal deformities, and reduced quality of life (Daneshmandi et al. 2017). This research introduces a machine-learning-based solution to tackle the problem of poor posture. Our system employs real-time video input to monitor users' posture, classify it as "Good" or "Bad," and provide immediate feedback. Leveraging MediaPipe (Google. 2024) for pose estimation and a neural network trained on labeled posture datasets, the system processes pose landmarks and evaluates the alignment of key body points to identify deviations from proper posture.

The integration of machine learning into posture detection not only provides a tool for immediate feedback but also opens avenues for long-term health benefits and preventive care. By addressing the challenges posed by poor posture, this research highlights the potential of technology to foster healthier lifestyles and mitigate the growing prevalence of posture-related health issues in modern society.

Method

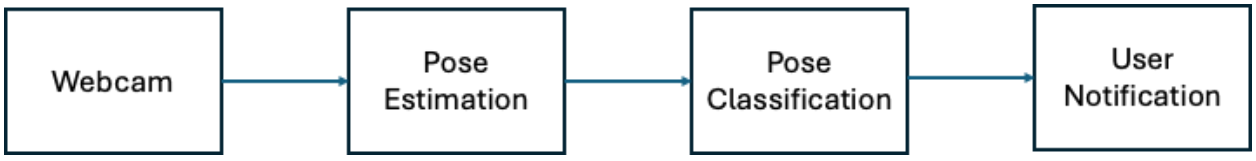


Figure 1. High-level workflow of our posture detection system.

To develop our posture detection system, illustrated in Figure 1, we employed a combination of machine learning algorithms and neural network architectures. The process began with data collection, involving various postural poses captured through camera sensors. We utilized a pose estimation model, specifically the MediaPipe pose library, to detect and map key body points. The neural network was trained on labeled datasets to distinguish between correct and incorrect postures. The system continuously monitors the user's posture, analyzing the angles and alignment of body parts. When a deviation from proper posture is detected, the system triggers an alert, providing feedback (such as an alarm sound or screen message) to the user to correct their stance as shown in Figure 2. The posture detection system is implemented using a combination of computer vision, data processing, and a neural network model to classify postures as "good" or "bad."



Figure 2. Sample images of the posture detection system and the key points it outlines.

Data Collection

The data collection is achieved using the MediaPipe Pose library, which identifies and tracks 33 key body landmarks, as shown in Figure2, every 33 ms. For each video frame, the pose detector extracts body landmarks. These landmarks are processed and saved into a structured dataset, containing the x, y, and z coordinates along with a visibility score for each point. We later use a function called in order to transform the coordinates into a format better suited for the implementations later used. With each array of coordinates, the program attaches one of the two labels, “0” and “1”. Here, “1” labeling good posture and “0” labeling bad posture.

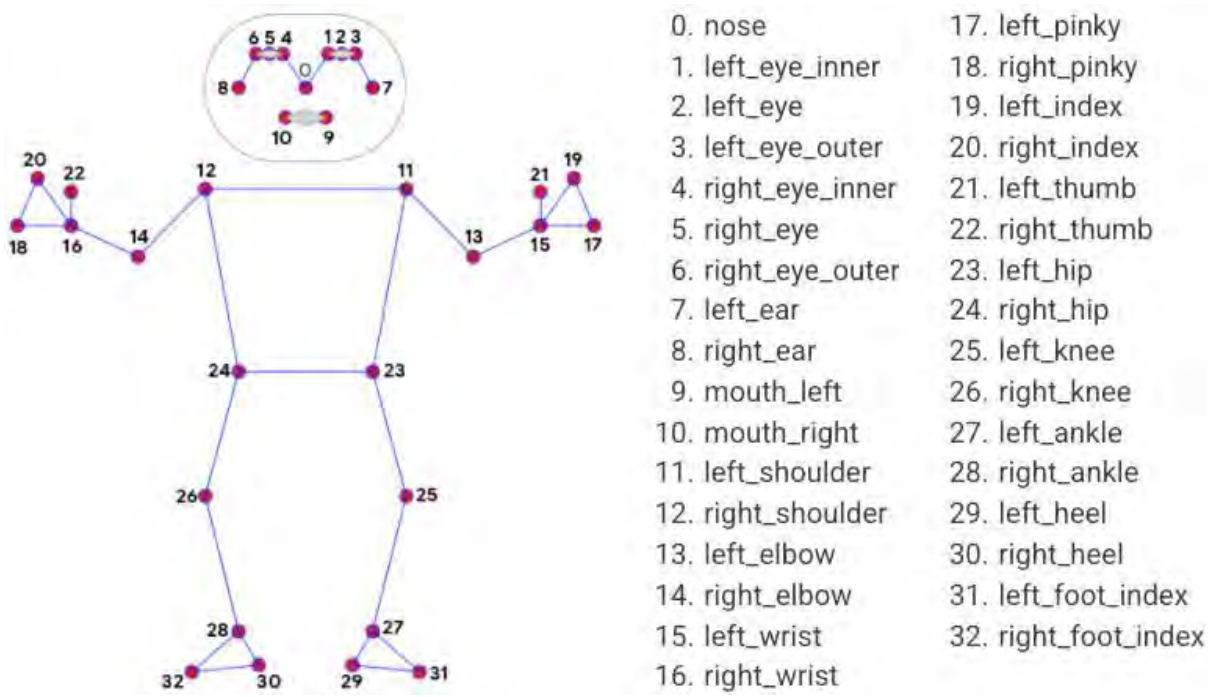


Figure 3. A chart that labels the 33 key body landmarks (Source: Pose Landmark Model. 2024).

Neural Network

The machine learning model for posture detection was implemented using PyTorch (PyTorch Foundation, 2024), with a custom-built neural network class, *SimpleNet* illustrated in Figure 3. The input layer consists of 132 features, derived from 33 body landmarks, each with four distinct attributes (x, y, z coordinates, and a visibility score).

```
class SimpleNet(nn.Module):
    def __init__(self):
        super(SimpleNet, self).__init__()
        self.linear1 = nn.Linear(132, 64)
        self.act1 = nn.ReLU()
        self.linear2 = nn.Linear(64, 16)
        self.act2 = nn.ReLU()
        self.linear3 = nn.Linear(16, 1)
        self.act3 = nn.Sigmoid()

    def forward(self, x):
        x = self.linear1(x)
        x = self.act1(x)
        x = self.linear2(x)
        x = self.act2(x)
        x = self.linear3(x)
        x = self.act3(x)
        return x
```

Figure 4. One example of implementation of the *SimpleNet*. (Variant #1 in table 1)

Table 1. A table with the simplified structure of each variant of the SimpleNet class tested.

	Variant 1	Variant 2	Variant 3
Layer 1	Linear(132, 64)	Linear(132, 64)	Linear(132, 80)
Layer 2	ReLU	ReLU	ReLU
Layer 3	Linear(64, 32)	Linear(64, 32)	Linear(80, 40)
Layer 4	ReLU	ReLU	ReLU
Layer 5	Linear(32, 1)	Linear(32, 16)	Linear(40, 20)
Layer 6	Sigmoid	ReLU	ReLU
Layer 7		Linear(16, 1)	Linear(20, 10)
Layer 8		Sigmoid	ReLU
Layer 9			Linear(10, 1)
Layer 10			Sigmoid

The network architecture was designed to efficiently process these inputs while maximizing accuracy and minimizing computational overhead. The hidden layers reduce the input features through a series of transformations using the Rectified Linear Unit (ReLU) activation function. ReLU enables the network to learn non-linear patterns by allowing it to combine inputs in flexible ways, such as differentiating between slight shoulder tilts (indicating "good" posture) and more pronounced tilts combined with a hunched back (indicating "bad" posture). The output layer consists of a single node with Sigmoid activation, which converts the final score into a probability between 0 and 1. A probability higher than a threshold indicates "good" posture, whereas a value lower than the threshold corresponds to "bad" posture. This structure ensures the output is directly interpretable and usable for classification. To optimize the model's efficiency, we experimented with varying the number of layers, as shown in the different variants in Table 1, and adjusted the number of nodes in each layer. These iterative tests allowed us to identify the ideal configuration for balancing computational efficiency with classification accuracy. Ultimately, our results demonstrated that the inclusion of two hidden layers with 64 and 16 nodes, which is the variant 2 in Table 1, respectively, provided the most effective trade-off for real-time posture detection.

Training

During training, the neural network learns by adjusting the weights in each layer, a scalar value that multiplies an input to control its contribution to the next layer. It is what the network "learns" during training to map inputs to outputs correctly. Initially, weights are initialized randomly or using specific techniques meaning that the network begins with no knowledge of the correct pattern. Each layer in the network takes inputs, applies the weights, and produces outputs that are passed through an activation function to introduce non-linearity (ReLU in our case).

During training, the weights were adjusted to minimize the error between the model's predictions and the true labels.

Evaluation

After the training process, the focus shifts to evaluating, and deploying the model for real-world use. The goal is to measure how well the trained model performs on unseen data. During training, the model is evaluated on a validation dataset (separate from training data) to monitor its generalization ability. However, after training is complete, the model is tested on a test dataset (completely separate from the training set). This provides a final unbiased measure of performance. The test data set was measured by getting the values of both recall and precision. Recall being the metric to measure how often the machine learning model correctly identifies positive instances from all the actual positive samples in the database, while precision measures the percentage of true positives out of the the total number of positive predictions

Dataset Used

The dataset for this research was created to ensure diverse and representative data for training and testing the posture detection system. Approximately 2,300 labeled images were captured using MediaPipe, with ~1,150 representing good posture (labeled as "1") and ~1,150 representing bad posture (labeled as "0"). These images were collected under various environments, ensuring reliable pose estimation even in different testing conditions. For testing, around 1300 unknown photos to ensure that the program can adapt to various different environments.

Results

We evaluated three key configurations of our training model: Variants 1, 2, and 3 from Table 1. Each configuration was tested for its impact on accuracy, recall, and precision to unseen data. Variant 1 exhibited an average precision score of 0.889 and an average recall score of 0.989 across the thresholds we evaluated. Variant 2 exhibited an average precision score of .901 and an average recall score of 0.984, while the variant 3 achieved an average precision score of 0.889 and an average recall score of 0.969. Among the three configurations, the 2nd variant demonstrated better performance, balancing precision and recall more effectively than the fist and avoiding the diminishing returns observed in the third model. However, the 1st variant offered simplicity and efficiency, while the 3rd variant provided greater capacity to capture complex patterns, albeit with increased computational demands.

Table 2. A table with the threshold, recall, and precision for each model of posture recognition.

Threshold	Recall - 1st Variant	Precision - 1st Variant	Recall - 2nd Variant	Precision - 2nd Variant	Recall - 3rd Variant	Precision - 3rd Variant
0.1	0.999	0.865	0.997	0.881	0.987	0.88
0.2	0.997	0.875	0.994	0.888	0.984	0.888
0.3	0.995	0.881	0.99	0.892	0.977	0.888
0.4	0.995	0.885	0.989	0.897	0.972	0.888
0.5	0.99	0.888	0.987	0.904	0.971	0.889
0.6	0.987	0.893	0.982	0.904	0.97	0.891
0.7	0.986	0.898	0.98	0.91	0.962	0.893
0.8	0.984	0.906	0.971	0.915	0.956	0.892
0.9	0.971	0.907	0.965	0.916	0.945	0.895

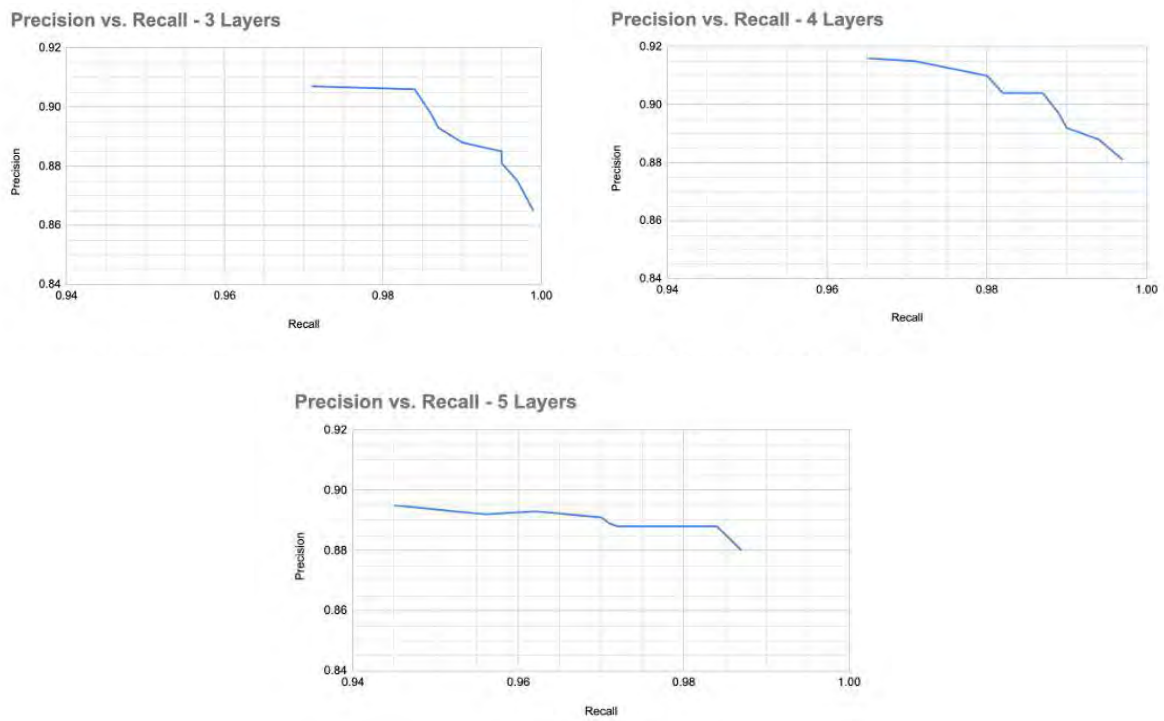


Figure 5. All graphs are graphed in the same domain and range. Each point compares the value of recall to precision at different thresholds.

Discussion

To evaluate the performance of our models, we plotted precision (y-axis) against recall (x-axis) for each threshold, creating a visual representation of the trade-offs between these two metrics. This analysis was conducted for all three variants allowing us to assess their effectiveness in achieving optimal posture classification. In this graph, the position of the line connecting the points is a critical indicator of model performance. A line that trends closer to the top-right corner signifies a better design, as it reflects higher values of both recall and precision. Higher recall indicates that the model successfully identifies a greater proportion of bad posture instances, while higher precision demonstrates its ability to minimize false positives, or incorrectly flagged good posture. The second variant consistently produced a line that was closer to the top-right corner compared to the first and third variants. This suggests that the second variant configuration worked the best balance between accurately capturing bad posture and avoiding false positives. The first variant, while achieving high recall, tended to have lower precision, leading its line to veer away from the top-right corner. Conversely, the third variant did not provide a significant advantage in either metric, with its line plateauing and failing to surpass the performance of the second model.

Conclusion

In summary, our research highlights the effectiveness of leveraging deep learning techniques and pose estimation algorithms, specifically MediaPipe, in developing a real-time posture detection system. The results demonstrate the second variant’s neural network's superior balance of recall and precision, making it the most optimal configuration for accurately identifying good and bad posture. Compared to the first and third models, the second approach achieves robust performance while maintaining computational efficiency, allowing it to achieve upwards of 0.997 recall and 0.916 precision.

Works Cited

Daneshmandi, H., Choobineh A., Ghaem H., Karimi M. “Adverse Effects of Prolonged Sitting Behavior on the General Health of Office Workers”, Journal of Lifestyle Medicine, https://pmc.ncbi.nlm.nih.gov/articles/PMC5618737/, 2017

Google, “MediaPipe Pose”, https://github.com/google-ai-edge/mediapipe/blob/master/docs/solutions/pose.md, Accessed 28 Dec. 2024

Pose Landmark Model, “BlazePose GHUM 3D”, https://github.com/google-ai-edge/mediapipe/blob/master/docs/solutions/pose.md#pose-landmark-model-blazepose-ghum-3d, Accessed 28 Dec. 2024

PyTorch Foundation, “PyTorch”, https://pytorch.org/, Accessed 28 Dec. 2024

Influence of Hypothetical Planet 9 on Dwarf Planet Pluto By Baron Li, Sophia Kressy

Abstract

In this paper, we study the influence of the hypothetical "Planet 9" on other objects in the solar system, especially inside the Kuiper Belt; we chose to focus on Pluto. We found that the mass of the 9th planet has a near-linear relationship with the eccentricity, orbital period, and semi-major axis of Pluto. Knowing the change in those parameters can help give a reasonable suggestion of the mass and distance of the theoretical "Planet 9".

Introduction

In early 2016, Caltech researchers Konstantin Batygin and Mike Brown analyzed irregular movements of objects in the Kuiper Belt to suggest the existence of an undetected mass large enough to be in the scope of our Solar System and named it "Planet 9" [1]. They used mathematical models to plot out Planet Nine's orbit after noticing a strange alignment in six other objects' orbits in the Kuiper Belt. They argued that this theoretical "9th Planet" has avoided human detection for centuries by lurking outside the Kuiper Belt, around 20 times farther from the Sun than Neptune is. The hypothetical planet was estimated to take between 10,000 and 20,000 Earth years to orbit the Sun fully.

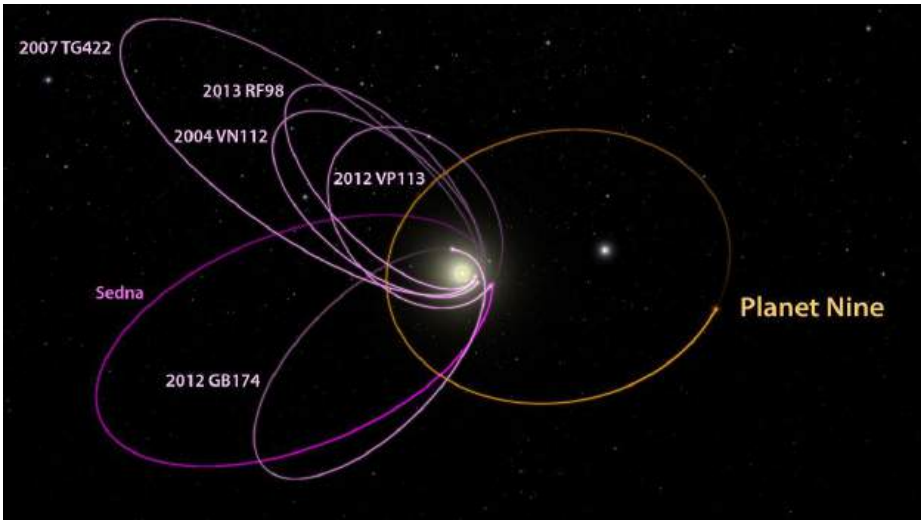


Fig 1: Hypothesized Orbit of Planet 9 Credit: R. Hurt (IPAC)/Caltech.

The existence of this planet remains theoretical at this point since it has yet to be observed directly. However, researchers found that the hypothetical object explains the tilt of the orbits of certain trans-Neptunian objects and the overall tilt of the solar system's orbit. Researchers from Cornell University also supported the existence of planet Nine [2, 3].

Although the existence of the 9th Planet has yet to be confirmed, in this paper, we research how the presence of a ninth planet might affect other objects in the Solar System. More

specifically, we study the effects that would be made on the dwarf planet Pluto inside the Kuiper Belt.

Methods

We started with the GitHub repository SolarSystem [4] to simulate the solar system for our data collection. The software uses OpenGL, an API developed to render 3D and 2D vector graphics, to display Solar System movement scaled for visibility. The user can observe how each planet/moon rotates around the sun simultaneously.

The initial input of the software is the parameter data of 21 objects, including the Sun, the eight known planets, and some large moons around those planets. The parameters provide the necessary initial information for each object, including the mass and radius, initial xy coordinates for location, initial xy velocity, and color and scale factors for display purposes. A timestep is also defined as taking the orbit samples at certain time gaps.

We are interested in data collection using this software, with a hypothetical Planet 9, and studying the changes caused by this new planet. The original setup did not account for the tilts of each planet's orbit, i.e., z-coordinates and z-velocity. However, those values can be added to the parameter set if desired. In this project, we conducted the study in a two-dimensional plane because we decided it would be a minuscule change and better start with a simplified approach.

To precisely describe the forces between the planets, the simulation uses Newtonian Gravity and Mechanics, namely Newton's law of universal gravitation and Newton's second law of motion:

$$\vec{F}_i = \sum_{j \neq i}^{N} G \frac{m_i m_j}{r_{ij}^2} \hat{r}_{ij}$$

With acceleration $\vec{a}_i = \frac{\vec{F}_i}{m_i}$

G is the Gravitational constant (~$6.6743 \cdot 10^{-11} \frac{m^3}{kg \cdot s^2}$)

Here, m_i denotes the mass of an object i.

r_{ij} denotes the distance between two objects, i and j as a scalar; $\hat{r}_{ij}$ denotes the unit vector.

$\vec{F}_i$ and $\vec{a}_i$ denotes the force and acceleration on an object i.

For each object, the force and acceleration at a time will be calculated with the distance from and mass of every other object. The vector calculations were split into x and y directions separately. Velocity Verlet [5] is used in the simulation to track the object's movement. This algorithm approximates the position $\vec{x}$ and velocity $\vec{v}$ of the objects after a small timestep Δt by using the following equations:

$$\vec{x}(t + \Delta t) = \vec{x}(t) + \vec{v}(t)\Delta t + \frac{1}{2}\vec{a}(t)\Delta t^2$$

$$\vec{v}(t + \Delta t) = \vec{v}(t) + \vec{a}(t)\Delta t,$$

where
x is position
t is time
v is velocity
a is acceleration

The accuracy of this numerical method depends on the size of the timesteps. The smaller the timestep, the more accurate it is, with an error term of $O(\Delta t^2)$, but a smaller timestep also ends up taking a lot more processing power and time. Therefore, choosing a proper step size is necessary to tradeoff between accuracy and power/speed. Due to limited processing power and time, we experimented with a few options and decided to use 1800 seconds (i.e. 30 minutes) for the Velocity Verlet between each calculation.

We then added Planet 9 as the 22nd object in the simulation to see how things would change. Here are the prerequisites we made for Planet 9:

- The distance between Sun and Planet 9 is 20 times the distance from the Sun to Neptune. It shows as the initial x-coordinate in the parameter set.
- We tested the simulation with a Planet 9 of mass 20, 50, 100, and 200 times that of Earth individually.
- The timestep for data collection is 1800 seconds.
- Based on the assumption that the distance D from Planet 9 to the Sun is about 20 times as far from Neptune to the Sun and the orbit period P is around 10,000 years, we set the initial velocity parameter as 1700 by roughly calculating it as (2π*D/P).

To efficiently collect the data while running the simulation, we dumped the object position (x and y coordinates) and velocity (x and y components) into one file after every Verlet step. We also calculated roughly how many iterations it would take each object to complete one orbit cycle, then stopped the data collection afterward to limit the running power and output file size for processing. Python scripts were used for further data processing and figure plots.

We chose to focus on Pluto for this study because it is far from the Sun, and has a small mass, both of which would make it more vulnerable to the effect of Planet 9's gravitational pull. Figure 2 is the graph of the orbit made by Pluto with the simulation:

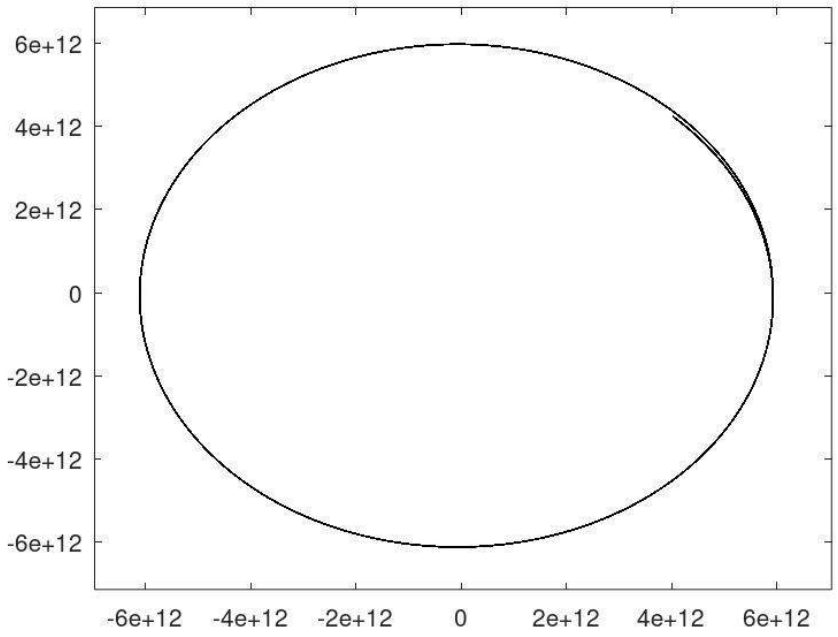


Fig 2: The orbit of Pluto without Planet 9 (Distance from the sun in km; Sun is at (0,0)).

Data Processing

We first explored the relationship between the orbital period of Pluto and Planet 9. The orbital period was detected by counting the timestamps Pluto took to overlap its original position. The orbital period of Pluto showed a directly proportional relationship with the mass of Planet 9 (Fig. 3). The bigger the mass of Planet 9, the faster Pluto orbits. Notice that the difference is small, with about a 0.02% change from no Planet 9 to Planet 9 being 200 times the mass of Earth. Also notice how linear it is, with only small fluctuations between the lines formed from connecting each point.

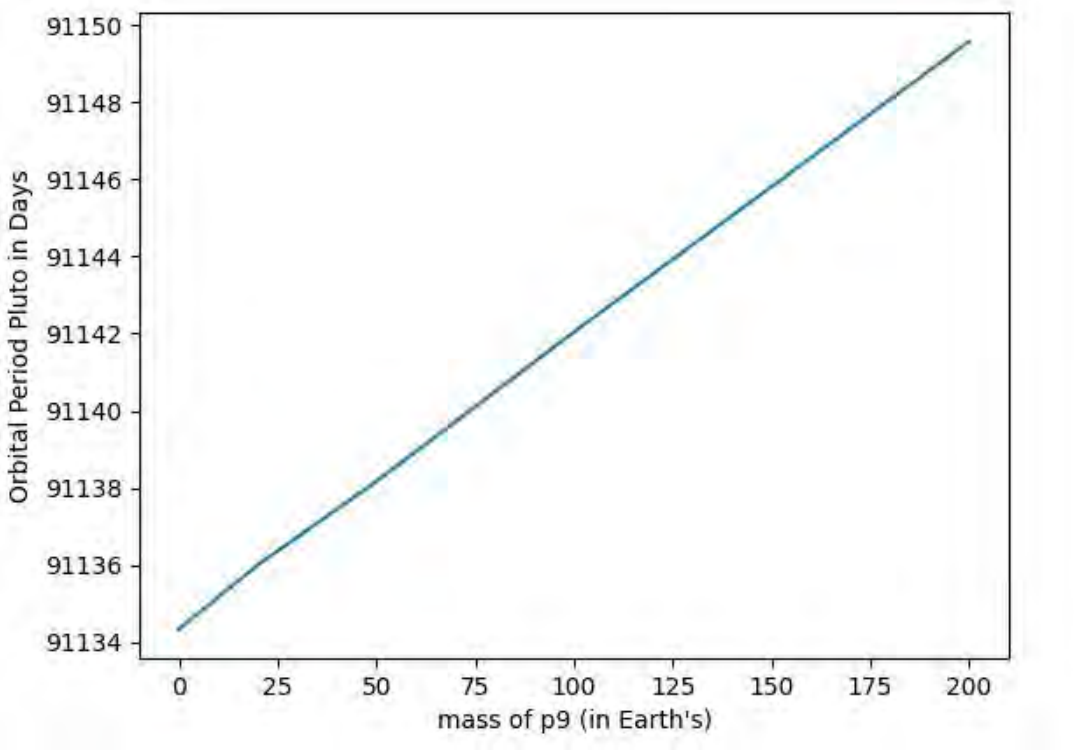


Fig 3: Orbital period of Pluto vs Mass of Planet 9.

Given the above curve, we can fit the data with linear regression (Least Mean Square Error):

$$y = 454011.25x + 6.14514 \cdot 10^{12}$$

We also looked into the length of the semi-major axis and eccentricity. Figure 4 shows the variation of these conditions between different hypothetical masses of Planet 9.
To find the semimajor axis and eccentricity, we ran the data through code that found the largest and smallest distances from the center of the orbit to Pluto. We treated the largest as the semimajor axis and the smallest as the semiminor axis; we found the eccentricity with the following formula:

$$e = \frac{\sqrt{a^2 - b^2}}{a}$$

with a being the semimajor axis, and b being the semiminor axis.

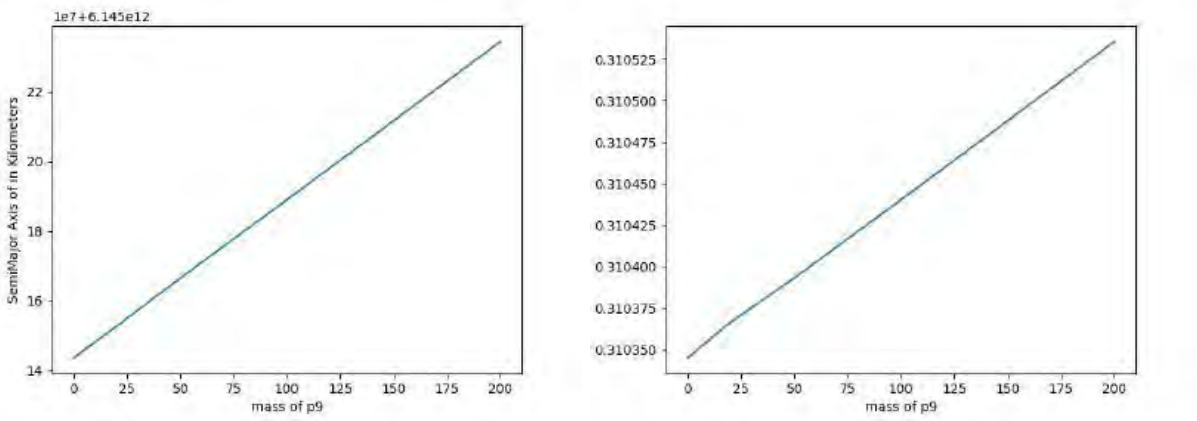


Fig 4: SemiMajor Axis (in km) and eccentricity of Pluto vs Mass of Planet 9 (in Earth's)

Our study showed that the mass of the ninth planet would be directly proportional to the eccentricity, orbital period of Pluto, and the Semi-Major axis of Pluto. All 3 of these conditions were near-linear and increased at very similar slopes.

Even though our data is not completely accurate, it still can represent the relationship between the mass of a hypothetical ninth planet and the objects in the Solar System. The results would help to understand the actual effect on Pluto and other objects in the Kuiper Belt, given any mass of Planet Nine.

Conclusion

We researched the effects of a hypothetical Planet 9, given its mass, on the objects of the Solar System. We used a simulation [4] to collect data for our study and processed it to find Pluto's orbital period, eccentricity, and semimajor axis with 5 separate masses of Planet 9. Our findings were that the mass of Planet Nine had a directly proportional relationship to the eccentricity, orbital period, and the Semi-Major axis of Pluto.

This research can help us redefine what we know about the solar system if the existence of a 9th planet is further considered or confirmed in the future. Further study can also use a similar approach to identify the mass and distance of Planet 9, given the real-time ground truth orbital information of other planets in the Kuiper belt.

Future Studies

Here is a list of some things that could be done to further this research:

- A more precise algorithm, such as Verlet Integration, can be used to calculate the movements of the Solar System objects.
- Add more objects to a simulation of the Solar System, such as the rest of the Kuiper Belt
- More accurate initial parameters should be added to the study, such as the z-axis.
- Use the ground truth of movements observed in our Solar System to reach further conclusions.

Works Cited

1. Batygin, Konstantin, and Michael E. Brown. “EVIDENCE for a DISTANT GIANT PLANET in the SOLAR SYSTEM.” The Astronomical Journal, vol. 151, no. 2, 20 Jan. 2016, p. 22, arxiv.org/pdf/1601.05438.pdf, https://doi.org/10.3847/0004-6256/151/2/22.
2. Brown, Michael E, et al. “A Pan-STARRS1 Search for Planet Nine.” ArXiv (Cornell University), 31 Jan. 2024, https://doi.org/10.48550/arxiv.2401.17977. Accessed 30 Dec. 2024.
3. Naess, Sigurd, et al. “The Atacama Cosmology Telescope: A Search for Planet 9.” The Astrophysical Journal, vol. 923, no. 2, 1 Dec. 2021, p. 224, https://doi.org/10.3847/1538-4357/ac2307. Accessed 25 May 2022.
4. aromanro. “SolarSystem/README.md at Master · Aromanro/SolarSystem.” GitHub, 2016, github.com/aromanro/SolarSystem/blob/master/README.md.
5. Verlet, Loup. “Computer “Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules.” Physical Review, vol. 159, no. 1, 5 July 1967, pp. 98–103, https://doi.org/10.1103/physrev.159.98.

An Overview of Congenital Mutations and Chromosomal Abnormalities, Effect on Cancer and Quality of Life By Nirukta Balaji

Abstract

Congenital mutations and chromosomal abnormalities can significantly influence cancer risk and prognosis. This paper focuses on key genetic alterations, including mutations in TP53 and WT1, and their roles in cancers like neuroblastoma and Wilms' tumor. It illustrates how these genes function under normal conditions, how their mutations disrupt processes like cell growth and death, and the broader impacts on development and quality of life. The study also examines detection methods, from prenatal genetic screening to postnatal sequencing, and how these tools shape outcomes by enabling earlier diagnosis and more precise treatments. By taking a closer look at how specific mutations affect survival rates and cancer progression, the paper sheds light on the biological, medical, and personal implications of living with these conditions. The discussion extends to environmental and lifestyle factors and the ethical and policy challenges of genetic testing. This work aims to provide a deeper understanding of how genetic changes shape health and life beyond the lab.

Understanding Mutations and Their Role in Disease

The human genome is a highly optimized and fine-tuned system, often recognized as a beautiful, machine-like mechanism. This finely tuned machine is sensitive in its construction, where even the smallest change can lead to significant effects. Genes, the building blocks of the genome, are made up of DNA which contains instructions for making RNA which is eventually translated to proteins. These proteins play crucial roles in the cell, such as moving molecules from place to place, building structures, and breaking down toxins. Any change in the nucleotide sequence of one or more genes is defined as a mutation. Variations in genetic structure can lead to observable changes in phenotype and overall health.

Minute changes can have a significant impact, but how does this happen? A textbook example of this is when there are mutations in genes called oncogenes. While oncogenes are commonly understood as a group of genes harboring genetic mutations with the potential to cause cancer, this definition is somewhat loose. Oncogenes are better understood as an individual or group of genes that play crucial roles in cell cycle elements, such as proliferation and cell death. These genes are especially critical in the context of cancer and other diseases. Mutations can be caused by errors in DNA replication or the effects of chemicals and radiation, known as mutagens, which react with DNA and alter or damage the structure of individual nucleotides. All cells possess enzymes whose sole purpose is to repair DNA, known as DNA repair enzymes. These enzymes attempt to minimize the number of mutations that occur. They work in one of two ways: some are pre-replicative, searching the nucleotides for unusual structures, while others are post-replicative, checking newly synthesized DNA for errors and correcting as many as possible. Therefore, a mutation can also be described as a failure of the DNA repair pathways.

This failure is what ultimately allows mutations to remain within a person's cell, otherwise explained as a mutation. When this mutation survives cell replication through the G1 checkpoint, the gene is capable of being passed down to offspring. This review will cover the intricacies of certain genetic abnormalities and their effects on what could manifest itself as childhood cancer.

Insights into Congenital Cancer, Genetics, and Hereditary Risk

Genetic sequencing has become a crucial tool for understanding congenital cancers and all cancers. Based upon technological breaks made in the past 20 years, referred to as Next-generation sequencing, this integral knowledge now allows us to sequence the DNA in cancer cells and comprehend which genes are mutated. It thereby improves our understanding of genetic disease, and furthers the development of genetically targeted medications. By analyzing an individual's genetic risk, preventative measures can be taken before cancer development, ergo-regular screenings and higher health measures.

Beyond their immediate health effects, mutations tied to congenital cancers hold significant hereditary consequences. When said genetic alterations are passed down, a persistent pattern of heightened cancer risk across generations can be created. Family risk assessments and genetic counseling have thus become increasingly valuable tools, providing families with the knowledge to understand and, where possible, reduce these inherited risks. Research focused on the precise mutations involved in congenital cancers has not only deepened insights into the genetic basis of these diseases but has also paved the way for precision medicine advances. By identifying these mutations and analyzing their effects on the genome, scientists can now create targeted therapies and proactive strategies for those at high risk. Through sequencing technology, the broader hereditary impact of congenital cancer becomes clearer, fostering early intervention possibilities and enhancing outcomes for future generations.

The Role of WT1 in Wilms' Tumor: Structure, Function, and Pathogenic Implications

Wilms tumor is a cancer that through greater sequencing technologies has a greater chance of prognosis and response to treatment, Nephroblastoma (Wilms tumor) is a pediatric kidney cancer that primarily affects children aged 2 to 4. It occurs in approximately 1 in every 10,000 children and accounts for 90% of malignant kidney tumors in children, making it one of the most frequently occurring solid tumors of this age group. With a timely diagnosis before the cancer has metastasized, Wilms tumor patients have an overall survival rate of above 90%. Over 80% of those children diagnosed are predicted to survive with an unimpaired quality of life post-treatment.

The gene Wilms' Tumor 1 (WT1), named after the pediatric cancer Wilms tumor, plays a critical role in a range of cellular and physiological processes. WT1 primarily regulates other genes and is involved in various biological processes, such as MAPK signaling, which is essential for cell growth, proliferation, and differentiation. WT1 is also expressed during mammalian embryonic development in systems such as the brain, spinal cord, limbs, diaphragm,

and mesothelial organs. According to, disruption of WT1 can result in abnormalities of the heart, lungs, spleen, and gonadal and renal agenesis. WT1 mutations are also known to play roles in a variety of diseases and syndromes such as Frasier syndrome, WAGR syndrome, Denys-Drash syndrome, breast cancer, lung cancer, and retinoblastoma.

Structurally, WT1 encodes for a transcription factor containing four zinc-finger motifs that preferentially bind to sequences similar to the EGR-1 consensus site. X-ray and NMR crystal structures have revealed that WT1's second, third, and fourth zinc-finger domains insert deeply into the major grooves of DNA, where base-specific interactions occur. The helix is non-canonical, and the fingertips project out of the groove. It is reasoned that the first zinc finger helps secure WT1 to the DNA by optimizing binding affinity. Although structurally similar, mutations in each zinc finger are associated with unique diseases. For example, specific mutations in fingers 2 and 3 are linked to Denys-Drash syndrome and nephrotic syndrome, respectively. Additionally, the presence of Wilms' tumor and Fragile X syndrome is associated with constitutional deletions within chromosome 11 band p13.

WT1's involvement in cell proliferation and growth can be represented through the increased stimulation or lack of regulation of other genes. This overstimulation can result in premature menstrual symptoms, fertility trouble, and other issues. The involvement of WT1 with the MAPK pathway should also be considered, in that the upregulation of said pathway can result in cellular apoptosis or the magnification of pro-oncogenic input from other signals.

As a rule, WT1 is a tumor suppressor in relation to nephroblastoma, that in certain cases can function as an oncogene. Promoter regions (A region of DNA where relevant proteins bind to initiate transcription of that gene) for WT1 resembled the sites for certain widely understood transcription factors such as SP1, MAZ, and ZNF219. WT1 is primarily involved in the function of other genes, apparent in numerous biological processes such as MAPK signaling (A crucial cellular pathway, involved in processes such as cell growth, proliferation, and differentiation). Wilm's Tumor 1 is also associated with Axon guidance (a process that has to do with the development of the nervous system, involving the precise pathfinding of axons to their target cells), suggesting the gene's involvement with neural development.

WT1 is a transcription factor, and much evidence supports that it is a transcriptional regressor specifically. Recent evidence has been gathered that WT1 is physically associated with splice factors, and one alternative splice isoform of WT1 with KTS inserted between fingers 3 and 4. This is the same alternative that contains the Lys-Thr-Ser amino acids, the one without said amino acids preferentially associates with splice factors. Regulation of splicing has been newly found to play a crucial role in the development of the genitourinary system, and tumors often arise due to aberrant splicing. The tumors we commonly associate with this phenomenon are, Wilm's Tumor, Glioblastoma, Uterine Sarcoma, and Leukimia (etc.)

Neuroblastoma and the Role of MYCN: Genetic Alterations and Prognostic Implications

A tumor that can often be mistaken for Wilm's Tumor in presentation, Neuroblastoma, can be discerned through the thorough study of constitutional symptoms. The prognosis can be poor but not always in regards to neuroblastoma, children diagnosed before 18 months of age have better chances of survival. There are also cases where the tumor spontaneously resolves without treatment. Despite being the 3rd most common type of childhood cancer, it is still a rare cancer, with only 600 to 1000 cases diagnosed in the U.S. annually.

A transcriptome (the protein-coding part of an organism's genome) enriched in cell cycle genes in neuroblastoma tumors has been associated with a worse prognosis. MYCN, a member of the MYC family of proto-oncogenes, plays a crucial role in this context. These proto-oncogenes can induce cancer in normal cells when mutated, typically in a dominant manner. As a transcription factor, MYCN is involved in controlling fundamental processes during embryonic development, including cell growth, proliferation, differentiation, apoptosis, angiogenesis, stem cell maintenance, and immune responses. Deregulation of MYCN signaling is associated with the development of several cancers, particularly childhood cancers such as neuroblastoma, medulloblastoma, and rhabdomyosarcoma.

In the case of neuroblastoma, patients with poor prognosis have been found to have tumor cell lines that contain several genetic alterations, such as the amplification of MYCN, a loss of heterozygosity on the short arm of chromosome one (indicating that both copies of a specific region in chromosome one have lost one of their alleles), extra copies of chromosomes 17 and 7, as well as abnormalities involving chromosome 11. MYCN is important for cell proliferation and migration, as well as holding stem cell homeostasis, with decreased levels being associated with terminal neuronal differentiation. These dangerous conditions can cause neuroblastoma tumors ranging from low-risk tumors with the ability to spontaneously differentiate and regress, to high-risk tumors causing the highest numbers of cancer-related deaths in Infants.

The MYC family of genes is expressed in three specific paralogs—c-MYC, MYCN, and MYCL. These genes have evolved from common ancestral backgrounds and share functional and structural similarities. Each paralog exists on a different chromosome, allowing the genes to be independently regulated and expressed. The MYC paralogs are expressed at different stages of an organism's embryonic development, with their expression levels varying widely depending on the stage. Despite their differences, the three oncogenes carry structural and functional similarities. For example, they all produce proteins with similar functional domains, specifically trans-activating and DNA-binding domains. Trans-activating domains enable proteins to activate the transcription of other genes.

Genetic Mutations in Oncogenes and Tumor Suppressors: Implications for Congenital Cancer Susceptibility and Patient Prognosis

In plain terms, TP53 is the gene that produces a protein that controls cell division and cell death. In what are less plain terms, but understandable nonetheless, TP53 is a gene that makes a protein found inside the nucleus of cells that plays a crucial role in controlling cell division, and

changes in said gene may cause cancer cells to grow and spread in the body, said changes affect conditions such as Li Fraumeni Syndrome and many types of cancers such as ovarian, esophageal, colorectal, head and neck, larynx, and lung cancers to about 5% in primary leukemia, sarcoma, testicular cancer, malignant melanoma, and cervical cancer. Now that it is sensible that this gene is one of the most popular in cancer research, we will discuss its regular methods of function. The gene is a sensor of numerous genotoxic, oncogenic, and non-genotoxic stress. It suppresses the replication and survival of stressed or malfunctioning cells and is the central point of selection pressure (an agent of differential mortality or fertility that tends to make a population change genetically) in tissues exposed to carcinogens.

While somatic mutations in the TP53 gene are one of the most frequent alterations in human cancers, we will discuss germline mutations with this gene and their significance regarding Li Fraumenii syndrome. Most mutations are single-base substitutions distributed through the coding sequence, with the inactivation of single-base substitution and loss of alleles, obtaining this inactivation through viral or cellular proteins playing a major role in specific cancers. TP53 is also highly polymorphic in coding and noncoding regions and some of these polymorphisms have been shown to increase cancer susceptibility and to modify cancer phenotypes in TP53 mutation carriers. TP53 is different from other tumor suppressors in that TP53 mutations are often missense and cause single amino acid changes in many different positions. Therefore mutations are diverse in sequence context and structural impact and it's nearly impossible to identify mutation patterns in relation with cancer type.

The nine most common mutations of TP53 are R175H, R248Q, R273H, R248W, R273C, R282W, G245S, R249S, and Y220. The mutations mentioned above account for about 30% of all P53 cancer-driving mutations, a majority of them being DNA contact mutants.

Li Fraumeni syndrome is inherited by the autosomal dominant pattern, meaning one copy of the altered gene in each cell is sufficient to increase the risk of developing cancer. It is a rare disorder that is thought to occur in one in every 5000 people to 1 in every 20,000 people worldwide. Nearly ¾ of families with Li Fraumeni syndrome have germline mutations in the TP53 gene. Most who inherit Li Fraumeni syndrome have an altered copy of the gene from an affected parent, whereas in 7-20% of cases, the altered gene is caused by a new mutation in the gene that occurred in the formation of reproductive cells or very early in developing fetuses. For cancer to develop with Fraumeni syndrome, a mutation involving the second copy of the TP53 gene must occur in the body's cells during a person's lifetime. Environmental factors, such as radiation or chemicals can cause this mutation. It can also occur spontaneously, without any external trigger. Cells with two altered copies of this gene do not make functional TP53 protein which causes tumors to develop. – Because this syndrome disables the ability of the body to identify mutations in their entirety it is understandable that often cancers metastasize and are harder to fight due to the body's compromised immunity.

Conclusion

Congenital mutations and chromosomal abnormalities are censorious in the understanding of cancer susceptibility and its impact on the quality of one's life. Through our summarization of works regarding TP53, WT1, and MYCN genes, it can be seen that small genetic alterations can lead to life-altering effects on cellular processes such as proliferation and apoptosis, causing elevated risk of aggressive cancers. As such, an understanding of these genetic changes should be delved into, and knowledge made clear to the common public. After all, medicine is made for all, and does it not help everyone boundlessly to have a basis of knowledge? Despite the beliefs all have in evidence-based research, if that evidence is not understood, and made comprehensible in an easily digestible manner, it often is the case it is not understood. A common belief that should be had is the pursuit of knowledge for those who need it, and where medicine is regarded, the demographic automatically becomes everyone.

Options for detecting genetic susceptibility are vital in giving adequate help and information to parents. Identifying at-risk individuals early, and allowing for timely actions, has been proven to help prognosis in most cases. This process is largely facilitated by advanced screening technologies that help identify mutations that would otherwise have remained undetected.

As science continues to disentangle this web of congenital alterations, it is transparent that showcasing genetic insights and targeted care for children is crucial in a field that requires patient care. Further research into these topics shows promise in improving treatment options to significantly improve the quality of life for affected individuals and their loved ones.

Works Cited

Wagner, Kay-Dietrich, et al. "The Complex Life of WT1." *Journal of Cell Science*, vol. 116, no. 9, 1 May 2003, pp. 1653–1658, https://doi.org/10.1242/jcs.00405.

https://www.pnas.org/doi/abs/10.1073/pnas.0901591106

Kim, M. K.-H., et al. "An Integrated Genome Screen Identifies the Wnt Signaling Pathway as a Major Target of WT1." *Proceedings of the National Academy of Sciences*, vol. 106, no. 27, 22 June 2009, pp. 11154–11159, https://doi.org/10.1073/pnas.0901591106.

https://www.ncbi.nlm.nih.gov/gene/7490

"WT1 WT1 Transcription Factor [Homo Sapiens (Human)] - Gene - NCBI." *Www.ncbi.nlm.nih.gov*, www.ncbi.nlm.nih.gov/gene/7490.

https://www.sciencedirect.com/science/article/abs/pii/S002228360700945X

Stoll, Raphael, et al. "Structure of the Wilms Tumor Suppressor Protein Zinc Finger Domain Bound to DNA." *Journal of Molecular Biology*, vol. 372, no. 5, Oct. 2007, pp. 1227–1245, https://doi.org/10.1016/j.jmb.2007.07.017. Accessed 27 May 2021.

https://www.pnas.org/doi/epdf/10.1073/pnas.91.9.3554 - protein structure WT1 (specific mutation)

Varanasi, R., et al. "Fine Structure Analysis of the WT1 Gene in Sporadic Wilms Tumors." *Proceedings of the National Academy of Sciences*, vol. 91, no. 9, 26 Apr. 1994, pp. 3554–3558, www.pnas.org/content/91/9/3554, https://doi.org/10.1073/pnas.91.9.3554. Accessed 14 Jan. 2021.

https://medlineplus.gov/genetics/gene/wt1/ - how Wt1 affects other genes

"WT1 Gene: MedlinePlus Genetics." *Medlineplus.gov*, medlineplus.gov/genetics/gene/wt1/.

https://biosignaling.biomedcentral.com/articles/10.1186/s12964-021-00741-y - nuclear localization

Lu, Juane, et al. "Types of Nuclear Localization Signals and Mechanisms of Protein Import into the Nucleus." *Cell Communication and Signaling*, vol. 19, no. 1, 22 May 2021, biosignaling.biomedcentral.com/articles/10.1186/s12964-021-00741-y, https://doi.org/10.1186/s12964-021-00741-y.

https://www.mdpi.com/2073-4425/8/4/113 – general function MYCN – protein structure

Ruiz-Pérez, María Victoria, et al. "The MYCN Protein in Health and Disease." *Genes*, vol. 8, no. 4, 30 Mar. 2017, p. 113, https://doi.org/10.3390/genes8040113.

https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2013.00001/full - general function tumor cell lines (genetic alteration)

He, Stanley, et al. "MYCN and the Epigenome." *Frontiers in Oncology*, vol. 3, 1 Jan. 2013, https://doi.org/10.3389/fonc.2013.00001. Accessed 24 Aug. 2024.

https://www.nature.com/scitable/topicpage/proto-oncogenes-to-oncogenes-to-cancer-883/ - proto oncogenes

Chial, Heidi. "Proto-Oncogenes to Oncogenes to Cancer | Learn Science at Scitable." *Nature.com*, 2008, www.nature.com/scitable/topicpage/proto-oncogenes-to-oncogenes-to-cancer-883/.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6390367/ - progenitor cells

Zakrzewski, Wojciech, et al. “Stem Cells: Past, Present, and Future.” *Stem Cell Research & Therapy*, vol. 10, no. 1, 26 Feb. 2019, stemcellres.biomedcentral.com/articles/10.1186/s13287-019-1165-5.
https://www.ncbi.nlm.nih.gov/gene/4613 - pathways MYCN
“MYCN MYCN Proto-Oncogene, BHLH Transcription Factor [Homo Sapiens (Human)] - Gene - NCBI.” *Www.ncbi.nlm.nih.gov*, www.ncbi.nlm.nih.gov/gene/4613.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4117150/- SHH pathways
Choudhry, Zia, et al. “Sonic Hedgehog Signalling Pathway: A Complex Network.” *Annals of Neurosciences*, vol. 21, no. 1, 1 Jan. 2014, https://doi.org/10.5214/ans.0972.7531.210109.
https://journals.biologists.com/dev/article/146/10/dev166892/48815/Biochemical-mechanisms-of-vertebrate-hedgehog — SHH biochemical
Kong, Jennifer H., et al. “Biochemical Mechanisms of Vertebrate Hedgehog Signaling.” *Development*, vol. 146, no. 10, 15 May 2019, p. dev166892, https://doi.org/10.1242/dev.166892.
https://www.frontiersin.org/articles/10.3389/fgene.2019.01103/full#:~:text=Among%20the%20three%20Hh%20genes,%2C%20facial%20structure%2C%20and%20limbs. - more effects of SHH pathway mutations
Sasai, Noriaki, et al. “Hedgehog Signal and Genetic Disorders.” *Frontiers in Genetics*, vol. 10, no. 1103, 8 Nov. 2019, www.ncbi.nlm.nih.gov/pmc/articles/PMC6856222/, https://doi.org/10.3389/fgene.2019.01103.
https://genesdev.cshlp.org/content/22/18/2454.short#:~:text=Germline%20mutations%20that%20subtly%20affect,multiple%20forms%20of%20human%20cancer. - germline/somatic?
Varjosalo, M., and J. Taipale. “Hedgehog: Functions and Mechanisms.” *Genes & Development*, vol. 22, no. 18, 15 Sept. 2008, pp. 2454–2472, https://doi.org/10.1101/gad.1693608.
https://www.frontiersin.org/articles/10.3389/fgene.2019.01103/full - biochemistry SHH
Sasai, Noriaki, et al. “Hedgehog Signal and Genetic Disorders.” *Frontiers in Genetics*, vol. 10, no. 1103, 8 Nov. 2019, www.ncbi.nlm.nih.gov/pmc/articles/PMC6856222/, https://doi.org/10.3389/fgene.2019.01103.
https://academic.oup.com/bjd/article-abstract/152/1/43/6636303?login=false - bcc (protein structure change)
https://medlineplus.gov/geneticdisorders.html#:~:text=Single%2Dgene%20disorders%2C%20where%20a,structures%20that%20hold%20our%20genes. - intro page
Medline Plus. “Genetic Disorders.” *Medlineplus.gov*, National Library of Medicine, 7 Dec. 2016, medlineplus.gov/geneticdisorders.html.
https://www.ncbi.nlm.nih.gov/books/NBK557691/#:~:text=A%20chromosomal%20abnormality%2C%20or%20chromosomal,%2C%20sex%20chromosomes%2C%20or%20both. - what is a chromosomal abnormality?
Queremel Milani, Daniel A., and Prasanna Tadi. “Genetics, Chromosome Abnormalities.” *PubMed*, StatPearls Publishing, 2020, www.ncbi.nlm.nih.gov/books/NBK557691/.

https://www.genome.gov/genetics-glossary/Gene#:~:text=Definition&text=The%20gene%20is%20considered%20the,differing%20functions%20within%20the%20body. - What is a gene?
Green, Eric. “Gene.” *Genome.gov*, National Human Genome Research Institute, 2019, www.genome.gov/genetics-glossary/Gene.
https://my.clevelandclinic.org/health/body/24949-oncogenes - what is an oncogene?
Cleveland Clinic. “Oncogenes.” *Cleveland Clinic*, 1 May 2023, my.clevelandclinic.org/health/body/24949-oncogenes.
https://www.ncbi.nlm.nih.gov/books/NBK21114/#:~:text=Mutations%20result%20either%20from%20errors,the%20structures%20of%20individual%20nucleotides. - how mutations?
Brown, Terence A. “Mutation, Repair and Recombination.” *Nih.gov*, Wiley-Liss, 2002, www.ncbi.nlm.nih.gov/books/NBK21114/.
https://www.sciencedirect.com/science/article/abs/pii/S0027510797001796 - How do chromosomal abnormalities arise before birth- in utero
McFadden, Deborah E, and J.M Friedman. “Chromosome Abnormalities in Human Beings.” *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, vol. 396, no. 1-2, Dec. 1997, pp. 129–140, https://doi.org/10.1016/s0027-5107(97)00179-6.
https://www.sciencedirect.com/topics/medicine-and-dentistry/constitutional-chromosome-abnormality#:~:text=Constitutional%20chromosome%20abnormalities%20are%20present,offspring%20with%20a%20chromosome%20abnormality. - chromosomal rearrangements vs chromosomal abnormalities
Giersch, A.B.S. “Introduction to Cytogenetics.” *Pathobiology of Human Disease*, 2014, pp. 3304–3310, https://doi.org/10.1016/b978-0-12-386456-7.06401-7. Accessed 7 Nov. 2020.
https://www.sciencedirect.com/science/article/abs/pii/S0027510797001796 - chromosomal abnormality
McFadden, Deborah E, and J.M Friedman. “Chromosome Abnormalities in Human Beings.” *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, vol. 396, no. 1-2, Dec. 1997, pp. 129–140, https://doi.org/10.1016/s0027-5107(97)00179-6.
https://aacrjournals.org/cancerres/article/59/7_Supplement/1747s/506060/Multiple-Roles-for-the-Wilms-Tumor-Suppressor-WT11 - Why does suppression of WT1 cause cancer?
Davies, Rachel, et al. “Multiple Roles for the Wilms’ Tumor Suppressor, WT11.” *Cancer Research*, vol. 59, no. 7_Supplement, Apr. 1999, pp. 1747s1751s, aacrjournals.org/cancerres/article/59/7_Supplement/1747s/506060/Multiple-Roles-for-the-Wilms-Tumor-Suppressor-WT11. Accessed 13 Dec. 2024.
https://medlineplus.gov/genetics/gene/tp53/ - how does tp53 usually work?
“TP53 Gene: MedlinePlus Genetics.” *Medlineplus.gov*, medlineplus.gov/genetics/gene/tp53.

https://www.cancer.gov/publications/dictionaries/cancer-terms/def/tp53-gene - What is TP53
National Cancer Institute.
"Https://Www.cancer.gov/Publications/Dictionaries/Cancer-Terms/Def/Tp53-Gene." *Www.cancer.gov*, 2 Feb. 2011, www.cancer.gov/publications/dictionaries/cancer-terms/def/tp53-gene.
https://www.sciencedirect.com/science/article/abs/pii/S0300908401013566 - normal function of TP53
Guimaraes, D.P., and P. Hainaut. "TP53: A Key Gene in Human Cancer." *Biochimie*, vol. 84, no. 1, Jan. 2002, pp. 83–93, https://doi.org/10.1016/s0300-9084(01)01356-6.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2827900/#:~:text=Somatic%20mutations%20in%20the%20TP53,distributed%20throughout%20the%20coding%20sequence. - germline/somatic?
Olivier, M., et al. "TP53 Mutations in Human Cancers: Origins, Consequences, and Clinical Use." *Cold Spring Harbor Perspectives in Biology*, vol. 2, no. 1, 11 Nov. 2009, pp. a001008–a001008, www.ncbi.nlm.nih.gov/pmc/articles/PMC2827900/, https://doi.org/10.1101/cshperspect.a001008.
https://medlineplus.gov/genetics/condition/li-fraumeni-syndrome/#:~:text=Li%2DFraumeni%20syndrome%20is%20inherited,gene%20from%20an%20affected%20parent%20. - Li Fraumeni syndrome
"Li-Fraumeni Syndrome: MedlinePlus Genetics." *Medlineplus.gov*, 1 June 2020, medlineplus.gov/genetics/condition/li-fraumeni-syndrome/.
https://www.icr.ac.uk/news-features/latest-features/how-the-human-genome-project-shook-the-world-of-cancer-research#:~:text=Today%2C%20we%20can%20sequence%20the,helping%20develop%20genetically%20targeted%20medicines. - how sequencing changed the understanding of cancer
The Institute of Cancer Research. "How the Human Genome Project Shook the World of Cancer Research - the Institute of Cancer Research, London." *Www.icr.ac.uk*, 15 Feb. 2021, www.icr.ac.uk/news-features/latest-features/how-the-human-genome-project-shook-the-world-of-cancer-research.
https://www.mdpi.com/2073-4425/15/2/219 - sequencing and why is it helpful?
Barili, Valeria, et al. "Genetic Basis of Breast and Ovarian Cancer: Approaches and Lessons Learnt from Three Decades of Inherited Predisposition Testing." *Genes*, vol. 15, no. 2, 8 Feb. 2024, pp. 219–219, https://doi.org/10.3390/genes15020219.
https://www.dana-farber.org/newsroom/news-releases/2023/dana-farber-cancer-institute-showcases-research-at-the-65th-american-society-of-hematology-ash-annual-meeting-and-exposition - hereditary (sequencing) '
"Dana-Farber Cancer Institute Showcases Research at the 65th American Society of Hematology (ASH) Annual Meeting and Exposition | Dana-Farber Cancer Institute." *Dana-Farber.org*, 6 Dec. 2023, www.dana-farber.org/newsroom/news-releases/2023/dana-farber-cancer-institute-showcases-research-at-the-65th-american-society-of-hematology-ash-annual-meeting-and-exposition. Accessed 13 Dec. 2024.
https://www.yalemedicine.org/conditions/neuroblastoma#:~:text=The%20prognosis%20is%20poor%20for,a%20greater%20chance%20of%20survival. - neuroblastoma prognosis
"Neuroblastoma." *Yale Medicine*, www.yalemedicine.org/conditions/neuroblastoma.

Proposing a CAR T-cell Therapy of Philadelphia Chromosome-Positive B-Cell Acute Lymphoblastic Leukemia (Ph+ B-ALL) via CD22 Apoptotic Pathway

By Nehal R. Revuri

Abstract

Philadelphia chromosome-positive B-cell acute lymphoblastic leukemia (Ph+ B-ALL) is a particularly aggressive subtype of B-ALL, originating in the lymphocytic cells of the blood and bone marrow. The excessive proliferation of these cancerous B-cell lymphoblasts disrupts the normal production of functional and healthy white blood cells (WBCs), red blood cells (RBCs), and platelets, compromising the immune system. Ph+ B-ALL accounts for approximately 25-30% of all adult B-ALL cases and is often associated with a poor prognosis, even after being treated with chemotherapy (Liam et al., 2022; Terwilliger & Abdul-Hay, 2017). Despite significant advancements in treatment (namely the CD22-targeting ADC, inotuzumab ozogamicin), hepatotoxicity and loss of CD22 expression over time remain formidable limitations to their effective use (Kantarjian, 2014). Effective solutions for this condition continue to elude scientists. In this proposal, we discuss a novel chimeric antigen receptor T-cell (CAR T-cell) system that targets the surface antigen CD22 to treat Ph+ B-ALL, increasing specificity, lowering the likelihood of cytokine release syndrome (CRS), reducing relapse rates, and potentially promoting complete remission in patients.

Introduction

Ph+ B-ALL is a type of bone marrow cancer characterized by the rapid proliferation of abnormal lymphoblasts. The translocation between chromosomes 9 and 22 results in the formation of the Philadelphia chromosome and, subsequently, the BCR-ABL1 fusion gene. This gene produces a tyrosine kinase protein that drives leukemogenesis, resulting in the rapid and uncontrolled division of lymphoblasts. Ph+ B-ALL is observed in 25-30% of adult ALL cases and around 5% of pediatric cases (Terwilliger & Abdul-Hay, 2017). Patients harboring this genetic translocation respond poorly to conventional therapies, namely chemotherapy, with a 15% complete remission (CR) rate, making it a challenging subtype to treat. This aggressive leukemia arises from the uncontrollable division of lymphoblasts, which invade the bone marrow and interfere with red blood cell production, posing challenges to the survival and growth of other types of cells. Ph+ B-ALL can be triggered by factors such as antineoplastic agents, ionizing radiation, benzene exposure, Hodgkin's lymphoma, and multiple myeloma. If at least 20% of the bone marrow consists of these immature cells, it is effectively considered leukemia. This leads to symptoms like fatigue, infections, and mucosal bleeding (Chennamadhavuni et al., 2023; Puckett & Chan, 2023; Liam et al., 2022).

Ph+ B-ALL is caused by chromosomal translocations involving the BCR-ABL1 gene, which result in the formation of a fusion protein capable of driving leukemogenesis. Chromosomal translocation drastically varies from point mutations, involving entire sections of chromosomes being switched out with each other. In this case, the resulting fusion genes produce

the dysfunctional BCR-ABL1 fusion protein that activates multiple signaling pathways, promoting the survival and proliferation of leukemic cells. Due to the implications caused by this chromosomal translocation and the fact that entire sections of chromosomes are switched, the effect of these translocations is currently irreversible. It is for these reasons that a permanent solution utilizing a retrovirus cannot be used. In the case of a simple mutation, a retrovirus packaged for mutated cells can effectively rewrite genetic material and spread to other infected cells, effectively correcting the mutation; however, in the case of full chromosomal translocations, a retrovirus is simply not enough as it cannot rewrite complete chromosomes (Byun et al., 2017; Canoy et al., 2022). This brings us to the conclusion that targeted immunotherapy, such as CAR T-cell therapy, may offer a promising approach to treating Ph+ B-ALL.

Overview of CAR T-cell Therapy

CAR T-cell (chimeric antigen receptor T-cell) therapy is a cutting-edge immunotherapy designed to treat cancer by harnessing the body's own immune system. The therapy involves a four-step process: collection, engineering, expansion, and infusion. First, T-cells are collected from the patient's blood through a procedure called leukapheresis. During this process, blood is drawn, T-cells are separated out, and the remaining blood components are returned to the patient. The isolated T-cells are then genetically modified in the laboratory to express chimeric antigen receptors (CARs) using viral vectors, such as retroviruses or lentiviruses, which integrate the CAR gene into the T-cells' DNA. This genetic modification allows the T-cells to recognize and attack cancer cells more effectively (*CAR T Cells: Engineering Immune Cells to Treat Cancer*, 2022).

Once engineered, the T-cells are expanded in culture to produce a large quantity of CAR T-cells. These expanded cells are tested to ensure successful CAR expression before being infused back into the patient's bloodstream. The infused CAR T-cells then target specific antigens present on the surface of cancer cells, marked by the CD, or "cluster of differentiation," nomenclature, which refers to a system used for the identification and characterization of cell surface molecules present on leukocytes (Major et al., 2006). By binding to these antigens, CAR T-cells can selectively destroy cancerous cells while minimizing harm to healthy cells, thus reducing off-target effects and enhancing treatment efficacy (Reddy et al., 2021).

A CAR T-cell therapy to tackle Ph+ B-ALL would be effective in that it would be much more selective than conventional therapies, as proved by various studies in the past decade. However, one limitation remains, although it may have been reduced. To begin, CAR T-cells may still target healthy cells as long as the antigen is present on them. This can lead to B-cell aplasia, a condition that is classified by a low B-cell count in a patient. Although optimal administration of intravenous immunoglobulin (IVIG) would help to tackle this problem by supporting patient immunity, challenges like volume overload and allergic reactions may still exist when using CAR T-cell and IVIG therapies in combination (Sterner & Sterner, 2021).

Additionally, the effectiveness of CAR T-cell therapy can be compromised if the cancer cells lose or downregulate the target antigen over time, a process known as antigen loss (Mishra et al., 2023). To address these challenges and improve the therapy, ongoing research is focused on refining CAR T-cell approaches, enhancing safety profiles, and increasing the applicability of this treatment to other types of cancer. Continued innovation is essential for overcoming these obstacles and expanding the therapeutic potential of CAR T-cell therapy for patients, especially those with Ph+ B-ALL.

Selecting an Antigen

CAR T-cell therapy is an emerging technology and is now part of two recently FDA-approved gene therapy products with acceptable levels of cancer cell toxicity (Wang & Wang, 2023). However, these therapies all target CD19 to initiate their response. Serious cytokine release syndromes and organ toxicities have been observed in patients with a high tumor burden (Murthy et al., 2019). Relatively high expression of CD22 is characteristic of B-ALL cancer cells, making it a viable candidate for most subtypes of B-ALL (Lanza et al., 2020).

A study identifying CD22 expression across various subtypes of B-ALL, including Ph+ B-ALL, found that InO demonstrated high efficacy in comparison to standard chemotherapy. Complete remission (CR), the term used to describe a complete disappearance of all signs and symptoms of cancer post-therapy, is characterized by normal counts of blood cells post-leukemia, including white blood cells, red blood cells, and platelets (*Definition of Complete Remission - NCI Dictionary of Cancer Terms*, n.d.). The study found that InO demonstrated superior CR rates when compared to chemotherapy (73% vs. 15%). Moreover, 63% of patients achieved minimal residual disease (MRD) negativity, indicating unmeasurable traces of cancer cells left behind and illustrating the efficacy of a CD22-targeting therapy. However, the study also uncovered that CD22 expression is variable and there remains a lack of literature regarding its expression on Ph+ B-ALL, meaning that CD22 expression should not be directly correlated to the significantly positive results of the study. Although the study indicates that CD22 intensity alone is not sufficient to predict therapeutic response, CD22 positivity still influences clinical response, making it a viable target for any therapy involving Ph+ B-ALL (Liu et al., 2024).

Due to the evasive abilities and limited data concerning this subtype of B-ALL, considering dual, tri, and quad CAR T-cell systems might also be crucial for increasing specificity and killing cancer cells more efficiently while also reducing the likelihood of off-target effects and cytokine release syndrome (Saleh et al., 2022; Shah et al., 2019). However, due to limited resources, it remains difficult to properly identify the distance between any given antigens on a Ph+ B-ALL cell. The current inability to identify the location of such antigens and achieve that level of specificity forces overgeneralized solutions to be developed. The “CD” nomenclature also does not indicate relative location, which may render multi-CAR therapies ineffective (Clark et al., 2016). Therefore, for the purposes of this proposal, let’s discuss the design of the CAR for the CD22 antigen alone without considering a dual, tri, or quad system.

There are various characteristics that make CD22 a viable candidate for the treatment of Ph+ B-ALL. The Inovate study highlighted that CD22 was expressed in over 90% of B-ALL (Lanza et al., 2020). In fact, there have already been established therapies targeting this antigen, namely antibodies and immunotoxins (Shah & Sokol, 2021; Sullivan-Chang et al., 2013). Upon ligation with an antibody, CD22 immediately internalizes, allowing cytotoxic attacks to take place in rapid succession (Du et al., 2008). As a surface antigen, it can be bound by both soluble antibodies and immobilized (membrane-bound) antibodies (Janeway et al., 2001).

Although CD19-targeted CAR T-cell therapies are effective against relapsed/refractory (R/R) B-ALL, resistance often arises due to CD19 antigen loss (Cappell & Kochenderfer, 2023). However, CD22, which is also expressed on most B-ALL cells, remains present even after CD19 is lost, even if it is in low quantities (Fry et al., 2017). Therefore, CD22-targeting therapies may be increasingly effective, especially where CD19-targeting therapies have failed. Its current establishment concerning all subtypes of B-ALL, ability to rapidly internalize, persistence compared to antigens like CD19, and efficacy in clinical contexts make it a viable candidate for treating Ph+ B-ALL regardless of its variable expression. Therefore, with its advantages as well as limitations noted, CD22 will be the primary antigen target moving forward.

Components of a CAR

The engineering of the chimeric antigen receptors (CARs) is selective. These modified receptors are capable of integrating into cytotoxic T-cells in order to detect and termiante cancer cells more efficiently. These modified T-cells have engineered components that allow them to detect the receptors of cancerous B-cells more efficiently, enabling their implementation as a potent strategy for the treatment of B-ALL. These receptors can be divided into three regions to interpret their multifaceted function (Zhang et al., 2022). The first is the exterior region, or the antigen-binding region, is known as the ectodomain. The ectodomain consists of the scFv and the hinge and is usually derived or reverse-engineered from an antibody. The transmembrane domain anchors the antigen-binding portion via the hinge, which acts as an intermediary between the scFv and the cell membrane. Finally, the endodomain contains the intracellular domains, which execute the immune response upon binding (Zhang et al., 2017).

The single-chain fragment variable (scFv) is the superficial component of a CAR that is specific in its antigen-binding capabilities. The scFv is composed of the heavy (VH) and light (VL) chains, which differ in their molecular weight. They work together to provide the scFv with its specificity and affinity by working in tandem with each other to form the antigen-binding site. Containing several complementarity-determining regions (CDRs) that directly interact with the antigen, these chains are linked by the linker peptide, which prevents the misfolding or aggregation of the scFv (Kozani et al., 2021). Together, these components form the initial antigen-antibody interaction to begin the signaling cascade that manages cancerous B-cells.

The space in between the scFv and the T-cell membrane is known as the hinge, which is the component that allows greater flexibility of the CAR and allows its integration into the transmembrane domain. Also known as spacers, hinges are particularly crucial in allowing

specificity (which is key in CAR T-cell treatment) because they overcome steric hindrance and provide greater flexibility for the scFv to attach to its corresponding antigen. Many hinge models vary in their size, so a longer hinge may allow the scFv to reach the target antigen and mediate recognition. Depending on the characteristics of the antigen, modifying the hinge is crucial in order to form the connection between the superficial binding portion and the cell membrane and to maintain the strength of that connection (Fujiwara et al., 2020; Zhang et al., 2021).

Because of the CAR's function as an extracellular protein, it must be fixed into the cell membrane for proper function. The transmembrane domain anchors the CAR onto the cell membrane and holds the hinge in place. The proper positioning of the CAR is essential for its recognition and binding to its corresponding antigen (Fujiwara et al., 2020).

Finally, the intracellular signaling domain receives the signal and triggers a signal transduction pathway deep into the cell's interior, thereby amplifying it in a phosphorylation cascade, activating genetic expression of T-cells, and inducing anti-tumor activity. This gene expression varies, often resulting in apoptosis or other methods for the management of cancerous B-cells. The intracellular domains also include the costimulatory domain(s) and the cytokine secretion module, which will be further discussed later (Zhang et al., 2017).

Engineering a CD22-Specific ScFv

A CAR's specificity is based on its scFv, so finding an antibody to begin this reverse-engineering process is critical. A group of researchers led by Dimiter S. Dimitrov discovered two antibodies: m971 and m972. Their study involved the testing of antibodies in different formats: Fab (fragment antigen-binding), IgG1 (immunoglobulin G1), and scFv. The IgG1 format evaluates the entire antibody's binding affinity, including the effects of the Fc (constant) region, which may influence the stability and solubility of the antibody. The Fab format evaluated only the antigen-binding portion as opposed to the entire antibody. Despite HA22's growing antitumor activity, it is not considered a viable antibody template due to its limited binding capability, primarily residing in domain 3. The epitopes containing domains 2 and 3 produced the best results for HA22 binding affinity. The m971 and m972 antibodies, on the other hand, exhibit more effective binding capabilities, exceeding HA22's epitope versatility. The two antibodies reacted to the CD22 antigen from domains 5 to 7. The study's ELISA technique confirmed that their episodes indeed do not overlap, which means that they do not interfere with one another. Therefore, only these two antibodies will be considered in this proposal (Xiao et al., 2009).

The study's results indicated no statistically significant results between either the m971 and m972 antibodies in the IgG1 and scFv formats, but specified a significantly greater binding affinity in m972 compared to m971 in the Fab format. The m972 antibody has also been expressed as a Fab, scFv, and IgG1, highlighting its flexibility in potential therapeutic applications. Furthermore, because they belong to non-overlapping epitopes, they can be used in dual, tri, or quad CAR T-cell systems. The study also confirmed that the antibodies do not interfere with other antibodies currently in clinical development (Xiao et al., 2009).

Now that an effective template has been decided, reverse engineering of the scFv is required to design an effective anti-CD22 antigen-binding portion of the CAR. Using reverse transcription, we would insert the gene of the scFv reverse-engineered from the m972 antibody into the newly transfected T-cells (Ahmad et al., 2012). For protein expression, the verified scFv expression vector will be transfected into a suitable mammalian cell line, namely HEK293 or CHO cells (Ahmad et al., 2012; O'Neill et al., 2023). The purified scFv will then be validated using functional assays like flow cytometry to confirm that it retains the binding specificity and affinity of the original m972 antibody (Gray et al., 2009).

The scFv is the most crucial aspect of this therapy. Therefore, modification and enhancement are necessary to refine the current scFv to be specific to the CD22 antigen. However, due to a lack of established literature, scFv protein design software is limited and does not have much accessibility, preventing the modeling and digital enhancement of this method.

Hinge and Transmembrane Domain Derivation

The following components are all derived from natural proteins, making them inherently compatible with CAR T-cell machinery. The "CD nomenclature" indicates that these molecules' natural evolution process optimizes their stability, signaling efficacy, and interactions with other immune components (Reis, 2022). This natural optimization reduces the need for artificial modification and makes them ideal for use in CAR constructs. Furthermore, the hinge and transmembrane domains both control CAR expression due to their characteristics' impact on overall CAR stability and cytotoxicity (Fujiwara et al., 2020).

The functionality of the hinge (otherwise known as the spacer) is dependent on its ability to provide flexibility and connect the antigen-binding portion (scFv) to the transmembrane domain (which anchors the CAR construct). The choice of the hinge greatly influences the overall movement of the CAR (Fujiwara et al., 2020). For example, implementing a longer hinge may be able to overcome steric hindrance and more effectively bind to larger antigens, while a shorter hinge may improve binding for larger antigens and promote functionality and expression over the long-term (Rafiq et al., 2019). Common choices for the hinge include the CD8α-derived hinge, the IgG1 hinge, and the CD28 hinge. Of these options, the most effective would be the CD8α natural protein-derived hinge. The CD8α hinge provides a balance between flexibility and rigidity, allowing the CAR to adopt different conformations and effectively bind to antigens in various spatial orientations (Chen et al., 2022). This balance is especially significant because there is evidence that the reduced flexibility of CD8α contributes to increased CAR survival and efficacy (Zhang et al., 2021).

Furthermore, there remains a lack of literature regarding CD22 expression in Ph+ B-ALL. Recent research has demonstrated that the CD8α hinge outperforms the CD28 hinge in killing CD22-positive leukemia cells with low antigen density. The study revealed that the CD8α hinge's dynamic exchange properties, including proline cis-trans isomerization, contribute to its effectiveness. The dynamic exchange properties of the CD8α hinge allow it to remain in a state of constant motion, enabling it to adjust to varying interactions within the cellular environment.

The ability of proline to switch between the cis and trans forms also adds flexibility to the hinge (Chen et al., 2022). Finally, CD8α's amino acid count (65) exceeded those for CD28, CD4, and CD3ζ, with amino acid counts of 36, 23, and 9, respectively. Due to the occurrence of a higher amino acid count positively correlating with CAR expression levels over the long-term, CD8α remains the most effective choice (Fujiwara et al., 2020).

Next, the transmembrane domain is a crucial component of the CAR, allowing the antigen-binding portion to become fixated onto the cellular membrane, acting as a bridge between the cytotoxic signal and the CD22 antigen, which has been internalized (Jayaraman et al., 2020). As a result, both the transmembrane domain and the intracellular domains are essential for CAR T-cell proliferation and survival. The gold standard for this component is either CD8a or CD28, which have been extensively researched and determined to be effective choices. Despite the efficacy of both in the literature, CD8α is known for its extensive usage within clinical trials, demonstrating its significant success, particularly in therapies targeting CD19 (Fujiwara et al., 2020). Furthermore, due to both, the transmembrane and hinge regions being adjacent to each other on the CAR construct, deriving both from the same natural protein (CD8α) would help to reduce steric hindrance, minimize the risk of structural incompatibility, and ensure functional efficacy of the CAR (Santos et al., 2022).

Engineering the Intracellular Domains

The intracellular domains (primary signaling, costimulatory, and cytokine-secreting) must be selected to effectively couple with the transmembrane domain (Ahmad et al., 2022). It is also important to note that many costimulatory domains require ligands to be presented by antigen-presenting cells (APCs), a standard that may not always be met, especially under Ph+ tumor burden (Weinkove et al., 2019). Therefore, the primary signaling, costimulatory, and cytokine-secreting domains should cooperate with the transmembrane domain, produce an effective cytotoxic signal, and improve the CAR T-cell's proliferation, survival, and signal activation. Therefore, the most suitable choice for the primary signaling domain is CD3ζ, a common choice for most CAR T-cell therapies due to its ability to couple well with various other domains and its capability to rapidly initiate phosphorylation cascades upon antigen recognition. The role of this domain is to provide the primary cytotoxic signal that does not require any ligand presence and effectively promotes cytokine production and T-cell activity. The CD3ζ signaling domain contains multiple immunoreceptor tyrosine-based activation motifs (ITAMs), leading to the activation of various downstream signaling pathways, such as the MAPK, NF-κB, and NFAT pathways. This increases the strength and duration of a signal while boosting T-cell survival (Sterner & Sterner, 2021).

Most costimulatory domains, however, do require a ligand to be expressed on APCs. This does not rule them out of the therapy because CD22 does not express these ligand families, but rather, APCs like dendritic cells, macrophages, and B cells would need to interact with CAR T-cells for the costimulatory domain to function. Given this, the best choice for the costimulatory domain is CD28. Aside from the primary cytotoxic signal initiated by CD3ζ, CD28 provides an

additional, costimulatory signal, requiring a CAR T-cell's efficient interaction with host APCs. The presence of CD28 in the CAR construct ensures that T-cells are effectively activated and persist longer, even if CD22 does not directly interact with CD28 or its ligands. This costumulatory domain is, however, usually coupled with the 4-1BB costimulatory domain, which will be discussed later (Harrison et al., 2021).

Finally, cytokine-secreting domains function in systems of autocrine or paracrine signaling to ensure homeostasis of the host and subsequent regulation of immune responses (Morán et al., 2013). Therefore, choosing an effective cytokine-secreting domain is crucial for enhanced anti-tumor activity. Of the options, the most effective would be to implement an IL-12 secretion domain. IL-12 is known for its ability to activate both T-cells and natural killer (NK) cells, promoting a strong anti-tumor response. Although there is no specific "domain" or "region" for cytokine secretion, their release can be activated downstream of the CD3ζ primary signaling domain's signal initiation, an event that can be replicated by including sequences that drive IL-12 secretion when engineering the CAR's intracellular domains. The sequences should include regulatory operators to control the release of cytokines and modify their sensitivity in response to CD3ζ signal initiation (Agliardi et al., 2021).

Linking the CAR

In order to link the components of the CAR, the gene sequences for each of the components in this approach (see Fig. 1) must be cloned into an expression vector, and expressed in the T-cells. Expressional analysis and functional assays would then be performed to ensure proper execution of vital CAR T-cell functions, including cytokine production and cytotoxicity (Srivastava & Riddell, 2015).

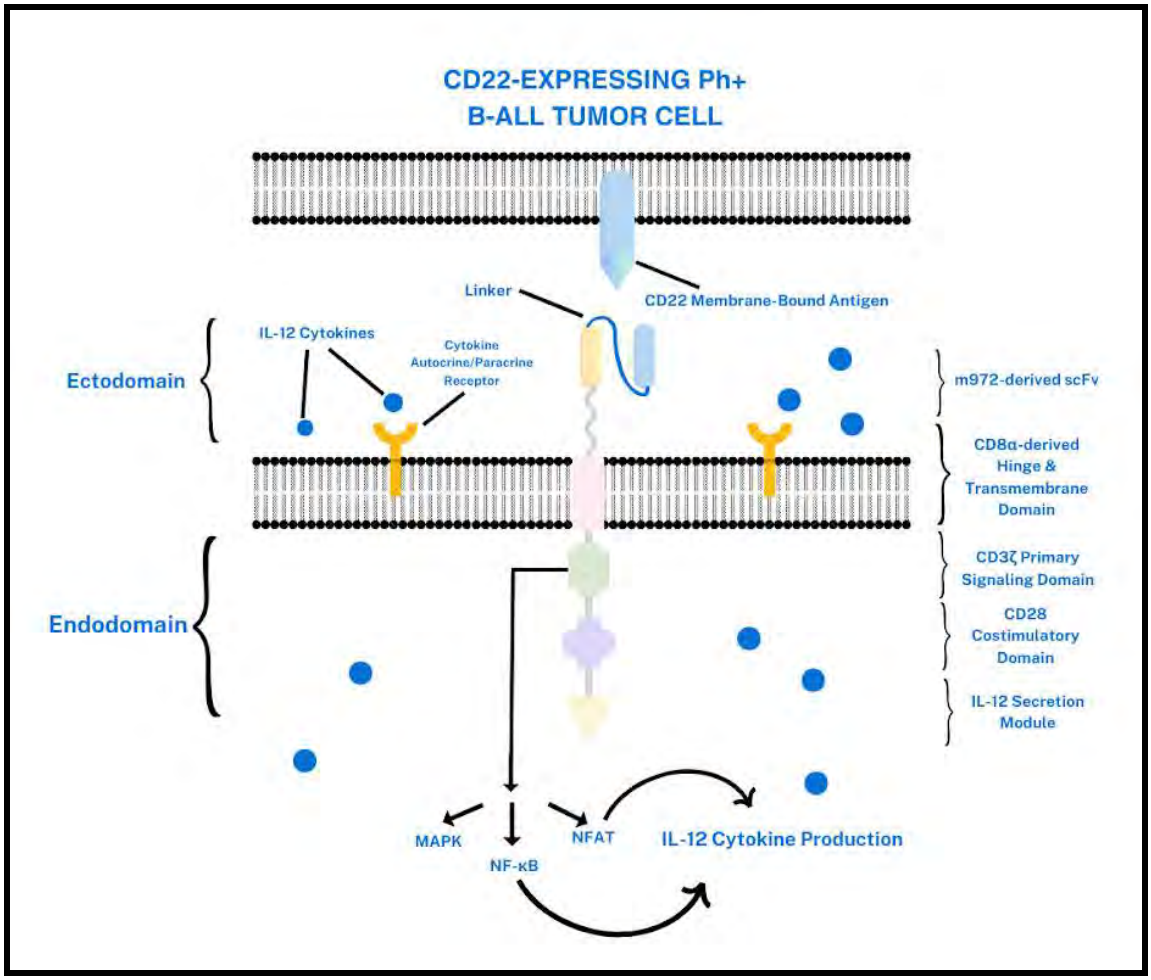


Figure 1: *Visual Model of Signal Initiation in the Membrane-Bound CAR*. This diagram depicts the initial signaling events triggered by the membrane-bound chimeric antigen receptor (CAR). Schematic inspired National Cancer Institute's CAR T-cell graphic (*CAR T Cells: Engineering Immune Cells to Treat Cancer*, 2022).

Apoptotic Activation: Pathway

The main cytotoxic pathway initiated by the CD3ζ primary signaling domain involves the simultaneous release of peforins and granzymes (Holzinger & Abken, 2019). Perforins are released to open pores within the target cell's membrane, and grazymes enter through these holes to carry out apoptosis (Voskoboinik et al., 2015). The binding of the m972-derived scFv initiates the signal that travels to the CD3ζ, initiating the mitogen-activated protein kinase (MAPK), nuclear factor-κB (NF-κB), and nuclear factor of activated T-cells (NFAT) pathways, each either increasing the proliferation and survival of CAR T-cells, improving immune response and communication, or enhancing IL-12 cytokine release (Oeckinghaus et al., 2011). CD28 costimulatory domain initiation leads to activation of the CAR T-cell's phosphatidylinositol 3-kinase/protein kinase B (PI3K/Akt) (see Fig. 2) (Weinkove et al., 2019).

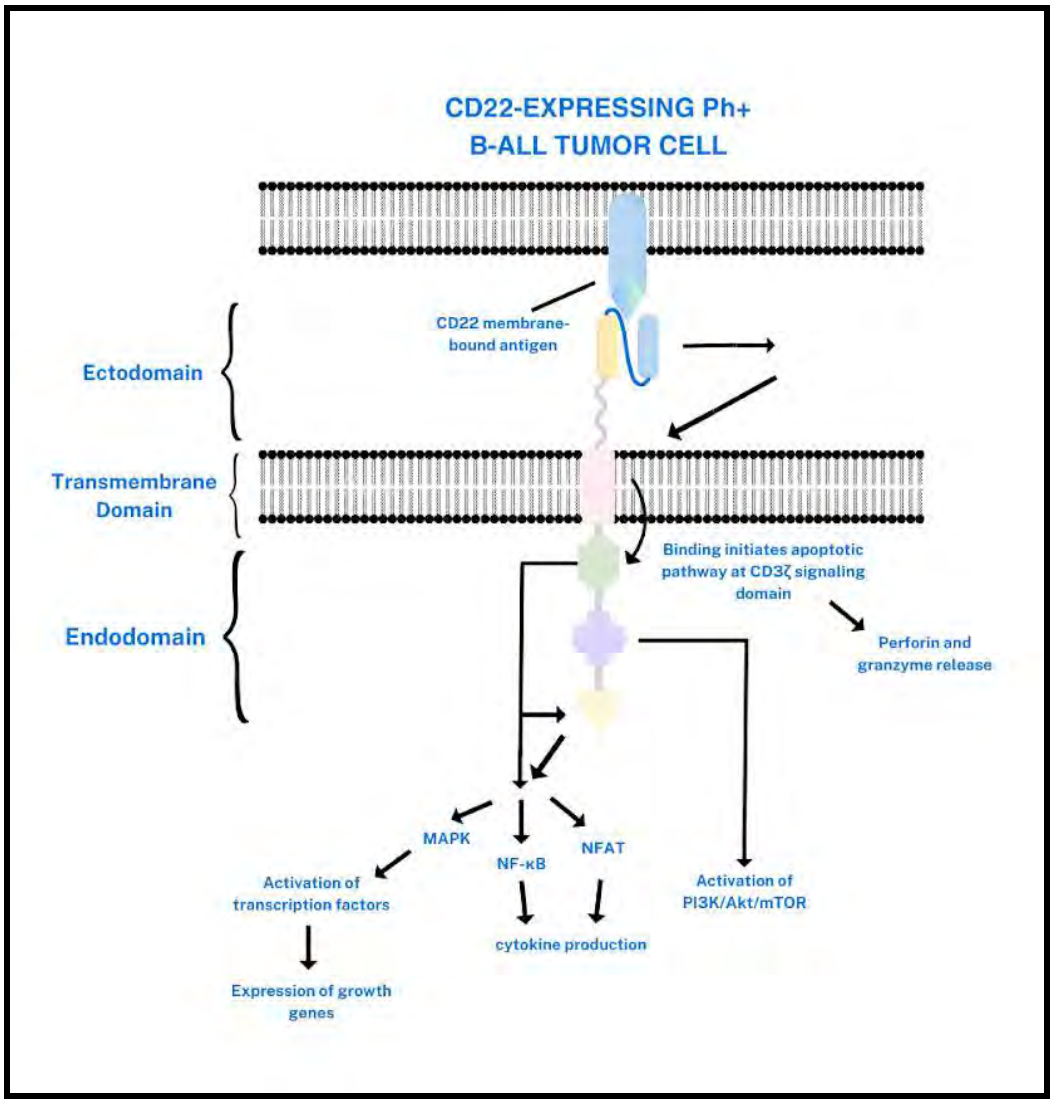


Figure 2: *Schematic of T-Cell Growth, Survival, Cytotoxic Pathways, and Cytokine Production*. This representation illustrates the key pathways involved in T-cell growth, survival, cytotoxic activity, and cytokine production.

In Ph+ B-ALL, the BCR ABL chromosomal translocation creates an abnormal chromosome called the Philadelphia chromosome. Resulting from the translocation between the breakpoint cluster region (BCR) gene on chromosome 9 and the ABL proto-oncogene 1 (ABL1) gene on chromosome 22, this mutation produces a fusion gene called BCR-ABL1, which creates a protein that causes leukemia cells to grow and divide uncontrollably (Kang et al., 2016). In cases of irreversible gene irregularities such as this, synthetic drugs or agents may be utilized to inhibit affected growth pathways (Anand et al., 2023). Although the inhibition of downstream over-proliferation pathways like PI3K/Akt/mTOR through synthetic drug ingestion has proven to be effective, CAR T-cell therapy utilizes the now enhanced T-cell mechanism using the ITAMs to flood tumor cells with granzymes, effectively initiating apoptosis cascades, leading to various effects like DNA fragmentation, cellular damage, and mitochondrial dysfunction. Granzymes also promote the pro-inflammatory environment, improving previously diminished

communication with immune cells like macrophages and NK cells (MacDonald et al., 1999). Cytokine production also has this effect, involving IL-12 cytokine binding to paracrine receptors, increasing immune cell activity, and promoting inflammation (Fig. 2) (Zhang & An, 2007).

Limitations

Studies have demonstrated the increased efficacy of coupling two different costimulatory domains together to avoid T-cell exhaustion and improve long-term survival. As a result, it may be more beneficial to couple the CD28 costimulatory domain in this therapy (or replace it) with the well-established 4-1BB domain (Philipson et al., 2020). However, due to limited testing and literature, it remains difficult to identify the combination's potential for incompatibility with the rest of the system. Furthermore, overcomplexity may increase the risk of overactivation or off-target effects (Sun et al., 2024). Although the synergy between these two costimulatory domains has been well-established and may provide benefits for long-term CAR T-cell survival and growth, their interaction with other components is still unknown, making it a risky choice. Due to 4-1BB's performance in synergy with the CD8α hinge and transmembrane domain in three different FDA-approved CAR T-cell therapies (Cappell & Kochenderfer, 2023), it may be more feasible to use 4-1BB rather than CD28.

Lack of software to test the feasibility of this CAR T cell approach (particularly to test the interaction between an m972-derived scFv and the CD22 antigen) has several implications, meaning that the theorized therapy may have to be iterated before in vivo testing.

Conclusion and Researcher's Notes

Due to limited access to software, the feasibility of this method still remains unknown. FDA approval may also take several decades due to the complexity of this proposed method. Due to the author's lack of experience in this field and his young age, there may be mistakes in the strategic formulation of this theory.

Works Cited

1. Terwilliger, T., & Abdul-Hay, M. (2017). Acute lymphoblastic leukemia: a comprehensive review and 2017 update. *Blood Cancer Journal*, *7*(6), e577. https://doi.org/10.1038/bcj.2017.53
2. Liam, C. C. K., Boo, Y. L., Chong, S. L., Sathar, J., Ong, T. C., & Tan, S. M. (2022). Philadelphia-positive (PH+) acute lymphoblastic leukemia (ALL): developing strategies for curing this stubborn disease. *Blood Research*, *57*(2), 158–161. https://doi.org/10.5045/br.2022.2020305
3. Kantarjian, H. (2014). Inotuzumab Ozogamicin in the treatment of acute lymphoblastic leukemia. *Frontiers in Bioscience-Elite*, *E6*(1), 40–45. https://doi.org/10.2741/e688
4. Puckett, Y., & Chan, O. (2023, August 26). *Acute lymphocytic leukemia*. StatPearls - NCBI Bookshelf. https://www.ncbi.nlm.nih.gov/books/NBK459149/
5. Chennamadhavuni, A., Lyengar, V., Mukkamalla, S. K. R., & Shimanovsky, A. (2023, January 17). *Leukemia*. StatPearls - NCBI Bookshelf. https://www.ncbi.nlm.nih.gov/books/NBK560490/
6. Byun, J. M., Koh, Y., Shin, D., Kim, I., Yoon, S., Lee, J., Bang, S., Kim, K. H., Jung, S., Lee, W. S., Park, Y., Jang, J. H., Han, J. J., Yhim, H., Kim, D. S., Lee, Y. J., Lee, H., Choi, Y., & Lee, S. (2017). BCR-ABL translocation as a favorable prognostic factor in elderly patients with acute lymphoblastic leukemia in the era of potent tyrosine kinase inhibitors. *Haematologica*, *102*(5), e187–e190. https://doi.org/10.3324/haematol.2016.159988
7. Canoy, R. J., Shmakova, A., Karpukhina, A., Shepelev, M., Germini, D., & Vassetzky, Y. (2022). Factors that affect the formation of chromosomal translocations in cells. *Cancers*, *14*(20), 5110. https://doi.org/10.3390/cancers14205110
8. *CAR T cells: Engineering immune cells to treat cancer*. (2022, March 10). Cancer.gov. https://www.cancer.gov/about-cancer/treatment/research/car-t-cells
9. Major, A. S., Joyce, S., & Van Kaer, L. (2006). Lipid metabolism, atherogenesis and CD1-restricted antigen presentation. *Trends in Molecular Medicine*, *12*(6), 270–278. https://doi.org/10.1016/j.molmed.2006.04.004
10. Reddy, S. R., Llukmani, A., Hashim, A., Haddad, D. R., Patel, D. S., Ahmad, F., Sneineh, M. A., & Gordon, D. K. (2021). The role of Chimeric Antigen Receptor-T cell therapy in the treatment of hematological malignancies: advantages, trials, and tribulations, and the road ahead. *Cureus*. https://doi.org/10.7759/cureus.13552
11. Sterner, R. C., & Sterner, R. M. (2021). CAR-T cell therapy: current limitations and potential strategies. *Blood Cancer Journal*, *11*(4). https://doi.org/10.1038/s41408-021-00459-7
12. Mishra, A., Maiti, R., Mohan, P., & Gupta, P. (2023). Antigen loss following CAR‐T cell therapy: Mechanisms, implications, and potential solutions. *European Journal of Haematology*, *112*(2), 211–222. https://doi.org/10.1111/ejh.14101

13. Wang, J., & Wang, L. (2023). CAR-T cell therapy: Where are we now, and where are we heading? *Blood Science*, *5*(4), 237–248. https://doi.org/10.1097/bs9.0000000000000173
14. Murthy, H., Iqbal, M., Chavez, J. C., & Kharfan-Dabaja, M. A. (2019). Cytokine Release Syndrome: Current Perspectives *ImmunoTargets and Therapy*, *Volume 8*, 43–52. https://doi.org/10.2147/itt.s202015
15. Lanza, F., Maffini, E., Rondoni, M., Massari, E., Faini, A. C., & Malavasi, F. (2020). CD22 expression in B-Cell acute lymphoblastic leukemia: Biological significance and implications for inotuzumab therapy in adults. *Cancers*, *12*(2), 303. https://doi.org/10.3390/cancers12020303
16. Liu, K., Li, M., Li, Y., Li, Y., Chen, Z., Tang, Y., Yang, M., Deng, G., & Liu, H. (2024). A review of the clinical efficacy of FDA-approved antibody–drug conjugates in human cancers. *Molecular Cancer*, *23*(1). https://doi.org/10.1186/s12943-024-01963-7
17. *Definition of complete remission - NCI Dictionary of Cancer Terms*. (n.d.). Cancer.gov. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/complete-remission
18. Shah, N. N., Maatman, T., Hari, P., & Johnson, B. (2019). Multi targeted CAR-T cell therapies for B-Cell malignancies. *Frontiers in Oncology*, *9*. https://doi.org/10.3389/fonc.2019.00146
19. Clark, G., Stockinger, H., Balderas, R., Van Zelm, M. C., Zola, H., Hart, D., & Engel, P. (2016). Nomenclature of CD molecules from the Tenth Human Leucocyte Differentiation Antigen Workshop. *Clinical & Translational Immunology*, *5*(1). https://doi.org/10.1038/cti.2015.38
20. Shah, N. N., & Sokol, L. (2021). Targeting CD22 for the treatment of B-Cell malignancies. *ImmunoTargets and Therapy*, *Volume 10*, 225–236. https://doi.org/10.2147/itt.s288546
21. Sullivan-Chang, L., O'Donnell, R. T., & Tuscano, J. M. (2013). Targeting CD22 in B-cell malignancies: Current status and clinical outlook. *BioDrugs*, *27*(4), 293–304. https://doi.org/10.1007/s40259-013-0016-7
22. Du, X., Beers, R., FitzGerald, D. J., & Pastan, I. (2008). Differential Cellular Internalization of Anti-CD19 and -CD22 Immunotoxins Results in Different Cytotoxic Activity. *Cancer Research*, *68*(15), 6300–6305. https://doi.org/10.1158/0008-5472.can-08-0461
23. Janeway, C. A., Jr, Travers, P., Walport, M., & Shlomchik, M. J. (2001). *The interaction of the antibody molecule with specific antigen*. Immunobiology - NCBI Bookshelf. https://www.ncbi.nlm.nih.gov/books/NBK27160/
24. Cappell, K. M., & Kochenderfer, J. N. (2023). Long-term outcomes following CAR T cell therapy: what we know so far. *Nature Reviews Clinical Oncology*, *20*(6), 359–371. https://doi.org/10.1038/s41571-023-00754-1
25. Fry, T. J., Shah, N. N., Orentas, R. J., Stetler-Stevenson, M., Yuan, C. M., Ramakrishna, S., Wolters, P., Martin, S., Delbrook, C., Yates, B., Shalabi, H., Fountaine, T. J., Shern, J. F., Majzner, R. G., Stroncek, D. F., Sabatino, M., Feng, Y., Dimitrov, D. S., Zhang, L., . . .

Mackall, C. L. (2017). CD22-targeted CAR T cells induce remission in B-ALL that is naive or resistant to CD19-targeted CAR immunotherapy. *Nature Medicine*, *24*(1), 20–28. https://doi.org/10.1038/nm.4441
26. Zhang, C., Durer, S., Thandra, K. C., & Kasi, A. (2022, October 3). *Chimeric Antigen Receptor T-Cell therapy*. StatPearls - NCBI Bookshelf. https://www.ncbi.nlm.nih.gov/books/NBK537294/
27. Zhang, C., Liu, J., Zhong, J. F., & Zhang, X. (2017). Engineering CAR-T cells. *Biomarker Research*, *5*(1). https://doi.org/10.1186/s40364-017-0102-y
28. Kozani, P. S., Kozani, P. S., & O'Connor, R. S. (2021). *Humanized Chimeric Antigen Receptor (CAR) T cells*. PubMed Central (PMC). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8813057/
29. Fujiwara, K., Tsunei, A., Kusabuka, H., Ogaki, E., Tachibana, M., & Okada, N. (2020). Hinge and transmembrane domains of chimeric antigen receptor regulate receptor expression and signaling threshold. *Cells*, *9*(5), 1182. https://doi.org/10.3390/cells9051182
30. Zhang, A., Sun, Y., Du, J., Dong, Y., Pang, H., Ma, L., Si, S., Zhang, Z., He, M., Yue, Y., Zhang, X., Zhao, W., Pi, J., Chang, M., Wang, Q., & Zhang, Y. (2021). Reducing hinge flexibility of CAR-T cells prolongs survival in vivo with low cytokines release. *Frontiers in Immunology*, *12*. https://doi.org/10.3389/fimmu.2021.724211
31. Xiao, X., Ho, M., Zhu, Z., Pastan, I., & Dimitrov, D. S. (2009). Identification and characterization of fully human anti-CD22 monoclonal antibodies. *mAbs*, *1*(3), 297–303. https://doi.org/10.4161/mabs.1.3.8113
32. Ahmad, Z. A., Yeap, S. K.,O'Neill, P., Mistry, R. K., Brown, A. J., & James, D. C. (2023). Protein-Specific signal peptides for mammalian vector engineering. *ACS Synthetic Biology*, *12*(8), 2339–2352. https://doi.org/10.1021/acssynbio.3c00157
33. Ali, A. M., Ho, W. Y., Alitheen, N. B. M., & Hamid, M. (2012). SCFV Antibody: principles and clinical application. *Clinical and Developmental Immunology*, *2012*, 1–15. https://doi.org/10.1155/2012/980250
34. Gray, S., Weigel, K. M., Miller, K. D., Ndung'u, J., Buscher, P., Tran, T. N., Baird, C. L., & Cangelosi, G. A. (2009). Flow cytometry‐based methods for assessing soluble scFv activities and detecting antigens in solution. *Biotechnology and Bioengineering*, *105*(5), 973–981. https://doi.org/10.1002/bit.22607
35. Reis, A., PhD. (2022, June 27). *The structure of CAR-T cells*. ProteoGenix. https://www.proteogenix.science/scientific-corner/car-t/car-t-structure/
36. Rafiq, S., Hackett, C. S., & Brentjens, R. J. (2019). Engineering strategies to overcome the current roadblocks in CAR T cell therapy. *Nature Reviews Clinical Oncology*, *17*(3), 147–167. https://doi.org/10.1038/s41571-019-0297-y
37. Chen, X., Mirazee, J. M., Skorupka, K. A., Matsuo, H., Youkharibache, P., Taylor, N., & Walters, K. J. (2022). The CD8α hinge is intrinsically disordered with a dynamic

exchange that includes proline cis-trans isomerization. *Journal of Magnetic Resonance*, *340*, 107234. https://doi.org/10.1016/j.jmr.2022.107234

38. Zhang, A., Sun, Y., Du, J., Dong, Y., Pang, H., Ma, L., Si, S., Zhang, Z., He, M., Yue, Y., Zhang, X., Zhao, W., Pi, J., Chang, M., Wang, Q., & Zhang, Y. (2021). Reducing hinge flexibility of CAR-T cells prolongs survival in vivo with low cytokines release. *Frontiers in Immunology*, *12*. https://doi.org/10.3389/fimmu.2021.724211
39. Jayaraman, J., Mellody, M. P., Hou, A. J., Desai, R. P., Fung, A. W., Pham, A. H. T., Chen, Y. Y., & Zhao, W. (2020). CAR-T design: Elements and their synergistic function. *EBioMedicine*, *58*, 102931. https://doi.org/10.1016/j.ebiom.2020.102931
40. Santos, M. H. D., Machado, M. P., Kumaresan, P. R., & Da Silva, T. A. (2022). Modification of Hinge/Transmembrane and Signal Transduction Domains Improves the Expression and Signaling Threshold of GXMR-CAR Specific to Cryptococcus spp. *Cells*, *11*(21), 3386. https://doi.org/10.3390/cells11213386
41. Ahmad, U., Khan, Z., Ualiyeva, D., Amissah, O. B., Noor, Z., Khan, A., Zaman, N., Khan, M., Khan, A., & Ali, B. (2022). Chimeric antigen receptor T cell structure, its manufacturing, and related toxicities; A comprehensive review. *Advances in Cancer Biology - Metastasis*, *4*, 100035. https://doi.org/10.1016/j.adcanc.2022.100035
42. Weinkove, R., George, P., Dasyam, N., & McLellan, A. D. (2019). Selecting costimulatory domains for chimeric antigen receptors: functional and clinical considerations. *Clinical & Translational Immunology*, *8*(5). https://doi.org/10.1002/cti2.1049
43. Harrison, A. J., Du, X., Von Scheidt, B., Kershaw, M. H., & Slaney, C. Y. (2021). Enhancing co-stimulation of CAR T cells to improve treatment outcomes in solid cancers. *Immunotherapy Advances*, *1*(1). https://doi.org/10.1093/immadv/ltab016
44. Morán, G. a. G., Parra-Medina, R., Cardona, A. G., Quintero-Ronderos, P., & Rodríguez, É. G. (2013, July 18). *Cytokines, chemokines and growth factors*. Autoimmunity - NCBI Bookshelf. https://www.ncbi.nlm.nih.gov/books/NBK459450/
45. Agliardi, G., Liuzzi, A. R., Hotblack, A., De Feo, D., Núñez, N., Stowe, C. L., Friebel, E., Nannini, F., Rindlisbacher, L., Roberts, T. A., Ramasawmy, R., Williams, I. P., Siow, B. M., Lythgoe, M. F., Kalber, T. L., Quezada, S. A., Pule, M. A., Tugues, S., Straathof, K., & Becher, B. (2021). Intratumoral IL-12 delivery empowers CAR-T cell immunotherapy in a pre-clinical model of glioblastoma. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-020-20599-x
46. Srivastava, S., & Riddell, S. R. (2015). Engineering CAR-T cells: Design concepts. *Trends in Immunology*, *36*(8), 494–502. https://doi.org/10.1016/j.it.2015.06.004
47. Holzinger, A., & Abken, H. (2019). CAR T Cells: A snapshot on the growing options to design a CAR. *HemaSphere*, *3*(1). https://doi.org/10.1097/hs9.0000000000000172
48. Voskoboinik, I., Whisstock, J. C., & Trapani, J. A. (2015). Perforin and granzymes: function, dysfunction and human pathology. *Nature Reviews. Immunology*, *15*(6), 388–400. https://doi.org/10.1038/nri3839

49. Oeckinghaus, A., Hayden, M. S., & Ghosh, S. (2011). Crosstalk in NF-κB signaling pathways. *Nature Immunology*, *12*(8), 695–708. https://doi.org/10.1038/ni.2065
50. Weinkove, R., George, P., Dasyam, N., & McLellan, A. D. (2019). Selecting costimulatory domains for chimeric antigen receptors: functional and clinical considerations. *Clinical & Translational Immunology*, *8*(5). https://doi.org/10.1002/cti2.1049
51. Kang, Z. J., Liu, Y. F., Xu, L. Z., Long, Z. J., Huang, D., Yang, Y., Liu, B., Feng, J. X., Pan, Y. J., Yan, J. S., & Liu, Q. (2016). The Philadelphia chromosome in leukemogenesis. *Chinese Journal of Cancer*, *35*(1). https://doi.org/10.1186/s40880-016-0108-0
52. Anand, U., Dey, A., Chandel, A. K. S., Sanyal, R., Mishra, A., Pandey, D. K., De Falco, V., Upadhyay, A., Kandimalla, R., Chaudhary, A., Dhanjal, J. K., Dewanjee, S., Vallamkondu, J., & De La Lastra, J. M. P. (2023). Cancer chemotherapy and beyond: Current status, drug candidates, associated risks and progress in targeted therapeutics. *Genes & Diseases*, *10*(4), 1367–1401. https://doi.org/10.1016/j.gendis.2022.02.007
53. MacDonald, G., Shi, L., Velde, C. V., Lieberman, J., & Greenberg, A. H. (1999). Mitochondria-dependent and -independent Regulation of Granzyme B–induced Apoptosis. *The Journal of Experimental Medicine*, *189*(1), 131–144. https://doi.org/10.1084/jem.189.1.131
54. Zhang, J., & An, J. (2007). Cytokines, inflammation, and pain. *International Anesthesiology Clinics*, *45*(2), 27–37. https://doi.org/10.1097/aia.0b013e318034194e
55. Philipson, B. I., O'Connor, R. S., May, M. J., June, C. H., Albelda, S. M., & Milone, M. C. (2020). 4-1BB costimulation promotes CAR T cell survival through noncanonical NF-κB signaling. *Science Signaling*, *13*(625). https://doi.org/10.1126/scisignal.aay8248
56. Sun, D., Shi, X., Li, S., Wang, X., Yang, X., & Wan, M. (2024). CAR-T cell therapy: A breakthrough in traditional cancer treatment strategies (Review). *Molecular Medicine Reports*, *29*(3). https://doi.org/10.3892/mmr.2024.13171

“Law, Language, and Legitimacy: The Importance of Speaking Legal"[75]
By Aytaj Gurbanova

Language is serving law, as both its vessel and its instrument. It conveys legal rules, rights, and responsibilities, shaping the framework within which societies operate. However, the intricacy of legal language often alienates the general public, creating barriers to understanding and access. This paper explores the multifaceted relationship between law and language, highlighting it defines, interprets, and legitimizes the legal system. By addressing both the challenges and opportunities posed by legal language, this paper emphasizes the importance of making it more accessible to strengthen trust, equity, and engagement within society.

1. Language as Law’s Tool: Precision, Interpretation, and Impact

Language is a foundational tool of human communication, serving as a structured system to convey ideas, emotions, and actions through words and syntax. In the legal field, language takes on a specialized role, becoming one of the most important instruments for interpretation, expression, and enforcement of rules and societal norms. Legal language is marked by its formal structure and precision, qualities that allow it to be interpreted differently depending on the specific context (Tiersma, 1999). These unique characteristics, such as standard terminologies and exact phrasing, are designed to ensure consistency in legal interpretation and facilitate accurate application. One of the primary purposes of legal language is to enable practitioners—lawyers, judges, and legal scholars—to navigate complex legal scenarios, as they are trained to interpret the law’s nuances within its structured language framework (Assy, 2011; Mellinkoff, 1963).

The overarching aim of legal language is to establish a system that sets clear behavioral expectations within society by defining and regulating proper and improper actions. By structuring rules and definitions, legal language forms the basis upon which laws are interpreted and enforced, providing a foundation for what society considers right or wrong (Ross, 2020). Thus, understanding legal terminology is essential to comprehend the legal standards expected of individuals, helping people to anticipate the legal consequences of their actions (Kimble, 1996). The consistent application of terminology—ranging from common terms like "witness" and "contract" to specialized terms like "voir dire" and "subpoena duces tecum"—not only ensures clarity but also aims for fairness in application. While the specificity of legal language might intimidate non-experts, this structured vocabulary allows for equitable treatment by defining clear standards (Tiersma, 2001).

The relationship between law and language is complex and foundational to the legal system itself. Law, as a set of rules, cannot exist without a language to convey, translate, and

[75] I would like to express my deepest appreciation to Dr. Carlos De la Rosa Xochitiotzi for guiding me through the entire process of this paper. This paper would not be possible without the resources and support given to me through Indigo Research. I attest that the ideas, graphics, and writing in this paper are entirely my own.

apply those rules. Legal language serves as a connection point between abstract legal principles and their real-world applications. It is by the use of language that laws get written, explained, and debated. Every single legal document—from fundamental law to contracts to court rulings—relies on precise language to define rights, obligations, and responsibilities regarding different procedures (Williams, 2004).

While legal language can seem complex or exclusive, this complexity serves a purpose: it enables the flexibility and adaptability required to meet varying situations. Legal language is not inherently negative; rather, it provides professionals with the latitude to interpret laws in ways that can align with societal progress and address unique case specifics (Danet, 1980). This interpretive flexibility is essential in contexts where rigid application may not be appropriate, allowing for legal decisions to reflect the nuances of individual circumstances. Furthermore, the open-ended nature of some legal language promotes social change by allowing room for evolving interpretations, which can adapt to shifting societal values over time (Charrow & Charrow, 1979). Although non-experts may sometimes find legal language challenging, its structured yet flexible nature allows legal practitioners to advocate for both justice and social responsiveness (Tiersma & Solan, 2002).

The interpretation of law is a dynamic and essential function within the legal system, shaped by the language used. Language commands considerable power in law, influencing how different cases are approached and resolved. Judges and other professionals wield knowledge of legal language that allows them to make informed decisions based on subtle distinctions. However, law is not merely a set of unyielding rules; it serves as a foundation for interpretation, enabling individuals to realize their rights and seek justice (Hart, 1961).

Interpretation by itself is not limited to legal professionals; instead, it is a process that involves all sectors of society, including ordinary citizens, granting everyone power to bring unique modifications into discussions. This subjectivity allows for ongoing debate and application of law in specific cases (Kelley, 2021). While clarity in law is essential to assist understanding, it often assumes individuals' familiarity with legal language that the general public may lack. In reality, many individuals do not have the capacity to navigate complex legal terminology, highlighting a significant flaw in the accessibility of the law. Philosophers such as H.L.A. Hart argued that law is a social practice rooted in language, indicating a need for a more inclusive approach to legal communications. Making legal language accessible to everyone is a matter of justice and equality. It is crucial to ensure that all members of society, regardless of literacy levels or educational background, can embrace their rights and responsibilities under the law (Solan, 2010).

2. Language, Legal Practice, and Systemic Legitimacy

Implementing these ideas, such as simplifying legal language, promoting widespread legal education, and fostering transparent communication between legal institutions and the public, would enhance the legitimacy of the legal system and improve the relationship between citizens and their governments (Rawls, 1971; Tyler, 1990). When individuals can engage with

the law on the same level as professionals and trust in legal institutions, a legal environment can emerge where everyone feels empowered to assert their rights and navigate the complexities of the law confidently (National Center for State Courts, 2014).

Building a legitimate legal system is essential for fostering public trust and cooperation, creating a more transparent relationship between individuals and authorities (Solum, 2004). When people view legal institutions as fair and credible, they're more likely to engage willingly with the law and navigate its complexities with confidence. This voluntary compliance, rooted in perceived legitimacy, minimizes the need for forced enforcement and encourages a cooperative spirit between the public and authorities—especially crucial in contexts like law enforcement and courtroom proceedings, where public trust can make the difference between community collaboration or resistance (Tyler, 2006). Research consistently supports that legitimacy enhances community-driven behavior, which stabilizes and strengthens society as a whole (National Center for State Courts, 2014; Tyler & Huo, 2002).

The concept of procedural justice underscores that legitimacy isn't inherent; it's earned through fair, consistent, and respectful interactions. Authorities build this perception by demonstrating neutrality, ensuring their actions reflect established laws over personal judgment, and by respecting the dignity of those they serve (Tyler, 2006). Meaningful communication, where people are given space to express their views and see those views considered, further bolsters trust. Accessible language plays a key role in this process, as it not only clarifies decisions made by authorities but also fosters understanding and transparency (National Center for State Courts, 2014; Plain Language Action and Information Network, 2011). As a result, people are more likely to trust legal authorities, thereby reinforcing the legitimacy and inclusiveness of legal institutions. This trust, in turn, strengthens the public's overall sense of justice and shared responsibility within the legal system (Rawls, 1971; Tyler, 1990).

The following subsections will explore specific examples of how accessible legal language supports the effective functioning of legal authorities while enhancing their legitimacy in the eyes of the public. By focusing on areas such as courtroom proceedings, policing practices, and interactions between defendants and prosecutors, these examples demonstrate how clear and comprehensible communication strengthens trust, fairness, and participation in the justice system.

a. The Role of Accessible Language in Courtrooms

The courtroom is a primary arena where accessible legal language significantly impacts justice. Simplifying legal language, as demonstrated by initiatives like the Plain Language Movement, enhances access for the general public by making legal documents and processes more comprehensible (Lloyd, 2023). Courts in jurisdictions such as the United States, Canada, and the United Kingdom have undertaken measures to increase public understanding of legal terminology, which promotes transparency and individual empowerment (Harris, 2020; Wexler, 2019).

Accessible language in courts enables individuals to better defend their rights and engage more actively in legal proceedings, bridging the gap between the general public and the legal system (National Center for State Courts [NCSC], 2022). For example, the NCSC reported that incorporating plain language into court documents has reduced feelings of overwhelm among self-represented litigants, thus enhancing their ability to participate effectively in their cases (NCSC, 2022). These changes not only empower the public by reducing their need for legal representation in basic matters but also alleviate the overall burden on the judicial system (Harris, 2020).

b. The Impact of Accessible Language in Policing

In policing, accessible language during interactions with the public is vital to maintaining trust and fostering perceptions of legitimacy. Studies show that individuals are more likely to view law enforcement as legitimate when their rights are explained to them in clear and comprehensible language (Tyler, 2006; Mastrofski et al., 2000). The use of plain language during arrests or rights explanations reduces the risk of misunderstandings and the potential for involuntary self-incrimination, particularly when individuals understand their rights to silence and legal representation (Agan & Strohl, 2020). When people have a clear grasp of their rights and the reasons for an arrest, they are less likely to resist, which leads to safer encounters and a stronger sense of fairness within the justice system (Kirk & Papachristos, 2011).

In multicultural and multilingual societies, translation services further enhance accessibility by ensuring that non-native speakers understand their legal rights (Dukes et al., 2016). For instance, police in areas with diverse linguistic backgrounds often provide translated rights explanations to bridge language barriers (Crawford et al., 2017). This commitment to equitable communication not only promotes procedural justice but also reduces language-based disadvantages, making the legal system more inclusive and accessible to everyone (Davis et al., 2020).

c. The Role of Plain Language for Legal Authorities and Defendants

The relationship between defendants and prosecutors is pivotal in the legal process, relying heavily on clear and accessible language to ensure fair proceedings. When defendants understand the legal language used during trials, they are better equipped to make informed decisions, such as whether to accept a plea bargain, challenge evidence, or testify. Complex legal jargon, however, often becomes a barrier, especially for individuals unfamiliar with legal processes (Harris, 2017; Muir, 2022). Accessible language allows defendants to grasp their rights, obligations, and the implications of each legal choice, fostering a sense of fairness and inclusion (NCSC, 2020).

For prosecutors, accessible language is equally crucial. Presenting charges, evidence, and arguments in plain terms allows them to communicate their case transparently and ethically. This approach not only helps defendants but also contributes to a courtroom environment rooted in

procedural justice, where both parties operate with a shared understanding of legal terms and processes (Tyler, 2016; McHugh, 2021).

Accessible language also promotes neutrality, a core principle in procedural justice. When defendants feel the language used is impartial and not intentionally convoluted, they are more likely to trust the fairness of the proceedings (Tyler, 2006). Neutral, respectful communication empowers defendants to exercise their voice in the process, enabling them to respond effectively to accusations, ask questions, and present their version of events (Harris, 2017). In turn, this interaction cultivates trust between defendants and the legal system, encouraging compliance and cooperation (Harris, 2017; Muir, 2022).

3. Improving the accessibility of the legal language

As shown, understanding legal language is essential for fostering trust and enabling meaningful participation in legal processes. The accessibility of legal language is critical for fostering trust and ensuring equitable access to justice. Currently, the complexity of legal documents often aliantes individuals who are unfamiliar with legal terminology. To address this issue, several strategies can be implemented, combining efforts from legal practitioners, educators, and designers.

Simplifying Legal Texts with Plain Language

One of the most effective methods for improving accessibility is adopting plain language principles. Plain language replaces technical jargon with clear, concise terminology that is easier for the general public to understand. For instance, courts in Michigan (Michigan Courts) have published plain-language guides for parenting time, incorporating simplified text and visual aids to address diverse family needs. Similarly, legal professionals can prioritize clarity by breaking down complex clauses into shorter, more comprehensible sentences.

Utilizing Visual Aids

Legal processes often involve intricate concepts that can be challenging to convey through text alone. Incorporating visuals such as diagrams, flowcharts, and infographics can significantly enhance understanding. For example, in California, visual aids like eviction process flowcharts have been introduced to help tenants and landlords navigate legal procedures effectively. These charts outline step-by-step actions, deadlines, and required forms, reducing confusion and improving compliance with legal requirements (California Courts). Similarly, a flowchart illustrating the steps of a legal process can make it easier for individuals to navigate their legal responsibilities. Visual aids not only improve comprehension but also make legal documents more engaging.

Streamlining Document Design

The design of legal documents plays a crucial role in their readability. Streamlined layouts featuring clear headings, bullet points, and logical flow help readers easily locate

information. For instance, the Legal Aid Society of New York has adopted simplified formatting for eviction notices by using bold section headers and bulleted lists to clarify tenants' options and deadlines (Legal Aid Society of New York). Consistent formatting, including standardized fonts and spacing, can further reduce barriers to comprehension. Features like interactive tables of contents or hyperlinks in digital documents also enhance navigation, as seen in government websites like the United States Citizenship and Immigration Services, which uses embedded links to guide users through application processes step by step (USCIS).

Promoting Legal Education

Improving the accessibility of legal language also requires raising public awareness about legal rights and terminologies. Community workshops, online tutorials, and partnerships with educational institutions can bridge the knowledge gap. For instance, the "Street Law" program, implemented in many countries, provides high school students with legal literacy training through interactive workshops and simulations, empowering them to understand and navigate the legal system ("Street Law"). Similarly, online initiatives like the "Law for All" platform offer free tutorials and resources aimed at educating individuals about contracts and dispute resolution. These programs focus on practical legal literacy, helping individuals interpret contracts, navigate disputes, or understand their rights more effectively ("Law for All").

Applying Legal Design Methodology

Legal design methodology offers a user-centric approach to drafting legal documents. This involves reevaluating traditional legal practices, incorporating plain language, and integrating visuals. For example, the Dutch Ministry of Justice's "Legal Aid Board" redesigned its public-facing documents to prioritize user accessibility by simplifying language and including visual elements like icons and infographics to guide users through legal aid applications (Dutch Ministry of Justice). Similarly, projects like "Hello Divorce" utilize legal design by presenting divorce procedures in an interactive, step-by-step format, helping users navigate complex legal processes with greater ease. By prioritizing the needs of the end-user, legal design can transform complex legal texts into tools for empowerment rather than barriers to justice.

The interplay between law and language lies in justice. Legal language, with its precision and formality, provides the foundation for interpreting and enforcing societal norms. Yet, its inherent complexity can alienate those without specialized training, diminishing the accessibility and legitimacy of the legal system. To bridge this gap, adopting plain language principles, integrating visual aids, and promoting legal education are critical steps toward empowering individuals to navigate legal structures confidently. By prioritizing inclusivity in legal communication, the justice system can foster greater trust and cooperation between citizens and authorities, creating a more equitable society. Ultimately, a legal system that speaks the language of its people strengthens its foundations, ensuring fairness, transparency, and shared accountability.

Works Cited

Mayerson, Denise, et al., editors. Procedural Justice and Relational Theory: Empirical, Philosophical, and Legal Perspectives. 2021, https://library.oapen.org/bitstream/id/ef03313b-0129-426b-9ac9-34155c843c70/9781000 2076 68.pdf

Butler**,** J. (2020). Legal language and procedural justice: The role of clear communication in law enforcement. Journal of Language and Law, 15(2), 103–118.

Hart, H. L. A. (2012). The concept of law (3rd ed.). Oxford University Press.

National Center for State Courts. (2021). Plain language in the courts: Enhancing access for all. https://www.ncsc.org/

Sarat, A., & Kearns, T. R. (Eds.). (2020). Law in the domains of culture. University of Michigan Press.

Tyler, T. R. (2006). Why do people obey the law? Princeton University Press.

United States Department of Justice. (2022). Community policing: Building relationships and trust through effective communication. https://www.justice.gov/

American Bar Association. (2018). Plain language in legal documents: Enhancing comprehension and accessibility. https://www.americanbar.org/

Baker, L. (2019). Law, language, and power: Exploring the relationship between legal language and authority. International Journal of Law and Society, 36(4), 112–129.

Fair Trials. (2022). Understanding your rights: The importance of accessible legal information in policing and courts. https://www.fairtrials.org/

National Institute of Justice. (2021). Procedural justice in policing: Building public trust and legitimacy. https://nij.ojp.gov/

Nielsen, M. (2020). Judicial language reform: Making courtrooms more accessible for non-lawyers. Legal Studies Journal, 45(3), 205–222.

Wexler, D. B. (2021). Therapeutic jurisprudence and the role of clear legal language. Law and Psychology Review, 44(1), 55–78.

Harris, R. (2020). The Impact of Plain Language in Law. Journal of Legal Education, 69(2), 203-215.

Lloyd, J. (2023). Simplifying Legal Language for Justice. Legal Language Journal, 12(1), 45-58.

National Center for State Courts (NCSC). (2022). Improving Access to Justice: The Role of Plain Language.https://www.ncsc.org

Wexler, D. (2019). Legal Language and Its Impact on Justice. Harvard Law Review, 132(5), 1403-1420.

Agan, A., & Strohl, J. (2020). The role of language in policing: Implications for justice. Journal of Criminal Justice, 48(3), 101-110.

Crawford, A., et al. (2017). Bridging the gap: Language access in policing. Police Quarterly, 20(2), 123-145.

Davis, M., et al. (2020). Procedural justice and language accessibility in law enforcement. Journal of Public Affairs, 20(1), e1988.

Dukes, R., et al. (2016). Language barriers in policing: Challenges and strategies. Criminal Justice Studies, 29(4), 445-463.

Kirk, D. S., & Papachristos, A. V. (2011). The role of language in police-community relations. American Journal of Sociology, 117(5), 1391-1430.

Mastrofski, S. D., et al. (2000). Police and community interactions: A case for procedural justice. Crime and Delinquency, 46(4), 446-469.

Tyler, T. R. (2006). Psychological perspectives on legitimacy and legitimation. Annual Review of Psychology, 57, 375-400.

Legal Aid Society of New York https://legalaidnyc.org/

United States Citizenship and Immigration Services https://www.uscis.gov/ Street Law program https://streetlaw.org/

Preliminary Investigation On the Visual Status of Older Adults In Community Care Center
By Sophie Wei Li

Abstract

Objectives

The purpose of this study is to assess the vision status of older adults (aged 70 years and above) residing in nursing homes, and to evaluate its impact on their quality of life, with the aim of providing more attention and support to the elderly population in the field of visual health.

Methods

At the Jinfurong Harvest Garden Center in Chengdu High-tech Zone, Sichuan Province, the vision of resident elderly individuals was evaluated for optimal corrected farsighted and nearsighted visual acuity. The impact of visual acuity on basic life needs was assessed using a simplified VF-14 questionnaire, and the medical history of other ocular and systemic diseases was documented.

Results

A total of 60 older adults resided at the center, of whom 47 were able to complete the vision screening and questionnaire. The average age of these 47 older adults was 87.36 ± 5.44 years (ranging from 73 to 98 years). Most of the older adults also suffer from systemic and ocular disorders, including hypertension, diabetes mellitus, cataracts, and macular degeneration, among others. The mean best corrected distance visual acuity (BCDV) of the right eye (OD) was logMAR 0.50 ± 0.34, with a median visual acuity of logMAR 0.4. The mean value of best corrected near vision (BCNV) in the right eye was logMAR 0.44 ± 0.26, with a median near vision of logMAR 0.5. The mean BCDV of the left eye (OS) was logMAR 0.51 ± 0.43, with a median distance vision of logMAR 0.4. The BCNV in the left eye was logMAR 0.43 ± 0.25, with a median near vision of logMAR 0.4. According to the VF-14 questionnaire, 36.17% of the elderly had no difficulty with fine reading, while 63.83% experienced difficulty; 38.92% had no difficulty with general reading and 61.08% experienced difficulty; 74.47% had no difficulty reading large print, whereas 25.53% faced difficulty. In daily life, 34.04% of the elderly had no difficulty identifying steps, stairs, or road edges, while 65.96% faced difficulty. Five (10.64%) of the elderly lost vision in one eye. The majority of older adults with better vision have undergone cataract surgery and/or wear appropriate glasses.

Conclusions

Different degrees of vision loss are common among the older adults at the investigated nursing care center. Due to the presence of multiple diseases, there are significant variations in vision among the elderly. Although the basic living conditions in care facilities have not been significantly affected, the safety of daily life and quality of life require more attention and support.

Keywords older adults, visual acuity

Introduction

At present, the world is facing the challenge of an aging population, and China is no exception. By the end of 2023, China had 296.97 million elderly people aged 60 and over, accounting for 21.1 percent of the total population, with 216.76 million aged 65 and over, representing 15.4 percent of the total[1]. This data indicates that China has officially entered a "moderately aging" society. Older adults deserve more attention in various areas, especially vision health. Approximately 90% of human information is obtained through the eyes, but with age, the function of human tissues and organs gradually deteriorates. This is compounded by the effects of genetics, environment, and various diseases, leading to reduced or even incapacitated vision. Age-related eye conditions, such as refractive errors, senile cataracts, glaucoma, age-related macular degeneration, and eye problems related to systemic diseases like diabetic retinopathy, are common afflictions impacting the visual function of the elderly[2]. However, due to the frailty associated with old age and economic constraints, it is more challenging to treat vision impairment in the elderly in a timely manner, imposing a significant burden on individuals, families, and society.

In response to this, the World Health Organization (WHO) and the International Agency for the Prevention of Blindness launched the "Vision 2020: The Right to Sight" initiative in 1999 to eliminate avoidable blindness in people aged 50 and over[2]. A study has updated global and regional estimates of the causes of moderate and severe visual impairment and blindness, noting that by 2020, cataracts remained the largest contributor to global blindness among adults aged 50 and over, affecting about 15 million people or approximately 45% of the 33.6 million cases of blindness worldwide. Uncorrected refractive errors continue to be the largest contributor to moderate and severe visual impairment (MSVI) in adults aged 50 and over globally, accounting for about 86 million people, or 42% of the 206 million MSVI cases worldwide[3]. Although glaucoma, diabetic retinopathy, and age-related macular degeneration are more difficult to treat, they collectively caused more than 19 million cases of moderate or more severe visual impairment in 2020, making them important targets for prevention and treatment. The study also highlights that despite the success of eye care services in reducing age-standardized sickness rates, they are not meeting the growing demands of an aging and increasing population[3]. This is particularly evident in China, where the health status of the large elderly population often goes unnoticed. This study aims to explore the vision status, related life capacity, quality of life, and associated diseases of older individuals through a small survey, with the hope of drawing the attention and intervention of relevant institutions to provide more care and support for the visual health of the elderly.

Method

On July 21, 2024, vision tests and questionnaires were conducted at a nursing home, the Jinfurong Harvest Garden Center in Chengdu High-tech Zone, Sichuan Province, for all elderly

individuals capable of participating in the screening. Some elderly individuals were excluded from the survey due to severe cognitive impairments caused by diseases such as Alzheimer's, or severe hearing loss that prevented them from responding to tests and questionnaires.

1.Visual Acuity Test

A Snellen visual acuity chart was used to measure the distant visual acuity of 47 elderly individuals under normal lighting conditions. If the largest letter on the chart could not be seen at 5 meters (corresponding to a vision of 0.1), the participant was tested at 4 meters. If the largest letter was visible, the vision was recorded as 0.08. If not visible at 4 meters, the test was conducted at 3 meters, with a vision record of 0.06 if the letter was visible, and so on, until the largest letter was seen at 1 meter, resulting in a visual record of 0.02. If the largest letter could not be seen at 1 meter, the tester would hold up fingers at a distance of 50 cm or less to count them, recording a vision of 0.01. A special near visual acuity chart was used to measure the near visual acuity of 47 elderly individuals. All participants were tested for both their unaided eyesight and corrected vision (using glasses), for both distant and near visual acuity.

2. Questionnaire

The VF-14 questionnaire is commonly used to determine the level of functional vision impairment in patients with cataracts. Considering the age of these older adults, a simplified version of the VF-14 was used. They were asked whether they had difficulties with different reading tasks/activities and were graded on a scale of 0-4, with 0 defined as unable to do the activity; 1, a great deal of difficulty; 2, moderate difficulty; 3, a little difficulty; and 4, no difficulty. Additionally, the participants were asked about major eye and systemic diseases.

3. Data Analysis

After data collection, a descriptive statistical analysis was performed. The fractional visual acuity obtained from the Snellen chart was converted to logMAR (minimum visual angle) for quantitative analysis using the best corrected visual acuity data (https://www.myvisiontest.com/index.php). The mean and standard deviation of the logMAR BCDV or BCNV were calculated. The mean age and standard deviation were also calculated. The elderly individuals were divided into different age groups, and the percentage of individuals in each age group was calculated and presented in a proportional chart. Based on the questionnaires, difficulty levels for "performing fine reading," "performing general reading," "conducting large-character reading," and "identifying stairs and aisles" were statistically analyzed and charted, respectively.

Results

1.General demography

A total of 47 elderly individuals were surveyed, comprising 32 women and 15 men. The age range was from 73 to 98 years, with an average age of 87.36 ± 5.44 years. Of these, 8.51% were under 80 years of age, 55.32% were aged 80 to 89 years, and 36.17% were over 90 years of age (see Figure 1). Among them, the vast majority were wearing either farsighted or nearsighted glasses. Six individuals had undergone cataract surgery in one or both eyes (four women and two men), while five had severe cataracts and had not yet undergone surgery (four women and one man). Additionally, two individuals had macular degeneration, two had dry eye disease, and one had chronic tear cystitis. Nine had hypertension, six had diabetes mellitus (DM), five had heart disease, four had both hypertension and DM, and two had all three conditions. Other conditions included cerebral infarction in three individuals.

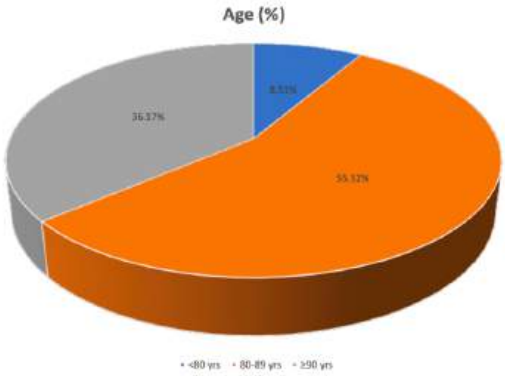


Figure 1. Pie chart depicting the age distribution of the 47 elderly individuals surveyed

2. Vision Condition

Forty-seven elderly individuals participated in visual acuity tests, with 42 and 44 having best corrected distance visual acuity (BCDV) and best corrected near visual acuity (BCNV) data for the right eye (OD), respectively. Among these 47 participants, five were unable to provide quantifiable BCDV data: one female had no light perception (NLP), one female had light perception (LP), one male could only perceive hand motion (HM/30 cm), and two females had cognitive impairment due to Alzheimer's disease. Three elderly individuals could not measure their BCNV. The number of participants with measurable distance and near vision in the left eye (OS) were 42 and 45, respectively, as five individuals could not have their distance vision quantified. This included two women with moderate cognitive impairment (Alzheimer's disease) and three women who could only perceive hand motion (HM/40 cm). Additionally, two individuals could not measure their near vision. Among the forty-seven participants, five had vision in only one eye, with the other eye being completely blind (without light perception).

The average BCDV and BCNV of the right eye (OD) were logMAR 0.50 ± 0.34 and logMAR 0.44 ± 0.26, respectively. The mean logMAR value of BCDV in the left eye (OS) was 0.51 ± 0.43, and the mean logMAR value of BCNV was 0.43 ± 0.25 (see Figure 2), indicating a general decline in both near and far vision among the elderly. The data also showed substantial

variability in visual acuity among different individuals, as indicated by the large standard deviation.

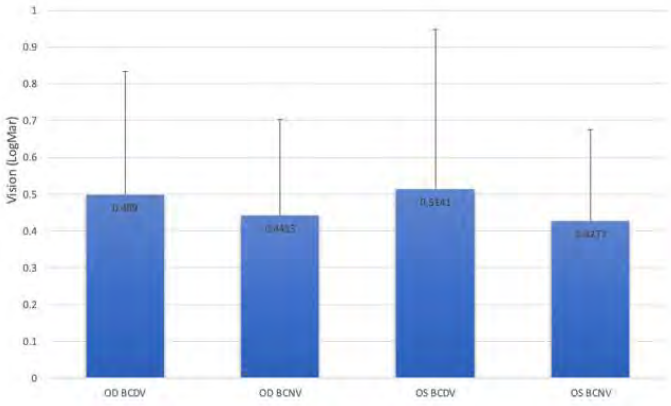


Figure 2. Mean and standard deviation of the BCDV and BCNV in the right (OD) and left (OS) eyes.

To investigate the reasons for the significant differences in individual vision, we further analyzed the percentage of individuals with different distance/near vision and the median distance/near vision in the right and left eyes, respectively. As shown in Figure 3, neither BCDV nor BCNV in either eye displayed a normal distribution in the studied population. The medians of BCDV and BCNV for the right eye were logMAR 0.4 and logMAR 0.5, respectively. The median BCDV for the left eye was logMAR 0.4, and the median BCNV was logMAR 0.4. Those with good near and far visual acuity (< logMAR 0.2) typically had undergone cataract surgery and/or wore suitable glasses, without other eye diseases. A minority with very poor vision (> logMAR 1.0) had untreated cataracts, macular lesions, or inappropriate glasses.

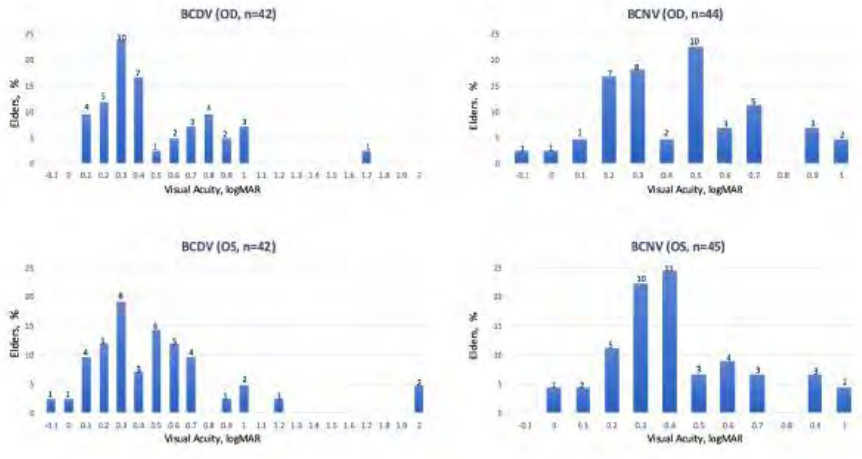


Figure 3. Distribution of the proportion of individuals with varying visual acuity

Additionally, most elderly individuals had better vision in one eye compared to the other. Therefore, the study also included data from the eye with the better vision for statistical analysis to accurately reflect the best visual function of the elderly in daily life. As shown in Figure 4, 45

individuals had measurable vision in the better eye (2 had cognitive impairment and were unable to measure far vision). The median BCDV and BCNV of the better eye were both logMAR 0.3. The number of individuals with good far and near vision in one eye (<logMAR 0.2) were 15 (33.33%) and 16 (34.04%), respectively. Furthermore, the number of individuals with both eyes having poor BCDV and BCNV greater than logMAR 1.0 was 2 and 1, respectively (Figure 4).

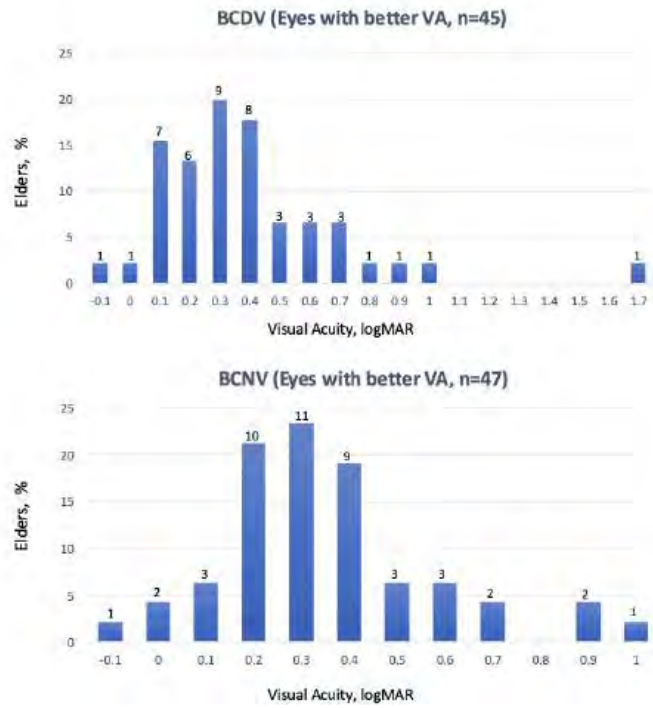


Figure 4. Distribution of the proportion of individuals with varying visual acuity in eyes with better visual acuity.

3. Questionnaire results

A total of 47 participants completed the VF-14 questionnaire survey. In terms of fine reading, approximately 36.17% of the elderly experienced no difficulty, while 63.83% encountered difficulties. Among those, 14.89% had slight difficulty, 14.89% had moderate difficulty, 29.79% had severe difficulty, and 4.26% were unable to perform the task. For ordinary reading, about 38.92% of the elderly had no difficulty, whereas 61.08% faced challenges, with 12.77% having slight difficulty, 12.77% moderate difficulty, 34.04% severe difficulty, and 2.13% unable to do it. Regarding large-character reading, around 74.47% of the elderly reported no difficulties, and 25.53% experienced slight difficulties. In daily life activities, roughly 34.04% of the elderly had no trouble recognizing steps or curbstones, while 65.96% had varying degrees of difficulty, with 46.81% facing slight difficulty, 10.64% moderate difficulty, 6.38% severe difficulty, and 2.13% unable to accomplish the task (see Figure 5).

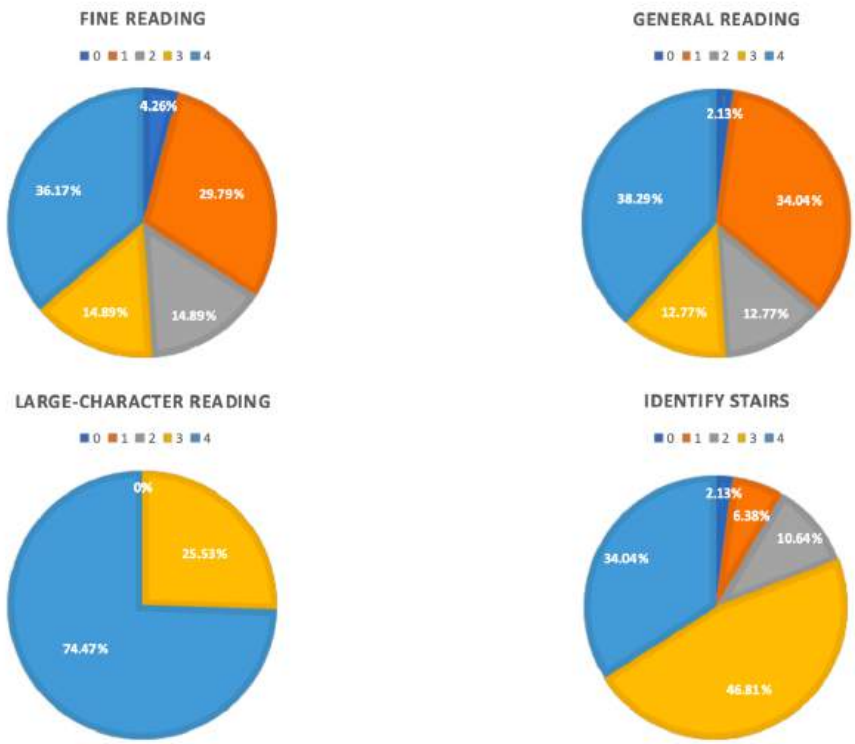


Figure 5. VF-14 questionnaire. The difficulty levels were categorized into four groups: "fine reading," "general reading," "large-character reading," and "identifying stairs." The percentage of older adults in each category is presented. 0 points: unable to do; 1 point: a great deal of difficulty; 2 points: moderate difficulty; 3 points: a little difficulty; 4 points: no difficulty.

Discussion

The 47 older adults in this group experienced general vision loss of varying severity. The large standard deviation, evident in the histogram, indicates that the visual acuity data are widely scattered, reflecting significant variability among the elderly. The median vision further confirms a noticeable decline in both distance and near vision. Considering that many elderly individuals also have a history of eye and other systemic conditions, such as diabetes mellitus (DM), it can be preliminarily inferred that their visual status may be related to these conditions and whether they have been adequately treated. However, this survey alone is insufficient to establish a definitive association, and further research is necessary. It is evident that older individuals who have undergone cataract surgery or who wear appropriate glasses exhibit significantly better vision. Therefore, for older adults with untreated cataracts, macular degeneration, or refractive errors, timely surgery, medical intervention, and proper eyewear can greatly enhance their vision, safety in daily life, and overall quality of life. This, in turn, may reduce the burden on their families and society.

The age range of the survey subjects is broad, with the majority of the elderly participants being over 80 years old. Given the increasing aging population globally and in China, more elderly individuals will likely face inconveniences in life due to vision decline. This could affect not only their ability to perform basic daily activities safely but also their autonomy and quality

of life. As highlighted by the VF-14 visual function questionnaire, while most older adults experience difficulties with general reading (such as books, newspapers, and mobile phones) and fine reading (like medication instructions), they encounter little difficulty with large-character reading, indicating that daily life activities may not be significantly impacted. Therefore, it is crucial to enhance community and social support. For instance, communities should provide materials that facilitate reading for the elderly, such as large-print newspapers and magazines. Since many older adults have difficulty recognizing steps or curbstones, which greatly impacts their safety, more visible signs should be installed on roads, steps, and curbstones. Additionally, there is a concern regarding the decreased ability to identify roadblocks: elderly people may spend less time outdoors to avoid walking accidents, potentially leading to reduced physical activity and social isolation, which in turn may cause or exacerbate other physical and psychological conditions.

Finally, it is worth noting that in the nursing home we surveyed, many older adults exhibited varying degrees of cognitive impairment. Except for a very few who could cooperate with certain assessments, the others were unable to participate in this visual condition and questionnaire survey. Numerous scientific studies have demonstrated that eye health and visual function are closely linked to cognitive impairment, and improving visual function positively impacts slowing the onset and progression of age-related cognitive impairment[4]. Focusing on the eye health of older adults with cognitive impairment will be our next area of focus.

Works Cited

2023 Annual Bulletin of National Aging Development by the Ministry of Civil Affairs & National Office for the Elderly in China. http://www.crca.cn/index.php/13-agednews/1113-2023-3.html

GBD 2019 Blindness and Vision Impairment Collaborators; Vision Loss Expert Group of the Global Burden of Disease. Study of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. *Lancet Glob Health.* 2021 Feb;9(2):e144-e160. doi: 10.1016/S2214-109X(20)30489-7. Epub 2020 Dec 1

Goodman L, Hamm L, Tousignant B, Black J, MisraS, Woodburn S, Keay L, Harwood M, Gordon I, Evans JR, Ramke J. Primary eye health services for older adults as a component of universal health coverage: a scoping review of evidence from high income countries. *Lancet Reg Health West Pac*. 2022 Aug 12;35:100560. Doi: 10.1016/j.lanwpc.2022.100560. eCollection 2023 Jun.

Vu TA, Fenwick EK, Gan ATL, Man REK, Tan BKJ, Gupta P, Ho KC, Reyes-Ortiz CA, TrompetS, Gussekloo J, O'Brien JM, Mueller-Schotte S, Wong TY, Tham YC, Cheng CY, Lee ATC, Rait G, Swenor BK, Varadaraj V, Brenowitz WD, Medeiros FA, Naël V, Narasimhalu K, Chen CLH, Lamoureux EL. The Bidirectional Relationship between Vision and Cognition: A Systematic Review and Meta-analysis. *Ophthalmology.* 2021 Jul;128(7):981-992. doi: 10.1016/j.ophtha.2020.12.010. Epub 2021 Feb 27. PMID: 33333104

Acknowledgements: I would like to express my deepest gratitude to my supervisor, Professor Ming Zhang, PhD Supervisor /Chief Physician from Department of Ophthalmology, West China Hospital, Sichuan University, for his invaluable guidance and support. I would like to express my gratitude to graduate students Guo Wanyue and Xiang Junhan from the West China Hospital School of Medicine, Sichuan University, for their participation and assistance in the visual acuity screening and part of the questionnaire survey. I also wish to thank my parents for their sincere encouragement and strong support, which enabled me to complete this project.

PTSD and Memory: A Comparative Study in a Child and Adult By Juliet Fox

Abstract

Memory and PTSD are deeply intertwined areas of psychological study. Memories form a core component of PTSD, as resurfacing memories are a source of trauma and can lead to persistent PTSD symptoms. At the same time, PTSD can lead to memory loss or memory alteration, in addition to difficulties with emotional regulation and overall mental health. Understanding these connections is vital for developing effective therapeutic interventions for PTSD. The purpose of this paper is to explore the relationship between memory and PTSD by analyzing case studies involving individuals affected by the disorder. By examining the neurological and psychological dimensions of PTSD, this paper aims to highlight the differences in how trauma impacts memory across different developmental stages and how understanding these impacts can improve therapeutic interventions. To begin, I will consider the different effects of memory on the mind and its mechanisms within the brain. I will discuss PTSD and its neuropsychology, considering how brain development alters from the previous discussion of memory and PTSD. I will then explore two case studies of an adult and child suffering from PTSD, comparing their therapies, symptoms, and outcomes. I will finish by looking at the conclusions reached by other studies of memory and PTSD to confirm my findings.

Introduction

Post-Traumatic Stress Disorder (PTSD) is a complex mental health condition deeply intertwined with the processes of memory. Individuals suffering from PTSD often experience distressing memories of traumatic events, which can resurface involuntarily and disrupt daily life. At the same time, PTSD can lead to memory alteration or loss, creating a nuanced relationship between memory and trauma. Understanding this relationship is crucial to the successful treatment of PTSD, as affected memories not only inform emotional regulation and overall mental health, but also the therapeutic interventions aimed at alleviating PTSD symptoms. This paper seeks to explore the intricate connection between memory and PTSD through an analysis of two case studies—one involving an adult and another, a child. By examining these cases, I aim to highlight differences in how trauma impacts memory across different developmental stages, offering insights into the distinct therapeutic approaches and outcomes associated with each. Grounded in existing literature, this study employs a combination of neurological and psychological frameworks to analyze memory's role within PTSD and proposes tailored treatment solutions to address the unique challenges observed in each case.

Short-Term Memory and Long-Term Memory

Memory—the continual process of information retention over the course of time—affects every person and animal on the planet. Humans have several different types of memory systems,

which utilize different brain regions of our minds; these systems include short-term memory and long-term memory.

Short-term memory, also known as working memory, plays a vital role in our daily cognitive functions. The prefrontal cortex and other brain regions use short-term memory to inform decision-making. Together with the hippocampus, which is located deep in the back of the brain, these areas help manage the temporary storage and processing of information, facilitating tasks such as problem-solving, decision-making, and language comprehension. The capacity of short-term memory is limited, typically holding around seven to nine items at one time (Passolunghi, 46). This phenomenon, known as Miller's Law, implies that our brain has a finite capacity to retain and process information simultaneously. However, the exact number of items that can be held in short-term memory may vary depending on factors such as the complexity and familiarity of information being retained and an individual's genetics (53).

Short-term memory is not merely a passive storage system; it actively processes and manipulates information. Working memory utilizes the short-term information to allow for executive decision making and processing of the environment. Several tools can improve working memory, such as chunking, rehearsal, and attention (Voss). Chunking involves grouping information into meaningful units, enhancing our ability to remember larger amounts of information. The hippocampus plays a vital role in encoding new information and forming memories, helping to organize related data into chunks. Rehearsal, the process of mentally repeating information, helps to maintain it in short-term memory. The prefrontal cortex is crucial for working memory, allowing for the manipulation of information during rehearsal and chunking. Attention is the process to selectively focus on relevant information while filtering out distractions. The parietal lobe is involved in attention and spatial awareness, aiding in the focus on relevant information for memory tasks. The anterior cingulate cortex contributes to attention regulation and error detection, supporting effective rehearsal and memory processes. Together, these regions facilitate the efficient functioning of short-term memory (Voss).

In contrast to short-term memory, long-term memory is a facet of human cognition that allows us to retain and recall information over years, even decades (Siegel). Encoding is the process by which information is transformed into a format suitable for storage in long-term memory. The hippocampus is most directly related to that, as it serves as the creator of new memories. Several situational factors influence the ability for a memory to be encoded, such as attention, elaboration, and organization. Firstly, attention directs our focus to the relevant information, while elaboration involves connecting new information to existing knowledge. Organization, on the other hand, involves structuring information in a coherent and logical manner, aiding in its retention. Once information is encoded, it is stored in long-term memory for future retrieval (Siegel). There is also emotional long-term memory, which usually lasts decades, and is most directly related to the brain region called the amygdala. The amygdala plays a significant role in emotional memory, helping to encode experiences that carry emotional significance (Voss).

Long-term memory is composed of different types, including episodic memory, semantic memory, and procedural memory. These memories are believed to be stored in various brain regions, with the hippocampus playing a crucial role in the initial consolidation of memories. The neocortex is involved in the storage of long-term memories, particularly for factual and procedural knowledge (Siegel). Retrieving information from long-term memory involves the process of recall or recognition. Recall refers to the ability to retrieve information without external cues, while recognition involves identifying previously encountered information among a set of options. The prefrontal cortex is important for the retrieval and organization of memories, especially in decision-making and planning processes. Cues show that stimuli aid in the retrieval of specific memories.

Memory plays a crucial role in mental health and therapeutic approaches, as it influences how individuals process experiences and emotions. For instance, the ability to recall past events is essential for understanding current behaviors and developing coping strategies. Many mental health conditions, such as PTSD and depression, also involve distorted or suppressed memories, which can affect a person's overall well-being (Du, 220). Therapeutic techniques, such as cognitive-behavioral therapy (CBT), often focus on helping clients reframe negative memories and develop healthier thought patterns. Additionally, interventions like mindfulness and memory-based therapies aim to strengthen memory function, promoting resilience and emotional regulation (236). By addressing memory's impact on mental health, therapists can create more effective treatment plans tailored to individual needs.

PTSD

PTSD is a severe psychiatric disorder caused by emotional trauma, usually a traumatic event. There are a number of things that are considered traumatic events, such as threatened death, sexual violence, and serious injury. Having a traumatic event does not guarantee PTSD, but 10% of those who have experienced a traumatic event develop PTSD (Sherin, 263). Individuals who have been diagnosed with this disorder often experience symptoms, such as increased anxiety and increased depressive episodes, and they frequently develop problems with substance abuse. PTSD requires urgent attention because of the severity of the disorder (265). When there are major events—including wars and pandemics, like the recently outbreak of COVID-19—the number of traumatic events increases for people across the population, and the number of PTSD cases rises accordingly.

From a neuropsychological perspective, PTSD has profound physical effects on the brain, influencing both its structure and function. Key areas impacted by PTSD include the amygdala, hippocampus, and various regions associated with fear and emotional regulation, similar to those found when looking into memory. The amygdala plays a crucial role in processing emotions, particularly fear. In individuals with PTSD, the amygdala often becomes hyperactive, leading to heightened emotional responses and increased sensitivity to potential threats (Voss). This hyperactivity contributes to the symptoms of anxiety and hyperarousal that characterize PTSD. Conversely, the hippocampus, which is essential for memory formation and context processing,

often shows reduced volume in PTSD patients. This alteration can lead to difficulties in distinguishing between past and present experiences, causing individuals to relive traumatic memories as if they were occurring in real time (Voss). Additionally, several brain regions involved in fear processing, such as the anterior cingulate cortex, insula, and orbitofrontal cortex, also demonstrate significant changes in PTSD. The ACC is important for regulating emotional responses and decision-making; its dysfunction may contribute to impaired emotional regulation in PTSD. The insula is associated with interoceptive awareness and emotional experience, often leading to heightened bodily sensations when confronted with reminders of trauma. Finally, the orbitofrontal cortex is critical for decision-making and the regulation of social behavior, and its involvement in PTSD can result in difficulties in regulating fear responses and managing stress (Sherin, 272). These neurobiological alterations reflect the complex interplay between different brain regions in the manifestation of PTSD symptoms. Understanding how PTSD affects the brain can inform therapeutic approaches, highlighting the need for interventions that address both the psychological and neurobiological aspects of this debilitating condition.

PTSD significantly impacts memory processes, affecting how individuals recall and interpret traumatic experiences. The interplay between PTSD and memory can be understood through the lens of neuropsychology, particularly in relation to key brain structures involved in memory formation and retrieval (274). One of the primary areas affected in individuals with PTSD is the hippocampus, which is crucial for forming new memories and contextualizing past experiences. Research indicates that PTSD is often associated with reduced hippocampal volume, which can lead to difficulties in distinguishing between past and present events. This impairment often results in fragmented or incomplete memories of the trauma, contributing to symptoms such as flashbacks and intrusive thoughts. These experiences can feel vivid and immediate, blurring the line between past trauma and current reality (Voss).

PTSD also alters the way memories are retrieved. The amygdala, responsible for processing emotions, becomes hyperactive in PTSD, especially in response to trauma-related cues. This heightened emotional response can influence memory retrieval, leading to an increased likelihood of recalling emotionally charged memories while suppressing neutral or less relevant information (Sherin, 277). Consequently, individuals may become trapped in a cycle of reliving traumatic memories, making it difficult to access more adaptive or positive memories. The ACC and orbitofrontal cortex are also involved in memory processes related to emotional regulation and contextual understanding. In PTSD, the dysfunction in these areas can impair an individual's ability to place memories within an appropriate context, further complicating their recovery. Without the ability to contextualize traumatic memories, individuals may struggle to integrate these experiences into their broader life narrative, exacerbating feelings of distress and disconnection.

Understanding the connection between PTSD and memory is important for developing effective treatment strategies. Treatment plans can vary from person to person and often do not have one specific plan that works in every case. These changes can exist in the context of two

people experiencing very different traumatic events, people having different symptoms, or people being in very different age ranges.

Case Studies

PTSD treatment and therapies range broadly. To explore the potential options, this paper will review the cases of Mustafa and Philip, two people suffering from PTSD. Mustafa, who faced trauma at a young age, engaged in a longer and slower form of therapy, taking into account the development that he still needed to undergo (Junod). Philip, who experienced trauma at a much older age, took a faster-paced and self-monitored form of therapy, where he was assigned homework and modules to read (Wild and Ehlers). By comparing both cases, we can better understand how developmental age and memory might inform PTSD therapy and treatment.

Mustafa was a ten-year-old Afghani immigrant who traveled to Switzerland because of complaints from his teachers and schoolmates about behavioral issues, including Attention Deficit Disorder. Mustafa started attending therapy sessions, and therapists noted things like restlessness and getting scared at sudden noises, even if they were small, such as a pencil dropping. There would be times in which Mustafa would be awake, but be absent minded, and soon after would claim he had a "nightmare." In children, this is a common symptom of PTSD called derealization and depersonalization (Scheeringa, 359). Doctors later learned that Mustafa, a member of the Hazara ethnicity, had been forced to flee Afghanistan with his family at the age of four due to persecution and the murder of his father by the Taliban. His father was shot at close range while trying to protect Mustafa's older brother. Although Mustafa likely did not witness the event, he was within earshot, and his aunt quickly took him to safety next door. Following the murder, Mustafa, his mother, older brother, and sister embarked on a three-month journey to Switzerland. His sister described their immigration as long, difficult, and filled with persistent fears of loss. Mustafa started going to school in Switzerland the following year, and started going to the doctor more frequently (Junod). After continued disruption in class, Mustafa was transferred to a therapeutic school that catered to his needs. In therapy, he admitted to envisioning "intrusive, frightening images of faceless men fighting, holding weapons, women and children crying, and nightmares of intruders, monsters." Mustafa also expressed fears about his scary childhood happening to him again. Things like re-experiencing, avoidance, hyperarousal, and derealization are all symptoms of PTSD, specifically in childhood cases. Mustafa went through three years of psychotherapy (Junod).

An adult PTSD case has both similarities and differences from childhood cases, as demonstrated by Philip, who developed PTSD after experiencing a traumatic car accident involving his daughter and grandson (Wild and Ehlers). In this case, Philip was driving and his car flipped onto its side, flinging both his grandson and his daughter from the car. Though Philip ended up being the most injured out of the three of them, he explained that he had intrusive thoughts involving car accidents and the deaths of his grandson and daughter. He had difficulty sleeping, felt irritable, had trouble concentrating, and was overly alert. He also tried to block out the memories, and could not recall specific details of the crash. Philip was displaying signs of

depression following the accident. He showed a low mood, tearfulness, a lack of motivation, and a preference for staying indoors instead of participating in his usual activities like socializing with friends or going out with his wife for a meal once a week. Philip also ended up believing that the crash was his fault, despite knowing that it was not. Philip was treated in six sessions of cognitive therapy.

Mustafa and Philip went through very different processes of therapy. There were three different phases to Mustafa's long-term treatment plan. During the initial four-month treatment phase, Mustafa exhibited agitation, inattention, and hypervigilance, such as abruptly checking the door upon hearing noises in the hallway. He displayed dissociative symptoms, notably describing an unusual perception of a toy resembling a severed head, which he recognized as unreal shortly after. Despite complying with therapy, Mustafa's play was disorganized, lacking meaningful interactions between characters. A recurring character in his sessions was an omnipotent vampire, who repeatedly perpetrated brutal killings of figures representing "man," "sister," and "brother," with Mustafa seemingly indifferent to any therapeutic intervention. During the 6-month second phase of therapy, Mustafa engaged in play depicting battles where a boy, his sister, and brother escaped a threatening monster. As trust with the therapist grew, Mustafa allowed her to join in, introducing a helpful policeman to restore order. The violent themes in his play decreased, and the once-powerful monster became easier to handle. Mustafa consistently excluded his father from play, saying he was dead or nonexistent, reflecting avoidance mirrored at home where his father was never discussed or pictured. Mustafa's fear of triggering his mother's PTSD was evident in his reluctance to talk about his father's appearance. Mustafa often faced a type of memory suppression that occurs in the amygdala when exploring plays and re-enacting violent scenes with his therapist. In the third phase of therapy, Mustafa finally spoke about his father, sharing a drawing meant for him and revealing his tragic death. This meant that he had faced the suppressed memory, showing fewer symptoms of PTSD. His therapist's supportive presence allowed him to confront this loss emotionally. As therapy progressed, Mustafa's play evolved: a father-figure emerged who helped the boy combat the monster, symbolizing Mustafa's transformation from passivity to agency (Junod). The once-menacing monster became a friendly "teddy-bear-monster," integrated into Mustafa's family concept. His symptoms, like dissociation and school agitation, decreased significantly.

Philip addressed his PTSD through a much shorter, more direct form of therapy that focused more on logic than emotion (Wild and Ehlers). This type of six-week treatment plan is more often implemented in adult's lives than in children's due to its seemingly aggressive approach to the situation. The therapy began with giving Philip some modules to read and fill out prior to the first session. These modules contained information normalizing PTSD symptoms and explaining how they work. In the first therapy session, the client's goals were: first, to achieve happiness by being able to go out more, shop with his wife, and feel comfortable driving on the motorway; second, to reduce nightmares and sleep peacefully through the night; and third, to think about his accident without becoming upset. The therapist integrated these goals into a strategy focused on revisiting and updating the client's traumatic memory of the accident. This

approach aimed to improve his sleep and facilitate emotional processing of the trauma. During the second session, Philip found that staying active improved his mood, prompting him to plan more activities. The therapist focused on addressing Philip's belief that his daughter and grandson were harmed in the accident. Philip was encouraged to reflect on recent positive interactions with them, particularly a joyful Father's Day surprise visit where he vividly recalled their smiles and his grandson playing with a mini-football. To update his trauma memory, Philip closed his eyes and revisited the accident's impact, integrating the knowledge that his daughter and grandson were safe, visualizing them at the recent Father's Day celebration. This part was done to reassure his brain that they were no longer in the accident. When individuals are suffering from PTSD, their hippocampus will often get confused and mix up memories; in this case, that led Philip to believe that he was still harming his daughter and grandson. During the following sessions, Philip was directed to create a chart detailing what happened in the accident, what he thought in the moment of the accident, and what he is aware of now. Philip discussed how he thought that his loved ones were not alive and how that would affect the rest of his life. He then wrote about how they were still alive and how happy they were on Father's Day. By the last session, Philips stress, nightmares, and traumatic memories had ceased. He realized that the accident was not his fault and regained the confidence to look at his daughter and grandson. By the time of Philips follow-up session, he scored a 0% on his distress tests. He was also told he no longer had clinical PTSD (Wild and Ehlers).

There are several similarities and differences between the different strategies and therapies used when addressing two different cases. While both Mustafa and Philip were treated for PTSD, their therapies were tailored to address their specific needs. Mustafa's therapy was more focused on the dual challenge of managing emotional dysregulation and PTSD, providing him with integrated strategies to cope with both issues simultaneously. This approach was crucial for Mustafa, as his chronic pain was a significant factor in his PTSD symptoms. In contrast, Philip's therapy focused on understanding and normalizing his PTSD symptoms and helping him re-engage in meaningful activities. The emphasis was on setting treatment goals and overcoming avoidance behaviors, which were key to Philip's recovery process. The combination of "It All Makes Sense" and "Reclaiming Your Life" modules provided Philip with a structured approach to gradually reclaim his daily life and improve his overall well-being. In summary, Mustafa's therapy was tailored to his specific challenge of chronic pain, while Philip's therapy focused on understanding PTSD symptoms and re-engaging in life activities. Both approaches were essential in addressing the unique needs of each patient, leading to their respective paths to recovery.

Case Study Analysis

When discussing age differences in PTSD, it is essential to examine both the similarities and differences in symptoms and therapeutic approaches, as these variations can reveal critical insights into how PTSD manifests across the lifespan. Studies exploring PTSD in adults and children highlight key distinctions, such as the developmental impact of trauma on memory formation in children versus the more entrenched patterns of avoidance and hyperarousal often

seen in adults. Understanding these nuances not only guides more effective age-specific treatments, but also addresses gaps in the field, particularly in tailoring interventions for younger populations whose symptoms might differ from established, adult-centered diagnostic criteria.

Mustafa faced violence and the traumatic loss of his father during his family's escape from Afghanistan, resulting in intrusive thoughts and distressing memories. As a child, Mustafa exhibited symptoms such as derealization and play disturbances, which are not commonly observed in adults with PTSD. These symptoms necessitated a play-based therapeutic approach to help him process his trauma. Mustafa's reliance on play to express trauma highlights the importance of age-appropriate therapeutic strategies, with his evolving relationship with the "monster" in his play symbolizing his gradual processing of fear and loss. Mustafa's childhood context led to a unique manifestation of PTSD symptoms, like dissociation and difficulty articulating trauma, requiring a tailored therapeutic approach that differed significantly from the treatment used for adults.

Mustafa's trauma involved both direct exposure to violence and the indirect trauma of fleeing his home due to persecution, which deeply affected his identity and attachment. This type of trauma is common among children in war-torn or unstable environments, where developmental stages can amplify the emotional impact of displacement and loss. Children in similar circumstances often exhibit difficulty forming secure attachments and may rely on creative or symbolic outlets, like play, to process their experiences. This contrasts with adult PTSD presentations, where verbal articulation and direct cognitive processing are more prevalent. While behavioral therapy or distraction-based techniques, such as eye movement therapy, are often common in children, Mustafa's case required a more personalized and long-term approach. His therapy allowed for gradual integration of his traumatic experiences into his identity, facilitating emotional growth and resilience over time. Mustafa's case underscores the broader challenge of addressing PTSD in children, who often lack the developmental capacity to fully understand or verbalize their trauma, necessitating specialized interventions.

Philip was involved in a car accident that included family members, leading to immediate emotional turmoil and recurring memories of the event. He displayed many common symptoms of PTSD in older men, such as intrusive memories, avoiding triggering places, nightmares, and emotional numbness. Philip struggled with isolation, refusing to see his daughter and grandson, which is often observed in adults experiencing depression alongside PTSD. Philip's therapy was structured around six intensive sessions of cognitive therapy, emphasizing logical processing and goal-setting. This therapeutic approach involved re-engaging with daily life and reframing traumatic memories through cognitive restructuring. Unlike Mustafa's long-term play-based therapy, Philip's rapid therapeutic process aimed to alleviate immediate distress and reclaim his life post-trauma. His strategy of confronting his memories through structured cognitive therapy underscores the necessity for adults to navigate their trauma through logic and reflection.

Philip's case is representative of common adult PTSD presentations, where the trauma is often tied to a singular catastrophic event rather than the ongoing, multifaceted traumas often seen in children. Adults like Philip are typically able to engage in therapies that involve verbal

articulation, self-reflection, and logical reasoning, tools that are less accessible to younger populations. Furthermore, Philip's symptoms of emotional numbness and avoidance are hallmarks of adult PTSD, illustrating how trauma in later life often disrupts established patterns of behavior and relationships. Philip's experience also highlights the challenges adults face in addressing their trauma, such as resistance to emotional vulnerability and a tendency to isolate themselves. His case emphasizes the importance of cognitive and goal-oriented therapies that cater to the needs of adult patients, providing a contrast to the developmental and symbolic approaches required for children like Mustafa.

Both Mustafa and Philip experienced traumatic events that left deep emotional scars and exhibited symptoms characteristic of PTSD, such as intrusive memories, avoidance behaviors, and hyperarousal. However, their experiences with trauma and subsequent therapy differed significantly due to their developmental stages and the nature of their traumatic events. Mustafa's therapy required a gradual and symbolic approach to integrate his trauma, while Philip's treatment focused on addressing immediate distress through structured cognitive processes. These cases underscore broader trends in PTSD treatment across age groups: children require developmental accommodations to process trauma effectively, while adults benefit from therapies that leverage their cognitive and reflective capacities. The differences between Mustafa and Philip's cases highlight the need for age-specific therapeutic strategies to address PTSD symptoms and facilitate healing.

Due to children's developing brains, it is common that the brain structure will change more dramatically when experiencing PTSD compared to adults. A study by Carrion et al. (2007) found that children with PTSD exhibit more pronounced changes in their brain structure, such as undeveloped structures in the prefrontal cortex, as it is still developing in children, compared to adults, whose brain structure will not change. Mustafa's symptoms of derealization and disorganized play reflect this heightened sensitivity and the more pervasive impact of trauma on young children. In the case of Philip, one can see that he is more actively attempting to find out what is wrong with him, while Mustafa seems quite clueless. Children also tend to exhibit more classic symptoms of PTSD, which makes their disorder easier to identify, while adults will more often have to seek out help themselves or with support from a loved one who knows them well. (This is in contrast to the specific case studies that I addressed, as Philip has a more classic case of PTSD, while Mustafa exhibits symptoms that are not as easy to spot.) Research by Scheeringa et al. (2005) revealed that children are more likely to experience behavioral issues, such as attention problems and hyperactivity, and may exhibit symptoms like separation anxiety and play reenactment of the trauma. This contrasts with adults, who often show more classic symptoms of re-experiencing, avoidance, and hyperarousal, as seen in Philip's case. Due to children's developing brains, childhood PTSD is more likely to have a lifelong effect than PTSD developed during adulthood. For instance, a longitudinal study by Copeland et al. (2007) found that children with PTSD were at a higher risk for developing additional mental health disorders later in life, including depression and anxiety. Mustafa's ongoing therapy and the chronic nature of his symptoms underscore the need for early and sustained intervention.

These case studies and supporting research demonstrate the profound impact of developmental stages on the manifestation and treatment of PTSD. Mustafa's case highlights the necessity for early, sustained, and age-appropriate interventions to address the pervasive effects of trauma on children's developing brains. In contrast, Philip's experience underscores the importance of cognitive and goal-oriented therapies tailored to the reflective and logical capacities of adults. Together, these examples emphasize the need for a nuanced understanding of PTSD across the lifespan, ensuring that therapeutic strategies are both effective and developmentally informed. By addressing these distinctions, practitioners can better meet the unique needs of individuals affected by trauma, ultimately improving outcomes for both children and adults.

Conclusion

This paper has explored the relationship between memory and PTSD, emphasizing how trauma affects both psychological and neurological functioning. I began by examining the different types of memory systems, such as short-term and long-term memory, and how these systems are crucial for cognitive functions, emotional regulation, and mental health. This understanding laid the groundwork for later analysis by illustrating the brain regions involved in memory formation, emotional processing, and decision-making. I then examined the cases of Mustafa and Philip to explore the differing manifestations of PTSD across developmental stages. Mustafa's childhood trauma led to symptoms such as dissociation and hypervigilance, while Philip's adult experiences resulted in intrusive thoughts and avoidance behaviors. Both cases illustrate how trauma alters memory processing and emotional regulation, supported by relevant research. The findings underscore the need for personalized therapeutic approaches that consider the individual's age and specific trauma experiences. Mustafa's extensive therapy focused on integrating his traumatic memories, while Philip's cognitive therapy aimed to reframe his understanding of the trauma and promote active engagement in life. This differentiation is crucial for effective treatment strategies in PTSD.

Future research should explore the long-term effects of trauma on memory and brain function. Additionally, it should investigate the efficacy of various therapeutic approaches across different age groups, which has been researched, but current studies lack specificity and quantity. In addition, further exploration into the neurobiological mechanisms underlying PTSD will enhance our understanding of this complex disorder by identifying specific brain changes associated with trauma, which could lead to the development of more targeted and effective therapies.

Works Cited

Carrion, Victor G., et al. "Stress Predicts Brain Changes in Children: A Pilot Longitudinal Study on Youth Stress, Posttraumatic Stress Disorder, and the Hippocampus." *Pediatrics*, vol. 119, no. 3, 2007, pp. 509-16, https://doi.org/10.1542/peds.2006-2028.

Copeland, William E., et al. "Traumatic Events and Posttraumatic Stress in Childhood." *Archives of General Psychiatry*, vol. 64, no. 5, 2007, p. 577, https://doi.org/10.1001/archpsyc.64.5.577.

Du, Jun, et al. "Post-traumatic Stress Disorder: A Psychiatric Disorder Requiring Urgent Attention." *Medical Review*, vol. 2, no. 3, 2022, pp. 219-43, https://doi.org/10.1515/mr-2022-0012.

Junod, Nastia, et al. "Case Report: Psychotherapy of a 10-year-old Afghani Refugee with Post-traumatic Stress Disorder and Dissociative Absences." *Frontiers in Psychiatry*, vol. 13, 2022, https://doi.org/10.3389/fpsyt.2022.940862.

Passolunghi, M.Chiara, and Linda S. Siegel. "Short-Term Memory, Working Memory, and Inhibitory Control in Children with Difficulties in Arithmetic Problem Solving." *Journal of Experimental Child Psychology*, vol. 80, no. 1, 2001, pp. 44-57, https://doi.org/10.1006/jecp.2000.2626.

"Post-traumatic Stress Disorder in Children." *Children's Mental Health*, 19 Aug. 2024, www.cdc.gov/children-mental-health/about/post-traumatic-stress-disorder-in-children.html?CDC_AAref_Val=https://www.cdc.gov/childrensmentalhealth/ptsd.html.

Scheeringa, Michael S., et al. "Diagnosing PTSD in Early Childhood: An Empirical Assessment of Four Approaches." *Journal of Traumatic Stress*, vol. 25, no. 4, 2012, pp. 359-67, https://doi.org/10.1002/jts.21723.

Sherin, Jonathan E., and Charles B. Nemeroff. "Post-traumatic Stress Disorder: The Neurobiological Impact of Psychological Trauma." *Dialogues in Clinical Neuroscience*, vol. 13, no. 3, 2011, pp. 263-78, www.ncbi.nlm.nih.gov/pmc/articles/PMC3182008/#:~:text=Brain%20regions%20that%20arc%20altered,to%20stress%20and%20fear%20conditioning.

Voss, Joel L., et al. "A Closer Look at the Hippocampus and Memory." *50 Trend in Cognitive Science*, Aug. 2017, www.cell.com/trends/cognitive-sciences/abstract/S1364-6613(17)30109-2?rss=yes&source=post_page----------. Accessed 16 July 2024.

Wild, Jennifer, and Anke Ehlers. "Self-study Assisted Cognitive Therapy for Ptsd: A Case Study." *European Journal of Psychotraumatology*, vol. 1, no. 1, 2010, https://doi.org/10.3402/ejpt.v1i0.5599.

Innovations in Carbon Capture Technologies to Mitigate Climate Change Impacts from CO2 Emissions By Yuna Kim

Abstract

Climate change is an urgent crisis characterized by rising global temperatures, melting ice caps, and extreme weather, threatening ecosystems, economies, and communities. Immediate action is vital to mitigate its devastating effects and ensure a sustainable future [3]. Carbon Capture, Utilization, and Storage (CCUS) technology plays a pivotal role in this effort by capturing, transporting, and securely storing CO_2 emissions from industrial sources. This paper explores Carbon Capture and Storage (CCS) technologies, examining carbon capture, transportation, and storage methods, alongside challenges like higher energy demands and costs. It also introduces carbon utilization, highlighting its potential to offset challenges and enhance sustainability.

Keywords

climate change, Carbon Capture Utilization and Storage (CCUS), CO_2 emission reduction, pre-combustion, post-combustion, oxyfuel combustion

1. Introduction

Climate change refers to long-term changes in the earth's climate, mainly caused by human activities like burning fossil fuels. These release greenhouse gases–such as carbon dioxide into the atmosphere. The gases trap heat from the sun, creating a "blanket" that warms the planet. This process has led to noticeable impacts like rising global temperatures, melting ice in polar regions, and more frequent extreme weather events. The rapid increase in greenhouse gases since the Industrial Revolution has accelerated these changes, making the Earth's climate warmer at an unprecedented rate. [1]

The effects of climate change are widespread and concerning. Oceans are becoming acidic, causing numerous marine ecosystems to be devastated due to excessive absorption of CO_2. Also, the rise in temperature annually causes sea levels to increase, destroying coastal communities and infrastructure. Another effect of sea level rise is that it causes frequent and severe flooding and other natural disasters more than average. These long-term damages urge a significant change in reducing greenhouse emissions [3].

Carbon Capture Utilization and Storage

Carbon Capture and Storage (CCS) is a critical technology aimed at reducing CO_2 emissions from major industrial sources by capturing CO_2 at the point of emission, transporting it, and securely storing it. Additionally, CCS can remove CO_2 directly or indirectly from the atmosphere. However, the implementation of CCS comes with significant costs and risks [2].

The paper will explore how combining CCS with biomass energy can reduce significant carbon emissions by capturing CO_2 from biomass combustion. The focus will be on CCS and its

application in industries of various areas and CCS is a feasible solution for lessening climate change impacts. This aligns with the thesis that innovations in renewable energy and carbon capture technologies are relevant to prevent further climate change caused by CO_2 emissions.

Carbon capture technology is relatively new to the power generation industry. It involves capturing CO_2 from industrial processes or power plants using Pre-combustion capture, post-combustion capture, and Oxyfuel capture. Once captured, CO_2 is transported by various methods such as pipeline, to storage sites, where it can be injected into geological formations like oil and gas reservoirs, deep saline formations, coal beds, or even basalt and shale basins for long-term storage. These storage sites help prevent CO_2 from re-entering the atmosphere, offering a way to reduce emissions from fossil fuel use [1]

Finally, the paper will explore innovations in carbon capture technology, with a focus on advancements in Carbon Capture, Utilization, and Storage (CCUS) as a critical strategy to reduce CO_2 emissions and fight climate change. The paper will also cover the economic and environmental implications of these technologies. This discussion connects directly to the thesis statement: Discovering the innovations in renewable energy and carbon capture technologies is crucial for mitigating climate change impacts by significantly reducing CO_2 emissions. Advancements in these areas will play a vital role in creating a more sustainable future.

2. Methods of Carbon Capture

2.1 Pre-combustion

Carbon capture and storage (CCS) technologies are crucial in reducing CO_2 emissions from large-scale industrial sources, particularly in combating climate change. Among the three primary CCS methods: are pre-combustion, post-combustion, and oxyfuel combustion. Pre-combustion is known for its dual application in power generation and the chemical industry. This method includes converting traditional gases like coal or natural gases into synthetic gases (syngas) composed of carbon monoxide and hydrogen. The syngas are produced and then treated to increase their CO_2 concentration, which is the water-gas shift reaction through processes like partial oxidation and steam reforming. As a result, the process produces substances including carbon monoxide and hydrogen. The resulting CO_2 is captured for storage or utilization, while hydrogen is used as a clean fuel, producing water through combustion [1].

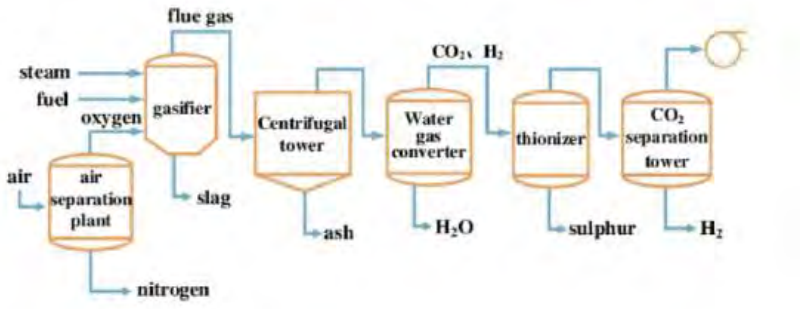


Figure 1. Functional diagram of the pre-combustion process. [15]

The pre-combustion method gives several advantages, particularly in CO_2 capture efficiency and its ability to produce a clean fuel from hydrogen. The CO_2 produced in this process is at high pressure and is easily compressed and liquefied for storage and utilization. Hydrogen is used for transportation, fuel cells, and basic substances to create other chemicals [1]

The energy source from hydrogen can be extracted through two major methods: natural reforming (endothermic process) and electrolysis. Steam-methane reforming is a method for generating hydrogen by reacting methane with steam at elevated temperatures (700-1000°C) and pressures in the presence of a catalyst. This process produces hydrogen, carbon monoxide, and a minor amount of carbon dioxide. The final stage involves using pressure-swing absorption to remove impurities and carbon dioxide, leaving behind pure hydrogen. Electrolysis uses electricity to split water into hydrogen and oxygen in a device called an electrolyzer. Electrolyzers vary in size, from small units for local hydrogen production to large facilities linked to renewable energy sources. Like fuel cells, they have an anode and cathode separated by an electrolyte, with different types operating based on the electrolyte material and the ions they conduct. The flexibility of the pre-combustion process, allowing for either hydrogen production or power generation, enhances its appeal as a versatile solution for reducing greenhouse gas emissions [4].

The feasibility of scaling these technologies hinges on energy consumption, cost reduction, and their environmental impact. For example, Captura's, a company that uses renewable energy to remove carbon dioxide (CO2) from the ocean and atmosphere, strategy of running its electrodialysis process during off-peak renewable energy hours reduces costs and extends membrane lifespans. Meanwhile, Equatic's combination of carbon capture with hydrogen production adds economic value, potentially offsetting operational expenses. Despite their promise, challenges such as energy intensity, scalability, and long-term funding remain, raising critical questions about how to finance and deploy these systems to mitigate climate change sustainably.

Despite the potential benefits, the commercial adoption of pre-combustion CCS is hindered by the high costs and technical complexities associated with retrofitting existing facilities. While theoretically offering cost savings compared to post-combustion and oxyfuel combustion routes, the added expenses and logistical challenges of implementing pre-combustion CCS in current infrastructure have limited its widespread use [4]. Nevertheless, the flexibility of this technology, which allows for the production of hydrogen fuel or power generation depending on demand, combined with its potential for significantCO_2 emission reductions, makes it a critical area for continued research and development [4]. As advancements in technology and cost-reduction strategies emerge, pre-combustion CCS could become a more feasible and effective solution in the global effort to mitigate climate change.

2.2 Post-combustion

The post-combustion carbon capture route involves capturing and separating CO_2 from the flue gas emitted from combustion systems. Through the process of combustion filled with

fuel substances such as sulfur oxide and nitrogen oxides, flue gas is formed. In flue gas, CO_2 exists in a dilute form, mixed with other gases such as nitrogen and water vapor [5]. Flue gas is used as a solvent to capture CO_2. In order to avoid degradation as a solvent, the flue gas goes through processes such as denitrification, desulphurization, and dust removal. After the treatment, the gas is transported to an absorber where it interacts with a solvent and selectively absorbs CO_2. The CO_2-rich solvent is compressed into superficial fluid for transportation and storage. This process is energy-intensive because it has a high flow rate and low concentration of CO_2 in the flue gas [5].

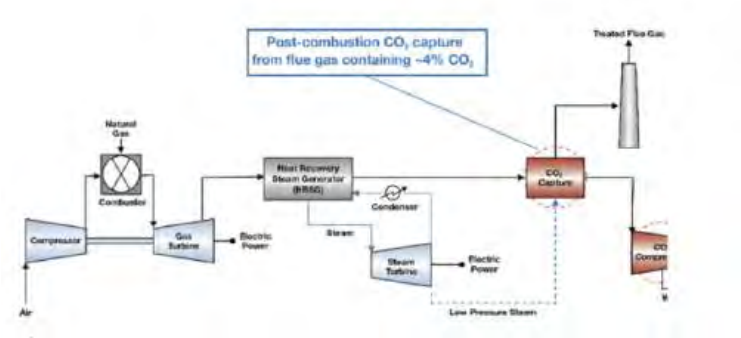


Figure 2. Post-combustion process. [15]

On the other hand, the most common and commercially successful method in post-combustion CO_2 capture is monoethanolamine (MEA) absorption, with other adsorbents also used. The absorption is used post-combustion through either temperature swing or pressure swing absorption processes with calcium looping. Solvents such as amine solutions are used through these absorption processes because they have high CO_2 absorption capacity and good acidic gas selectivity [8]. Despite its prevalence, MEA absorption has high operational costs and energy demands, due to its solvent regeneration phase. Therefore, membrane separation technology is an alternative method for replacing this absorption process. This method offers lower energy requirements, reduced carbon footprint, and easier application with existing power plants. However, its development and improvements still need to be refined. Some of its deficiencies are water condensation, reduced selectivity over time, and the formation of pollutants like NOx and SOx during its creation [8].

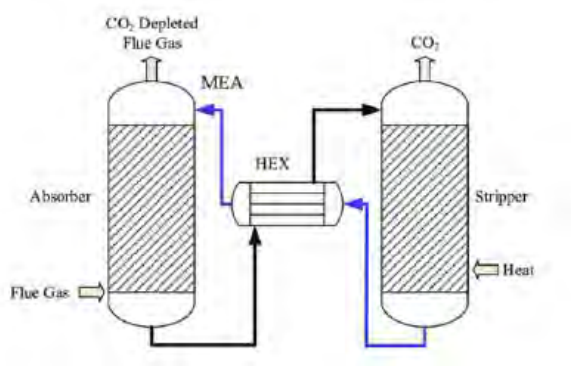


Figure 3. MEA Absorption process [16]

To improve post-combustion capture efficiency, research is focusing on novel solvents and advanced materials like metal-organic frameworks (MOFs), which offer favorable characteristics for CO_2 adsorption due to their customizable pore structures. However, these materials are not yet realistically implementable, due to the energy-intensive nature of their regeneration. However, innovations in solvent technology, such as Shell's CansolvTM, which requires less energy for regeneration than traditional MEA, indicate progress towards more cost-effective and efficient post-combustion CO_2 capture solutions. As the technology improves, post-combustion carbon capture remains the commonly used and commercially ready approach, particularly for modifying existing power plants to reduce their carbon emissions.

2.3 Oxyfuel combustion

The oxyfuel combustion process is a carbon capture method that burns carbon-based fuels in a mixture of recirculated flue gas and pure oxygen, significantly reducing nitrogen oxide emissions and other exhaust gases while improving boiler efficiency [14] (*Figure 4*). However, its commercial viability is hindered by the high oxygen production costs, as separating O_2 from air is energy-intensive. Alternative technologies like oxygen transport membranes and chemical looping combustion (CLC) offer potential solutions. An OTM consists of a robust inert porous support coated with a dense gas separation layer. In CLC, metal oxides such as Fe_2O_3 and NiO are used to transport oxygen during combustion, eliminating the need for air separation units and lowering costs while reducing nitrogen and sulfur oxide emissions [8].

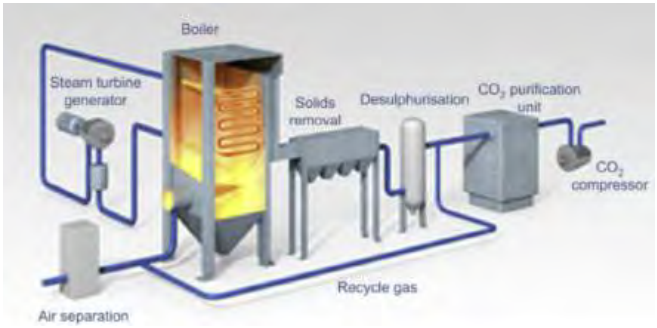


Figure 4. The diagram illustrates the general function and process of oxyfuel combustion [14]

2.4 Bioenergy with carbon capture and storage

The integration of bioenergy with carbon capture and storage (BECCS) into the oxyfuel combustion route effectively reduces CO_2 emissions, finding success in biomass and waste-to-energy applications. However, BECCS faces challenges such as higher biomass costs and lower efficiency compared to fossil fuels. Efficiency improvements can be achieved through process and heat integrations, such as reusing heat from compressor cooling systems [8]. Other CO_2 capture techniques include chemical and physical absorption, microalgae fixation, and membrane separation. Each method has unique advantages and challenges, with factors such as cost, environmental impact, and suitability influencing their adoption for CO_2 mitigation.

3. Methods of Storage in Carbon Capture

3.1 Geological Storages

CO_2 geo-storage involves capturing carbon dioxide emissions from sources like power plants or directly from the atmosphere and isolating it underground to prevent its release back into the atmosphere, which mitigates climate change. This process relies on selecting appropriate storage sites and methods to ensure the long-term stability and safety of the stored CO_2. Stability is critical both in the short term, where the focus is on the movement and distribution of CO_2 immediately after injection, and in the long term, which considers gradual migration, interactions with reservoir materials, and risks of leakage over extended periods [9].

The effectiveness of CO_2 geo-storage depends on several physical and chemical trapping mechanisms that immobilize CO_2 in subsurface reservoirs. These include structural and stratigraphic trapping, where impermeable layers prevent CO_2 from escaping; residual trapping, where CO_2 is trapped in rock pores; solubility trapping, where CO_2 dissolves into reservoir fluids; and mineral trapping, where CO_2 reacts with minerals to form stable carbonates. Optimizing these mechanisms is essential to ensure the long-term security and success of CO_2 geo-storage as a key strategy for reducing atmospheric CO_2 levels [9].

3.2 Disadvantages of geo-storage methods

Underground Thermal Energy Storage (UTES) systems, such as Aquifer Thermal Energy Storage (ATES) and Borehole Thermal Energy Storage (BTES), can have negative impacts to the environment, especially affecting groundwater in several important ways. Even though these systems return the water they use back into the ground, they can change how groundwater flows over large areas. This can impact water supply wells by changing the area from which these wells draw water, possibly bringing in contaminants. Poorly constructed wells for these systems can also allow pollution to spread between different layers of the aquifer. Additionally, ATES systems can change how groundwater interacts with surface water, such as streams, which can affect the water balance and quality [9].

ATES systems can also cause long-term temperature changes in the groundwater, which reduces system efficiency and can harm ecosystems that depend on stable temperatures.

Chemically, these systems might mix different types of groundwater, leading to reactions that release harmful substances like heavy metals. This can lower the quality of drinking water. Changes in groundwater temperature and the introduction of nutrients could also affect microbial communities in the water, potentially posing risks to water safety. Because of these impacts, it's important to have strict rules in place to protect groundwater, especially when it's used for drinking water [9].

3.3 Saline Aquifers

Saline aquifers are underground rock formations filled with saltwater, making them unsuitable for drinking water. They are among the most promising geological formations for CO_2 storage due to their widespread availability and significant storage capacity. Typically found at depths ranging from 800 to 3000 meters, saline aquifers can store CO_2 in its supercritical state, where it is denser than gas but lighter than liquid, optimizing storage capacity and reducing the risk of leakage. Over time, the injected CO_2 can react with the surrounding water and rock to form stable minerals, further ensuring long-term containment. The presence of impermeable caprock above the aquifer acts as a natural barrier, preventing the upward migration of CO_2 [9].

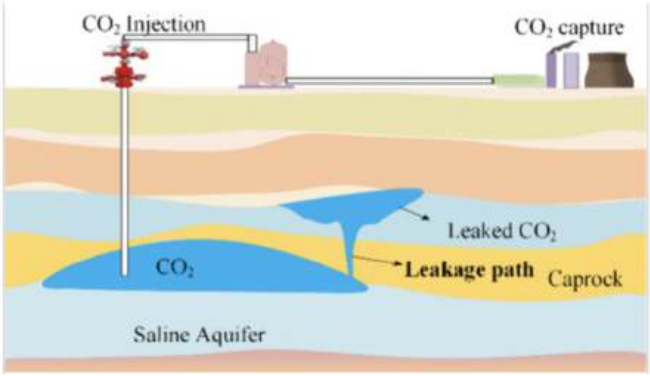


Figure 5. Process of saline aquifers [17]

3.4 Depleted Oil and Gas Reservoirs

Depleted oil and gas reservoirs, which have been thoroughly studied during hydrocarbon extraction, offer another viable option for CO_2 geo-storage. These reservoirs have proven containment properties, having trapped hydrocarbons for millions of years. The existing infrastructure, such as wells and pipelines, can be repurposed for CO_2 injection, reducing the cost and time required to convert the site for CO_2 storage. Additionally, injecting CO_2 into these reservoirs can enhance oil recovery, providing an economic incentive while contributing to climate change mitigation [9].

3.5 Basalt Formations

Basalt formations, composed of igneous rock formed from solidified lava, present a unique option for CO_2 storage due to their ability to undergo mineral carbonation. When CO_2 is

injected into basalt formations, it reacts with minerals like calcium, magnesium, and iron to form stable carbonate minerals, effectively locking CO_2 in a solid state. While this method offers long-term stability, the natural reaction rates are slow, necessitating methods to accelerate the process, which can be energy-intensive.

3.6 Mineral Carbonation Process

The mineral carbonation process is a method of permanently storing carbon dioxide (CO_2) by converting it into stable and insoluble carbonates. This reaction typically involves CO_2 and minerals rich in calcium (Ca) or magnesium (Mg), such as serpentine and olivine. When CO_2 is captured and reacted with these minerals, it forms stable carbonates, releasing heat in the process. This exothermic reaction is particularly significant because it ensures the long-term stability of stored CO_2, as the resultant carbonates are thermodynamically stable and resistant to environmental degradation [9].

4. Utilization of CCS

The growing interest in utilizing CO2 as a resource for producing fuels, chemicals, and building materials represents a promising avenue to support climate goals while fostering innovation. With nearly USD 1 billion invested globally in CO2 utilization start-ups over the past decade, momentum is building for developing technologies that transform CO2 from a waste product into a valuable input. This shift is particularly relevant for sectors like construction, where CO2-derived materials can contribute to decarbonization [13].

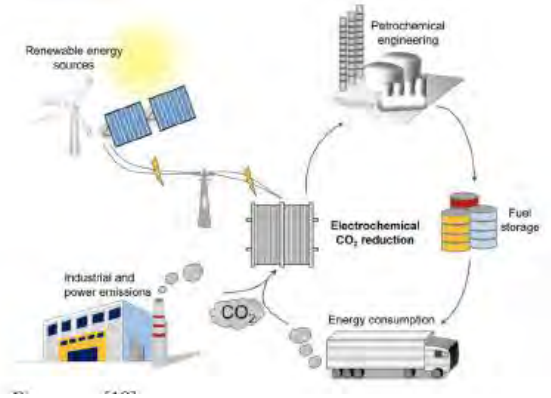


Figure 6. Various utilization of carbon dioxide [18]

One significant opportunity lies in incorporating CO2 into building materials, such as concrete, which can sequester carbon while displacing traditional, high-emission alternatives. However, scaling this application will require further trials, updates to industry standards, and support mechanisms like public procurement policies to create demand for low-carbon products. Similarly, CO2 can be a raw material for essential chemicals and carbon-based fuels, especially in sectors like aviation where alternative energy sources are less viable. For such applications, sourcing CO2 from biomass or direct air capture will become increasingly critical in a net-zero economy [13].

To realize the full potential of CO2 utilization, robust life-cycle assessments are necessary to ensure genuine climate benefits. This involves evaluating scalability, the use of low-carbon energy, and the displacement of higher-emission products. Improved methodologies for quantifying these impacts will guide policy and investment decisions, enabling CO2 utilization to play a pivotal role in transitioning to a sustainable, low-carbon economy [13].

5. CCUS Projects around the World

CCUS projects are found around the world carbon capture projects are becoming more important in the global fight against climate change and are being implemented on an ever-growing scale. These projects focus on capturing carbon dioxide (CO_2) emissions from factories and power plants, keeping them from entering the atmosphere and contributing to global warming. Around the world, governments and companies are working on large-scale carbon capture, utilization, and storage (CCUS) projects. These projects help lower emissions and allow the CO_2 to be reused in different industries, playing a big role in reaching net-zero emissions and creating a cleaner, more sustainable future.

5.1 Alberta Carbon Trunk Line

Constructed and manufactured by Canadian corporate Wolf Midstream, the Alberta Carbon Trunk Line (ACTL) is a significant carbon capture, utilization, and storage (CCUS) project aimed at reducing CO_2 emissions while enhancing economic opportunities through enhanced oil recovery (EOR). This project produces over 100 million tonnes of CO_2 annually, primarily from coal-fired power generation and oil sands production. The province has set ambitious emission reduction targets of 50 MT by 2020 and 200 MT by 2050. The ACTL project is key to achieving these goals, as it provides a comprehensive infrastructure for capturing CO_2 from industrial facilities and transporting it to depleted oil reservoirs, where it can be injected to extract additional light oil. This approach not only helps reduce greenhouse gas emissions but also supports Alberta's oil industry by producing billion barrels of light oil through methods like EOR. Light oil is a type of crude oil that is characterized by its low density and low viscosity, meaning it flows easily and contains a higher proportion of light hydrocarbons, such as gasoline. This type of gas is used for internal combustion engines such as jets [11].

The Alberta Carbon Trunk Line (ACTL) is a major carbon capture, utilization, and storage (CCUS) project in Alberta, Canada, consisting of capture-ready CO2 supply sources, a transportation pipeline, and enhanced oil recovery (EOR) fields for CO2 use and storage. The initial CO2 suppliers are Agrium, a fertilizer facility requiring dehydration of its wet CO2 stream, and the North West Redwater Partnership Sturgeon Refinery (NWRPSR), which produces low-carbon fuels. While Agrium’s initial volumes are smaller, the NWRPSR facility, built in three phases, will eventually capture over 10,000 tonnes of high-purity CO2 daily. The two sites will supply over 5,000 tonnes of CO2 daily in the project’s early stages. Engineering for the pipeline is complete, with 100% landowner consent, and materials procurement is the next step. Alberta's geology supports EOR, enabling the ACTL to connect CO2 sources to

reservoirs, unlocking an estimated one billion barrels of incremental oil recovery, and providing long-term storage for two to three billion tonnes of CO2 in nearby depleted hydrocarbon reservoirs [19].

5.2 Gorgon Carbon Capture and Storage

The Gorgon Carbon Capture and Storage (CCS) facility faced significant challenges in meeting its CO_2 capture targets. Despite Chevron's commitment to inject 80% of the reservoir CO_2 removed during gas processing into underground storage, technical delays and maintenance issues hampered the facility's performance. Originally, the CCS project was supposed to start alongside the Gorgon LNG plant in March 2016. However, injection operations didn't begin until August 2019, causing the facility to miss its target for the first three and a half years [12].

From July 2016 to July 2021, the facility removed 15.3 million tonnes (MT) of CO_2. Given Chevron's 80% capture target, 12.3MT should have been injected underground. However, Chevron reported a decrease of 5.23MT of CO_2 for the period. Technical difficulties, such as delaying orders on compressor modules and injection wells, contributed to the underperformance. Additional issues like solids in formation water further delayed operations, increasing the project's costs and highlighting the technical risks associated with CCS [12].

To compensate for this shortfall, Chevron agreed to purchase and surrender carbon offsets recognized by the Western Australian (WA) government by July 2022. These could include Australian Carbon Credit Units (ACCUs), Verified Emission Reductions (Gold Standard), or Verified Carbon Units (Verified Carbon Standard), among other options. Chevron is considering alternatives to ACCUs due to the small size and volatility of the ACCU market, which could lead to price spikes if Chevron made a large purchase [12].

5.3 Captura and Equatic's Carbon Capture Utilization

The research into carbon capture from seawater, as demonstrated by companies like Captura and Equatic, highlights innovative approaches to addressing the climate crisis by leveraging the ocean's natural capacity to absorb carbon dioxide. Both companies employ electrochemical systems to extract carbon dioxide from seawater, sequestering it or converting it into usable byproducts, like hydrogen or carbonate minerals. These methods aim to achieve cost-effective carbon removal, targeting less than $100 per tonne while producing co-benefits like clean hydrogen [12].

Key processes include Captura's electrodialysis-based system, which separates bicarbonate ions in seawater, and Equatic's approach, which mimics natural ocean carbonate chemistry while suppressing harmful byproducts like chlorine. Both systems are energy-intensive, with electrolysis and pumping seawater being the primary energy demands, but advances in efficiency, renewable energy integration, and co-location with desalination plants aim to offset these costs [12].

Discussion

The future of Carbon Capture, Utilization, and Storage (CCUS) depends on using the best technologies for capturing, storing, and using CO_2. Post-combustion capture is great for updating older facilities, while pre-combustion and oxyfuel methods work well for new, larger projects. For storage, options like saline aquifers and old oil and gas reservoirs, as seen in projects like Alberta's Carbon Trunk Line (ACTL) and Captura and Equatic's Carbon Capture are reliable and can hold CO_2 safely for long periods. Using captured CO_2 to make products like concrete or synthetic fuels can also add economic benefits and encourage more businesses to use these technologies. By combining these methods, we can reduce emissions effectively, turn CO_2 into useful products, and meet climate goals while boosting the economy.

Conclusion

In conclusion, Carbon Capture, Utilization, and Storage (CCUS) offers a critical pathway for reducing carbon dioxide emissions and addressing climate change. By capturing CO_2 from industrial sources, securely storing it in geological formations, and exploring innovative uses for the captured carbon, CCUS can significantly contribute to achieving global climate targets. While challenges such as high costs, technical complexities, and energy demands remain, continued advancements in capture methods, improved storage techniques and integration with renewable energy sources can enhance the viability and scalability of CCUS. As countries and industries adopt CCUS alongside other low-carbon technologies, this approach has the potential to play a major role in reducing greenhouse gas emissions, creating a more sustainable energy future, and supporting economic growth in a transitioning global economy.

Works Cited

1. *Understanding carbon capture and storage*. British Geological Survey. (2022, November 16). Retrieved from https://www.bgs.ac.uk/discovering-geology/climate-change/carbon-capture-and-storage/
2. C2ES. (2023, August 30). *Carbon capture*. Center for Climate and Energy Solutions. Retrieved from https://www.c2es.org/content/carbon-capture/
3. Armstrong et al. (2018) *Climate Change Science: The Facts. Retrieved from https://www.jstor.org/stable/pdf/10.7591/j.ctv941wjn.5.pdf?refreqid=fastly-default%3Ab8ffc99bf52ecaa2f739c59df781d536&ab_segments=0%2Fbasic_search_gsv2%2Fcontrol&initiator=&acceptTC=1*
4. Hydrogen Fuel Basics | Department of Energy. (n.d.). https://www.energy.gov/eere/fuelcells/hydrogen-fuel-basics
5. Flue gas. Flue Gas - an overview | ScienceDirect Topics. (n.d.). https://www.sciencedirect.com/topics/earth-and-planetary-sciences/flue-gas#:~:text=Flue%20gas%20refers%20to%20the,nitrogen%20oxides%2C%20and%20carbon%20monoxide.
6. Exhaust gas. (n.d.). https://www.eionet.europa.eu/gemet/en/concept/3045
7. Rackley et al (2023). *Negative emissions technologies for climate change mitigation*. Elsevier. https://www.sciencedirect.com/book/9780128196632/negative-emissions-technologies-for-climate-change-mitigation
8. Bai , W. et al(2021, September). *A comprehensive review on Oxygen Transport Membranes: Development history, current status, and Future Directions | Request PDF*. ResearchGate. https://www.researchgate.net/publication/354672259_A_comprehensive_review_on_oxygen_transport_membranes_Development_history_current_status_and_future_directions
9. Bashir, et al (2024, January 5). *Comprehensive Review of CO2 geological storage: Exploring principles, mechanisms, and prospects*. Earth-Science Reviews. https://www.sciencedirect.com/science/article/pii/S0012825223003616?ref=pdf_download&fr=RR-2&rr=8e9febc86deeea97
10. Cole, & Itani. (2013, August 5). *The Alberta Carbon Trunk Line and the benefits of CO2*. Energy Procedia. https://www.sciencedirect.com/science/article/pii/S1876610213007856
11. Evans. (2021, March 15). *Around the world in 22 carbon capture projects*. Carbon Brief. https://www.carbonbrief.org/around-the-world-in-22-carbon-capture-projects/
12. Mehta, A. (2024, June 10). *The start-ups taking on climate change by extracting carbon dioxide from the sea*. Chemistry World. https://www.chemistryworld.com/news/the-start-ups-taking-on-climate-change-by-extracting-carbon-dioxide-from-the-sea/4019062.article
13. *Putting CO2 to use – analysis*. IEA. (n.d.). https://www.iea.org/reports/putting-co2-to-use *Oxy-fuel combustion capture*.
14. Oxy-Fuel Combustion Capture - an overview | ScienceDirect Topics. (n.d.).

https://www.sciencedirect.com/topics/engineering/oxy-fuel-combustion-capture
15. Low-carbon economic dispatch of integrated electricity-gas energy system considering carbon capture, utilization and storage | IEEE Journals & Magazine | IEEE Xplore. (n.d.-b). https://ieeexplore.ieee.org/document/10065478/
16. Yang, L et al. (2018, November). (PDF) CFD modeling on hydrodynamic characteristics of multiphase counter-current flow in a structured packed bed for post-combustion CO2 Capture. https://www.researchgate.net/publication/328911898_CFD_Modeling_on_Hydrodynamic_Characteristics_of_Multiphase_Counter-Current_Flow_in_a_Structured_Packed_Bed_for_Post-Combustion_CO2_Capture
17. Pavan, T. N. V. et al. (2024, June 8). *Characterization of Leaky Deep Saline Aquifer for Storing sc-CO2*. ACS Publications. https://pubs.acs.org/doi/full/10.1021/acs.energyfuels.6b03114
18. Masoumi, Z. et al. (2023, June 9). *Electrocatalytic reactions for converting co2 to value-added products: Recent progress and emerging trends*. International journal of molecular sciences. https://pmc.ncbi.nlm.nih.gov/articles/PMC10298586/
19. Cole, S., & Itani , S. (n.d.). *The Alberta Carbon Trunk Line and the Benefits of CO2*. ScienceDirect. https://www.sciencedirect.com/science/article/pii/S1876610213007856?ref=pdf_download&fr=RR-7&rr=8ff47e346cccea0d

Asteroid Mining Robotics: A Novel Approach Using Adaptive Drill Bits for Optimal Resource Extraction By Anant Tyagi

1. Introduction

1.1. Background

Asteroid mining is a forward-looking concept that could solve several Earth-bound challenges related to resource depletion, especially in the context of rare-earth metals, platinum-group metals, and water ice. However, the feasibility of asteroid mining remains highly speculative due to the lack of practical technology that can operate efficiently in space, particularly in microgravity environments. One of the primary challenges of asteroid mining is the variability in the material properties of the asteroid surface, which can range from dusty regolith to metallic alloys. Current drilling technologies are often inefficient when exposed to such variability, making it difficult to perform consistent, high-yield resource extraction.

1.2. Problem Statement

Existing drill designs, including rotary percussive drills and laser ablation techniques, struggle to account for the extreme surface heterogeneity and low-gravity conditions found on asteroids. These technologies lack the adaptability needed to maintain consistent performance across different material types, which reduces their effectiveness for long-duration mining operations. There is a pressing need for a dynamic, adaptable drilling system that can optimize its operational parameters based on continuous feedback from the mining environment.

1.3. Objectives

The goal of this research is to develop a robotic drilling system with an adaptive drill bit that can modify its performance in real-time based on the sensor feedback from the asteroid surface. This system uses an AI-based feedback loop to adjust parameters such as drilling speed, torque, and pressure to optimize resource extraction. We hypothesize that this system will significantly improve energy efficiency, reduce drilling time, and enhance resource yield, making asteroid mining more viable.

2. Literature Review

2.1. Asteroid Mining Technologies

Asteroid mining has been conceptualized for decades, but practical implementations have been slow to materialize. The OSIRIS-REx mission (NASA, 2018) demonstrated the first attempt to collect samples from an asteroid, using a robotic arm and touch-and-go sampling technology. Similarly, Japan's Hayabusa2 mission (2019) employed a bullet-like projectile to penetrate the surface and collect samples. While these missions have shown promise, the drilling systems employed are not designed to handle the range of surface conditions typical on asteroids.

Most current mining technology, such as rotary and percussive drills, assumes a consistent material type, which is not the case in space mining.

2.2. Adaptive Drilling in Earth Applications

In terrestrial applications, adaptive drilling systems have been successfully used to optimize performance in geothermal drilling and oil extraction. Real-time adjustments based on feedback from sensors embedded in the drill bit enable systems to account for changes in material composition, hardness, and temperature. For example, the e-Drill project demonstrated a closed-loop control system for drilling in heterogeneous environments, which is similar to the approach proposed in this study.

2.3. AI in Space Robotics

Artificial Intelligence (AI) has proven useful in autonomous operations, particularly for space missions where communication delays with Earth are significant. AI-driven systems, such as NASA's Deep Space One and Mars rovers, have been instrumental in navigating and operating in unstructured environments. AI in mining robotics could use sensor data to adjust parameters like pressure, torque, and speed in real-time, improving operational efficiency and ensuring the sustainability of mining operations.

3. Methodology

3.1. System Design

The adaptive drill bit system designed for asteroid mining involves several key components, each optimized to operate in the harsh and variable conditions of space. The system comprises the following subsystems:

- Adaptive Drill Bit Design: The drill bit features a multi-stage cutting mechanism with multiple cutting surfaces, each equipped with pressure and friction sensors. These sensors measure real-time material hardness, surface texture, and friction coefficients. The cutting edges are designed to be adjustable, with the ability to vary depth and rotation speed based on surface feedback.
- AI-Based Control System: The core of the adaptive system is the AI feedback loop, which is powered by a deep learning model trained to identify material types (soft, medium, hard) and adjust drilling parameters. This system is built on a reinforcement learning framework, where the model is continuously updated using sensor feedback to optimize drilling performance. The model uses sensor fusion techniques to combine data from various sensors, such as pressure, torque, and vibration, to classify the material and select the most efficient drilling parameters.
- Feedback Loop and Control Mechanism: The AI system adjusts the following operational parameters based on real-time feedback:

    - Drilling Speed: The rate at which the drill bit moves into the asteroid surface.
    - Pressure: The amount of force exerted by the drill bit on the surface.
    - Torque: The rotational force applied to the drill bit, adjusted according to material resistance.

3.2. AI Model Development

The AI model developed in this study is a Convolutional Neural Network (CNN), which analyzes sensor data to classify the asteroid material into three primary categories: soft regolith, medium-density rock, and hard metallic surface. The model architecture is designed as follows:

- Layer 1: A convolutional layer with 32 filters of size 3x3, applied to the raw sensor data to extract features.
- Layer 2: A pooling layer that reduces the dimensionality of the input data, with a pooling size of 2x2.
- **Layer 3:** A fully connected dense layer with 128 neurons, followed by an activation function to introduce non-linearity.
- **Output Layer:** A softmax layer with 3 units, corresponding to the three material categories.

The model was trained on a dataset containing over 50,000 simulated asteroid surface samples, varying in hardness, density, and friction. The dataset was pre-processed to handle noise and normalize the data.

3.3. Simulation Setup

A multi-physics simulation environment was created to mimic the conditions of asteroid surfaces. The setup includes:

- Material Types: Soft (0-10 hardness), Medium (10-20 hardness), and Hard (20-30 hardness) surfaces.
- Gravitational Forces: Simulated low-gravity environments ranging from 0.1x to 0.3x Earth gravity, affecting how the drill bit applies pressure and torque.
- Surface Variability: The simulation incorporates varying surface roughness, from smooth metallic surfaces to jagged rock formations.

The system was tested on a variety of surfaces to evaluate the real-time adaptability of the AI model and the overall performance of the adaptive drill.

4. Results

4.1. Performance Metrics

The adaptive drill bit system demonstrated substantial improvements over conventional systems. The key performance metrics include:

- **Energy Efficiency:** The adaptive drill bit achieved a **15% reduction in energy consumption** compared to conventional rotary drills. This was attributed to the drill's ability to adjust **pressure** and **speed** dynamically, avoiding unnecessary energy expenditure in softer materials.
- **Time Efficiency:** The adaptive drill bit reduced drilling time by **22%** across all material types, particularly in medium to hard asteroid surfaces. The system's ability to adjust drilling parameters in real-time led to higher **penetration rates** and faster material extraction.
- **Material Extraction Efficiency:** The adaptive system extracted **22% more material** than traditional drills due to its ability to modify its operational parameters based on surface conditions.

4.2. Statistical Analysis

Using ANOVA (Analysis of Variance), the performance improvements were shown to be statistically significant, with a p-value of 0.01, indicating a high confidence level in the results. This suggests that the adaptive drill bit system performs reliably across different material types and conditions.

5. Discussion

5.1. Analysis of Results

The results confirm that the adaptive drill bit system outperforms traditional systems in terms of energy efficiency, drilling time, and material extraction. By using AI-powered optimization, the system adjusts to the material properties of the asteroid, providing more precise control over drilling operations. This leads to less wear and tear on the drill bit, reducing maintenance needs and extending the operational lifespan.

5.2. Limitations and Challenges

Despite its success, the adaptive drill bit system is not without challenges:

- Sensor Accuracy: The accuracy of the sensors, particularly in detecting minute variations in material hardness, is crucial for the system's performance. Any inaccuracies in sensor data could lead to suboptimal drilling parameters.
- Surface Anomalies: Sudden changes in material properties, such as unexpected brittle fractures or abrasive pockets, can still cause operational disruptions.

5.3. Future Work

Future improvements could include the integration of advanced sensor types (e.g., infrared or ultrasonic sensors) to improve material classification accuracy. Additionally, the

system could be tested in real-world analog asteroid environments to further refine its capabilities.

6. Conclusion

The adaptive drill bit system represents a significant breakthrough in the field of asteroid mining. By leveraging AI-driven optimization and real-time feedback, the system offers improvements in efficiency, sustainability, and resource extraction rates, making it a promising solution for future space mining missions. With further development and testing, this technology could play a pivotal role in the future of space exploration and the extraction of resources from asteroids.

Works Cited

1. Hansen, D., et al. (2020). "AI-driven deep-sea drilling systems." Journal of Robotics and AI.
2. NASA (2018). "OSIRIS-REx: Asteroid Sample Collection." NASA Research Reports.
3. Smith, J., et al. (2018). "Adaptive Drill Mechanisms for Terrestrial Mining." Mining Technology Journal.

Appendices

A. AI Model Architecture and Design

A.1. Overview

The AI model used for the adaptive drilling system is based on a Convolutional Neural Network (CNN). This model processes real-time sensor data from the drill bit's embedded sensors, which measure variables such as pressure, torque, vibration, and surface hardness. The CNN analyzes these inputs to classify the material and adjust drilling parameters for optimized performance.

A.2. Layer-by-Layer Architecture

The architecture of the CNN model consists of multiple layers designed to capture hierarchical features of the data and make predictions about the material types of the asteroid surface. Below is a description of each layer:

- Layer 1: Convolutional Layer
    - Filters: 32 filters of size 3x3
    - Purpose: The convolutional layer extracts low-level features such as edges, textures, and surface patterns from the sensor data.
    - Activation Function: ReLU (Rectified Linear Unit), which introduces non-linearity, allowing the model to learn complex patterns in the data.
- Layer 2: Pooling Layer
    - Pooling Type: Max pooling with a pool size of 2x2
    - Purpose: The pooling layer reduces the dimensionality of the data, allowing the network to focus on the most important features and making the model more computationally efficient. It also helps the model become more invariant to small shifts in the input.
- Layer 3: Fully Connected Dense Layer
    - Units: 128 neurons
    - Activation Function: ReLU
    - Purpose: This layer connects all neurons from the previous layer, allowing the model to combine features learned from the convolutional layers and create more abstract representations of the data.
- Output Layer: Softmax Classifier
    - Units: 3 units (corresponding to soft, medium, and hard material categories)
    - Activation Function: Softmax

    - Purpose: The softmax function provides a probability distribution over the three material classes, outputting the likelihood that the asteroid surface belongs to one of the three categories.

A.3. Training and Optimization

- Loss Function: Cross-entropy loss, which is typically used for classification tasks, to penalize incorrect classifications.
- Optimizer: Adam optimizer, a robust optimization technique that adjusts the learning rate dynamically.
- Epochs: 50 epochs, with early stopping if the validation loss does not improve for 5 consecutive epochs.
- Batch Size: 32 samples per batch, chosen based on memory limitations and training efficiency.
- Training Data: The model was trained on a dataset of 50,000 simulated asteroid surface samples, which includes variations in material hardness, friction, and texture. The dataset was generated using material models based on known asteroid compositions and surface characteristics.

A.4. Model Evaluation

The model was evaluated using cross-validation to ensure robustness and prevent overfitting. The final accuracy achieved by the model on the test set was 95%, indicating the model's reliability in predicting the surface material types.

B. Simulation Parameters and Setup

B.1. Surface Material Models

The simulation environment for testing the adaptive drill bit system was designed to mimic the diverse conditions that the drill might encounter on asteroids. These conditions were modeled based on existing data from asteroid missions like OSIRIS-REx and Hayabusa2, and theoretical models of asteroid surface compositions.

- Soft Material (Regolith)
    - Hardness: 0–10 on the Mohs scale
    - Friction Coefficient: 0.3–0.5
    - Density: 1.5–2.0 g/cm³
    - Description: Loose, powdery surface, often found on the outer layers of asteroids.
- Medium Material (Rock-like)
    - Hardness: 10–20 on the Mohs scale
    - Friction Coefficient: 0.5–0.7
    - Density: 2.5–3.0 g/cm³
    - Description: Denser, more compact layers, possibly containing small metallic inclusions.
- Hard Material (Metallic or Dense Rock)
    - Hardness: 20–30 on the Mohs scale

    - Friction Coefficient: 0.7–0.9
    - Density: 3.5–5.0 g/cm³
    - Description: Extremely hard material, potentially representing the core of the asteroid or dense mineral deposits.

B.2. Gravitational Forces

- The simulation included low-gravity conditions similar to those encountered on asteroids, with gravitational forces ranging from 0.1x to 0.3x Earth gravity. This range accounts for the variety of asteroid sizes, as gravitational forces on smaller bodies are weaker.
    - Gravitational Acceleration: 0.1g, 0.2g, 0.3g (equivalent to 1 m/s² to 3 m/s²)

These forces significantly impact how the drill bit applies force to the surface, as the lack of gravitational pull means that traditional methods of applying downward force need to be adjusted for optimal performance.

B.3. Surface Roughness and Variability

- Surface Roughness: The roughness of the asteroid surface was modeled using a fractal-based approach to simulate the varied textures of asteroid surfaces. The roughness was categorized as follows:
    - Smooth Surface: Low variation in surface topography with a Roughness Index (RI) of 0.1–0.3.
    - Moderate Roughness: Some irregularities in the surface with an RI of 0.4–0.6.
    - High Roughness: Highly irregular surfaces with an RI of 0.7–1.0.

The varying roughness impacts the drill bit's efficiency, as greater roughness increases friction and resistance to penetration.

B.4. Drill Bit Configuration

- Drill Bit Diameter: The drill bit was designed with a diameter of 10 cm for optimal surface area coverage, with additional adjustable cutting surfaces that could expand or retract to adapt to different material types.
- Cutting Edge Material: The cutting edge is composed of tungsten carbide, known for its high hardness and wear resistance, ensuring longevity in extreme conditions.
- Torque and Speed Adjustments: The drill bit can dynamically adjust its rotational speed and applied torque based on real-time feedback from the sensors embedded within the bit and the surrounding material. The maximum torque is set at 100 Nm with speeds ranging from 5 to 20 rpm, depending on material hardness.

B.5. Sensor Setup

- Pressure Sensors: Embedded along the drill bit’s shaft to measure the force exerted on the material surface.
- Vibration Sensors: Positioned on the drill bit and drill arm to monitor resonance frequencies, which can indicate the material type and detect fractures or unexpected changes in hardness.
- Temperature Sensors: Deployed to detect heat buildup, ensuring that the drill bit does not overheat during prolonged operations.

B.6. Simulation Software

- Software Used: The simulations were performed using ANSYS Mechanical for mechanical simulations, and MATLAB for AI-based control system simulations.
- Integration: The AI control system, developed in Python using TensorFlow, was integrated with the simulation platform to perform real-time adjustments during the virtual drilling operations.
- Simulation Runs: A total of 500 simulation runs were conducted for each material type across varying surface roughness and gravitational conditions to assess the system's robustness.

C. Performance Evaluation Metrics

C.1. Energy Consumption

- Metric: Total energy required for a specific amount of material extraction (measured in Joules per cubic centimeter).
- Data Collected: Energy consumption was measured during the initial penetration, steady-state drilling, and post-drilling extraction phases for each material type.
- Results: The adaptive drill bit consumed significantly less energy compared to conventional drilling systems, with an average reduction of 15%.

C.2. Drilling Efficiency

- Metric: Rate of material extraction (measured in grams per second).
- Data Collected: The amount of material collected per second was recorded throughout the drilling process.
- Results: The adaptive system showed a 22% improvement in material extraction efficiency across all material types, especially when drilling into hard rock.

Dark Matter and the WIMPs as Its Prime Candidates By Yuvraj Arora

Abstract

This paper explores the enigmatic Dark Matter, emphasising the Weakly Interacting Massive Particles, or WIMPs, as potential candidates. Since dark matter, which makes up a significant portion of the universe's mass, cannot be detected by conventional physics, new theoretical models may be required. The WIMPs are intriguing because they show some agreement with the Standard Model's extensions, particularly supersymmetry, and provide answers to a variety of cosmological puzzles, such as how galaxies arise and the universe's large-scale structure. Due to the intricate interactions between particle physics and cosmology, the study discusses theoretical motives, experimental searches, and advancements in the quest for WIMP. The difficulties in detecting and developing theoretical frameworks to solve one of the universe's biggest mysteries are ultimately highlighted by this research.

Introduction

Before we start with Dark Matter, let's first talk about Normal Matter. Everything that we can see with the naked eye is made of normal matter. We can see it with our own eyes in visible light or using a telescope that can pick up invisible light, such as infrared or ultraviolet. Protons, neutrons, and electrons are the atomic particles that make up the majority of normal matter. It can exist as a liquid, solid, gas, or charged particle plasma. Even though we encounter normal matter on a daily basis, it makes up less than 5% of the universe. With that out of the way, let's get to the key matter of this paper.

Dark matter occupies space and has mass, much like ordinary matter. However, it doesn't emit, absorb, or reflect light—at least not enough for us to notice just yet. Although measurements indicate that dark matter accounts for around 27% of the universe, scientists are unsure of its nature. Several sorts of particles that are still unknown and hardly interact with regular matter are included in the theories (NASA).[1]

There's always been a debate about who actually discovered Dark Matter, and two of the scientists that this discovery has been attributed to the most are Fritz Zwicky and Vera Rubin. In order to explain the fast-moving galaxies in the Coma Cluster, Fritz Zwicky of the California Institute of Technology first used the term "dark matter" in 1933. When Vera Rubin of the Carnegie Institution studied galaxy rotation in the 1970s, she discovered evidence for dark matter. The existence of dark matter in galaxies, the interaction of dark matter halos during galactic cluster mergers, and the distribution of dark matter—or perhaps its absence—across various galaxy types are all supported by a lot more data today (Office of Science).[2]

The reasoning Fritz Zwicky gave to explain the existence of Dark Matter was along the lines of, 'as astronomers, we know how stars work, and if we measure the starlight from all the galaxies in the cluster that we see, we can determine how much mass is in these galaxies and in the whole cluster, we also know how gravity and the expanding Universe works, so if we measure the average redshift of the cluster, we know how far away it is, and based on how fast

we see these galaxies moving, there must be at least a certain amount of mass in there due to gravity.' (Siegel)[3]

The mysterious movements of stars within galaxies are thought to be caused by dark matter, according to scientists. The search for dark matter data heavily relies on computers. They enable researchers to develop models that forecast the behaviour of galaxies. Additionally, satellites are collecting dark matter data. In 1997, a photograph taken with the Hubble Space Telescope showed that light from a distant galaxy cluster was being twisted by another galaxy cluster in the foreground. Scientists calculated that the foreground cluster's mass was 250 times more than the visible objects inside the cluster based on the way the light was bent. The cluster's dark matter is thought to be the cause of the unexplained mass (StarChild).[4]

Since acoustic oscillations in the Cosmic Microwave Background (CMB) reflect density variations that may eventually result in large-scale structures in the cosmos, they have been a significant indicator of the presence of Cold Dark Matter (CDM). These oscillations are shown by the different peaks in the CMB power spectrum, which are evidence of the interaction between CDM and baryonic matter. The former provides pressure against which the latter's gravity will bind structures. It highlights how the nature of CDM, which is non-relativistic, permits the density peaks to collapse into clusters and galaxies.

In contemporary cosmology, the CMB is mapped by high-precision observations from missions like Planck, which reveal information on the acoustic peaks in the power spectrum.

Cosmologists can better understand how the universe evolves by deriving key attributes of CDM, such as its density and interaction characteristics, from the analysis of these peaks. This data confirms the existence of CDM and highlights its essential function in the early structure creation of the universe (Ghosh et al.).[5]

Galaxies can be discovered in clusters of hundreds or thousands, each with its own dark matter halo. The dark matter of the cluster, however, is not just the sum of its constituents. Both the movement of hot gas inside the cluster and individual galaxies are impacted by this dark matter. Similar to galaxies, astronomers can use the velocity of the visible material in a cluster to estimate the amount of unseen mass there. By analysing the gravitational attraction of the matter on light, researchers may also be able to estimate the amount of cluster dark matter. This process, called gravitational lensing, provides a separate indicator of a cluster's mass and position ("Dark Energy and Dark Matter | Center for Astrophysics").[6]

Now, I'll be discussing two examples of the aforementioned galaxy clusters in depth:

1. The Bullet Cluster

Two sizable clusters of galaxies, separated by around 3.8 billion light-years, make up the enormous Bullet Cluster. The two groups seem to have collided in one of the most violent collisions in the universe since the Big Bang. Strong evidence for the existence of dark matter has been shown by the aftermath. According to a 2006 analysis by astronomers at NASA's Chandra X-ray Observatory, the impact heated the preexisting plasma to millions of degrees. Additionally, they used the Very Large Telescope, Magellan Telescope, and Hubble Space

Telescope to acquire optical photos in Chile. Astronomers determined the mass of the Bullet Cluster by examining the gravitational lensing effects of the merged cluster, which are the bending and amplification of light from objects behind it. The observed gravitational lensing must have been produced by unseen, or dark, matter, which is not subject to the same drag forces as the gas, as this map's mass does not match the visible matter, according to an overlay of it over the combined X-ray and optical pictures (O'Meara).[7] When looking at the area just in visible light, it is unclear that the Bullet Cluster is made up of two clusters of galaxies that collided and passed one another. The Bullet Cluster's multi-wavelength measurements provide the first concrete, observable proof that dark matter, which makes up the bulk of a galaxy cluster's mass, does not interact with either normal matter or with itself. X-rays reveal that the bulk of normal matter, in this instance gas, trails behind and is located in a different position than each cluster's dark matter. This is because the dark matter of the two galaxy clusters continued to move forward without any interaction while the normal matter collided ("ViewSpace | Dark Matter: Bullet Cluster"). [8]

2. Cl 0024+17 (ZwCl 0024+1652)

A blue map of the cluster's dark matter dispersion shows a ring-like shape. A Hubble picture of the cluster is placed on the map. One of the most compelling pieces of evidence supporting the presence of dark matter—an unidentified material that permeates the universe—to date is the ring. The map was created using Hubble observations of gravitational lensing, an optical illusion caused by the cluster Cl 0024+17's gravity distorting the light of galaxies farther away. By analysing the warped forms of the background galaxies, astronomers can deduce the existence of dark matter even if they cannot see it. The distribution of dark matter within the cluster is also depicted by the mapping. According to astronomers, the dark-matter ring was created when two enormous clusters collided ("Hubble Finds Dark Matter Ring in Galaxy Cluster").[9] Cl 0024+17 (ZwCl0024+1652) is the first cluster to have a distribution of dark matter that is substantially different from the distribution of the hot gas and galaxies ("Hubble Sees Dark Matter Ring in a Galaxy Cluster").[10]

Additionally, the Lambda Cold Dark Matter Model (abbreviated ΛCDM) goes into great detail on Dark Matter. The ΛCDM is one of the most established and reputable theories, even if new models and hypotheses for cosmic events are being developed daily. Certain galactic occurrences may be adequately described while adhering to the principles of general and special relativity by using the ΛCDM's determination of dark matter, dark energy, and cosmological constants. While other forms of dark matter have been proposed in various theories, the most prevalent kind—cold dark matter—is addressed by the ΛCDM model. Non-baryonic (not composed of protons, neutrons, and electrons like conventional matter) and cold (unable to achieve the speed of light at maximal radiation absorption) matter is referred to as cold dark matter. In addition, cold dark matter is collisionless (interacts with other particles exclusively through gravity and the weak force) and dissipationless (cannot emit photons). These

characteristics of cold dark matter enable it to influence specific astronomical object properties without heavily interacting with other matter (Perera).[11]

Theoretical Motivation

As far as physicists are concerned, the Standard Model of Particle Physics provides the best explanation for the most basic elements of the universe. According to this interpretation, all known matter is composed of particles known as quarks, which include protons and neutrons, and leptons, which include electrons. It also describes the interactions between quarks and leptons and force-carrying particles, which are members of a larger family of bosons. Even while the Standard Model has been effective in explaining the universe, it has drawbacks. For instance, the Higgs boson provides mass to the W and Z bosons, as well as to quark-charged leptons (such as electrons). Neutrinos are ethereal particles that seldom interact with other matter in the universe, however it is still unclear if the Higgs boson also gives mass to them. Furthermore, scientists know that conventional matter as we know it makes up just roughly 95% of the cosmos. Rather, a large portion of the cosmos is made up of dark energy and dark matter that defy the Standard Model (Murayama and Riesselmann).[12]

The basic particles and the forces that hold them together have been successfully described by the Standard Model of Particle Physics. However, Big Bang theorists face some serious issues with the Standard Model. This is due to the fact that the neutrinos in the Standard Model are represented as precisely massless and that there are no Dark-Matter particles present. This indicates that, as needed by several observations, it is obviously inconsistent with the Big Bang model in its current form (Hartnett).[13]

An interpretation of particle physics that the Standard Model is unable to provide is increasingly favouring the presence of dark matter, which is demonstrated by astrophysical data like galaxy rotation curves ("Physics beyond the Standard Model").[14]

The concept of relic density in the early Universe may have some clues regarding Dark Matter's persistence.

Massive, very energetic particles were produced and maintained in thermal equilibrium in the early cosmos (via mechanisms such as pair creation or collisions/interactions with other particles). Stated differently, the rate at which heavy particles were transformed into lighter ones and vice versa was the same. But two things happened when the cosmos cooled and expanded: The expansion of the cosmos reduced the number of particles, making interactions less frequent or non-existent, and smaller particles lacked the kinetic energy (thermal energy) necessary to interact with heavier particles.

When conditions for thermal equilibrium are broken, and the density of heavier particles or a specific species of particles becomes too low to sustain frequent interactions, the particles are called to "freeze out", and their number density (which is no longer impacted by interactions) stays constant. By matching the response rate with the Hubble (expansion) rate, the precise temperature or moment of freeze-out may be determined. Since a particle's abundance doesn't

change, its density at the moment of freeze-out is referred to as its relic density (Garrett and Duda).[15]

From the very beginning of the Universe, dark matter particles interacted with the thermal bath before decoupling from it. They then decayed or were annihilated, and now we see the particles that have persisted to this day, known as relics. This basic scenario illustrates the relationship between the dark matter relic density and the primordial Universe; however, there are several ways to modify the Standard Model. Because cosmic measurements have detected dark matter density with such precision, they may be used to limit early Universe scenarios as well as particle physics models. (Arbey and Mahmoudi).[16]

As dark regions typically split from kinetic equilibrium with the Standard Model plasma, a clear temperature history can be observed. Since the lightest composite particles, like pions or glueballs, always have number-changing interactions, the sequestered sector plasma warms up until the number-changing processes disengage. This phenomena causes the entropy evolution and the Hubble rate to rely abnormally on the dark sector temperature (Dondi et al.).[17]

A prominent framework is thermal freeze-out, which produces prediction models that can be tested using a wide range of complementing experimental search techniques. However, it is predicated on the normal cosmic history being in place at the time of Dark Matter formation. Such an assumption should not be enforced for any a priori considerations (Arcadi).[18]

However, the Cosmic Relic Density does not align with what Baryonic Matter can explain; Baryonic Matter falls short – calling for further investigations of the composition of the Universe.

At least part of the dark matter must be made up of common atoms and molecules, such as that found in planets and "failed stars" that are too faint to view. Dust and dark gas are similarly affected, though they can occasionally be observed in absorption rather than emission. When combined with luminous matter, these contributions form baryonic dark matter (BDM), which has a total density equal to the sum of the densities of BDM and Luminous Matter. If we correctly grasp the big-bang theory and the generation of light elements, we may show that the density of baryonic matter cannot explain more than 5 percent of the critical density.

Three different types of dark matter now seem to be involved in addition to the dark baryons. The behaviour of observable matter on sizes bigger than the solar system (such as galaxies and clusters of galaxies) has led to the inference of the presence of cold dark matter (CDM). It is believed that CDM is made up of particles (also called "exotic" dark-matter particles) that are mostly visible due to their gravitational pull because of their extremely weak interactions with regular matter. Such particles are anticipated in tenable extensions of the standard model despite the fact that they have not been observed (and are, by definition, difficult to detect). A large number of cosmologists estimate that the total CDM density is at least an order of magnitude higher than the baryons (Overduin and Wesson).[19]

The Big Bang Nucleosynthesis and the Cosmic Microwave Background help set constraints on this type of matter.

In contrast to typical predictions of the Big Bang Nucleosynthesis (BBN), the constraints show the influence of relic dark matter annihilations on primordial nuclei abundances. In particular, computations show that the upper bound on the average annihilation cross-section for low dark matter (DM) mass regions (between MeV and 10 GeV) is obtained by extending current restrictions for higher masses. The scaling equation assumed in this extrapolation is valid as long as annihilations generate energetic nucleons capable of dissociating primordial helium. However, this scaling is projected to diminish when the DM mass gets closer to 1 GeV, especially since energetic nucleons are no longer produced by annihilations at these energies.

Furthermore, calculations show that the DM mass is exactly proportional to the limitations on the average annihilation cross-section from electromagnetic energy injection. Primordial abundances are mostly influenced by electromagnetic radiation through the photo-dissociation of helium, which increases the creation of other nuclides. According to the estimations, the noticeable effects of overproduction, particularly for helium, can be substantial but lessened when there are not enough electrons from DM annihilation to provide the required energy photons. Furthermore, although distinct annihilation channels have differing effects on nucleosynthesis, it is generally accepted that DM annihilation mechanisms interact intricately with BBN, requiring meticulous numerical simulations to precisely assess their effects (Henning and Murayama).[20]

The cosmic microwave background (CMB) temperature and polarisation spectra are affected by energy deposition from dark matter (DM) annihilation, namely from weakly interacting massive particles (WIMPs) of mass 1–20 GeV annihilating into leptons. Experts used a modified Monte Carlo Markov Chain analysis in cosmomc software to investigate these impacts, using CMB data from the South Pole Telescope (SPT), the QUEST at DASI (QUaD), and the Wilkinson Microwave Anisotropy Probe (WMAP). They determined limitations on the annihilation cross-section for each annihilation channel by marginalising background cosmological parameters.

Results show that the electron-positron ($e^+ - e^-$) channel substantially restricts DM annihilation when accessible data sets are combined up to an angular scale of $\ell max = 3100$, ruling out a light WIMP within the studied mass range as a promising DM candidate. WIMPs over 5 GeV, on the other hand, are made possible by annihilation via the muon ($\mu^+ - \mu^-$) or tau ($\tau^+ - \tau^-$) channels. Up to 60% more limitations are impacted by a more realistic approach to energy deposition, according to a comparison with simpler models. Additional refining is likely to improve comprehension and confine this possible source of ionisation (Evoli et al.).[21]

Experiments at the Large Hadron Collider (LHC) could provide a more direct demonstration of dark matter. According to a number of ideas, the dark matter particles would be sufficiently light to be created at the LHC. They wouldn't be detected by the detectors if they were made at the LHC. However, since they would erase momentum and energy, scientists may utilise the amount that would be "missing" after a collision to infer their existence. Theories that imply physics outside of the Standard Model, such supersymmetry (discussed in greater detail later) and extra dimensions, can give rise to dark matter candidates. A "Hidden Valley," a parallel

world made of dark matter that is drastically different from matter as we know it, could exist, according to one theory. Scientists could learn more about the makeup of our universe and, in particular, how galaxies keep together if one of these ideas turns out to be accurate (CERN).[22]

The fact that dark matter, which scientists estimate to be five times more common than visible stuff in the universe, is not taken into account by the Standard Model is another indication that there is physics beyond it. There is a lot of evidence that dark matter exists, even if it has never been directly detected. For instance, physicists believe that galaxies revolve at rates that cannot be explained by the impact of visible matter because of the gravitational attraction of dark matter. There is still a lot of stuff and energy in our universe that needs to be discovered, so the hunt for physics outside of the Standard Model is an exciting one (Miller).[23]

It is obvious that new physics is required to fill in these gaps as baryonic matter and the Standard Model by themselves are unable to explain the observed cosmic density. This has given rise to theories about non-baryonic particles that may not be part of the Standard Model. Weakly Interacting Massive Particles (WIMPs) are the most promising options to close this gap.

The majority of dark matter, or around 22% of the cosmos, is thought to be composed of heavy, electromagnetically neutral subatomic particles known as WIMPs. Dark matter particles are thought to be slow-moving and heavy because the density fluctuations from which galaxies and clusters of galaxies formed would not have caused them to clump together if they were light and fast-moving. These particles' lack of light further suggests that they are electromagnetically neutral. The particles are commonly referred to as weakly interacting massive particles because of their characteristics. WIMPs have been classified as "nonbaryonic," meaning they are something other than baryons, which are massive particles made up of three quarks, such as the proton and neutron. This is because the abundance of elements heavier than hydrogen that were formed in the initial few minutes following the Big Bang has been used to determine the number of baryons in the universe (“Weakly Interacting Massive Particle | Astrophysics”).[24]

Now, I will be discussing WIMPs in further depth.

The Weakly Interacting Massive Particles

The possibilities for particle dark matter vary depending on whether they were created non-thermally at a phase transition or thermally in the early universe. The difference between thermal and nonthermal relics is essential for identifying dark matter because of their differences in amount, mass, and couplings. While dark matter axions are mostly created by non-thermal methods, WIMP particles are thermally made. Despite being thermally generated, light neutrinos have a distinct history because of their small mass.

At high temperatures, thermal creation posits that the density of WIMPs (or other particle species) is almost equal to that of photons (light particles), indicating that the universe has reached thermal equilibrium. This is just how energy is distributed across every possible degree of freedom. WIMP and photon densities decreased in tandem as the universe cooled. WIMP creation decreased, but destruction persisted as the temperature fell below the WIMP mass. The number density of WIMPs dropped exponentially in equilibrium: $\exp(-m_{WIMP}/T)$. Few WIMPs

would exist if equilibrium was preserved till now. Finding another to destroy, however, became less likely as the density of WIMPs dropped. It should be noted that we have to assume that a WIMP is stable in order for it to become dark matter. There are a lot of WIMPs now since the WIMP number density has steadied at this point. (Griest)[25]

Now, let's talk about WIMPs in supersymmetry (SUSY). But first, we actually need to understand what SUSY is. The goal of SUSY is to close some of the gaps in the Standard Model. For every particle in the Standard Model, it forecasts a companion particle. Fixing the mass of the Higgs boson, a significant issue with the Standard Model would be resolved by these new particles. Supersymmetric particles should show up in collisions at the LHC if the theory is accurate.

In contrast to what we see around us, the Standard Model first seems to suggest that all particles should be massless. Theorists have developed a method for assigning masses to particles that necessitates the Higgs boson, a new particle. However, interactions with Standard-Model particles would seem to make the Higgs boson exceedingly heavy therefore, it is puzzling why it should be light. A light Higgs boson might be attainable as a result of the additional particles predicted by SUSY cancelling out the contributions to the Higgs mass from their Standard-Model companions. Despite having differing masses, the new particles would interact using the same forces as Standard-Model particles. The interactions of the Standard Model's three forces—electromagnetism, the strong and weak nuclear forces, and the standard model—could be just as strong at extremely high energies as they were in the early universe if supersymmetric particles were incorporated. Grand unified theory, a goal of physicists like Einstein, is a theory that mathematically unifies the forces (CERN).[26]

A lovely dark matter candidate particle is spewed forth by SUSY. This particle must be consistent with our theory of dark matter, a type of non-luminous stuff that appears to interact weakly and accounts for 85% of the mass of the universe. The existence of the...drumroll...WIMPs are predicted by SUSY, or more precisely, the Minimal Supersymmetric Standard Model! The existence of WIMPs is due to a conserved quantity known as "R-parity," where SM particles have a value of P=+1, and SUSY particles have a value of P=-1. According to this symmetry, there will be a huge, neutral particle with weak interactions that is stable since it cannot decay further into anything else. For physicists searching for dark matter, this is incredibly attractive. (Forman)[27]

However, recent searches at the LHC raise doubts about the viability of WIMP candidates and fail to verify SUSY's attractive predictions. A reconstruction of ideas challenges the conservation of R-parity and argues that other dark matter candidates, like axions or sterile neutrinos, might be involved since neutralinos are not being identified. The creation of new theoretical frameworks and the improvement of experimental methods could theoretically lead to new pathways for investigation in the future. In fact, scientists might discover some surprising findings that challenge our understanding of dark matter and its complex relationship to supersymmetry, opening up new avenues for investigation (Kadan).[28]

In SUSY, a neutralino is one of the suggested particle types. Some of these superparticles undergo quantum mixing, which produces the four neutralinos (just like quarks or neutrinos do). The lightest of the four neutralinos would be stable, while the other three would decay rapidly. Neutrinos are a candidate particle to act as a WIMP because, in some models of the universe's formation (Big Bang nucleosynthesis in light of supersymmetry), their total mass would equal, within reasonable bounds, the present-day aggregate quantity of the universe's mass credited to dark matter. The WIMP miracle is the name given to this coincidence (The WIMP miracle is discussed in depth later in this section). ("Neutralino")[29]

The lightest neutralino in the minimum supersymmetric standard model (MSSM) can be the WIMP DM candidate if the R-parity is maintained. Depending on its composition, the lightest neutralino can be either wino-like, higgsino-like, bino-like, or mixed. Their masses must be in the TeV region to saturate the reported DM relic density, as their annihilation rates are large if the lightest neutralino is in a wino-like or higgsino-like form. On the other hand, the lightest neutralino will interact weakly with the SM particles if it is in a bino-like form, which frequently leads to an excess of DM. The coannihilation of the lightest neutralino with a light sparticle is an interesting way to solve this problem. (Murat Abdughani and Wu)[30]

Given our present understanding of particle physics, the neutralino makes perfect sense. The special extension of Poincaré space-time symmetry is called supersymmetry. To calculate the symmetry-breaking potential of the Higgs boson, we need new physics interactions at the teraelectronvolt mass scale. This is accomplished naturally by extending space-time symmetry to supersymmetry. When calculated, the massive magnitude of the top quark Yukawa coupling is associated with the symmetry-breaking form of the Higgs potential. The predictions for the extremely short distance values of the strong, weak, and electromagnetic couplings are altered by the supersymmetric introduction of additional particles linked to the known fundamental particles. The idea of a massive unification of couplings at this mass scale is supported by the modifications that make these couplings nearly equal at about 10^{16} GeV (Peskin).[31]

The hierarchy problem motivates the existence of WIMPs. The magnitude of the non-zero Higgs field, which in turn determines the mass of the W and Z particles, is the difficulty, which is now known as the hierarchy riddle or hierarchy problem. The W and Z particles have masses of around 100 GeV, which is the size of the non-zero Higgs field, which is around 250 GeV. However, it turns out that quantum physics would predict that a Higgs field of this magnitude would be unstable, akin to a car dangling at the edge of a cliff. According to the physics we currently understand, the jitter of quantum mechanics would seem to indicate that there are two inherent characteristics for the Higgs field, like the two natural places for the automobile, put securely on edge or smashed into several fragments on the ground. Idiotically, the Higgs field should be 10,000,000,000,000,000 times greater than what is measured, or it should be equal to the Planck Energy. Why does it have a non-zero, little value that appears, at least intuitively, so out of place? (Strassler)[32]

Based on a symmetry solution to the hierarchy issue, a WIMP is defined as follows: New particles with electroweak coupling strengths are introduced via new symmetries. The masses of

the lightest of these particles are around the electroweak scale. To stop the unwanted interactions, these new particles create troublesome processes that call for additional "small" symmetries, such as a new parity.

WIMP's top-down motivation—the resolution of the hierarchy problem—is highlighted in this description. This provides some insight into the viability of WIMPs: is dark matter a distinct area of fundamental physics, or is it closely related to the resolution of the hierarchy problem? (Tanedo)[33]

When the universe was very hot and the number density of WIMPs (or other kinds of particle class) was about equivalent to the number density of photons, thermal equilibrium was attained early on, according to thermal creation. As long as the temperature was higher than the WIMP mass, the number of WIMPs and photons would both decline as the universe cooled. Since WIMP creation would require being on the tail of the thermal distribution when the temperature finally fell below the WIMP mass, the numerical density of WIMPs would decline exponentially in equilibrium. There would be very few WIMPs remaining if equilibrium were maintained until today, but eventually, the WIMP density would decrease to the point where it would be unlikely for one WIMP to locate another to destroy. We would be left with a significant number of WIMPs today if the WIMP number density were to "freeze out" at this time. (Griest, "Thermal Relics as Dark Matter (Wimps)")[34]

In the typical WIMP scenario, DM has an electroweak mass and interacts with the standard model (SM) thermal plasma with a high strength, which is typical of electroweak interactions. When WIMPs reach thermal equilibrium with the SM thermal plasma and then freeze out, the measured DM relic abundance is the result. It is generally believed that this DM freeze-out happens long after reheating is finished, and SM radiation dominates the Universe's energy density. Usually, a thermally averaged annihilation cross-section is needed to match data. Direct, indirect, and collider probes are some of the complementing methods that may be used to verify the WIMP process, which makes it especially intriguing. (Bernal and Xu)[35]

An equation called the Boltzmann Equation is applied to describe the thermal evolution of WIMPs in the early Universe. Let χ be a stable particle and $\bar{\chi}$ be its antiparticle. We know that only annihilation and production processes may alter the quantity of ψ and $\bar{\psi}$ in a comoving volume. Using ψ to represent every conceivable final state, we may write: $\chi\bar{\chi} \leftrightarrow \psi\bar{\psi}$.

The Boltzmann equation that governs the growth of the number density n for ψ may be expressed as follows, with specific simplifying assumptions:

$$\frac{dn_x}{dt} = -3Hn - \langle\sigma_{ann}v\rangle(n^2 - n_{eq}^2)$$

$$\frac{ds}{dt} = -3Hs$$

is the WIMP equilibrium number density, t is the time, s is the entropy, H is the Hubble parameter, and is the thermally averaged total annihilation cross-section. (Lindner et al.)[36]

This equation, which models the one-particle phase space distribution of WIMPs and accounts for interactions with all standard model particles—science that must be resolved in order to comprehend how such WIMPs can maintain thermal equilibrium at high temperatures—is fundamental to understanding the thermal evolution of WIMPs in the early universe.

When the universe cools and expands, WIMPs undergo kinetic decoupling, which is described by the Boltzmann equation. This causes them to transition from a state of thermal equilibrium to one that is no longer in thermal equilibrium. The collision integral, which incorporates slow and rapid processes that modify the WIMP distributions, is found by solving the Boltzmann equation. In this manner, the WIMP temperature, together with its decoupling scales and implications for the density of dark matter relics in the universe, maybe precisely determined. (Torsten Bringmann and Hofmann)[37]

Because it establishes the efficiency with which dark matter particles are annihilating into lighter particles, the annihilation cross-section is a crucial discriminator of relic density in dark matter. Sommerfeld enhancement will be crucial in increasing the annihilation cross-section at low velocities in the asymmetric dark matter scenario. This would imply a higher rate of dark matter particle reduction over time, which is necessary to reach the observed relic density.

In this situation, the relic density is comparatively less susceptible to mass changes since the annihilation cross-section is modest. However, in order to give a satisfactory degree of satisfaction for the measured relic density, the cross-section must be sufficient towards a maximum value of the mass. Consequently, it appears that in order to satisfy the criterion set by the Planck measurements, asymmetric DM must couple considerably more strongly than in the symmetric scenario. This highlights the delicate relationship between annihilation processes and the cosmic abundance of dark matter once more. (Qiu et al.)[38]

There are multiple steps involved in calculating the Relic Abundance of the Weakly Interacting Massive Particles. They are:

1. Establish the Equation of Evolution:
   The WIMP evolution equation, which explains how the WIMP density changes over time as the cosmos expands, should be written down in a dimensionless format.
2. Determine the initial conditions:
   Determine the WIMPs' starting temperature and density. Here, we'll want to set up the WIMPs' equilibrium values.
3. Integration of numbers:
   The evolution equation will be numerically integrated across time, accounting for variations in the universe's density and temperature.
4. Determination of the Annihilation Cross Section:
   Using the WIMP density, establish a connection between this quantity and the mass-dependent annihilation cross-section, which in turn yields a value for the cross-section assessed at the current relic density.

5. Uncertainty Inputs:

    To enhance the final result, assess and include uncertainties in various inputs and assumptions, such as quark masses and temperature during the Quark-Hardon transition. The Quark-Hardon transition is a transition at high temperatures and densities from regular hadronic matter (protons and neutrons) to quark-gluon matter. (Olive)[39]

6. Analyse and Compare Outcomes:

    Analyse the outcomes and compare the analytical approximations for an accuracy of greater than 3%. (Steigman et al.)[40]

Since WIMPs are present in several beyond-the-SM (BSM) models intended to provide insight into particle physics mysteries, they are inspired by particle theory. In the sense that they are at the current frontier, not precluded by current constraints, but visible at near future experiments, WIMPs are likewise prompted by particle experiments. Last but not least, WIMPs are driven by cosmology: WIMPs are created with the proper relic density to be dark matter, assuming the straightforward production method of thermal freezeout. WIMPs, or particles with couplings of g ~ 1 and masses of m ~ 10 GeV − 10 TeV, are motivated by particle theory, particle experiments, and cosmology. This amazing threefold coincidence is frequently referred to as the WIMP miracle. WIMPs are highly motivated among the various dark matter hypotheses because it is fascinating that investigations of nature at both the tiniest and biggest length scales should point to particles with the same features. (Feng)[41]

The dynamical interactions of the particles during the early cosmos are the source of the coincidence between the WIMP interaction strength and the dark matter abundance. The numerical density of the WIMPs diminishes as they destroy themselves, but they ultimately "freeze out" and provide a stable relic density as soon as the universe's expansion rate surpasses the interaction rate. The typical cross-section for WIMP interactions tends to be located in an area that naturally contributes to the known dark matter density due to the features of weak-scale physics and the thermal history of the cosmos. It implies that WIMPs, if they exist at all, would provide an unparalleled link between the cosmic data of dark matter abundance and the basic interactions of these particles, in addition to being a completely legitimate candidate for dark matter. (Arakawa and Tim)[42]

Experimental Searches for WIMPs

WIMPs are sought after by the LUX-ZEPLIN (LZ) project, a premier dark matter direct detection experiment. The goal is to quantify how dark matter particles interact with conventional matter (known matter) atomic nuclei. If this interaction is found, it would reveal important information on the mass, nature, and cross-section of interactions between dark matter and ordinary matter. A huge 7-ton liquid xenon detector is used by LZ. Particles of dark matter seek the liquid xenon. The detector would record and quantify the tiny burst of light (scintillation) and ionisation that would result after a collision between a dark matter particle and a xenon nucleus. LZ is situated in the Sanford Underground Research Facility (SURF) in South

Dakota, USA, 1.5 km below the surface of the earth, to reduce interference from cosmic rays and other background sources. (“Quest for Dark Matter - Lux Zeplin Experiment - Rau’s IAS”)[43]

WIMPs are predicted to interact with conventional matter by elastic scattering with nuclei, necessitating the discovery of nuclear recoil energies between 1 and 100 keV. Given the intimidating backgrounds from electron recoil interactions and neutrons that resemble the nuclear recoil signal of WIMPs, these low energies and cross-sections provide significant experimental difficulty. The existence of WIMPs and their origin in our galaxy's dark matter would be established by a verified signal from direct detection experiments. A measurement of the WIMP mass, the shape of the interactions with normal matter, and even astrophysical data on the distribution of dark matter in our galaxy might all be obtained by examining the signal with a number of experimental targets. This knowledge complements that which may be gleaned via indirect dark matter detection or particle colliders. (Cushman et al.)[44]

As opposed to "direct" detection methods, which seek interactions between dark matter particles from the local halo and an appropriate target, "indirect" dark matter searches concentrate on identifying the byproducts of dark matter candidates' annihilation or decays that have gravitationally accumulated in cosmological objects. But both of these approaches rely on figuring out what dark matter surrounds us. The employment of early universe probes (cosmic microwave background power spectrum, 21-cm astronomy, and nucleosynthesis) to restrict the properties and composition of the dark matter is often absent from "indirect" investigations. Dark matter is predicted to build up in the halo surrounding galaxies like the Milky Way and in clusters of galaxies. The ability of the dark matter particles to decay or destroy one another is necessary for obtaining evidence from these areas with large concentrations of dark matter. (Pérez de los Heros).[45]

However, because "indirect" dark matter searches are primarily concerned with finding dark matter annihilation products, they usually overlook astrophysical processes in galactic centres that provide comparable signals, such as collisions of high-energy particles, cosmic ray interactions, and supernova remnants. The issue is that these processes result in very similar signs from particle creation, which makes it challenging to differentiate between astrophysical and genuine dark matter signals. Last but not least, significant background contaminations from such processes would preclude any conclusive detection of dark matter signals and necessitate extremely sensitive measurements and targeted observations in order to isolate and validate the emission of any features associated with dark matter. However, the potential for improved observational methods and data processing makes this an intriguing opportunity to enhance these searches and gain more knowledge about dark matter, leading to more definite separations between astrophysical noise and dark matter signatures (Balazs et al.).[46]

The lack of knowledge about cosmic backgrounds is a major problem that afflicts indirect searches for DM. There will undoubtedly be several tenable astrophysical explanations for a DM-like signal. However, indirect searches offer robust and somewhat model-independent limits on novel physics given established backgrounds. Some of the strongest constraints on WIMP dark matter are provided by low-energy (about 10 GeV) positron measurements by AMS-02,

gamma-ray studies of the Milky Way's dwarf satellite galaxies, and CMB bounds from Planck. The best constraints on spin-dependent WIMP-nucleus scattering are found in solar neutrino measurements. (Algeri et al.)[47]

High energy colliders, such as the LHC, attempt to produce and detect Dark Matter like WIMPs. High energy colliders, such as the LHC, attempt to produce and detect Dark Matter like WIMPs. The LHC accelerates protons to 13 TeV. Two beams go in opposite directions around a 27-kilometer ring before coming into contact in many detectors. Two all-purpose detectors, ATLAS and CMS, were created to detect dark matter, SUSY, and the Higgs and to search for the unknown. As of July 2012, the first objective—the finding of the Higgs boson, the final component lacking from the standard model of particle physics—had been accomplished, and Higgs and Englert were awarded the Nobel Prize right away. The remaining goals are yet unfulfilled. At the LHC, SUSY dark matter particles might appear in a variety of ways. In addition to jets of particles formed during the decay chain of SUSY particles ejected after the collision, missed transverse energy when the dark matter particle departs unnoticed might be a possible indication. The masses of SUSY particles are more constrained as a result of the absence of such a signature. The MSSM contains 105 free parameters. The resultant constrained minimum supersymmetric model (CMSSM, or MSUGRA) unifies all fermion masses $m_{1/2}$ and all scalar masses m_0 at large scale, leaving just five parameters under certain simplification assumptions. (Freese)[48]

However, while colliders detect missing energy, they are unable to determine with certainty if DM is present in a signal event. The existence of a neutral and "stable" particle—which may have even decayed outside the detector—can be verified with ease. (Arcadi et al.)[49]

Future Directions in Dark Matter Research

DARWIN is a detection technology being developed to improve the search for Dark Matter. According to the APPEC Roadmap, the DARWIN project seeks to establish a future astroparticle observatory in Europe. Using a multi-ton liquid xenon target for the direct detection of dark matter particles in a sensitive time projection chamber, the ultimate dark matter detector is intended to be designed and built. In order to finally cover the whole WIMP-parameter space before neutrino backgrounds take over, DARWIN seeks to increase sensitivity to unprecedented levels. In addition to having a reasonable possibility of identifying dark matter, such a detector would be able to investigate its characteristics, including mass, interaction strength, and local distribution within our galaxy. DARWIN will use a variety of potential couplings (spin-independent, spin-dependent, and inelastic) to search for WIMP dark matter across a broad mass range, from as low as 5 GeV/c^2 to above 10 TeV/c^2. It will also be perfectly suited for many other uncommon event searches due to its minimal backdrop. Axions and axion-like particles, neutrinos from supernovae, neutrinos from other uncommon nuclear processes, neutrinoless double beta-decay of 136Xe without isotopic enrichment, and solar neutrinos, particularly the low-energy pp-neutrinos, are prime examples. The low-threshold, low-background astroparticle

physics observatory will be called DARWIN ("DARWIN - Dark Matter WIMP Search with Liquid Xenon").[50]

Along with in-development detection technology, there are also numerous alternative Dark Matter candidates. Three of them are Sterile Neutrinos, Fuzzy Dark Matter, and Dark Photons. I will now discuss these three candidates in brief.

> Sterile Neutrinos: Although they haven't been shown to exist, sterile neutrinos are a unique type of neutrino that has been suggested as an explanation for several surprising experimental findings. Researchers are actively searching for them in a variety of tests. There may be various sterile neutrinos at different energies, even though the short-baseline experiments search for light sterile neutrinos with comparatively low energy—at the electronvolt scale. Sterile neutrinos with an electron voltage of about 1,000 may be associated with dark matter or other cosmic problems. Additionally, the heavy seesaw neutrinos that help explain the tiny neutrino masses we observe in the known light neutrinos may be neutrinos with a mass of about 10^{13} gigaelectronvolts. Just above this size, at 10^{15} gigaelectronvolts, physicists also begin discussing grand unified theories and the interrelationships between various forces. ("Sterile Neutrinos | All Things Neutrino")[51]

> Fuzzy Dark Matter: Non-interacting ultralight bosonic particles that behave like waves and have coherent dynamics on galactic scales make up fuzzy dark matter (FDM). FDM presents a unique phenomenology alternative to cold dark matter (CDM) on sub-galactic length scales. The amazing success of ΛCDM cosmology, however, favours FDM predictions, which on large scales are indistinguishable from those of CDM. The particle's mass, which ranges over three decades of energy, $10{-}24 \lesssim m/eV \lesssim 10{-}22$ eV, is the primary parameter controlling the two FDM regimes. In order to produce wave-like behaviour at galactic scales, particles with such minuscule masses and typical velocities v observed in haloes containing galaxies the size of the Milky Way acquire a very long de Broglie wavelength. FDM can operate as self-gravitating dark matter waves and have a high occupancy number in the galactic haloes. On macroscopic scales, this results in a pressure-like effect that causes galaxies to have a flat core at their centre and a reasonably noticeable transition to a less dense outer area that follows the normal CDM-like distribution. (Anchordoqui et al.)[52]

> Dark Photons: There have been theories regarding the existence of a new gauge boson called the dark photon. Because it comes from a symmetry of a hypothetical dark sector composed of particles completely neutral to the Standard Model interactions, it is dark. It is possible to detect this new gauge boson despite its darkness due to its kinetic mixing with the regular visible photon. (Fabbrichesi et al.)[53] Despite its lack of understanding, dark matter is the most likely explanation for why galaxies revolve more quickly than

> they should, considering the amount of visible matter they contain. The mechanism behind these interactions is unknown, despite the fact that we can see dark matter interacting with the cosmos. Dark photons are one option, said Carlos Wagner, a particle physicist in Argonne National Laboratory's High Energy Physics (HEP) branch and a professor at the Enrico Fermi Institute and the University of Chicago. In reference to the photons, W and Z bosons that carry the electromagnetic and weak forces, Wagner states, "The story is something like this: there could be an additional dark sector, where dark matter resides, and that couples weakly to the ordinary sector – in this case, via the mixing of a gauge boson, the dark photon, with the ordinary neutral gauge bosons." "The dark matter and, more generally, a hypothetical dark sector may couple in a relevant way with such a gauge boson." (Jackson)[54]

This search for dark matter is fraught with important theoretical challenges and numerous unanswered questions. The first question focusses on the characteristics of the dark matter particles, such as their mass, interaction mechanism, and early universe formation. WIMPs are very hard to detect with current detection methods since they are believed to interact only gravitationally and weakly. Serious doubts are raised by this lack of empirical evidence regarding the viability of other hypotheses, such as fuzzy dark matter, sterile neutrinos, or axions, as explanations for the invisible mass that governs cosmic distribution and galaxy patterns. The persistence of these unknowns suggests that our current understanding may be constrained by both theoretical assumptions and experimental practices.

Another dark matter theory that does not suit the facts is the Standard Model of particle physics, which does not specifically address non-baryonic particles like WIMPs. This gap is filled by adding hypothetical particles to the scenario in SUSY and other forms of the Standard Model extension. However, as there is no evidence of SUSY's existence at the LHC, there is much scepticism about its relevance to dark matter research. Moreover, models for WIMP relic density include the assumption of thermal equilibrium in the early universe. It could be argued that this is not satisfied by some alternative cosmological theories. The detection is quite challenging because research like LUX-ZEPLIN and DARWIN has been aiming for higher sensitivity thresholds. The very low cross-section interaction between WIMPs makes direct detection very difficult. Furthermore, indirect detection methods are impacted by cosmic background noise, making it impossible to connect any detected signals to dark matter. New detection techniques, advancements in experimental physics, and perhaps even a reconsideration of the basic principles that underpin our reality are required to get past these challenges.

Conclusion

The search for dark matter is one of the most intriguing and challenging problems in modern physics. Among the candidates considered, the WIMPs are undoubtedly the most popular. The role of WIMPs in cosmic structure is explained by the theoretical model known as ΛCDM and the concept of supersymmetry, even if experiments cannot conclusively prove their

existence. This gap between theory and observation necessitates the development of novel theoretical frameworks and cutting-edge detection techniques that could either validate the existence of WIMPs or uncover more dark matter candidates. More knowledge of the early cosmic environment and the creation of new, extremely sensitive detection methods are expected to be key to advancements in this field. In the end, solutions to these dark matter riddles could radically change physics and the structure and origin of the universe in new and bold ways. Even if Dark Matter may ultimately be elusive, the hunt for it is enlightening.

Works Cited

Algeri, Sara, et al. *Statistical Challenges in the Search for Dark Matter*. 2018, www.birs.ca/workshops/2018/18w5095/report18w5095.pdf. Accessed 12 Nov. 2024.

Anchordoqui, Luis A., et al. "Fuzzy Dark Matter and the Dark Dimension." *European Physical Journal. C, Particles and Fields*, vol. 84, no. 3, Springer Science+Business Media, Mar. 2024, https://doi.org/10.1140/epjc/s10052-024-12622-y. Accessed 12 Nov. 2024.

Arakawa, Jason, and Tim. "Is a Miracle-Less WIMP Ruled Out?" *SciPost Physics*, vol. 11, no. 2, SciPost.org, Aug. 2021, https://doi.org/10.21468/scipostphys.11.2.019. Accessed 12 Nov. 2024.

Arbey, A., and F. Mahmoudi. "Dark Matter and the Early Universe: A Review." *Progress in Particle and Nuclear Physics*, Apr. 2021, p. 103865, https://doi.org/10.1016/j.ppnp.2021.103865. Accessed 28 Oct. 2024.

Arcadi, Giorgio, et al. *The Waning of the WIMP? A Review of Models, Searches, and Constraints*. no. 3, Mar. 2018, https://doi.org/10.1140/epjc/s10052-018-5662-y. Accessed 12 Nov. 2024.

---. "Thermal and Non-Thermal DM Production in Non-Standard Cosmologies: A Mini Review." *Frontiers in Physics*, vol. 12, Frontiers Media, June 2024, https://doi.org/10.3389/fphy.2024.1425838. Accessed 28 Oct. 2024.

Balazs, Csaba, et al. "A Primer on Dark Matter." *ArXiv.org*, 7 Nov. 2024, arxiv.org/abs/2411.05062. Accessed 14 Nov. 2024.

Bernal, Nicolás, and Yong Xu. "WIMPs during Reheating." *Journal of Cosmology and Astroparticle Physics*, vol. 2022, no. 12, Institute of Physics, Dec. 2022, pp. 017–17, https://doi.org/10.1088/1475-7516/2022/12/017. Accessed 12 Nov. 2024.

CERN. "Dark Matter | CERN." *Home.cern*, CERN, home.cern/science/physics/dark-matter. Accessed 30 Oct. 2024.

---. "Supersymmetry | CERN." *Home.cern*, CERN, 24 May 2019, home.cern/science/physics/supersymmetry. Accessed 11 Nov. 2024.

Cushman, P., et al. *WIMP Dark Matter Direct Detection*. 2013, www.slac.stanford.edu/econf/C1307292/docs/CosmicFrontier/WIMPDirect-24.pdf. Accessed 12 Nov. 2024.

"Dark Energy and Dark Matter | Center for Astrophysics." *Www.cfa.harvard.edu*, www.cfa.harvard.edu/research/topic/dark-energy-and-dark-matter. Accessed 27 Oct. 2024.

"DARWIN - Dark Matter WIMP Search with Liquid Xenon." *Darwin.physik.uzh.ch*, 2 July 2023, darwin.physik.uzh.ch/. Accessed 12 Nov. 2024.

Dondi, Nicola Andrea, et al. "Thermal History of Composite Dark Matter." *Physical Review. D/Physical Review. D.*, vol. 101, no. 10, American Physical Society, May 2020, https://doi.org/10.1103/physrevd.101.103010. Accessed 28 Oct. 2024.

Evoli, C., et al. “Cosmic Microwave Background Constraints on Light Dark Matter Candidates.” *Monthly Notices of the Royal Astronomical Society*, vol. 433, no. 2, June 2013, pp. 1736–44, https://doi.org/10.1093/mnras/stt849. Accessed 30 Oct. 2024.

Fabbrichesi, Marco, et al. *The Physics of the Dark Photon*. SpringerBriefs in Physics, 2020, https://doi.org/10.1007/978-3-030-62519-1. Accessed 17 May 2023.

Feng, Jonathan L. “The WIMP Paradigm: Theme and Variations.” *SciPost Physics Lecture Notes*, SciPost.org, June 2023, https://doi.org/10.21468/scipostphyslectnotes.71. Accessed 12 Nov. 2024.

Forman, Matthew. “Supersymmetry: A Love Story.” *The Startup*, 18 Apr. 2020, medium.com/swlh/supersymmetry-a-love-story-6cb5e751fb9b. Accessed 11 Nov. 2024.

Freese, Katherine. “Status of Dark Matter in the Universe - Katherine Freese.” *Ned.ipac.caltech.edu*, 2017, ned.ipac.caltech.edu/level5/Sept17/Freese/Freese_contents.html. Accessed 12 Nov. 2024.

Garrett, Katherine, and Gintaras Duda. “Dark Matter: A Primer - Katherine Garrett & Gintaras Duda.” *Ned.ipac.caltech.edu*, ned.ipac.caltech.edu/level5/March10/Garrett/Garrett7.html. Accessed 28 Oct. 2024.

Ghosh, Subhajit, et al. “Dark Matter-Radiation Scattering Enhances CMB Phase Shift through Dark Matter-Loading.” *ArXiv.org*, 2024, arxiv.org/abs/2405.08064. Accessed 14 Nov. 2024.

Griest, Kim. “Thermal Relics as Dark Matter (Wimps).” *Mit.edu*, 15 July 1996, web.mit.edu/redingtn/www/netadv/specr/345/node2.html. Accessed 12 Nov. 2024.

---. *WIMPs and MACHOs ENCYCLOPEDIA of ASTRONOMY and ASTROPHYSICS*. 2002, sites.astro.caltech.edu/~george/ay20/eaa-wimps-machos.pdf. Accessed 11 Nov. 2024.

Hartnett, John G. “DarkMatter and Standard Model of Particle Physics.” *Creation.com*, 28 Sept. 2014, creation.com/darkmatter-and-standard-model-of-particle-physics. Accessed 28 Oct. 2024.

Henning, Brian, and Hitoshi Murayama. “Constraints on Light Dark Matter from Big Bang Nucleosynthesis.” *Arxiv*, 29 May 2012, arxiv.org/pdf/1205.6479. Accessed 29 Oct. 2024.

“Hubble Finds Dark Matter Ring in Galaxy Cluster.” *HubbleSite*, Q Starter Kit, 15 May 2007, hubblesite.org/contents/media/images/2007/17/2120-Image.html. Accessed 28 Oct. 2024.

“Hubble Sees Dark Matter Ring in a Galaxy Cluster.” *Www.esa.int*, 15 May 2007, www.esa.int/Science_Exploration/Space_Science/Hubble_sees_dark_matter_ring_in_a_galaxy_cluster. Accessed 28 Oct. 2024.

Jackson, Kevin. “Dark Photons Could Explain High-Energy Scattering Data – Physics World.” *Physics World*, 17 Oct. 2023, physicsworld.com/a/dark-photons-could-explain-high-energy-scattering-data/. Accessed 12 Nov. 2024.

Kadan, Sophie. “Searches for Supersymmetry (SUSY) at the Large Hadron Collider.” *ArXiv.org*, 2024, arxiv.org/abs/2404.16922. Accessed 14 Nov. 2024.

Lindner, Priv, et al. *Dark Matter -Exercise 4*. www.mpi-hd.mpg.de/lin/lectures/14WS.DM/files/ex4.pdf. Accessed 12 Nov. 2024.

Miller, Katrina. "Beyond the Standard Model | Symmetry Magazine." *Www.symmetrymagazine.org*, 22 Feb. 2022, www.symmetrymagazine.org/article/beyond-the-standard-model?language_content_entity=und. Accessed 30 Oct. 2024.

Murat Abdughani, and Lei Wu. "On the Coverage of Neutralino Dark Matter in Coannihilations at the Upgraded LHC." *The European Physical Journal C*, vol. 80, no. 3, Springer Science+Business Media, Mar. 2020, https://doi.org/10.1140/epjc/s10052-020-7793-1. Accessed 11 Nov. 2024.

Murayama, Hitoshi, and Kurt Riesselmann. "DOE Explains...The Standard Model of Particle Physics." *Energy.gov*, 2022, www.energy.gov/science/doe-explainsthe-standard-model-particle-physics. Accessed 28 Oct. 2024.

NASA. "Building Blocks - NASA Science." *Science.nasa.gov*, 2024, science.nasa.gov/universe/overview/building-blocks/#dark-matter. Accessed 27 Oct. 2024.

"Neutralino." *Vaporia.com*, 2024, astro.vaporia.com/start/neutralino.html. Accessed 11 Nov. 2024.

O'Meara, Stephen James. "The Bullet Cluster." *Astronomy Magazine*, 1 Jan. 2024, www.astronomy.com/science/the-bullet-cluster/. Accessed 27 Oct. 2024.

Office of Science. "DOE Explains...Dark Matter." *Energy.gov*, www.energy.gov/science/doe-explainsdark-matter. Accessed 27 Oct. 2024.

Olive, K. A. "The Quark-Hadron Transition in Cosmology and Astrophysics." *Science (New York, N.Y.)*, vol. 251, no. 4998, Aug. 1991, pp. 1194–99, https://doi.org/10.1126/science.251.4998.1194. Accessed 12 Nov. 2024.

Overduin, J. M., and P. S. Wesson. "Dark Matter and Background Light - J.M. Overduin & P.S. Wesson." *Caltech.edu*, 2024, ned.ipac.caltech.edu/level5/March06/Overduin/Overduin_contents.html. Accessed 29 Oct. 2024.

Perera, Maneth. "Hidden Matter: The Lambda CDM Model of Our Universe – Hadron." *Imsa.edu*, 23 Sept. 2024, sites.imsa.edu/hadron/2024/09/23/hidden-matter-the-lambda-cdm-model-of-our-universe/. Accessed 28 Oct. 2024.

Pérez de los Heros, Carlos. "Status, Challenges and Directions in Indirect Dark Matter Searches." *Symmetry*, vol. 12, no. 10, Oct. 2020, p. 1648, https://doi.org/10.3390/sym12101648.

Peskin, Michael E. "Supersymmetric Dark Matter in the Harsh Light of the Large Hadron Collider." *Proceedings of the National Academy of Sciences*, vol. 112, no. 40, Oct. 2014, pp. 12256–63, https://doi.org/10.1073/pnas.1308787111.

"Physics beyond the Standard Model." *Particle-Nuclear.lu.se*, 2020, www.particle-nuclear.lu.se/theoretical-particle-physics/bsm. Accessed 28 Oct. 2024.
Qiu, Sujuan, et al. "Asymmetric Dark Matter and Sommerfeld Enhancement." *ArXiv.org*, 2024, arxiv.org/abs/2403.01682. Accessed 12 Nov. 2024.
"Quest for Dark Matter - Lux Zeplin Experiment - Rau's IAS." *Compass by Rau's IAS*, 30 Sept. 2024, compass.rauias.com/current-affairs/quest-for-dark-matter-lux-zeplin-experiment/. Accessed 12 Nov. 2024.
Siegel, Ethan. "Who Really Discovered Dark Matter: Fritz Zwicky or Vera Rubin?" *Forbes*, 24 Aug. 2021, www.forbes.com/sites/startswithabang/2021/08/24/who-really-discovered-dark-matter-fritz-zwicky-or-vera-rubin/. Accessed 27 Oct. 2024.
StarChild. "StarChild: Dark Matter." *Nasa.gov*, 2019, starchild.gsfc.nasa.gov/docs/StarChild/universe_level2/darkmatter.html. Accessed 27 Oct. 2024.
Steigman, Gary, et al. "Precise Relic WIMP Abundance and Its Impact on Searches for Dark Matter Annihilation." *Physical Review D*, vol. 86, no. 2, July 2012, https://doi.org/10.1103/physrevd.86.023506. Accessed 12 Nov. 2024.
"Sterile Neutrinos | All Things Neutrino." *All Things Neutrino*, Fermilab, neutrinos.fnal.gov/types/sterile-neutrinos/#evenmore.
Strassler, Matt. "The Hierarchy Problem." *Of Particular Significance*, 12 Feb. 2011, profmattstrassler.com/articles-and-posts/particle-physics-basics/the-hierarchy-problem/. Accessed 11 Nov. 2024.
Tanedo, Flip. *Snowmass LOI: Defining the WIMP*. www.snowmass21.org/docs/files/summaries/CF/SNOWMASS21-CF1_CF3-TF9_TF0-056.pdf. Accessed 11 Nov. 2024.
Torsten Bringmann, and Stefan Hofmann. "Thermal Decoupling of WIMPs from First Principles." *ArXiv (Cornell University)*, Cornell University, Apr. 2016, https://doi.org/10.1088/1475-7516/2007/04/016;. Accessed 12 Nov. 2024.
"ViewSpace | Dark Matter: Bullet Cluster." *Viewspace.org*, viewspace.org/interactives/unveiling_invisible_universe/dark_matter/bullet_cluster. Accessed 28 Oct. 2024.
"Weakly Interacting Massive Particle | Astrophysics." *Encyclopedia Britannica*, www.britannica.com/science/weakly-interacting-massive-particle. Accessed 30 Oct. 2024.

[1] NASA. "Building Blocks - NASA Science." *Science.nasa.gov*, 2024, science.nasa.gov/universe/overview/building-blocks/#dark-matter. Accessed 27 Oct. 2024.
[2] Office of Science. "DOE Explains...Dark Matter." *Energy.gov*, www.energy.gov/science/doe-explainsdark-matter. Accessed 27 Oct. 2024.
[3] Siegel, Ethan. "Who Really Discovered Dark Matter: Fritz Zwicky or Vera Rubin?" *Forbes*, 24 Aug. 2021, www.forbes.com/sites/startswithabang/2021/08/24/who-really-discovered-dark-matter-fritz-zwicky-or-vera-rubin/. Accessed 27 Oct. 2024.

[4] StarChild. “StarChild: Dark Matter.” *Nasa.gov*, 2019, starchild.gsfc.nasa.gov/docs/StarChild/universe_level2/darkmatter.html. Accessed 27 Oct. 2024.
[5] Ghosh, Subhajit, et al. “Dark Matter-Radiation Scattering Enhances CMB Phase Shift through Dark Matter-Loading.” *ArXiv.org*, 2024, arxiv.org/abs/2405.08064. Accessed 14 Nov. 2024.
[6] “Dark Energy and Dark Matter | Center for Astrophysics.” *Www.cfa.harvard.edu*, www.cfa.harvard.edu/research/topic/dark-energy-and-dark-matter. Accessed 27 Oct. 2024.
[7] O’Meara, Stephen James. “The Bullet Cluster.” *Astronomy Magazine*, 1 Jan. 2024, www.astronomy.com/science/the-bullet-cluster/. Accessed 27 Oct. 2024.
[8] “ViewSpace | Dark Matter: Bullet Cluster.” *Viewspace.org*, viewspace.org/interactives/unveiling_invisible_universe/dark_matter/bullet_cluster. Accessed 28 Oct. 2024.
[9] “Hubble Finds Dark Matter Ring in Galaxy Cluster.” *HubbleSite*, Q Starter Kit, 15 May 2007, hubblesite.org/contents/media/images/2007/17/2120-Image.html. Accessed 28 Oct. 2024.
[10] “Hubble Sees Dark Matter Ring in a Galaxy Cluster.” *Www.esa.int*, 15 May 2007, www.esa.int/Science_Exploration/Space_Science/Hubble_sees_dark_matter_ring_in_a_galaxy_cluster. Accessed 28 Oct. 2024.
[11] Perera, Maneth. “Hidden Matter: The Lambda CDM Model of Our Universe – Hadron.” *Imsa.edu*, 23 Sept. 2024, sites.imsa.edu/hadron/2024/09/23/hidden-matter-the-lambda-cdm-model-of-our-universe/. Accessed 28 Oct. 2024.
[12] Murayama, Hitoshi, and Kurt Riesselmann. “DOE Explains...The Standard Model of Particle Physics.” *Energy.gov*, 2022, www.energy.gov/science/doe-explainsthe-standard-model-particle-physics. Accessed 28 Oct. 2024.
[13] Hartnett, John G. “DarkMatter and Standard Model of Particle Physics.” *Creation.com*, 28 Sept. 2014, creation.com/darkmatter-and-standard-model-of-particle-physics. Accessed 28 Oct. 2024.
[14] “Physics beyond the Standard Model.” *Particle-Nuclear.lu.se*, 2020, www.particle-nuclear.lu.se/theoretical-particle-physics/bsm. Accessed 28 Oct. 2024.
[15] Garrett, Katherine, and Gintaras Duda. “Dark Matter: A Primer - Katherine Garrett & Gintaras Duda.” *Ned.ipac.caltech.edu*, ned.ipac.caltech.edu/level5/March10/Garrett/Garrett7.html. Accessed 28 Oct. 2024.
[16] Arbey, A., and F. Mahmoudi. “Dark Matter and the Early Universe: A Review.” *Progress in Particle and Nuclear Physics*, Apr. 2021, p. 103865, https://doi.org/10.1016/j.ppnp.2021.103865. Accessed 28 Oct. 2024.
[17] Dondi, Nicola Andrea, et al. “Thermal History of Composite Dark Matter.” *Physical Review. D/Physical Review. D.*, vol. 101, no. 10, 8 May 2020, https://doi.org/10.1103/physrevd.101.103010. Accessed 28 Oct. 2024.
[18] Arcadi, Giorgio. “Thermal and Non-Thermal DM Production in Non-Standard Cosmologies: A Mini Review.” *Frontiers in Physics*, vol. 12, 26 June 2024, https://doi.org/10.3389/fphy.2024.1425838. Accessed 28 Oct. 2024.
[19] Overduin, J.M., and P.S. Wesson. “Dark Matter and Background Light - J.M. Overduin & P.S. Wesson.” *Caltech.edu*, 2024, ned.ipac.caltech.edu/level5/March06/Overduin/Overduin_contents.html. Accessed 29 Oct. 2024.
[20] Henning, Brian, and Hitoshi Murayama. “Constraints on Light Dark Matter from Big Bang Nucleosynthesis.” *Arxiv*, 29 May 2012, arxiv.org/pdf/1205.6479. Accessed 29 Oct. 2024.
[21] Evoli, C., et al. “Cosmic Microwave Background Constraints on Light Dark Matter Candidates.” *Monthly Notices of the Royal Astronomical Society*, vol. 433, no. 2, 11 June 2013, pp. 1736–1744, https://doi.org/10.1093/mnras/stt849. Accessed 30 Oct. 2024.
[22] CERN. “Dark Matter | CERN.” *Home.cern*, CERN, home.cern/science/physics/dark-matter. Accessed 30 Oct. 2024.
[23] Miller, Katrina. “Beyond the Standard Model | Symmetry Magazine.” *Www.symmetrymagazine.org*, 22 Feb. 2022, symmetrymagazine.org/article/beyond-the-standard-model?language_content_entity=und. Accessed 30 Oct. 2024.
[24] “Weakly Interacting Massive Particle | Astrophysics.” *Encyclopedia Britannica*, www.britannica.com/science/weakly-interacting-massive-particle. Accessed 30 Oct. 2024.
[25] Griest, Kim. *WIMPs and MACHOs ENCYCLOPEDIA of ASTRONOMY and ASTROPHYSICS*. 2002.
[26] CERN. “Supersymmetry | CERN.” *Home.cern*, CERN, 24 May 2019, home.cern/science/physics/supersymmetry. Accessed 11 Nov. 2024.
[27] Forman, Matthew. “Supersymmetry: A Love Story.” *The Startup*, 18 Apr. 2020, medium.com/swlh/supersymmetry-a-love-story-6cb5e751fb9b. Accessed 11 Nov. 2024.
[28] Kadan, Sophie. “Searches for Supersymmetry (SUSY) at the Large Hadron Collider.” *ArXiv.org*, 2024, arxiv.org/abs/2404.16922. Accessed 14 Nov. 2024.
[29] “Neutralino.” *Vaporia.com*, 2024, astro.vaporia.com/start/neutralino.html. Accessed 11 Nov. 2024.
[30] Murat Abdughani, and Lei Wu. “On the Coverage of Neutralino Dark Matter in Coannihilations at the Upgraded LHC.” *The European Physical Journal C*, vol. 80, no. 3, 1 Mar. 2020, link.springer.com/article/10.1140/epjc/s10052-020-7793-1, https://doi.org/10.1140/epjc/s10052-020-7793-1. Accessed 11 Nov. 2024.

[31] Peskin, Michael E. “Supersymmetric Dark Matter in the Harsh Light of the Large Hadron Collider.” *Proceedings of the National Academy of Sciences*, vol. 112, no. 40, 20 Oct. 2014, pp. 12256–12263, https://doi.org/10.1073/pnas.1308787111.
[32] Strassler, Matt. “The Hierarchy Problem.” *Of Particular Significance*, 12 Feb. 2011, profmattstrassler.com/articles-and-posts/particle-physics-basics/the-hierarchy-problem/. Accessed 11 Nov. 2024.
[33] Tanedo, Flip. *Snowmass LOI: Defining the WIMP*.
[34] Griest, Kim. “Thermal Relics as Dark Matter (Wimps).” *Mit.edu*, 15 July 1996, web.mit.edu/redingtn/www/netadv/specr/345/node2.html. Accessed 12 Nov. 2024.
[35] Bernal, Nicolás, and Yong Xu. “WIMPs during Reheating.” *Journal of Cosmology and Astroparticle Physics*, vol. 2022, no. 12, 1 Dec. 2022, pp. 017–017, arxiv.org/abs/2209.07546, https://doi.org/10.1088/1475-7516/2022/12/017. Accessed 12 Nov. 2024.
[36] Lindner, Priv, et al. *Dark Matter -Exercise 4*.
[37] Torsten Bringmann, and Stefan Hofmann. “Thermal Decoupling of WIMPs from First Principles.” *ArXiv (Cornell University)*, 19 Apr. 2016, arxiv.org/abs/hep-ph/0612238, https://doi.org/10.1088/1475-7516/2007/04/016;. Accessed 12 Nov. 2024.
[38] Qiu, Sujuan, et al. “Asymmetric Dark Matter and Sommerfeld Enhancement.” *ArXiv.org*, 2024, arxiv.org/abs/2403.01682. Accessed 12 Nov. 2024.
[39] Olive, K A. “The Quark-Hadron Transition in Cosmology and Astrophysics.” *Science (New York, N.Y.)*, vol. 251, no. 4998, Aug. 1991, pp. 1194–9, pubmed.ncbi.nlm.nih.gov/17799279/, https://doi.org/10.1126/science.251.4998.1194. Accessed 12 Nov. 2024.
[40] Steigman, Gary, et al. “Precise Relic WIMP Abundance and Its Impact on Searches for Dark Matter Annihilation.” *Physical Review D*, vol. 86, no. 2, 3 July 2012, https://doi.org/10.1103/physrevd.86.023506. Accessed 12 Nov. 2024.
[41] Feng, Jonathan L. “The WIMP Paradigm: Theme and Variations.” *SciPost Physics Lecture Notes*, 13 June 2023, https://doi.org/10.21468/scipostphyslectnotes.71. Accessed 12 Nov. 2024.
[42] Arakawa, Jason, and Tim. “Is a Miracle-Less WIMP Ruled Out?” *SciPost Physics*, vol. 11, no. 2, 3 Aug. 2021, https://doi.org/10.21468/scipostphys.11.2.019. Accessed 12 Nov. 2024.
[43] “Quest for Dark Matter - Lux Zeplin Experiment - Rau’s IAS.” *Compass by Rau’s IAS*, 30 Sept. 2024, compass.rauias.com/current-affairs/quest-for-dark-matter-lux-zeplin-experiment/. Accessed 12 Nov. 2024.
[44] Cushman, P, et al. *WIMP Dark Matter Direct Detection*. 2013.
[45] Pérez de los Heros, Carlos. “Status, Challenges and Directions in Indirect Dark Matter Searches.” *Symmetry*, vol. 12, no. 10, 8 Oct. 2020, p. 1648, https://doi.org/10.3390/sym12101648.
[46] Balazs, Csaba, et al. “A Primer on Dark Matter.” *ArXiv.org*, 7 Nov. 2024, arxiv.org/abs/2411.05062. Accessed 14 Nov. 2024.
[47] Algeri, Sara, et al. *Statistical Challenges in the Search for Dark Matter*. 2018.
[48] Freese, Katherine. “Status of Dark Matter in the Universe - Katherine Freese.” *Ned.ipac.caltech.edu*, 2017, ned.ipac.caltech.edu/level5/Sept17/Freese/Freese_contents.html. Accessed 12 Nov. 2024.
[49] Arcadi, Giorgio, et al. *The Waning of the WIMP? A Review of Models, Searches, and Constraints*. Vol. 78, no. 3, 10 Mar. 2018, https://doi.org/10.1140/epjc/s10052-018-5662-y. Accessed 12 Nov. 2024.
[50] “DARWIN - Dark Matter WIMP Search with Liquid Xenon.” *Darwin.physik.uzh.ch*, 2 July 2023, darwin.physik.uzh.ch/. Accessed 12 Nov. 2024.
[51] “Sterile Neutrinos | All Things Neutrino.” *All Things Neutrino*, Fermilab, neutrinos.fnal.gov/types/sterile-neutrinos/#evenmore.
[52] Anchordoqui, Luis A, et al. “Fuzzy Dark Matter and the Dark Dimension.” *European Physical Journal. C, Particles and Fields*, vol. 84, no. 3, 15 Mar. 2024, https://doi.org/10.1140/epjc/s10052-024-12622-y. Accessed 12 Nov. 2024.
[53] Fabbrichesi, Marco, et al. *The Physics of the Dark Photon*. SpringerBriefs in Physics, 4 May 2020. Accessed 17 May 2023.
[54] Jackson, Kevin. “Dark Photons Could Explain High-Energy Scattering Data – Physics World.” *Physics World*, 17 Oct. 2023, physicsworld.com/a/dark-photons-could-explain-high-energy-scattering-data/. Accessed 12 Nov. 2024.

Tamil-English Code-Mixed POS Tagging By Mohnish Sivakumar

Abstract

This paper addresses the task of Part-of-Speech (POS) tagging for Tamil-English code-mixed text. This paper compares the performance of three models on monolingual Tamil data—a Hidden Markov Model (HMM), an Averaged Perceptron model, and a Bi-Directional Encoder Representation from Transformers model (BERT). It finds that the Averaged Perceptron model performed the best on the Tamil data, with 77.6% accuracy. This model was then applied to the task of POS tagging Tamil-English Code mixed data. This paper attempts to find an approach that deals with the limited resources for Tamil language processing, the rise of code-mixing in multilingual communities, and the scarcity of Tamil-English code-mixed corpora. This work presents alternate, possibly more viable approaches for POS tagging in Tamil-English code-mixed scenarios and shows the potential that simpler models have in working with scenarios with limited data, both monolingual and multilingual. Additionally, this work proposes a way to create datasets for use in NLP in the absence of high-quality, annotated data.

1. Introduction

POS tagging is crucial for many natural language processing (NLP) processes. It involves receiving input, breaking down the sentences into tokens, and determining the part of speech for each token (verb, noun, adjective, etc.). The results of POS tagging can be used for tasks such as named entity recognition (NER), grammar checking, and machine translation (Anbukkarasi and Varadhanapathy).

Tamil, one of the world's oldest languages, originates from the Dravidian family and has a documented history stretching back over two millennia. Early Tamil inscriptions, found in Tamil-Brahmi script, date to around 300 BCE, making it one of the few classical languages with continuous use (Wikipedia Contributor). Over time, Tamil evolved through three main periods: Old Tamil (300 BCE – 700 CE), during which the Sangam literature was produced; Middle Tamil (700 – 1600 CE), marked by significant religious and literary developments; and Modern Tamil (1600 CE – present), which saw the language's spread among diaspora communities (Krishnamurti). Tamil remains vibrant today as a spoken language and in rich literary traditions maintained across Tamil Nadu, Sri Lanka, and Tamil communities worldwide.

அவர் புதிய திட்டத்தை உருவாக்கும் பணியில் ஈடுபட்டுள்ளான்.
He new project creating to work is involved
He is involved in creating a new project.

Fig 1: This is an example of a sentence in Tamil. This is followed by a word-by-word translation in English and the intended meaning.

A lot of work has been done in the past to POS tag many different languages, especially European languages. However, for most Indian languages, accurate POS taggers and high-quality annotated corpora are still being developed (Abirami et al. 73), so creating accurate POS taggers is a challenge. Increasingly, with the rise of social media, people in multi-lingual communities generate a lot of code-mixed text instead of formal writing in a single language. This is why the selected task is building a POS tagger for Tamil-English code-mixed data. Since there is so little work in this field, this paper aims to improve on early approaches to accomplish the tasks and consider alternate approaches. The inputs of this model are code-mixed sentences taken from a variety of different sources.

2. Related Work

Initially, rule-based approaches were used for the task of POS tagging. They would assign the most likely POS tags to words using a dictionary derived from a tagged corpus, and then use rules created by experts to change the POS tags as needed based on the word's context (Wei et al. 2). However, these models would not perform so well because of the ambiguity and complexity of the language (Wei et al. 2). Later, statistical-based approaches involving models such as HMMs became more common, and quickly overtook the rule-based models. Making the words in a sentence the observations and the POS tags hidden states, the HMM aims to learn the transition and emission probabilities—the probability of one POS tag appearing after another and the probability of a word being observed given a POS tag, respectively—to be able to POS tag efficiently (Wei et al. 2). Statistical models such as conditional random fields were also utilized. More recently, deep learning methods became the most viable for POS tagging. They have been able to find patterns from large amounts of training data and yield high performance, especially when combined with techniques such as character embeddings (Wei et al. 2).

Lots of efforts have been made to annotate different languages and develop syntactic treebanks. The Brown Corpus was one of the first major corpora for computer analysis, with about 1,000,000 words of English text taken from a variety of publications (Francis and Kucera). It has 87 POS tags and was annotated by hand and reviewed over multiple years and been used for many important studies. One of the earliest and most influential syntactic treebanks is the Penn Treebank, featuring around 4.5 million words used for POS tagging sourced from the Wall Street Journal (Marcus et al. 313). Although it was based on the Brown tagset, the Penn treebank has a notably smaller tagset with 36 POS tags (Marcus et al. 316). It was large and well-annotated enough to enable statistical models to be viable (Marcus et al. 313), and was instrumental in the advancement of NLP at the time. There has also been work creating a gold-standard corpus and a model for Tamil-English sentiment analysis, one of the first of its kind (Chakravarthi et al. 202). This corpus consists of social media comments written in 'Tanglish,' which means it is code-switched Tamil and English sentences. They are written in English script. A very significant project in syntactic treebank creation across multiple languages is the Universal Dependencies project. The Universal Dependencies project was created to establish cross-linguistically consistent annotations and to build treebanks for different languages

that supported this (Nivre et al. 1659). Initially, the project had treebanks for 33 languages but has expanded to over 150 different languages (Nivre et al. 1659), with more to come. It combined the Stanford Dependencies project, the Universal Google Dependencies project, the Google universal POS tags, and the CONLL-U format (Nivre et al. 1659). Recently, a project called Aalamaram was created to create a large treebank for Tamil. This is based largely on UD specifications, with improvements to account for Tamil language differences (Abirami et al. 73). However, this could not be accessed to use for this paper.

Code-switching is the alternation between two or more languages or dialects in a conversation. According to a paper about the progress of code-switching research in NLP, for a long time, code-switching was only prevalent in spoken language, making it not relevant to NLP. However, with the recent rise of social media, code-switching has become much more common in writing, and the need to process code-switched language has increased (Winata et al. 2936). The most popular tasks related to code-switching are language identification, sentiment analysis, NER, and POS tagging (Winata et al. 2939). Language identification is important when preprocessing text for other NLP tasks. Sentiment analysis has applications in market research and analysis of public opinion. With code-mixing being prevalent in social media in bilingual and multilingual communities, it is important to be able to develop sentiment analysis for code-mixed text to continue with these applications. NER, or the identification and extraction of information like names, places, organizations, dates, etc. from text, is useful for turning raw text into structured data, and the advent of code-mixing in written text makes NER in code-mixed text important. POS tagging is important for other NLP tasks, and being able to POS tag code-mixed data helps make those other tasks easier.

Recently, there has been a lot of related work about POS tagging for Indian languages. For example, there has been work attempting to create a Tamil POS tagger using deep learning models (Anbukkarasi and Varadhanapathy 1), in which the researchers concluded that out of the models they tried, a Bi-Directional Long-Short Term Memory model was the best, achieving 94% accuracy. For Tamil, in the past, neural network-based approaches and deep learning models have been the most successful, as well as statistical model approaches such as the Hidden Markov Model (HMM) (Anbukkarasi and Varadhanapathy 3). However, there is still a discrepancy for POS-tagged Tamil-English corpora. There have been attempts to POS tag Tamil-English code-mixed texts using a framework involving a CRF, achieving results of 64.83% (Ghosh et al. 95). Tasks such as Tamil-English code mixed POS tagging are difficult because of the scarcity of quality resources (Abirami et al. 75), and the results shown were attributed to that issue (Ghosh et al. 95). Despite this, advancements are still being made. One experiment used a Cross-Lingual Transformation Learning Technique to create a hybrid tagger, which achieved 73.51% accuracy (Senthamizh Selvi and Anitha).

3. Our System

3.1 Models

In this study, three different models were employed to examine their performance in monolingual Part-of-Speech (POS) tagging.

The first model utilized was the Natural Language Toolkit (NLTK) Averaged Perceptron tagger model (Bird et al.). The earliest Perceptron model, the first artificial neural network, was created in 1958 by Professor Frank Rosenblatt (Banoula). It is one of the simplest methods available today, but is still very effective in certain applications (Banoula). The Averaged Perceptron Tagger is one of the most optimized Python taggers available, able to process large amounts of data in fast times, with high accuracy (Honnibal).

Next, a BERT-based model pre-trained for Tamil POS tagging by researchers at L3Cube Pune was implemented to leverage its contextual understanding of language (Joshi 1). BERT is a very powerful model, and a pre-trained model can be fine-tuned to perform almost any NLP task well, even with multiple languages ("BERT 101"). This is one of the reasons this model was chosen for this experiment.

The third approach involved a Hidden Markov Model (HMM) based tagger from the NLTK library, chosen for its application in sequence labeling tasks (Bird et al.). It was one of the best methods for POS tagging before deep learning approaches became viable, so it was chosen here.

4. Experiments and Results

4.1 Data

The monolingual datasets used for English and Tamil were taken from Universal Dependencies (Nivre et al. 1665). Specifically, I used UD English GUM and UD Tamil TTB, as well as TamilTB v0.1 (Ramasamy and Zabokrtsky 1888). These datasets were in the CONLL-U format. The Tamil data was in Tamil script, and the English data was in Latin script. Two code-mixed datasets were used to evaluate the model. One of the datasets consists of parts of the Brown corpus (Francis and Kucera), accessed through NLTK. Only a portion of it is used, containing a little under 12,000 sentences. The second one is the English Web Treebank (EWT), found in Universal Dependencies (Bies et al.). The English corpus was transformed into a code-mixed corpus by randomly replacing nouns, verbs, and adjectives with their Tamil-translated variant. A very small dataset used early in the experiment is also included. The sentiment analysis dataset mentioned in the related work section was manually annotated myself, and consists of twenty POS-tagged code-mixed sentences taken from social media (Chakravarthi et al. 202). It is in English script, but the sentences have Tamil and English words. Sometimes, the sentences are structured like Tamil sentences but have some words in English, and sometimes it is the opposite (Chakravarthi et al. 202). Additionally, there are no labeled POS tags, so a small sample was annotated. The UPOS tagset (Nivre et al. 1661) was used to tag the data. An example of a sentence with pos tags can be found in the figure below. However, the data is in the form of a tab-separated-values (TSV) file.

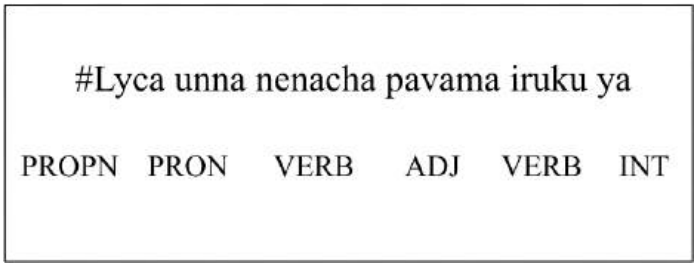


Fig 2: This is an example of a code-mixed tagged sentence from the dataset.

4.1.1 Preprocessing

For the CONLL-U formatted datasets, the Conllu library was used to parse the data. For the Perceptron tagger, the data was transformed into a list of sentences, each sentence being a list of (word, POStag) tuples.

For the BERT model, the AutoTokenizer from the Hugging Face transformers library was used to tokenize the sentences. This tokenizer was specific to the pre-trained BERT model being used. POS tags were mapped to numerical IDs, and vice-versa, to facilitate processing by the BERT model. Sentences were padded to ensure a consistent sentence length.

For the HMM, sentences were split into individual tokens using whitespace as the delimiter. A vocabulary was created where each unique word in the dataset was mapped to a numerical index. This was essential for the HMM model, which works with numerical sequences. The sentences were converted into sequences of numerical indices based on the vocabulary. The pre-processing steps were specific to the models being used.

4.2 Experiment Setup

The three models were tested on a Tamil monolingual dataset to compare their POS tagging capabilities. After evaluating the results, the model that showed the best performance in the monolingual task was selected. This top-performing model was then applied to code-mixed POS tagging to explore its effectiveness in handling multilingual text within the same context.

4.2.1 Training

The HMM model used the preprocessed data in the form of numerical sequences. The model was then initialized with a fixed number of hidden states corresponding to the POS tags. A trainer method was used to train the model. This involved estimating the transition probabilities between states and the emission probabilities of observations (words) given a state. Then, Maximum Likelihood Estimation (MLE) is used to estimate these probabilities based on the labeled data. In addition, a smoothing technique (Lidstone probability distribution) is used with a smoothing parameter to handle unseen words or transitions in the training data by assigning small probabilities to them, preventing zero probabilities (Jurafsky and Martin).

The Perceptron model used the CONLL-U formatted data as tuples. It was initialized with random weights for its features, which were extracted from the words and their context. An

inbuilt training method was used to train the model, which iterated through the training data multiple times (epochs). In each epoch, the model adjusted its weights based on the errors made. In the end, the average weights were used to smooth out the effect of large updates for the purpose of better performance on unseen data (Honnibal).

The BERT model used the CONLL-U formatted data after it had been tokenized, and padded, with tags mapped to IDs. It started with a pre-trained model. Then, a trainer method was used to fine-tune the model on the task of POS tagging with the given dataset. This involved adjusting the model's weights using backpropagation and gradient descent to minimize the difference between predicted and actual POS tags ("BERT 101"). The training process used the Hugging Face Trainer class, which handled optimization and evaluation ("Trainer").

4.2.2 Evaluation

For calculating the accuracy of the tests, the sentences tagged by the models were compared against the gold-standard sentences. The accuracy returned was the percentage of correct tags over the total. This allowed a consistent accuracy calculation across all of the models, so the comparisons could be on even grounds.

4.3 Results

Table 1: Model Accuracy

Model	Accuracy
NLTK HMM	62.1%
NLTK Perceptron	77.6%
L3Cube Tamil BERT	35.6%

Table 1 depicts the performances of the models on the datasets given. Even though BERT is one of the most powerful models, models that complex need a lot of data to train, and even to fine-tune as was done here ("BERT 101"). The simpler HMM and Perceptron models did far better, achieving results around double BERT's. This is likely because they are relatively simpler, and did not require as much data to train. The Perceptron model, being the simplest of the three, did the best, with over 77% accuracy. Since Perceptron was the best, it was the model that was used to attempt code-mixed POS tagging.

4.3.1 Experiments for Code-Mixed Data

The Perceptron tagger was then tested on the code-mixed datasets in varying ways, to try and achieve higher accuracies. A new Perceptron tagger was initialized and trained in the Tamil monolingual data. The code-mixed sample of 20 sentences was transliterated and combined with the Tamil data for retraining, and then was tested purely on the code-mixed segment. A different, new instance of the model was trained on the dataset created from the EWT. Then, a new model

was used to train on the dataset created from the Brown Corpus. Finally, one model was used to train on all of these datasets sequentially. Despite concern that the Brown Corpus's usage of different POS tags would skew the data, the results did not seem concerning. The accuracies based on these datasets can be found in the table below.

Table 2: Code-Mixed Accuracies

Dataset	Accuracy
Tamil+Sample	77.4%
Sample	72.4%
EWT	78.1%
Brown	93.9%
Tamil & Sample & EWT	93.7%
Tamil & Sample & EWT & Brown	94.0%

4.4 Analysis

The model performed well, with accuracies similar to or higher than previous models. This is likely due to the simplicity and robust nature of the Perceptron model, as it was able to learn the patterns and achieve good accuracy without much data to work with. Then, when it was introduced to large amounts of data through the Brown Corpus, it was able to achieve higher accuracies. The model adapted to the Brown POS tags well, with accuracy increasing. It is possible that the sheer volume of data used from the Brown Corpus allowed for the tagger to switch tagging systems. Additionally, there might have been some transfer from the UPOS tagset to the Brown tagset, since both sets have tags for the same parts of speech. This may have led to it simply being a matter of changing the tag, but the tagging process would have been similar. The model proved its capability to process different types of texts, since the different datasets were taken from different contexts. The sentences in EWT were taken from the internet roughly a decade ago (Bies et al.), but the Brown Corpus was created in the 1960s (Francis and Kucera). The code-mixed datasets, however, were created somewhat crudely. The method of creating the datasets by replacing words of parts of speech that were commonly code-switched with the equivalent words in a different language has some merit and can be used for early work in low-resource languages, but it is not the same as a fully annotated and review, authentic code-mixed dataset. Additionally, this is difficult to do with more complex languages. The agglutinative nature of Tamil may not have been fully captured in this experiment and will have to be expanded on in the future. However, until sufficient datasets are introduced, methods such as this one can be considered.

Conclusion

The Averaged Perceptron Tagger emerged as the most effective model for POS tagging of Tamil language data, outperforming both the HMM and BERT-based models. Its simplicity and ability to perform well with limited data are likely the causes of this success. When applied to Tamil-English When applied to Tamil-English code-mixed data, the model showed promising results. However, it is important to note that the results achieved on the code-mixed data may not be fully representative, due to the crude nature of the dataset. Future work should focus on creating an annotated code-mixed dataset to enable experiments and training on representative, authentic data. This would allow for a more thorough and representative evaluation of the model's performance on Tamil-English code-mixed text as well as the testing of new approaches in the future. Overall, this study takes a step forward for POS tagging in Tamil-English code-mixed scenarios and highlights the potential of the Averaged Perceptron Tagger in handling multilingual contexts with limited resources. Approaches like this could be used for other languages with limited resources, and could greatly benefit NLP tasks for those languages.

Works Cited

Abirami, A. M., et al. ‘Aalamaram: A Large-Scale Linguistically Annotated Treebank for the Tamil Language’. *WILDRE*, 2024, api.semanticscholar.org/CorpusID:269950692.

Anbukkarasi, S., and S. Varadhaganapathy. ‘Deep Learning Based Tamil Parts of Speech (POS) Tagger’. *Bulletin of the Polish Academy of Sciences Technical Sciences*, 01 2024, pp. 138820–138820, doi.org/10.24425/bpasts.2021.138820.

Banoula, Mayank. ‘What Is Perceptron? A Beginners Guide for Perceptron’. *Simplilearn. Com*, May 2021, www.simplilearn.com/tutorials/deep-learning-tutorial/perceptron.

“BERT 101 - State Of The Art NLP Model Explained”. *Hugging Face*, huggingface.co/blog/bert-101.

Bies, Ann, et al. *English Web Treebank*. Linguistic Data Consortium, Aug. 2012, doi.org/10.35111/M5B6-4M82.

Bird, Steven, et al. *Natural Language Processing With Python*. O’Reilly Media, Inc., 2009, www.oreilly.com/library/view/natural-language-processing/9780596803346/.

Chakravarthi, Bharathi Raja, et al. ‘Corpus Creation for Sentiment Analysis in Code-Mixed Tamil-English Text’. *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-Resourced Languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*, edited by Dorothee Beermann et al., European Language Resources association, 2020, pp. 202–210, aclanthology.org/2020.sltu-1.28.

Francis, W. N., and H. Kucera. *Brown Corpus Manual*. Department of Linguistics, Brown University, Providence, Rhode Island, US, 1979, http://gandalf.aksis.uib.no/icame/brown/bcm.html.

Ghosh, Souvick, et al. ‘Part-of-Speech Tagging of Code-Mixed Social Media Text’. *Proceedings of the Second Workshop on Computational Approaches to Code Switching*, edited by Mona Diab et al., Association for Computational Linguistics, 2016, pp. 90–97, doi.org/10.18653/v1/W16-5811.

Honnibal, Matthew. ‘A Good Part-of-Speech Tagger in about 200 Lines of Python · Explosion’. *Explosion*, Sept. 2013, explosion.ai/blog/explosion.ai.

Joshi, Raviraj. ‘L3Cube-HindBERT and DevBERT: Pre-Trained BERT Transformer Models for Devanagari Based Hindi and Marathi Languages’. *arXiv [Cs.CL]*, 2023, arxiv.org/abs/2211.11418.

Jurafsky, Daniel, and James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*. 3rd ed., Prentice Hall, 2024.

Krishnamurti, Bhadriraju. *The Dravidian Languages*. Cambridge University Press, 2003. Cambridge Language Surveys.

Marcus, Mitchell P., et al. ‘Building a Large Annotated Corpus of English: The Penn Treebank’. *Comput. Linguist.*, vol. 19, no. 2, MIT Press, June 1993, pp. 313–330.

Nivre, Joakim, et al. ‘Universal Dependencies v1: A Multilingual Treebank Collection’. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, edited by Nicoletta Calzolari et al., European Language Resources Association (ELRA), 2016, pp. 1659–1666, aclanthology.org/L16-1262.

Ramasamy, Loganathan, and Zdeněk Žabokrtský. ‘Prague Dependency Style Treebank for Tamil’. *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12)*, edited by Nicoletta Calzolari et al., European Language Resources Association (ELRA), 2012, pp. 1888–1894, http://www.lrec-conf.org/proceedings/lrec2012/pdf/456_Paper.pdf.

Senthamizh Selvi, S., and R. Anitha. ‘Bilingual Corpus-Based Hybrid POS Tagger for Low Resource Tamil Language: A Statistical Approach’. *Journal of Intelligent & Fuzzy Systems*, vol. 43, no. 6, Jan. 2022, pp. 8329–8348, doi.org/10.3233/JIFS-221278.

“Trainer”. *HuggingFace*, huggingface.co/docs/transformers/en/main_classes/trainer.

Wei, Chengwei, et al. ‘GWPT: A Green Word-Embedding-Based POS Tagger’. *arXiv [Cs.CL]*, 2024, arxiv.org/abs/2401.07475.

Wikipedia contributors. *Old Tamil --- Wikipedia, The Free Encyclopedia*. 2024, en.wikipedia.org/w/index.php?title=Old_Tamil&oldid=1261393932.

Winata, Genta, et al. ‘The Decades Progress on Code-Switching Research in NLP: A Systematic Survey on Trends and Challenges’. *Findings of the Association for Computational Linguistics: ACL 2023*, edited by Anna Rogers et al., Association for Computational Linguistics, 2023, pp. 2936–2978, doi.org/10.18653/v1/2023.findings-acl.185.

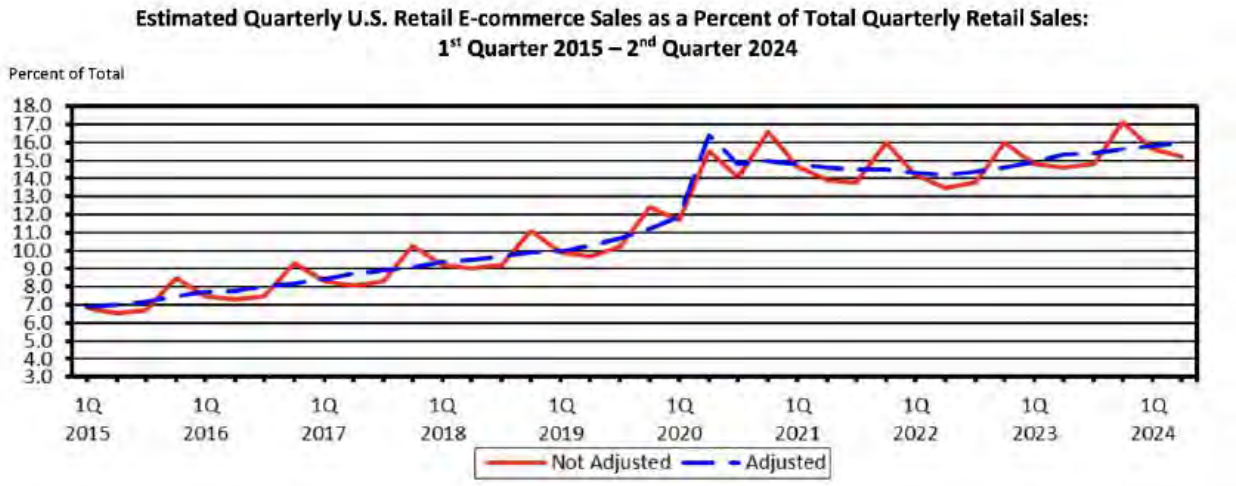
How Can Policymakers Guide Online Retailers to Nudge Consumption Towards Sustainable Products? By Eda Baser

Abstract

As a new normal for organizations in the context of social responsibility, the impact of encouraging consumers towards environmentally friendly initiatives cannot be overemphasized. Behavioral sciences have been used to impact consumer decision-making towards more sustainable choices in various domains of everyday life like transportation, utility consumption, and recycling. Therefore, with their wide exposure and influence, online retail stores also hold this potential commanding position in guiding consumers toward making environmentally conscious decisions. As there are numerous examples of successful environmental policies imposed on companies, industry-wide enforcement of online retailers could also be impactful in leading consumers to sustainable choices through standardized nudging applications. To convince consumers, in particular, to embrace sustainability, it is necessary first to understand the many biases and heuristics in decision-making. In this paper, we examine how anchoring, status-quo, and herding biases can be used to help policymakers guide online retailers to drive consumption towards sustainable products.

1. Introduction

As a new normal for organizations in the context of social responsibility, the impact of encouraging consumers towards environmentally friendly initiatives cannot be overemphasized. Online retail stores, with their extensive reach and impact, have an advantage in directing consumers on how to make more environmentally friendly decisions. As more and more of our societies' purchases shift towards online channels, certain key proven nudges could be enforced to help drive consumers attention and choice towards sustainable products in online purchases.



The Department of Commerce's Census Bureau stated that the estimate of U.S. retail e-commerce sales for the second quarter of 2024, on an unadjusted basis, was $282.3 billion, up 5.3 percent (± 0.7) from the first quarter of 2023. In the second quarter of 2024, e-commerce sales made up 16.0 percent of total retail sales. Over the last decade, the share of e-commerce sales has been steadily increasing from 7.0% in the first quarter of 2015 to 16% in the second quarter of 2024 (as shown in the above table). With online resources being used more day by day, persuading the consumer in an online environment, especially towards sustainability, involves understanding and appreciating the biases and heuristics that exist in decision-making.

Behavioral sciences have been successfully used to impact consumer decision-making towards more sustainable choices in various domains of everyday life like utility consumption, energy-efficient appliance choices, and transportation. As most of these are applied through policies enforced on companies and further policies like Extended Producer Responsibility (EPR), Plastic Bag bills and the like have been instrumental in driving change in this field, industry-wide enforcements on online retailers could also play an enormous role in affecting sustainable consumer choice through standardized nudging applications. Understanding how these cognitive biases may affect consumer behavior may be of interest to policymakers and online retailers who want to promote sustainable purchasing.

This paper explores how anchoring, status quo bias, and herding can be applied to promote sustainable consumption in online retail environments. The first section explains the - concept of anchoring, going further into how the initial presentation of information can significantly influence consumer decisions. The next section addresses status quo bias, discussing how making sustainable products the default option can shift consumer behavior. Finally, it examines herding, illustrating how highlighting the popularity of sustainable products can create a social norm that encourages more consumers to follow. By the end of this exploration, the potential for policymakers to harness these cognitive biases to foster a more sustainable future will be made clear.

2. Heuristics

In this section, a number of research papers will be presented that show how heuristics have been applied successfully to solve sustainability-related problems. The examples provided show several ways in which simplified decision-making procedures led to more sustainable consumption choices. We illustrate how heuristics promote environmental initiatives and encourage sustainable behavior at the human and organizational levels.

2.1 Anchoring and Framing

The anchoring phenomenon demonstrates how initial information influences decision-making processes, providing a starting point from which people make judgments and estimates. “In many situations, people make estimates by starting from an initial value that is adjusted to yield the final answer. The initial value, or starting point, may be suggested by the formulation of the problem, or it may be the result of a partial computation. In either case,

adjustments are typically insufficient. That is, different starting points yield different estimates, which are biased toward the initial values. We call this phenomenon anchoring." (Tversky and Kahneman, Science, 1974.)

A study by Heinzle and Wüstenhagen (2012) investigates how design variations in energy efficiency labels influence consumer choice, focusing on the product category of TV sets. When energy labels changed their traditional scale model ranging from A to G to an extended model ranging from A+++ to D, the way consumers perceived energy efficiency changed. It became obvious that all categories labeled A started to be perceived by consumers as similar, which made them value improvements in energy efficiency less than before.

The experiment carried out with Danish consumers, exposed them to different TV options using these two labeling systems. Participants were asked to make personal choices among virtual TVs with several attributes like brand, price, and energy efficiency. It is expected that, when the new label-which ranges from A+++ to D is used, the consumer would choose with less priority given to energy efficiency. The findings suggest that the redesign of the label reduced the perceived importance of the energy efficiency attribute and resulted in a less green choice. This indicates that minor changes to labeling can have great impacts on consumer responses. In other words, research underlines that label design is the key point in the process of bringing consumers closer to energy-efficient behavior.

"The European energy labeling is used to illustrate the principle of anchoring (Tversky and Kahneman 1974), how it systematically biases consumer perceptions of messages, including eco-labels, and its implications for environmentally relevant consumer choices." This is likely to produce an anchoring effect that can reduce the effectiveness of the label. Returning to the example above, consumer perceptions and evaluations of the new labels may still be anchored to the old energy label categories and ratings that they had grown accustomed to. This is to say that they may have a very poor comprehension or appreciation of the change, thus leading them to confusion or misinterpretation of the energy efficiency of appliances. For the reason that people tend to rely on their first impression or initial reference point, a new label does create a potential consequence of inadvertently causing consumers to discount the importance of the label or make mistakes in judging the energy efficiency ratings of products. The broader understanding of "choice architecture" in this case is that framing information can be profoundly influential in making decisions. Knowledge of anchoring therefore helps policy makers come up with labels that are less confusing and promote better choices from consumers. In other words, anchoring is essential in this experiment, which evidences the biases in cognition impeding the efficiency of environmental policies. One should be very careful while structuring information and presenting it to the public.

2.2 Status-quo

The status quo bias, often manifested as people's tendency to choose default options, is a key heuristic in designing effective nudges to influence decision-making. "The status quo bias and the related "path of least resistance"—which often leads people to select a default option that

will apply if the chooser does nothing—is one of the most utilized heuristics in the choice architecture associated with nudging." (Folke Ölander, John Thøgersen (2014)). People's preference to maintain things as they are is known as the status quo. By understanding the status quo, policymakers can identify gaps in retail practices and certain consumer behaviors.

This anchoring and default setting is further manifested in a study involving Danish homeowners. In this case, the participants of the study were requested to imagine that their electricity company asked for permission to install, at no cost, a remote-controlled Smart Meter in their homes. A Smart Meter is a digital device that measures and records real-time gas, electricity, or water consumption. It automatically sends readings to your energy supplier and provides information about your energy usage and costs. In one condition, there was an opt-in option: "Tick the box below if you accept installation. If not, please continue to the next question." The prompt in the opt-out condition read: "Tick the box below if you do not accept installation. If you accept installation, please continue to the next question." The results of this survey indicated that the wording of the question greatly influenced the number of people who opted for the installation of Smart Grid technology at home. Indeed, the below Figure 2 in the presented paper shows that the highest Smart Grid technology installation acceptance rate was when the framing was made as an opt-out framing.

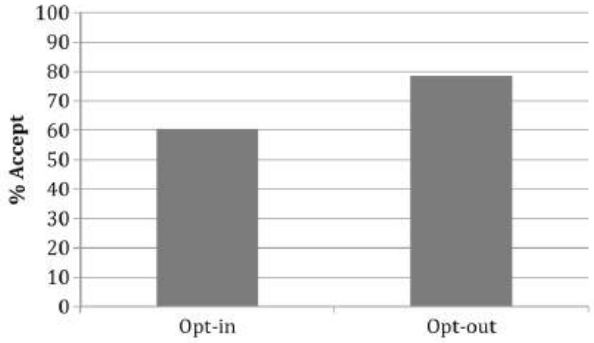


Fig. 2 Percentage accepting to have Smart Grid technology installed under two framing conditions, opt-in (N= 345) and opt-out (N=332)

This is to say that the way one frames a question with the default option available might make a difference in the result. Generally speaking, too many people procrastinate on decisions, and hence an opt-in framing - wherein inaction yields a decision not to participate - can generate results that will not serve society or the individual with the best long-term interests. An opt-out framing is, nevertheless, not always applicable in every situation. The best takeaway from this research is not that opt-out framing should always be employed over opt-in framing but, rather, it is that the framing of choices and the defaults those present can have enormous effects on outcomes. Therefore, campaigners should think hard about which framing to use to achieve the desired outcome.

2.3 Herding

Social norms, whether consciously or unconsciously followed, are powerful influencers of behavior that can be effectively leveraged as nudges to guide people's choices in predictable

ways. “Social norms which influence behavior due to a general tendency to “follow the herd” are often included in the nudges inventory. Undoubtedly, choosing to do what others do is sometimes the result of a conscious process, but in other cases, the social norm effect can be obtained even when no conscious thought is associated with it.” (Folke Ölander, John Thøgersen (2014)) Herding heuristic can be explained within multiple instances where when consumers observe that others are choosing certain products, they are likely to follow suit, assuming that the majority's choice reflects the best or most popular option. For instance, if online retailers highlight that a particular eco-friendly product is frequently purchased or highly rated by other customers, it can create a social norm that encourages more people to choose it, even if they might not have initially considered it.

“The human inclination to follow the crowd" can lead to environmental challenges, but it also offers potential for "social nudging", Cialdini et al.'s (1990) focus theory of normative conduct has inspired significant research showing that "herding behavior," or following "descriptive norms," often results in environmentally harmful actions like littering or collecting petrified wood from national parks (Cialdini et al. 2006). However, this body of research has also revealed that simple social nudges can be highly effective in curbing harmful behaviors and encouraging environmentally positive actions, such as reusing hotel towels and conserving household energy (Goldstein et al. 2008; Schultz et al. 2008).

In this experiment, the researchers aimed to test two hypotheses regarding consumer behavior toward ecological products. First, they hypothesized that participants exposed to a strong descriptive norm would purchase and spend more on ecological products compared to those in a control group. Second, strong norms were expected to be more effective than weak norms in promoting green consumption. 122 students from the University of Toulouse-II participated in this experiment. These students were approached on campus and later participated in a lab setting. On average, they spent 24 euros (with a standard deviation of 1 euro) in a simulated online store called the "Greenshop." For the experiment, participants were placed in front of a computer where the "Greenshop" program was set up. They were then randomly assigned to one of four different groups: Control Condition: Participants in this group were only provided with basic information about the store, specifically that "This shop sells several daily usage products." Weak Norm Condition where participants received the same basic information as the control group, but were additionally informed that only 9% of previous participants had purchased an ecological product, suggesting a low rate of green shopping behavior. Strong Norm 1 Condition was given the control group information, along with an additional statement indicating that 70% of previous participants had purchased at least one ecological product, implying a high rate of green shopping behavior. This setup was designed to observe how different social norms, presented as information about previous participants' behavior, would influence the participants' purchasing decisions in the Greenshop.

Table 1
Descriptive statistics from Experiment 1.

Percentage of participants who bought at least one eco-product	Means (*SD*) of purchased eco-products	Percentage (*SD*) spent on eco-products
58.6	2 (2.45)	24.41 (28.29)
87.1	2.84 (2.38)	36.40 (26.55)
83.3	2.77 (2.31)	33.47 (25.27)
96.9	3.25 (2.24)	43.83 (28)

As shown above in Table 1 of the presented paper, the first hypothesis was supported; participants who were presented with a norm added, one more ecological product to their purchases and spent approximately 13% more on these products compared to the control group. The paper's analysis focused on the number of ecological (eco-labeled) products purchased and the percentage of the budget allocated to them. This analysis revealed a significant impact of norms on both the quantity of ecological products purchased and the percentage of the budget spent on these products.

Further pairwise comparisons highlighted significant differences between the control condition and all three norm conditions (weak and two strong norms) for both measures, except for a marginally significant difference in the number of products purchased between the control and one strong norm condition. The other comparisons did not reach significance. Interestingly, contrary to their second hypothesis, strong norms did not significantly outperform weak norms in encouraging greater purchases or expenditures on eco-products.

This experiment is the first to highlight how descriptive norms can effectively drive the purchase of genuine ecological products. This finding is notable, especially given that real, information-based norms were used regarding an uncommon behavior. However, contrary to expectations, there was no significant advantage of strong norms over weak norms. One possibility for this explanation is that merely activating any norm, regardless of its strength, can make the norm more prominent and influence consumer decisions, particularly in a student demographic known for its pro-environmental attitudes. However overall, it was stated that "The results of the first experiment indicate that descriptive norms can indeed influence shopping for "green" products in a realistic online shopping environment." (Christophe Demarque a, *, Laetitia Charalambides b , Denis J. Hilton a , Laurent Waroquier, 2015)

3. Relation to Topic

In addressing the research question, "How can policymakers leverage behavioral economics principles at online retailers to promote sustainable product consumption?", three key behavioral concepts have been examined: anchoring, status quo bias, and social herding. Policymakers aiming to foster sustainable consumption through e-commerce platforms could utilize these concepts to design effective policies. Specifically, understanding how these biases and heuristics operate in online retail environments is crucial for developing strategies that nudge consumers towards more sustainable choices in these platforms.

3.1 Potential Anchoring Applications

Anchoring is a bias related to people relying too heavily on an initial piece of information (the "anchor") when trying to make a decision. In an online retail context, this bias may significantly lead to more sustainable choices being made. Anchoring could be exploited by using the environmental benefit of sustainable products in addition to their price, as the first piece of information that is viewed or read by the consumer. This might be, for example, the long-term cost savings associated with using a sustainable product; for example, an energy-efficient appliance may show an estimated dollar savings during its lifetime as directly derived from the reduced energy consumption. That would become an anchor that changes their perception and raises the desirability of the sustainable product. Durable goods, similarly, could be sold by informing people that they would last longer and therefore decrease how often they would need to be replaced, saving money in the long run for the consumer. By anchoring the price comparison on these long-term financial benefits rather than just the initial purchase cost, the consumer may begin to see sustainable products as not only environmentally responsible but also economically advantageous.

Another use of anchoring would be to point out the negative impacts of a product that is not sustainable; by doing so, it would make the sustainable alternative seem much better in comparison. Policymakers could also introduce an extreme policy that would force online retailers to flash environmental impact labels as the first information shown to consumers upon viewing a product. These would include full details of the carbon footprint for the product, water usage, energy efficiency, recyclability, and total environmental impact throughout its life cycle. By making this the first point of reference, the policy effectively anchors the consumer's perception of the product on its sustainability credentials. This could lead to anchoring, in which green products would appear as a much more responsible and better choice directly next to their non-sustainable alternatives. The strategic use of the anchoring bias may make this a very influential policy in consumer decision-making for the wide diffusion of sustainable products. The powerful anchor would point out not only the benefits of sustainable consumption but also would shift consumer perceptions in favor of environmentally responsible choices, created by mandatory environmental impact labels, emphasis on long-term cost savings, and the use of comparison tools.

3.2 Potential Status Quo Applications

Status quo bias maintains that the status should remain unchanged or that choices should be made to keep things running as they are. This leads to an important change in consumer behavior towards sustainable products because many consumers are anti-change. The reason could be a comfort or simply habit that consumers are resistant to shift to other sustainable brands/products even after being sensitized about those products. Policymakers might want to counteract the status quo bias by guiding online retailers to apply the default option for sustainable products. One pivotal policy recommendation will be to have online retailers make sustainable product default (first-listed or recommended) options in relevant product categories.

For instance, whenever consumers are searching or browsing for any particular product category, such as household cleaning supplies, personal care items, and/or clothing, the first products shown should be sustainable alternatives (e.g. re-fill options).

The policy could even go further by requiring that shopping cart default selections or pre-filled recommendations also be sustainable products. For example, if a consumer puts an unsustainable product into their cart, a website could automatically suggest a sustainable alternative and indicate why it's a better choice given the environmental impact or long-term value.

This "default" positioning capitalizes on status quo bias by making the sustainable choice the path of least resistance, therefore lowering the level of effort consumers must exert in making environmentally friendly choices. Such a shift could also be facilitated by policymakers' support through adopting an "opt-out" rather than "opt-in" principle. Under this mechanism, consumers would automatically be pre-selected for sustainable products; they would have to opt-out if they preferred other choices. This would make it easier for consumers to stick with the sustainable default, rather than taking the extra steps needed to select a less environmentally friendly option. It might work quite well, particularly in subscription-based services, whereby the default reorder option is set to sustainable products unless actively changed by the consumer.

Another policy that could be used to help get people out of the status quo bias would be to incentivize consumers to try out sustainable products. Policymakers can engage in loyalty programs online that, together with retailers, reward customers through special loyalty rewards, especially for buying sustainable products. Rewards can come in the form of accrued points with every purchase, which can, in turn, be redeemed as discounts on any future purchases of sustainable products or even other eco-friendly services. The incentive structure might award a consumer who consistently purchases sustainable products with points redeemable against discounts on renewable energy subscriptions, public transport passes, or donations to environmental causes. Such positive reinforcement may trigger more frequent repurchases of sustainable products but also contribute to a long-term shift in consumers' habits by making sustainable choices part of the regular pattern of their shopping behavior.

Financial incentives, on the part of policymakers, besides loyalty programs would encompass discounts and tax credits for consumers making sustainable purchases. There could be the possibility of a policy that would guarantee that retailers give discounts to consumers on sustainable products, provided by the subsidy of government programs specifically for environmental sustainability. Otherwise, one can have consumers get tax rebates or credits that depend on the proportion of one's annual expenditure on sustainable products. This could be even more effective for higher-cost categories, such as electronics or major home appliances, where the comparative financial burden of a non-sustainable option might otherwise be very high.

These complex and multi-dimensional policies, when implemented, would allow the proper application of status quo bias by policymakers and make sustainable consumption the default choice of consumers. A supportive policy that doesn't just minimize friction about the

adoption of sustainable products but builds an ecosystem to incentivize consumers toward choices benefiting the well-being of the planet would be impactful. The more people shop with a conscience, the faster the societal norms shift.

3.3 Potential Herding Applications

Herding refers to the fact that individuals often choose based on what the crowd is doing or someone else's opinion, rather than making an independent choice. Herding, in online retail, will be driving consumers to popular products or those with many positive reviews, not necessarily those most sustainable. Herding can be used to the advantage of policymakers by cooperating with online retailers to make sustainable products appear popular. For example, labeling products with tags like "Best Eco-Seller" or "Most Popular Eco-Friendly Products Today" creates a vision that these items are widely accepted and preferred by all others. Herding will help influence more people toward sustainable options by influencing consumer perceptions of what others buy.

To harness the power of herding, policymakers might work with online retailers to create and display badges or labels that identify the relative popularity of various sustainable products, such as "Best-Selling Eco-Friendly Choice." Prominent uses of social proof like this will get people to follow the crowd toward more sustainable purchases. Social media sharing features that allow consumers to easily show others their sustainable purchases also normalize eco-friendly consumption and encourage others to do the same. Beyond product labeling, policymakers could push online retailers to develop dynamic features showing in real-time the popularity of sustainable products. For instance, websites could show in real time the number of consumers who have bought a sustainable product within the last hour, or how many users are currently viewing it. Real-time data would amplify this herding effect even more since it could make the consumer perceive there is yet a greater movement toward sustainability.

This awareness of being in collective action sends a powerful message that will make them more willing to follow through with sustainable products to keep in tune with what everybody else is doing. More importantly, this suggests a perhaps policy interest in the integration of social media sharing into online retail platforms: it should be seamless for consumers to share their sustainable purchases with their networks. In facilitating this visibility, we normalize and socially reward sustainable consumption. For example, at the very act of buying a sustainable product, consumers can be requested to post a message on social media with a pre-prepared text: "Just made an eco-friendly choice! Join me in supporting sustainable products." Such messages could also include hashtags or links to connect the posts to other larger-scale campaigns for sustainability; which would further extend and intensify the effects of individual consumer choices.

4. Conclusion

As the share of online consumer sales grows rapidly over the years and as retailers increasingly embrace both sustainable practices as well as behavioral economics principles, the

field of online shopping poses great opportunities to increase consumption of environmentally friendly options. The focus of this paper is to bring forth cognitive biases, namely anchoring, status quo bias, and herding, as tools that policymakers and online retailers can use to enhance sustainable consumption. If these biases are addressed strategically, businesses can then lead consumer choices toward eco-friendly products with ease.

The anchoring bias shows how the first piece of information provided acts as a strong determinant of the shaping of consumer perceptions and choices. Policymakers should leverage this by making sustainable products prominent along with beneficial comparisons and environmental impact that guide consumers toward greener choices. Offering sustainable products as a default, further helps an organization reduce the discomfort associated with change and makes environmentally responsible habits easier to acquire for consumers. Besides, the herding effect used to underline the popularity of eco-friendly merchandise could also contribute to the establishment of social norms to be in line with this trend. This highlighting of the best-selling sustainable products and promoting features for social sharing creates a social norm for one to buy in an environmentally friendly way.

Taken together, these approaches could open up ways to a more sustainable future, showing that lessons learned from behavioral science are vital in setting trends for improving environmental stewardship. Like in other environmental policies, policymakers' guidance on utilizing these practices would enable these biases to be harnessed in the years ahead and bring the desirable changes in sustainable consumption.

Works Cited

US Department of Commerce. Ec Current. 2016.

Heinzle, Stefanie Lena, and Rolf Wüstenhagen. "Dynamic Adjustment of Eco-Labeling Schemes and Consumer Choice - the Revision of the EU Energy Label as a Missed Opportunity?" Business Strategy and the Environment, vol. 21, no. 1, 10 July 2011, pp. 60–70, https://doi.org/10.1002/bse.722. Accessed 2 Nov. 2019.

Tversky, Amos, and Daniel Kahneman. "Judgment under Uncertainty: Heuristics and Biases." Science, vol. 185, no. 4157, 27 Sept. 1974, pp. 1124–1131, www2.psych.ubc.ca/~schaller/Psyc590Readings/TverskyKahneman1974.pdf.

Ölander, Folke, and John Thøgersen. "Informing versus Nudging in Environmental Policy." *Journal of Consumer Policy*, vol. 37, no. 3, 8 Mar. 2014, pp. 341–356, https://doi.org/10.1007/s10603-014-9256-2.

Cialdini, Robert B., et al. "A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places." Journal of Personality and Social Psychology, vol. 58, no. 6, 1990, pp. 1015–1026, https://doi.org/10.1037/0022-3514.58.6.1015.

Cialdini, Robert B., et al. "Managing Social Norms for Persuasive Impact." Social Influence, vol. 1, no. 1, Mar. 2006, pp. 3–15, www.tandfonline.com/doi/full/10.1080/15534510500181459, https://doi.org/10.1080/15534510500181459.

Schultz, P. W., Khazian, A. M., & Zaleski, A. C. (2008). Using normative social influence to promote conservation among hotel guests. Social Influence, 3(1), 4–23.
https://doi.org/10.1080/15534510701755614.

Göckeritz, Susanne, et al. "Descriptive Normative Beliefs and Conservation Behavior: The Moderating Roles of Personal Involvement and Injunctive Normative Beliefs." European Journal of Social Psychology, 2009, p. n/a-n/a, https://doi.org/10.1002/ejsp.643.

Demarque, Christophe, et al. "Nudging Sustainable Consumption: The Use of Descriptive Norms to Promote a Minority Behavior in a Realistic Online Shopping Environment." *Journal of Environmental Psychology*, vol. 43, Sept. 2015, pp. 166–174, www.sciencedirect.com/science/article/abs/pii/S0272494415300219?via%3Dihub, https://doi.org/10.1016/j.jenvp.2015.06.008.

Using AI Models to Help Banks Save Over 170 Billion Dollars By Prediction Loan Defaults
By Prashanth Venkatesh

1. Abstract

Commercial banks all around the world are grappling with the problem of loan defaults that occur each year and the write off they need to do each year: According to Statista.com, “In the first quarter of 2024, roughly 2.68 percent of all consumer loans at commercial banks in the United States were delinquent”(1). In this paper we have developed a machine learning model that helps banks predict if a customer will default on his loan. We estimate that this can save banks $178.43 billion each year. We hypothesized that if the AI model receives some input parameters about the person who seeks a loan, it will be able to predict whether the said person will default on his loan. From this model we managed to outperform random guessing and choice: Our weights algorithm that combined the results of all models (Logistic Regression, Decision Tree, Random Forest, Multi-Layer perceptron, and K-Nearest Neighbor) was able to predict with an incredibly high accuracy of about 70%. The accuracy of random guessing is about 50% and our accuracy of 70% shows that the algorithm outperforms it by a huge margin. This indicates that our model is much more accurate than simple guessing. This was developed to help banks minimize the number of defaults that occur on loans handed out by them by giving them more information to base their decisions on. This research will help provide banks with the required tools to decide whether they should hand out a loan or not.

2. Introduction

2.1 References

The inspiration for this project was from several key papers published by other individuals:

- Tech Giants in Banking: The implications of a new market power, by Nicola Bilotta and Simone Romano
    - It discusses the influence of technology giants in the banking industry underscoring the competitive advantage but also mentioning the disruption to traditional banking models and the regulatory challenges faced due to their entry into financial services.
- The Expansion of Big Data Companies in the Financial Services Industry, and EU Regulation by Martina Scopsi.
    - This paper discusses the impact of the entrance of big data companies into the Financial Services Industry and highlights the importance of assessing their impact created in the EU as a sample market.

These helped us follow the right track and come up with a real life problem to solve: Loan Defaults. By using some simple features like Gender, Marital Status, Living Area,

Employment status, and Education Level, we can accurately predict whether a person will default on his loan or not.

2.2 AI in Finance

Artificial Intelligence and Machine Learning are often used to analyze data in the field of finance and economics. AI helps in risk management, improving personalized services, automation and ensuring compliance. The finance sector is dealing with an enormous amount of data, and such data has been manually analyzed (even with assistance of various computer systems) so far; however, now, AI can automate and accelerate such analysis, revolutionizing the industry.

AI has many possible uses in the industry like:

- developing derivative trading algorithms by analyzing patterns within data sets
- Determining credit score and worthiness
- Predicting loan risks and defaults by analyzing historical performances, trends in existing customers and their cash flow projections etc.
- Acting as an investment advisor by predicting stock movements
- Processing documents in finance firms
- Eliminating repetitive tasks

Further with the application of generative AI, machines can teach themselves continuously and thereby interpret and predict with better accuracy.
While AI can be useful in any field of finance from personal finance to corporate finance its use is currently widely popular amongst financial institutions like commercial banks, credit unions, hedge funds, investment banks etc.

This paper focuses on the use of AI in predicting if a client will pay off or default on a loan he/she takes from a bank.

2.3 Loans

The dictionary defines loan “as a thing that is borrowed, especially a sum of money that is expected to be paid back with interest”. In the finance industry it is a sum of money that is lent by an individual, an institution like a Bank, or a government to another individual, institution or government for a profit (usually in the form of interest). Banks for example collect deposits from customers and pay some interest to customers. They then use such funds to lend to individuals and institutions at an interest rate (higher than the one they provide to their customers for deposits) thereby enabling the cash of those who have a surplus to be made available for those who need cash. There are various types of loans like car loans, personal loans, home loans, project lending, corporate lending etc.

2.4 Hypothesis

Identifying factors, such as Income, Personal Life Factors, and Financial History, that predict loan default chances using machine learning.

2.5 This paper and its possible impact

This research paper includes a structured analysis of various machine learning models that work on a publicly accessible dataset to predict, based on a set of factors, whether an individual will default on his loan.

The average personal loan debt per borrower is $11,652 as of Q3 2024, in the United States (2).

The implications of the total amount of loans handed out by banks and the rate of delinquency, could be significant. Loan defaults, by both small and large customers, are a huge problem that has plagued the finance sector for countless decades. Notwithstanding the efficiency of these institutions, many loan defaults can lead to their failure and even large-scale financial crisis. All one needs to do is to look at various banks' financials year after year and the amount of loans they write off in their books. This paper demonstrates how AI can greatly assist in predicting potential defaults. This can in turn enable banks to proactively take measures to avoid loaning when the chances of defaulting are high. Even a minor step in the direction of avoiding loan defaults can help banks save millions of dollars. Banks can also reduce the number of employees doing the manual checking and help further reduce costs incurred by banks.

3. Results

In this results section we can see whether our models will provide sufficient evidence, validating or invalidating our hypothesis. It will help us see whether we can truly use a few simple parameters (Gender, Marital Status, Living Area, Employment status, and Education Level) to predict whether an individual will default on his loan or not.

3.1 Hyperparameter Tuning

Each Machine Learning Model has its own training parameters: These parameters are called hyperparameters. When coding an Artificial Intelligence, we modify, or tune, these hyperparameters until we get the most optimal results from the model. In this paper, we chose the most critical hyperparameter(s) for each model and tuned them. For hyperparameter tuning in this model we created a block of code which consists of a for loop that will run the parameter from values of 1 to 1000 and at each value it will analyze the performance metrics of the model for all values of '*X_validate*' 100 times, then average the results. At the end it will calculate the hyperparameter value which gives the best results. After calculating the hyperparameter values with the best precision, recall, and accuracy respectively, we still used the hyperparameter value providing the best accuracy because at the end, accuracy represents the percentage of correct predictions and this is the most important from a real-life standpoint.

3.1.1 Decision Tree - Max Depth

The first parameter we will look at is the '*max_depth*' parameter for the decision tree model. The maximum depth of the tree is the number of layers of nodes in the tree. A deeper tree can fit the training data better but if the max depth becomes too high, it can lead to overfitting. After running the code segment for the hyperparameter tuning of the '*max_depth*' parameter. These were the final results:

"Highest Testing Accuracy: 0.73 at parameter value: 4"
Hence we used the '*max_depth*' value 4

3.1.2 Random Forest - Max Depth

'*max_depth*' is a parameter for the random forest machine learning model. The max_depth of a random forest is the maximum depth of the tree: that is the number of layers in each decision tree (5). After running the code segment for the hyperparameter tuning of the '*max_depth*' parameter. These were the final results:

"Highest Testing Accuracy: 0.72 at parameter value: 4"
Hence we used the '*max_depth*' value of 4

3.1.3 Multi-Layer Perceptron - Max Iterations

In the Multi-Layer Perceptron machine learning model, the '*max_iter*' hyperparameter is the maximum number of times or epochs that the neural network should perform. For the MLP model, the operation only converges after the '*max_iter*' value is above 300, so we will set our parameters to range from 300 to 1000. After running the code segment for the hyperparameter tuning of the '*max_iter*' parameter, These were the final results

"Highest Testing Accuracy: 0.72 at parameter value: 308"
Hence we will use the '*max_iter*' value of 308

3.1.4 K-Nearest Neighbors - Number of Neighbors and Wights

In the K-Nearest Neighbors machine learning model, the 'n_neighbors' hyperparameter is the number of closest neighbors considered when predicting a value for a data value using the k-nearest neighbors algorithm. Additionally, we can only use hyperparameter values for '*n_neighbors*' from 1 to the length of the number of neighbors which is equal to 334.

The K-Nearest Neighbors machine learning model also has a '*weights*' hyperparameter which is the weight or influence of each neighbor. There are only two possible values for this : '*uniform*' and '*distance*'. In '*uniform*' all neighbors have equal weights. In '*distance*' all the weights are equal to the inverse of their distance from the data value : Greater the distance, lesser the influence (weight).

In this case because one of the two parameters has only two possible values, we will simply run the code twice, one for '*uniform*' and one for '*distance*'. We will then calculate when

we get the higher performance metric and along with which weight for each performance metric. Now we will have three values: Highest accuracy and which performance metric it had that accuracy and with which value of n_neighbor, highest precision and which performance metric it had that precision and with which value of '*n_neighbors*', and highest recall and which performance metric it had that recall and with which value of *n_neighbors*.

Finally we will calculate the best value of *n_neighbors* by averaging the three. For the best weight, we will simply use the weight that gave the majority of best results. So if '*uniform*' was the best for two or more of the three performance metrics, we will use 'uniform'. After running the code segment for the hyperparameter tuning of the '*n_neighbors*' and '*weights*' parameter, these were the final results:

"Highest Testing Accuracy: 0.7 at parameter value: 19 and with uniform weights" Hence we will use the '*n_neighbors*' value of 19 and we will use the '*weights*' value of 'uniform'

3.2 Evaluation Metrics Results

All the models were tested finally using the 'X_finaltest' and 'y_finaltest' variables in order to see their final performance after hyperparameter tuning. In the case of models with hyperparameters, we have included the performance metric for a few different hyperparameters other than the best hyperparameter chosen from the Hyperparameter tuning. This is in order to help show that the chosen hyperparameter yields the best accuracy. In all cases, the higher the performance metric, the better the model. The weight column in each models' results is the final weight of the model after the weights algorithm runs. The higher the weight of the model, the better it is. This weights algorithm will be further explained in later sections of the paper.

This shows that our algorithm can help provide extra data on the chance of an individual defaulting on his/her loan with which banks can decide whether they will hand out a loan to a customer or not. This algorithm has an accuracy of around 70% and hence can provide rather accurate data that can help banks in achieving lower loan default rates.

As discussed earlier, the average loan taken by customers in the US is $11,652 as of Q3 2024. As per Equifax.com, $12.65 trillion was the total amount of loan taken in the US in 2024. Thus if we divide it, we can see that there were approximately 1.0857 billion loans taken in the US in 2024. The current loan default rate in the US is 4.7% and thus we can estimate that currently 51,029,211.08 loans that are defaulted. Our model with its 70% accuracy can thus help reduce this number of defaults to approximately 15.3 million defaults, a reduction of around 70%. This can help save an approximate amount of $178.43 billion (3)(4).

3.2.1 Logistic Regression

Testing Accuracy	Training Accuracy	Precision	Recall	Weight
0.70	**0.68**	**0.70**	**1.00**	**59**

3.2.2 Decision Tree

Max Depth	Testing Accuracy	Training Accuracy	Precision	Recall	Weight
1	**0.67**	**0.67**	**0.67**	**1**	**57**
5	**0.69**	**0.72**	**0.73**	**0.90**	**58**
25	**0.57**	**0.72**	**0.66**	**0.75**	**48**
50	**0.66**	**0.73**	**0.70**	**0.88**	**56**
4 (Ideal)	**0.70**	**0.70**	**0.71**	**0.96**	**59**

This table shows that the testing accuracy is highest for the '*max_depth*' '4' chosen from the hyperparameter tuning.

3.3.3 Random Forest

Max Depth	Testing Accuracy	Training Accuracy	Precision	Recall	Weight
1	**0.60**	**0.68**	**0.60**	**1.00**	**51**
5	**0.66**	**0.73**	**0.71**	**0.90**	**56**
25	**0.67**	**0.72**	**0.67**	**0.80**	**51**
50	**0.68**	**0.73**	**0.68**	**0.88**	**54**
4 (Ideal)	**0.67**	**0.70**	**0.69**	**0.95**	**57**

This table shows that the testing accuracy is highest for the 'max_depth' '4' chosen from the hyperparameter tuning.

3.3.4 Multi-Layer Perceptron

Max Iterations	Testing Accuracy	Training Accuracy	Precision	Recall	Weight
300	**0.64**	**0.73**	**0.68**	**0.89**	**51**
310	**0.67**	**0.72**	**0.71**	**0.92**	**52**
335	**0.58**	**0.72**	**0.66**	**0.79**	**49**
350	**0.68**	**0.73**	**0.68**	**0.88**	**55**
308 (Ideal)	**0.67**	**0.70**	**0.69**	**0.96**	**58**

This table shows that the testing accuracy is highest for the ‘max_iterations’ ‘308’ chosen from the hyperparameter tuning.

3.3.5 K-Nearest Neighbors

Weights (Parameter)	Max Iterations	Testing Accuracy	Training Accuracy	Precision	Recall	Weight
Distance	**1**	**0.60**	**0.70**	**0.67**	**0.82**	**54**
	8	**0.61**	**0.69**	**0.74**	**0.71**	**57**
	50	**0.58**	**0.72**	**0.66**	**0.77**	**49**
	19	**0.65**	**0.73**	**0.69**	**0.88**	**54**
Uniform	**1**	**0.70**	**0.63**	**0.79**	**0.78**	**61**
	8	**0.67**	**0.69**	**0.70**	**0.93**	**59**
	30	**0.67**	**0.68**	**0.67**	**1.00**	**54**
	19 (Ideal)	**0.72**	**0.69**	**0.72**	**1.00**	**61**

This table shows that the testing accuracy is highest for the ‘uniform’ ‘weights’ with ‘max_iterations’ ‘18’ chosen from the hyperparameter tuning.

3.3.6 Weights Algorithm

The weights algorithm resulted in the following:

No. of Correct Predictions: 57
No. of Total Predictions: 83
Accuracy = 0.69

4. Discussion

Through this paper we were able to develop a strong machine learning model that accurately tracks whether individuals will default on their loans or not based on a set of input parameters provided to them. The model that resulted with the highest accuracy was the K-Nearest Neighbours model.

In the end, our hypothesis is validated. By using 5 simple parameters (input values regarding the individual), we are able to predict with an average accuracy of 70% whether an individual is going to default on his loan or not. Our models also clearly do not have the problem of overfitting because the training and testing accuracies in all cases are relatively similar. However, there is a gap left for future performance developments: Our current accuracy on average is about 70% and we would like to develop it in the future to reach around 90%. This could be done by maybe adding more parameters or training the models on an even larger dataset.

4.1 Possible Future Developments

There are a few ways in which this project could be further developed by others wanting to take up and continue the research. For one, the model's testing has so far been only in a short time frame and we have not been able to test whether the model will continue to work with this dataset accurately in the future as trends in the finance industry change. An individual could test the model over a long period of a few years or longer in order to test how the model performs in the long-term.

Additionally, an individual could also check if the model's performance varies with respect to the country. From one country to another, the local demographics changes and many factors including mean income, type of occupation etc. changes and all this has the possibility of resulting in different outcomes. The performance of the model across various countries (Developing compared to developed for example) would need to be tested.

We would also like to add some natural language elements to this code. This could be able to then simplify the job of the end-users of the application : Bankers. They would now need to simply provide the individual's bank account details file and the code could be able to directly find out the details it requires from this file.

4.2 Impact

As mentioned earlier in the paper, this has a possibly huge impact that can help reduce the global rate of defaults by providing additional information to banks(the model's predictions) allowing them to make more informed choices on whether to grant a loan or not to an individual.

5. Materials and Methods

5.1 The Dataset used

The dataset for this machine learning model is sourced from Kaggle. It provides information related to the chance of a person defaulting on his loan : Gender, Marital Status, Area, Education, Type of Employment, Income, No. of dependents etc. The dataset provides real life results for over 600 people stating their details(the factors) and if the status of their loan is paid or defaulted. The dataset has a slight bias towards people who don't default their loans, but this can be easily overcome by most machine learning models including the ones used in the code.

This dataset is publicly accessible on Kaggle (link). The machine learning algorithm is trained based on the *load_train.csv* dataset.

5.2 Preparing the dataset for training

5.2.1 Replacing necessary values to make the entire dataset into integers

Most values in the dataset were strings. This would result in an added complexity while training the model. To simplify the dataset and thereby make the coding of the model more

streamlined and also simpler, the values in the dataset under the following columns were changed:

- Gender Column: *'Male'* was replaced with the integer value 1 and *'Female'* was replaced with the integer value 0
- Marital Status ['Married'] Column: *'Yes'* was replaced with the integer value 1 and *'No'* was replaced with the integer value 0
- Education Column: *'Graduate'* was replaced with the integer value 1 and *'Not Graduate'* was replaced with the integer value 0
- Self_Employed Column: *'Yes'* was replaced with the integer value 1 and *'No'* was replaced with the integer value 0
- Area Column: *'Urban'* was replaced with the integer value 2, *'Semiurban'* was replaced with the integer value 1, and *'Rural'* was replaced with the integer value 0
- Dependents Column: *'3+'* was replaced with the integer value 3 , *'2'* was replaced with the integer value 2, *'1'* was replaced with integer value 1, and *'0'* was replaced with integer value 0
- Loan Status Column: *'Y'* was replaced with the integer value 1 (Those who paid their loans) and *'N'* was replaced with the integer value 0 (Those who defaulted on their loans)

5.2.2 Removing NaN values

The dataset also had some values that were '*NaN* ' or 'Not a Number.' We cannot replace these with random integers : doing so will result in the data values possibly not matching correctly and can lead to issues with the models' performance metrics later. Therefore, the rows containing the NaN values in any column were simply removed. The new dataset, once the rows containing NaN values were removed, was stored in a new variable '*data*_cleaned'. Before cleaning the dataset by removing NaN values, the dataset had 614 values; After cleaning the dataset, there were 499 values.

5.2.3 Hyperparameters and Splitting the dataset

Before training, we had to split the dataset into training and testing sets. The kaggle link provided a testing dataset however we did not use this because the testing set did not have the final ['Status'], the status of the loan, column thereby preventing us from calculating precision metrics like accuracy, precision, and recall for the models as we did not have access to the final correct result to compare with in each case.

We split the dataset into a training and testing set of X and Y values. The dataset was split such that 33.3% of the entire dataset was allocated for testing the models into variables '*X_test* ' and '*y_test* '; the remaining 66.6% was allocated for training the models into variables '*X_train*' and '*y_train*'.

Some of the models used such as the Decision Tree, Random Forest, Multi-Layer Perceptron, and K-Nearest Neighbors have certain hyperparameters that are tuned to provide the

ideal output from the model. Since modifying the hyperparameters based on the evaluation metrics obtained from the models predictions on the testing set would result in violating the use of the testing set : The testing set is to ensure that the model is able to provide accurate predictions for new data, but if we keep tuning hyperparameters to get the best result from the testing set, we are almost training the model again on the testing set.

Therefore '*X_test*' and '*y_test*' were further split in half into: 'X_validate' and 'y_validate', representing the validation set that would be used for hyperparameter tuning; and 'X_finaltest' and 'y_finaltest', representing the final testing set containing new unseen values that the model would predict on. In other words, the testing set was further split in half into a validation set and a final testing set: 50% to the final testing set and 50% to the validation set. This was done to ensure the models had a sufficiently large training dataset to allow them to train effectively, while also making sure that the models undergo hyperparameter tuning to obtain the best possible results and the models are tested on new unseen data.

5.3 The Models used

The models used for the code in this paper were from the python library scikit-learn. Scikit-learn or sk-learn for short is a python library dedicated to machine learning models. It contains pre written codes for classification, regression, clustering, dimensionality reduction, model selection, and preprocessing machine learning models. It also provides prewritten codes in the form of methods for calculating various performance metrics for the models: Accuracy, Precision, and Recall. There are also functions that were used to split the dataset. This library also provided the functions needed to access, train, and test all the 5 models used in the code: Logistic Regression, Decision Tree, Random Forest, Multi-Layer Perceptron, and K-Nearest Neighbour. (6)

5.3.1 Logistic Regression

Logistic regression is the first machine learning model used in the code of the paper. Logistic Regression is a method that estimates the probability of a result based on a provided set of independent variables. In machine learning, logistic regression classifies as a supervised machine learning model: It is provided data and learns based on the provided labeled data. Logistic Regression is also a linear model: That means the prediction is based on different weights assigned to parameters during the training phase. (10)(12)

5.3.2 Decision Tree

Decision tree is also a supervised learning algorithm that is utilized in machine learning for both classification and regression tasks. It works by using a divide and conquer technique to identify the best points to classify data within a tree into specific class labels. This is a non-parametric model: The model does not create any closed-form formula for the connection between independent and dependent variables. Here we see an example of a decision tree on whether a person should go surfing or not. (10)(11)

5.3.3 Random Forest

Random forest is a machine learning model that combines the outputs of multiple decision trees, finally obtaining a single result. Each decision tree is trained on a sample with an out of bag sample to cross-validate: A third of the data is set aside for testing also known as out-of-bag data. The random forest averages the results of decision trees for regression and takes the majority vote for classification. The random forest model also has a lower likelihood to overfit compared to the decision tree. The latter is used in this paper. The random forest in the code for this paper consists of 100 decision trees. (13)

5.3.4 Multi-Layer Perceptron

Multi-Layer Perceptrons are a type of feedforward neural network: The simplest form of artificial neural networks where information flows only in one direction. The MLP has an input layer with nodes or neurons that receive input data. The next layer is the hidden layer which can have one or more sub-layers within it; Each layer produces an output that is carried forward to the next layer. Each connection between neurons of a layer has a weight that determines the strength of the connection. These weights are learned by the model during training. Finally the output layer receives the result from the hidden layers. The neural network has non-linear dependencies between the input and the output. (8)(9)

5.3.5 K-Nearest Neighbour

K-Nearest Neighbour is a machine learning algorithm that uses the proximity of input values to pre-provided labeled data in order to make classifications or predictions. It is a non-parametric, supervised learning model: It does not create any relationship between independent and dependent variables and is trained on labeled data. (7)(14)

5.4 Evaluation Metrics

Evaluation Metrics are quantitative measures that provide information on the performance/effectiveness of a machine learning model. There are different evaluation metrics that provide different performance-related information about the machine learning model. This paper uses three performance metrics: Accuracy, Precision, and Recall. (15)

5.4.1 Accuracy

Accuracy is a machine learning evaluation metric that measures how often a machine learning model predicts the correct result.

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions}$$

5.4.2 Precision

Precision is a machine learning evaluation metric that measures how often a machine learning model predicts true positives correctly

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

Taking the example of a model that outputs 1s and 0s the below grid explains true and false positives and negatives:

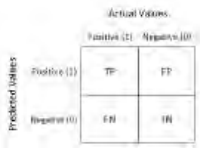


Fig 1 : Confusion Matrix

- True positives occur when the machine learning model predicts 1 and the actual correct value is 1
- True negatives occur when the machine learning model predicts 0 and the actual correct values is 0
- False positives occur when the machine learning model predicts 1 but the actual correct value is 0
- False negatives occur when the machine learning model predicts 0 but the actual correct value is 1

5.4.3 Recall

Recall is a machine learning evaluation metric that measures how often a machine learning model correctly predicts positive values out of all the actual positive values in the dataset. In other words Recall is the number of True positives divided by the sum of True Positives and False Negatives

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

5.5 Overfitting

Overfitting occurs when the results of a machine learning algorithm fits too closely, sometimes exactly, to its training dataset. This results in the model not being able to make accurate predictions on new unseen data. Overfitting generally occurs when the training accuracy is comparatively much higher than the testing accuracy. The models in this paper with the best parameters from hyperparameter tuning do not overfit as the testing and training accuracy are relatively similar. (16)

5.6 Weights Algorithm - Combining all models

In order to combine the 5 machine learning models used in this paper, an advanced weights algorithm was developed. The models all start with a weight of 1. The code is then looped for iterations equal to the length of the final testing dataset/validation set (depending on whether hyperparameter tuning or testing is ongoing). A variable '*sum*' is reset to 0 at the beginning of each iteration of the loop. Then all the models are made to predict the result for the '*i* ' th (current iteration of the loop) row of the dataset.

Following this if a model predicts the result correctly, then its weight is increased by 1. After this, the sum of weights of all models is calculated and stored in the variable '*sum*'; and the sum of all model predictions multiplied by their respective weights is calculated and stored in the variable '*result*'. After this a conditional statement is run in which :

- If the result is greater than half the sum, the final result is stored as 1
- Else the final result is stored as 0

We also have counter variables for the number of total iterations and the number of times the final result is correct. After running the loop, the final weights of the models show how many times the models predicted correctly minus one, and can be used as another evaluation metric for the performance of the models. The number of correct predictions and total predictions is also printed to provide another metric: This metric is not for any of the specific models but for the weights algorithm, and thus for all the models working together.

Works Cited

Venkatesh, Prashanth. "Personal Loan Statistics: 2025." *LendingTree*, www.lendingtree.com/personal/personal-loans-statistics/#:~:text=To%20compare%2C%20Americans%20owe%20%241.166,%2411%2C652%20as%20of%20Q3%202024. Accessed 11 Jan. 2025.

Venkatesh, Prashanth. "March 2024 U.S. National Consumer Credit Trends Reports." *Equifax*, www.equifax.com/newsroom/all-news/-/story/march-2024-u-s-national-consumer-credit-trends-reports/#:~:text=Consumer%20Credit%20Observations%3A,95.8%25%20of%20mortgage%20debt%20outstanding. Accessed 18 Jan. 2025.

Venkatesh, Prashanth. "Under Pressure: Leveraged Loan Defaults Are Spiking: Debtdynamics North America." *ION Analytics*, 19 Nov. 2024, ionanalytics.com/insights/debtwire/under-pressure-leveraged-loan-defaults-are-spiking-debtdynamics-north-america/#:~:text=Annual%20loan%20defaults%20are%20soaring,when%20defaults%20peaked%20at%2010.5%25. Accessed 18 Jan. 2025.

Venkatesh, Prashanth. "Delinquency Rate on Consumer Loans in the U.S. 2024." *Statista*, 5 Nov. 2024, www.statista.com/statistics/1325074/delinquency-rate-on-consumer-loans-at-commercial-banks-in-the-us/#:~:text=In%20the%20first%20quarter%20of,the%20United%20States%20were%20delinquent. Accessed 07 Dec. 2024.

Venkatesh, Prashanth. "Tuning Random Forest Hyperparameters." *KDnuggets*, www.kdnuggets.com/2022/08/tuning-random-forest-hyperparameters.html#:~:text=max_depth%3A%20The%20maximum%20depth%20of,a%20decision%20tree%20can%20have.. Accessed 07 Dec. 2024.

Venkatesh, Prashanth. "Learn." *Scikit*, scikit-learn.org/stable/index.html. Accessed 07 Dec. 2024.

Venkatesh, Prashanth. "Kneighborsclassifier." *Scikit*, scikit-learn.org/dev/modules/generated/sklearn.neighbors.KNeighborsClassifier.html. Accessed 07 Dec. 2024.

Venkatesh, Prashanth. "MLPClassifier." *Scikit*, scikit-learn.org/dev/modules/generated/sklearn.neural_network.MLPClassifier.html. Accessed 07 Dec. 2024.

Venkatesh, Prashanth. "Multilayer Perceptrons in Machine Learning: A Comprehensive Guide." *DataCamp*, 28 July 2024, www.datacamp.com/tutorial/multilayer-perceptrons-in-machine-learning. Accessed 07 Dec. 2024.

Venkatesh, Prashanth. "Non-Parametric Model." *DeepAI*, 17 May 2019, deepai.org/machine-learning-glossary-and-terms/non-parametric-model. Accessed 07 Dec. 2024.

Venkatesh, Prashanth. "What Is a Decision Tree?" *IBM*, 15 Aug. 2024, www.ibm.com/topics/decision-trees. Accessed 07 Dec. 2024.

Venkatesh, Prashanth. “What Is Logistic Regression?” *IBM*, 1 Nov. 2024, www.ibm.com/topics/logistic-regression. Accessed 07 Dec. 2024.
Venkatesh, Prashanth. “What Is Random Forest?” *IBM*, 25 Oct. 2024, www.ibm.com/topics/random-forest. Accessed 07 Dec. 2024.
Venkatesh, Prashanth. “What Is the K-Nearest Neighbors Algorithm?” *IBM*, 28 Oct. 2024, www.ibm.com/topics/knn?mhsrc=ibmsearch_a&mhq=KNN. Accessed 07 Dec. 2024.
Venkatesh, Prashanth. “Accuracy vs. Precision vs. Recall in Machine Learning: What’s the Difference?” *Evidently AI - Open-Source ML Monitoring and Observability*, www.evidentlyai.com/classification-metrics/accuracy-precision-recall. Accessed 07 Dec. 2024.
Venkatesh, Prashanth. “What Is Overfitting?” *IBM*, 2 Sept. 2024, www.ibm.com/topics/overfitting. Accessed 07 Dec. 2024.

Appendix

The code used in this research paper can be found in the following GitHub Repository: https://github.com/PrashanthVenkatesh2024/Artificial_Intelligence_In_Loan_Default_Detection

Empowering Tomorrow's Voters:
Evaluating How Access to Pre-registration Affects Youth Voter Turnout By Anna Gagne

Abstract

With the increase in state pre-registration laws that allow youths as young as 16 in some states to register to vote before being eligible, youth can now approach voting registration from a more convenient and accessible standpoint. This paper examines the inequities of access to voter pre-registration across the US and its impact on youth voter engagement. I use data from the Movement Advancement Project (MAP), the National Conference of State Legislatures (NCSL), and US News to determine pre-registration access and relative education rankings across US states. Through the analysis of state voting patterns and pre-registration accessibility, this study will evaluate potential correlations between access to pre-registration at a younger age and voter turnout rates. I find that states with pre-registration for 16-year olds have an approximately 5% higher youth turnout than states that do not, and while all states allow pre-registration, some only allow it at older ages. My results suggest that states should lower the pre-registration age in order to increase youth turnout. The positive correlation between a lower pre-registration age and higher youth voter turnout demonstrates the importance of encouraging the younger generation to vote by lowering pre-registration ages across states and implementing further educational resources.

Introduction and Literature Review

With growing tensions approaching the 2024 Presidential elections, youth voter participation may be more significant than ever. Pre-registration,which allows ineligible voters to register early to vote, has increased youth turnout. Ever since voting became available to 18-year olds across the U.S in 1972, a larger voter turnout gap has existed; in 1972, 50% of 18-24 year olds voted as compared to 70% of 25 and older citizens (Holbein and Hillygus 2016). The participation gap has widened from 20% to 24% through 2012 as 41% of youth voted compared to 65% of individuals 25 and older.

By implementing pre-registration into states' voting practices, youth between the ages of 16-18 have the opportunity of taking the first step towards being involved in what may appear to be an intimidating process. With inequities surrounding civic education access across the US as well as voter resources, newly eligible voters may find it difficult to navigate the process of registering to vote and casting a ballot, causing them to miss certain requirements or registration deadlines (McDonald 2009). Pre-registration may minimize these errors by "allowing young people to take care of the paperwork ahead of time so that they don't have anything standing in their way on Election Day" (Markey 2004). By expanding access to pre-registration within public facilities such as Department of Motor Vehicles (DMVs), states gain an "automatic" aspect in which youth may pre-register to vote during a necessary trip to the DMV in order to gain one's license. Public facilities offer an accessible and convenient way to register to vote

which can greatly benefit youth who may look forward to voting in the future but are unsure how or where to get started.

Prior Research and Effects

When looking into the effects of pre-registration from prior studies, researchers have evaluated the positive outcomes of pre-registration implementation on youth voting. For instance, after examining individual state implementations of pre-registration programs in states like Florida, research has identified that access to pre-registration has resulted in increased voter turnout. After Florida implemented pre-registration access to 17 year olds in 2000, voting turnout for 18-year olds jumped from 11% to 34% between 1992 and 2008 (McDonald 2009).

Similarly, while the National Voter Registration Act (NVRA) of 1993 was first applied in order to expand pre-registration to state motor vehicle agencies, it was largely ignored by public assistance agencies. This eventually led to lawsuits that influenced substantial increase in voter registration Cherry 2012). A study on states that allow pre-registration showed that states with the lowest/no amount of 16-year old drivers have little effect on youth voter turnout compared to states with the highest number of 16-year old drivers with an increased youth voter turnout of 4 percentage points (Fowler 2017). Since the majority of U.S states allow issuing a driver's license as early as 16 years old (Witmer 2022), it could be more convenient for 16 year-old drivers to pick up their license and pre-register to vote at the DMV simultaneously compared to states which allow pre-registration only at ages 17+.

Despite pre-registration correlating with voter turnout, it also has a relationship with expenditures on higher education. In fact, pre-registration programs have raised higher education expenditures by 5.1%, or $25, in states that have adopted pre-registration (Bertocchi, Dimico, Lancia, Russo 2020). Pre-registrations helps ensure that youth are registered to vote by the time they are eligible, removing possible hassles and wasted time in the future, thus leading to potential increased youth voter turnout. Since many youths may value spending money on education, politicians might try to persuade youth voters by campaigning for increased value in education spendings such as loan forgiveness or student debt in order to secure a vote.

Methodology

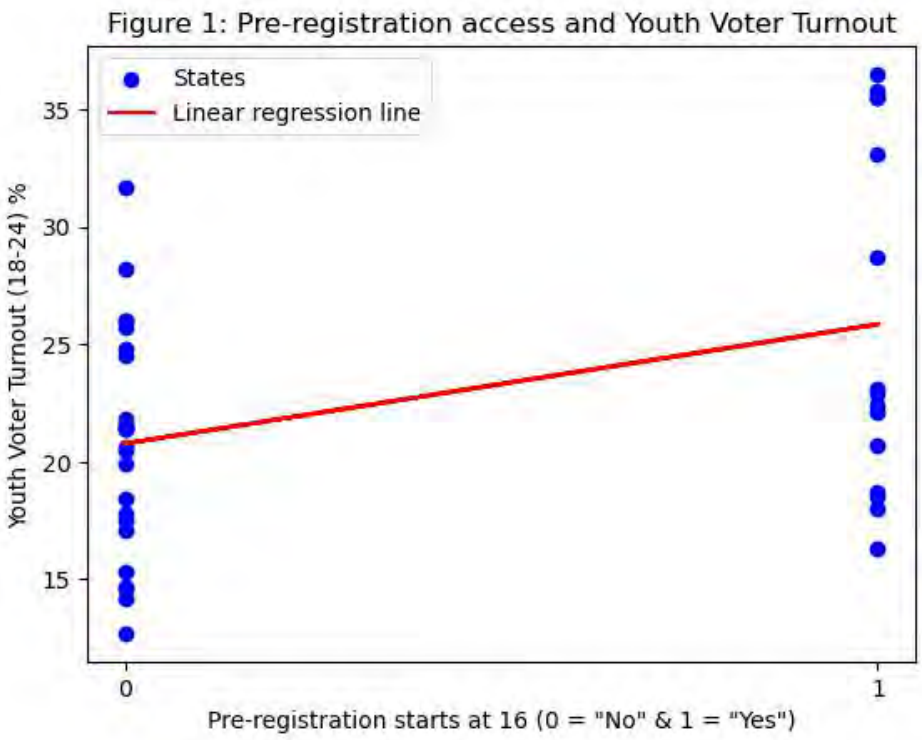


Figure 1

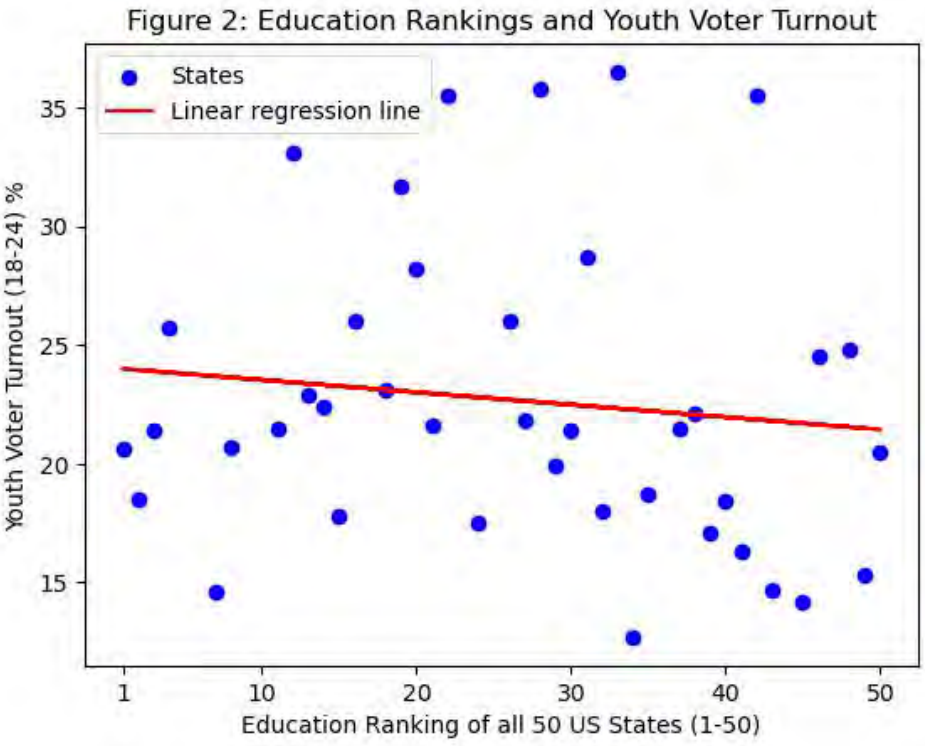


Figure 2

Using the Movement Advancement Project (MAP) and the National Conference of State Legislatures (NCSL), I determined the earliest pre-registration age available in each state with the following outcomes: 16, 17, turns 18 before the next election, and 6 months before turning 18. From there, I categorized each state using an indicator of 0 or 1. A value of 0 indicates that the state does not offer pre-registration at 16 years old whereas 1 indicates that the state *does* offer pre-registration as early as 16. Through Tufts Circle, I determined the youth voting percentage in the 2022 midterm election with voters aged 18-29. By reviewing the US News Pre-K-12 education rankings by state, I ranked each state correspondingly from 1-50 (excluding Washington D.C).

To understand the general patterns in the data, Figures 1 and 2 show the relationship between a state's 2022 youth voter turnout and the presence of a pre-registration law and its education ranking. Using Python and Matplotlib, I created linear regression graphs highlighting these relationships. Figures 1 and 2 help to identify potential factors that may contribute to a higher youth voter turnout. I noticed that pre-grestiation laws and better education rankings are positively associated with youth turnout.

By evaluating the pre-registration accessibility, the 2022 youth voter turnout, and the 2023 pre-k-12 education rankings, I developed models used to predict the approximate change in the youth voter turnout percentage based on an individual state's pre-registration access and relative education ranking. By adding a state's education ranking as a variable, I was able to analyze potential correlations between education quality and youth initiative to vote.

However, there may be confounding factors that influence both law and turnout. One concern is education. To control for this I used a multivariate regression to evaluate the impact of a state's education ranking on youth voter turnout. I controlled for education to analyze potential motives that could lead to the implementation of pre-registration laws, and how education rankings could correlate with youth voter turnout. Although I was controlling for education, there may also be other factors that confound the analysis. For example, a state's civic education requirement may influence passing a pre-registration law due to the state's politicians being more concerned with civic education at a younger age and thus supporting pre-registration access.

Results

	coef	std err	t	P>|t|	[0.025	0.975]
const	21.7678	2.179	9.992	0.000	17.354	26.182
Indicator Variable	4.9773	1.956	2.544	0.015	1.014	8.941
Education Ranking	−0.0362	0.067	−0.544	0.590	−0.171	0.099

Figure 3

I show that there is a positive correlation between states starting pre-registration at 16 years old and youth voter turnout percentage as compared to all other states who do not start pre-registration at 16 (Figure 3). According to the coefficient value 4.9773 of the indicator

variable (0 representing age ≠ 16 and 1 = 16), the policy is associated with an increase in youth voting turnout percentage of 4.9773 . We found that the p-value, which is the probability of obtaining data as extreme if the null hypothesis was true, shows us that the coefficient (4.9773) is statistically different from 0. Since the p-value of 0.015 is less than 0.05, the threshold of which a value is statistically significant, the indicator variable is deemed significant, showing a positive correlation between a state's pre-registration age equaling 16 and that state's youth voter turnout percentage. The confidence interval is the way of estimating the uncertainty of the relationship between indicator variable and outcome. While there is a true value/relationship between indicator variable & youth turnout, there is still a bound of uncertainty due to random noise. However, we are fairly confident that the truth is somewhere between 1 and 8.9 for our indicator variable.

Additionally, the coefficient value -0.0362 of the education ranking indicates that for every number increase in education ranking–ranking moves lower on list–the voter turnout percentage decreases by (coefficient=0.0362, p=0.59). We found that the p-value–although being significant for the indicator variable–was actually insignificant for the education ranking with a value of 0.59.

Discussion

In this study, I evaluated whether or not there exists a correlation between a lower pre-registration age, a state's education ranking, and the youth (18-24) voting turnout percentage. Using data from the Movement Advancement Project, US News, and the 2022 Midterm Election voting turnouts, I found that states that begin voting pre-registration at 16 years old have an approximately 5% higher youth voting percentage than states who do not. Additionally, when implementing educational ranking as an additional variable, we discovered that as the educational ranking of a state decreases, the corresponding youth voting percentage decreases by approximately 3.6%; however, this correlation is statistically insignificant and will require further evidence and analysis to demonstrate a causative relationship.

Future possible research directions should include evaluating the relationship between civic education access and youth voter turnout. While this study finds an insignificant correlation between education ranking and youth voter turnout, civic education requirements across states can offer a new perspective to the possible correlation between a student's access to civic education resources and initiative to pre-register to vote. Another suggested research direction is to evaluate how pre-registration age impacts youth voter turnout at major elections. Although the voting data within this research paper comes from the 2022 midterm election, it would be interesting to see how previous and upcoming major elections such as the 2016, 2020, and 2024 presidential elections show potential trends of youth voting with the help of prior pre-registration.

Works Cited

Bertocchi, Graziella, et al. "Youth Enfranchisement, Political Responsiveness, and Education Expenditure: Evidence from the US." *American Economic Journal: Economic Policy*, vol. 12, no. 3, Aug. 2020, pp. 76–106, https://doi.org/10.1257/pol.20180203.

Cherry, Ceridwen. *Increasing Youth Participation: The Case for a National Voter Pre-Increasing Youth Participation: The Case for a National Voter Pre- Registration Law Registration Law*. p. 2012, repository.law.umich.edu/cgi/viewcontent.cgi?article=1074&context=mjlr.

"Democracy Maps | Voter Pre-Registration." *Mapresearch.org*, 2024, www.mapresearch.org/democracy-maps/voter_preregistration.

Fowler, Anthony. "Does Voter Preregistration Increase Youth Participation?" *Election Law Journal: Rules, Politics, and Policy*, vol. 16, no. 4, Dec. 2017, pp. 485–94, https://doi.org/10.1089/elj.2017.0436.

"Making Young Voters: The Impact of Preregistration on Youth Turnout on JSTOR." *Jstor.org*, 2016, www.jstor.org/stable/24877627?read-now=1&seq=2#page_scan_tab_contents.

Mcdonald, Michael, and Matthew Thornburg. *REGISTERING the YOUTH through VOTER PREREGISTRATION**. www.nyujlpp.org/wp-content/uploads/2012/11/Michael-P.-McDonald-Matthew-Thornburg-Registering-the-Youth-Through-Voter-Preregistration.pdf.

"Preregistration for Young Voters." *Ncsl.org*, 2024, www.ncsl.org/elections-and-campaigns/preregistration-for-young-voters.

Pre-Registration. www.mapresearch.org/img/maps/citations-voter-preregistration.pdf.'

"State-By-State Youth Voter Turnout Data and the Impact of Election Laws in 2022." *Tufts.edu*, 6 Apr. 2023, circle.tufts.edu/latest-research/state-state-youth-voter-turnout-data-and-impact-election-laws-2022.

How Black Holes Kill People By Sunny Shin

Abstract

The immense mass of black holes creates gravitational forces powerful enough to be lethal. In this work, we emphasize the extreme dangers of black holes by comparing them with "Hanged, drawn and quartered," a historical execution method. To do so, we introduce a concept of tidal forces mathematically and present the corresponding formula.

Keywords black holes, tidal force, execution

1. Introduction to black holes

Black holes possess immense masses, making them regions of extreme gravitation influence. Due to their enormous mass, even the light cannot escape their gravitational pull, which is why black holes appear dark and cannot be observed directly via reflected light, as illustrated in Figure 1.

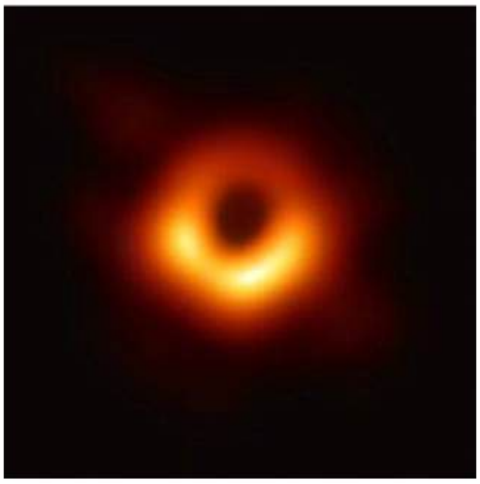


Figure 1 Black hole at the core of Messier 87, credit: [1]

The light around the black hole came from behind the black hole, and the black hole refracted it. This is why the light surrounding the black hole seems like a donut. According to Wikipedia [1], black holes are classified into five classes. Since their masses are too heavy, their masses were compared with the masses of the Sun and the Moon ($M_{Sun} = 2 \times 10^{30}$kg and $M_{Moon} = 7 \times 10^{22}$kg), and their radii were compared with the distance between the Sun and the Earth (1AU$= 1.5 \times 10^{11}$m), as in the following table.

Class	Approximated Mass	Approximated radius
Ultramassive black hole	$(10^9 - 10^{11}) \times M_{Sun}$	Greater than 10^3AU
Supermassive black hole	$(10^6 - 10^9) \times M_{Sun}$	0.001 - 400AU
Intermediate-mass black hole	$(10^2 - 10^5) \times M_{Sun}$	10^3km (almost radius of the Earth)
Stellar black hole	$(2 - 150) \times M_{Sun}$	30km
Micro black hole	up to M_{Moon}	Up to 0.1mm

Table 1 Classification of black holes

As we can see in the table above, black holes are incredibly dense. For instance, a micro black hole may have a mass comparable to that of the Moon, yet its radius remains below 0.1mm. To accurately interpret black holes, we need to use general relativity developed by Albert Einstein. However, since general relativity is highly complex, the simplified approach can be used by considering the escape velocity condition. If we use the fact that a particle cannot escape from black holes even if its speed is similar to that of light, we can derive a relationship between the mass and radius of black holes.

Let a black hole be given and the mass and radius of the black hole be set by M and R, respectively. Assume that a mass m particle is rotating the black hole and its radius is r. From the formulas of gravitational force and centripetal force, we get

$$\frac{GmM}{r^2} = \frac{mv^2}{r}$$

where G is the gravitational constant $G = 6.67 \times 10^{-11}$N · m2 · kg^{-2}. Here, the left-hand side represents the gravitational force, while the right-hand side represents the centripetal force. The relationship above can be simplified into

$$\frac{GM}{r} = v^2$$

We know that v is the minimum velocity to maintain the rotational motion. If the speed of the object is less than v, then the particle no longer rotates and it will fall down to the black hole. Since v cannot exceed the speed of light c, we know $v \leq c$. Also, the radius of the rotational motion r should be greater than or equal to R. Therefore, we also get $r \geq R$. Finally, if we substitute these two relations $v \leq c$ and $r \geq R$ into the above relationship, then we get

$$\frac{GM}{R} \leq c^2$$

This provides an approximation of the relationship between a black hole's mass and radius. However, this calculation does not apply to photons since the mass of the photon is zero ($m = 0$) and interacts with gravity only through their momentum. To derive a more precise relationship, we need to use general relativity, as we mentioned. If we calculate the range of $\sqrt{\frac{GM}{R}}$ for each class of black holes, then we get the following table.

Class	Mass	Radius	Value of $\sqrt{\frac{GM}{R}}$
Ultramassive black hole	$10^9 \times M_{Sun}$	10^3 AU	2.98×10^7m/s
Supermassive black hole	$10^6 \times M_{Sun}$	0.001 AU	9×10^8m/s
Intermediate-mass black hole	$10^2 \times M_{Sun}$	10^3 km	1.15×10^8m/s
Stellar black hole	$2 \times M_{Sun}$	30 km	9.43×10^8m/s
Micro black hole	M_{Moon}	0.1 mm	2.16×10^8m/s

Table 2 The values of $\sqrt{\frac{GM}{R}}$ for various types of black holes

As we can see in the table above, the values of $\sqrt{\frac{GM}{R}}$ for each type are not that different to the speed of light 3×10^8m/s. We just used the minimum mass and radius (except for the micro black hole case, which used maximum mass and radius) to choose the mass and radius for each type of black hole. This approximation is one of the possible reasons for errors.

2. Tidal force

Just like gravitational forces, tidal forces are also induced by massive objects like the Sun, Earth, and Moon. In this section, we introduce the tidal force mathematically.

To interpret the tidal force, we consider a system in which one object rotates another object. We consider two objects. The first object has mass M_1 and its radius is R. The second object has mass M_2 and we do not set the radius of this object since it is not needed to express the tidal force. We also set the distance between two objects by d and assume that this distance is much larger than the radius R, i.e., $\frac{d}{R}$ is a small number compared to 1. Furthermore, we

consider two points A (the furthest point from the second object) and B (the nearest point from the second object) on the surface of the first object.

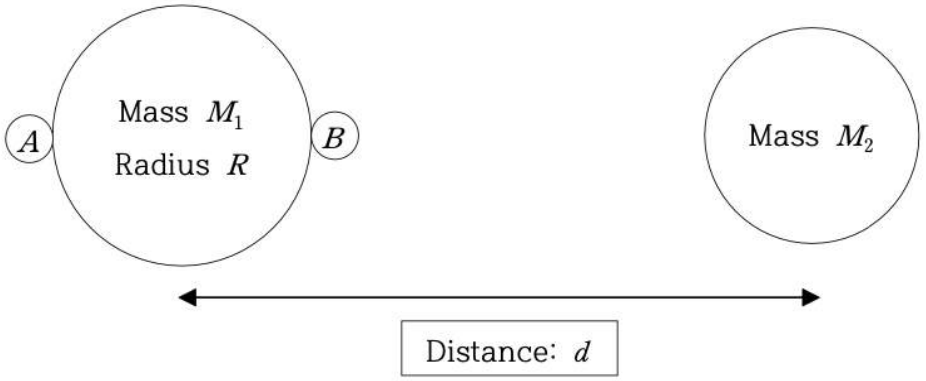


Figure 2 Two objects and two locations A and B on the first object

Now, we are ready to calculate the tidal force. To express the tidal force, we assume that mass m objects are placed at both A and B. If we calculate the force between the mass M_1 object and the mass M_2 object, then it is

$$\frac{GM_1M_2}{d^2}$$

Due to this force, the acceleration of the mass M_1 object is

$$\frac{GM_2}{d^2}$$

and the direction of this acceleration is the right. Since the distances from the mass M_2 object to the mass m objects at A and B are $d + R$ and $d - R$, respectively, the accelerations of mass m objects at A and B due to M_2 object are

$$\frac{GM_2}{(d+R)^2} \text{ and } \frac{GM_2}{(d-R)^2}$$

and the directions of these accelerations are both on the right side. Since these accelerations are not equal to $\frac{GM_2}{d^2}$, the deviation from the above accelerations and the acceleration of M_1 yields the tidal forces. Below, we calculate the tidal forces at A and B separately.

<u>Case 1: mass m object at A</u>

The acceleration of the mass *m* object at *A* is $\frac{GM_2}{(d+R)^2}$ and the acceleration of the mass M_1 object is $\frac{GM_2}{d^2}$ (directions are both on the right side). Since the acceleration of the mass *m* is smaller, this object experiences an apparent acceleration acting to the left (towards M_1) and the magnitude, the difference between two accelerations, is calculated as follows:

$$\begin{aligned}\frac{GM_2}{d^2}-\frac{GM_2}{(d+R)^2} &= \frac{GM_2}{d^2(d+R)^2}\left((d+R)^2-d^2\right).\\ &= \frac{GM_2}{d^2(d+R)^2}(2dR+R^2)\\ &= \frac{GRM_2}{d^3}\times\frac{2+\left(\frac{R}{d}\right)^2}{\left(1+\frac{R}{d}\right)^2}\end{aligned}$$

If we use the fact that $\frac{R}{d}$ is much smaller than 1, then the above difference can be approximated as

$$\frac{2GRM_2}{d^3}$$

Case 2: mass *m* object at *B*

The acceleration of the mass *m* object at *B* is $\frac{GM_2}{(d+R)^2}$ and the acceleration of the mass M_1 object is $\frac{GM_2}{d^2}$ (directions are both on the right side). Since the acceleration of the mass *m* is larger, this object experiences an apparent acceleration acting to the right (towards M_2) and the magnitude, the difference between two accelerations, calculated as follows:

$$\begin{aligned}\frac{GM_2}{(d-R)^2}-\frac{GM_2}{d^2} &= \frac{GM_2}{d^2(d-R)^2}\left(d^2-(d-R)^2\right).\\ &= \frac{GM_2}{d^2(d-R)^2}(2dR-R^2)\\ &= \frac{GRM_2}{d^3}\times\frac{2-\left(\frac{R}{d}\right)^2}{\left(1-\frac{R}{d}\right)^2}\end{aligned}$$

If we use the fact that $\frac{R}{d}$ is much smaller than 1, then the above difference can be approximated as

$$\frac{2GRM_2}{d^3}$$

For both cases, the magnitudes of the apparent accelerations are both $\frac{2GRM_2}{d^3}$ and the directions of them are the outwards of the mass M_1 object. Since we assumed mass m objects are given at A and B, we know the magnitude of the tidal force acting on mass m is

$$\frac{2GRmM_2}{d^3}$$

We will use the above formula of the tidal force and compare it with 'Hanged, Drawn and Quartered' in the next section.

3. Comparison

To compare tidal force induced by a black hole with the force involved in the historical execution method known as “Hanged, Drawn, and Quartered,” we first introduce and describe this method, as illustrated in the figure below.

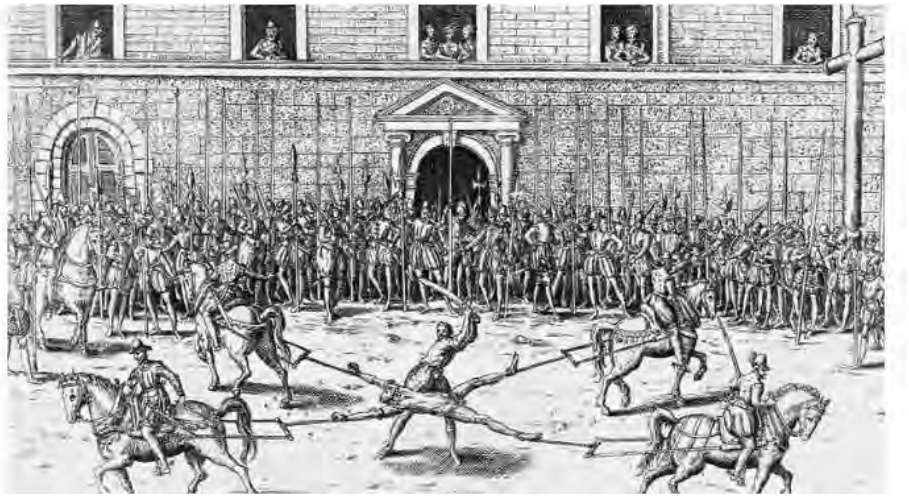


Figure 'Hanged, Drawn and Quartered' In this way of execution, a person is hanged on four horses, and these holes run in four different directions.

In this execution method, as shown in the figure above, a person's hands and feet are tied to four horses, which are then made to run in opposite directions. To simplify this scenario for our analysis, we assume two horses pull the person’s hands in one direction, while the other two horses pull the feet in the opposite direction.

According to Wikipedia page [4], each horse can exert a pulling force of approximately 800N. Since two horses are pulling in each direction, we can assume that the total force acting in opposite directions is estimated at 1600N forces.

Now, let us consider a person interacting with the tidal force of a black hole. Assume the black hole has a mass *M* and the person has a mass *m*. Further, we assume the person is oriented such that their feet are closer to the black hole while their arms point away from it. For simplicity, we approximate the person's body as a rectangular shape with a length of *h*(distance between the feet and hands), and we assume their center of mass is located at the geometric center of this rectangle.



Figure 4

We divide the person's body into two parts: the left of the center of mass and the right of the center of mass. Then, each part's mass should be $\frac{m}{2}$ since the center of mass of the body is located at the center of mass of the rectangle. Furthermore, the distance from the center of mass of the whole body to center of mass of each part is $\frac{h}{4}$ since each part is rectangular in shape. Now, assume that the distance from the center of mass of the person to the black hole is *d*. If we substitute $\frac{m}{2}$ into *m*, *M* into M_2, and $\frac{h}{2}$ into *R* of the tidal force formula, then we obtain

$$\frac{GhmM}{2d^3}.$$

We want to find the distance *d* at which makes the above value 1600N. To do so, we substitute real data of an average man, $h = 2\text{m}$ and $m = 80\text{kg}$, into the formula above. Then, we can simplify $\frac{GhmM}{2d^3} = 1600\text{N}$ as follows

$$\frac{M}{d^3} = 3 \times 10^{11}\ \text{kg} \cdot \text{m}^{-3}.$$

If we substitute mass of each class of black hole into *M*, then we can fill the table as follows.

Class	Mass	Radius
Ultramassive black hole	$10^9 \times M_{Sun}$	$1.88 \times 10^9\text{m}$
Supermassive black hole	$10^6 \times M_{Sun}$	$1.88 \times 10^8\text{m}$

Intermediate-mass black hole	$10^2 \times M_{Sun}$	8.74×10^6m
Stellar black hole	$2 \times M_{Sun}$	2×10^6m
Micro black hole	M_{Moon}	6156m = 6.156km

Table 3

As shown in the table above, approaching the black hole within a distance *d* results in a tidal force on the body that is as strong as or stronger than the force applied in 'Hanged, Drawn and Quartered'. For instance, even a micro black hole with the mass of the Moon would exert a dangerous tidal force at a distance of approximately 6 km, which is relatively far. This demonstrates that black holes, regardless of their size, can pose a severe threat to the human body.

Works Cited

Wikipedia: "Black hole." https://en.wikipedia.org/wiki/Black_hole

Ramamurti Shankar: "Fundamental of Physics I: Mechanics, Relativity, and Thermodynamics", (2019)

Dave Roos: "The 'Hanged, Drawn and Quartered' Execution was Even Worse than you Think," Howstuffworks, (2024). https://history.howstuffworks.com/history-vs-myth/hanging-drawing-and- quartering .htm

Wikipedia: "Horsepower." https://en.wikipedia.org/wiki/Horsepower

Biomedical Imaging Use Metastatic Cancer Research in Underdeveloped Regions By Jessica Wang and Ti'Aair Riggins

Abstract

Cases of cancer in developing countries continue to rise daily, with many of these countries unable to decrease the rate due to a lack of funds or poor treatment plans. To counteract this problem, researchers are aiming to potentially unveil causes for these disparities by conducting studies and tests to measure cancer growth. In developed countries, current tools to detect cancer comprise of the ever-growing field of biomedical imaging, involving tools such as magnetic resonance imaging (MRI) scans, positron emission tomography (PET) scans, and computed tomography (CT) scans. These tools are used to take images of a tumor and/or region of interest (ROI), and can aid physicians in studying cancer metastasis, and cancer growth. Radiomics, or the acquisition of images and converting them into quantitative and qualitative data, has been used to identify features needed by physicians for identification of cancer growth. Combining radiomics and biomedical images proves a promising future for increasing efficiency of cancer detection. This same concept can be applied to developing countries. By using biomedical imaging, cancer can be detected quicker, reducing the death rate of cancer due to late diagnosis and metastasis. By applying radiomics, this process of detection can be increased in efficiency, helping detect patterns in cell metastasis quicker. This would decrease both the amount of cell metastasis related deaths in developing countries and the amount of undiagnosed people who are susceptible to dangerous undiagnosed cancer growth.

Introduction

Cancer in lower and middle income countries is very prevalent, posing as an issue mainly caused by healthcare disparities. Countries in Africa are expected to face an increase in cancer cases to 2.1 million by 2040 without proper management and care (Huang et al.). Most cancers are caused by viral infections, making up 1 out of 4 cancer deaths (Jones, "Cancer in the Developing World: A Call to Action"). Furthermore, most patients diagnosed with cancer are diagnosed at a stage too late, where their disease is incurable due to cell metastasis. Cell metastasis is a process in which cancerous cells travel through the bloodstream and invade other tissues (Pouliot et al., "Investigating Metastasis Using In Vitro Platforms"). The term itself is very broad, as there are multiple ways for cancer cells to migrate, including kinesis and taxis, which has its own categories including chemotaxis, haptotaxis, electrotaxis, and durotaxis. Kinesis is when cancer cells migrate randomly due to growth factors or stick to surfaces for cells to cling to anywhere in the body. Taxis refers to directed movement, or when cancer cells migrate towards a specific direction due to a signal, demonstrated in *figure 1*, which could include a gradient change or any other change in the environment. Chemotaxis is the moving of cancer cells away from a signal, haptotaxis is the attraction of a cancer cell to a sticky surface, whereas electrotaxis is the moving towards an electrical signal, and duotaxis is when cells migrate to areas that are more physically stable and firm, making the attaching process quicker

(Pouliot et al., “Investigating Metastasis Using In Vitro Platforms”). Metastasis causes the microenvironment, extracellular matrix, and junctions that hold cells together to change. At first, building blocks may break down, but eventually cell signaling is key to allowing the tumor to spread to other parts of the body.

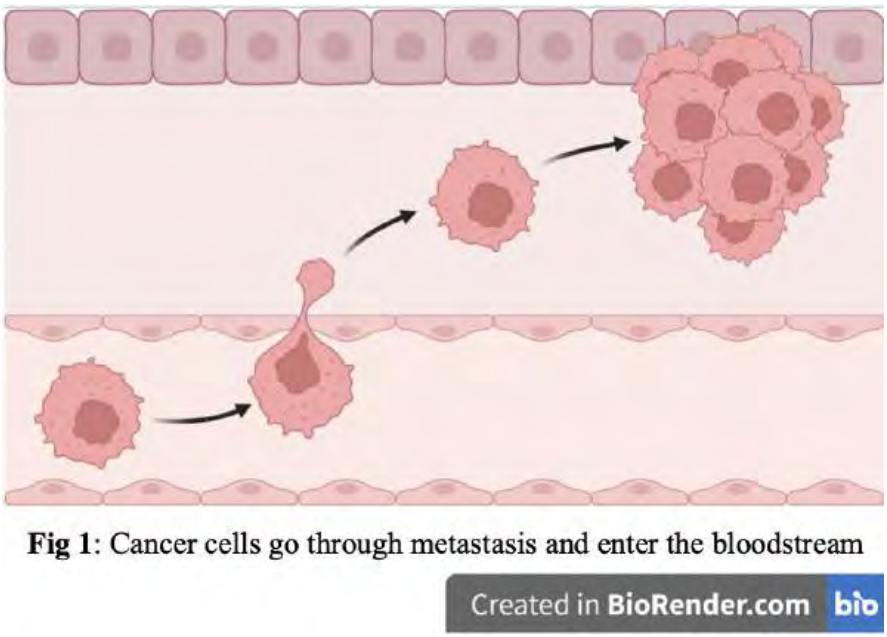


Fig 1: Cancer cells go through metastasis and enter the bloodstream

3D lab models are being studied to better understand how to fight cancer, as changes in tumors can lead to no or different reactions to medicine. Cost and limited accessibility to a research laboratory are some of the main limitations, while others being the difficulty of adapting 3D cultures to screens and the lengthy process of creating a functional organoid, can take up to 10 weeks at a time (Pouliot et al., “Investigating Metastasis Using In Vitro Platforms”).

Without proper technology or healthcare funds, many LLMICS are unable to detect cancer in its earlier stages. And when patients reach a late stage where metastasis has progressed significantly, they face the threat of a fatal disease, where death can occur in weeks to months (c=AU and o=The State of Queensland). However, in more developed countries, scientists have used biomedical imaging to detect cancer early to allow potential treatment (Fass, “Imaging and Cancer: A Review”). Physicians have used biomedical imaging combined with radiomics to more accurately predict clinical outcomes and treatment plans for patients who were early diagnosed (*Website*). However, these tools are often limited to access in developing countries, causing an increasing disparity gap. A key cause for these growing gaps include the lack of government funding for technology, poor financial resources, and geographical barriers to implement necessary tools (*Website*). Many developing regions lack an adequate safety zone for scanning technology needed for radiomics such as MRI scans, a form of technology that is key to implementing the technology (Murali et al., “Bringing MRI to Low- and Middle-Income Countries: Directions, Challenges and Potential Solutions”). Compared to developed countries,

who have MRI machines that contain a high magnetic field and are greater in availability, developing countries have a significantly fewer number of MRI machines . Due to these disparities, researchers have focused on interventions and finding ways to limit preventable cancer. They have also implemented increased studies to determine causes for disparities, comparing statistics of socio-economic factors to compare how gross domestic product (GDP) affects healthcare in a country (Murali et al., "Bringing MRI to Low- and Middle-Income Countries: Directions, Challenges and Potential Solutions"). Scientists and physicians have been increasingly working on developing new tools and imaging techniques, adding to and improving the current biomedical models.

Current Biomedical Models



Fig 2: Current role of biomedical imaging found in cancer management (Fass, "Imaging and Cancer: A Review")

Biomedical imaging serves as one of the most essential tools in cancer care. It allows for physicians to monitor patients without damaging tissue through procedures that can be invasive. The process for cancer screening through early imaging, *figure 2*, is fundamental to improving mortality outcomes. Imaging systems are commonly based on interactions between electromagnetic radiation and tissues or fluids found in the body (Fass, "Imaging and Cancer: A Review"). These systems permit the depiction of light and dark areas through pictures, utilizing contrast agents. Such agents are tools used to make biomedical images more clear and are often used to help aid the process of detection and diagnosis. These agents enter through openings in the human body or veins, and can be natural or manmade. Biomedical imaging comes in many forms, one being x-ray based systems including CT, which has its own subset of systems: computed radiography (CR) and digital radiography (DR). CR incorporates plates and outweighs DR in terms of portability, but recent advances have been made to make DR portable as well. DR has the ability to slice apart images similar to a 3D scan, speeding up the process doctors take to

find cancerous spots commonly based on interactions between electromagnetic radiation and tissues or fluids found in the body (Fass, “Imaging and Cancer: A Review”).

Tomosynthesis, a 3D mammogram used for detecting breast cancer, creates many low-dose X rays, captured at different angles, incorporating an iodine-based contrast agent. Another technological advancement for the 3D mammogram includes the cone beam CT, a system that is currently being developed where flat panel detectors with higher contrast, sensitivity, and no breast compression are used. This tool is preferable over the mammogram, as mammograms can be uncomfortable due to compression (Fass, “Imaging and Cancer: A Review”).

Additionally, MRI is used to measure signals coming from atoms in the human body. MRI is used for both cancers and cardiovascular disease, and scientists have estimated that the increasing use of MRI imaging could cause cancer deaths to decrease by 2.46 million worldwide (Murali et al., “Bringing MRI to Low- and Middle-Income Countries: Directions, Challenges and Potential Solutions”), which could also decrease the amount of radiation patients are exposed to compared to CT and x-ray scans. MRI has multiple subsets, including diffusion-weighted imaging (DWI), magnetic resonance elastography (MRE), stimulated raman scattering (SRS) (Wang et al.) and dynamic contrast-enhanced elastography (DCE). DWI has been in use for over 23 years and aids physicians in seeing how blood and other liquids move in the body. This can lead to the detection of pathophysiology such as tumors or strokes. DWI has been used in detecting breast cancer as well, as it can determine whether a breast lump is benign or malignant. MRE uses MRI and sound waves to capture images of tissues in the body and separates them from hard and soft tissue. SRS can form histologic images without the use of labels or dyes by using lasers to interact with molecular vibrations of the sample of a substance. DCE involves injecting a contrast agent that allows the MRI to capture movement of the contrast throughout the body. DCE is considered flexible because of its ability to analyze blood flow in multiple types of tissue, analyzing perfusion and permeability. Thus, the abilities of DCE provides data to determine characterization and possible treatment plans. A transrectal ultrasound (TRUS) serves the purpose of checking for prostate cancer by guiding a needle to take small samples such as biopsies. Endoscopic ultrasounds (EUS) checks to see if metastasis of cancer has occurred and has the ability to take biopsies for further research (Fass, “Imaging and Cancer: A Review”).

Introduction to Radiomics

Radiomics can be defined as the conversion of images and pictures taken by scans into high-dimensional data. It is designed to take large amounts of quantitative data from these images and place them in databases for testing, prediction, or forming a hypothesis (Gillies et al.). Quantitative features extracted from radiomics include shape, volume, size, texture, or any other physical attributes of a tumor and its environment. Paired with radiomics is radiogenomics, or the collection of radiomic data to find genomic patterns. For example, a singular three dimensional image can offer researchers hundreds of quantitative features describing shape and texture. The process of sample collection, *illustrated in figure 3*, occurs in 6 steps.

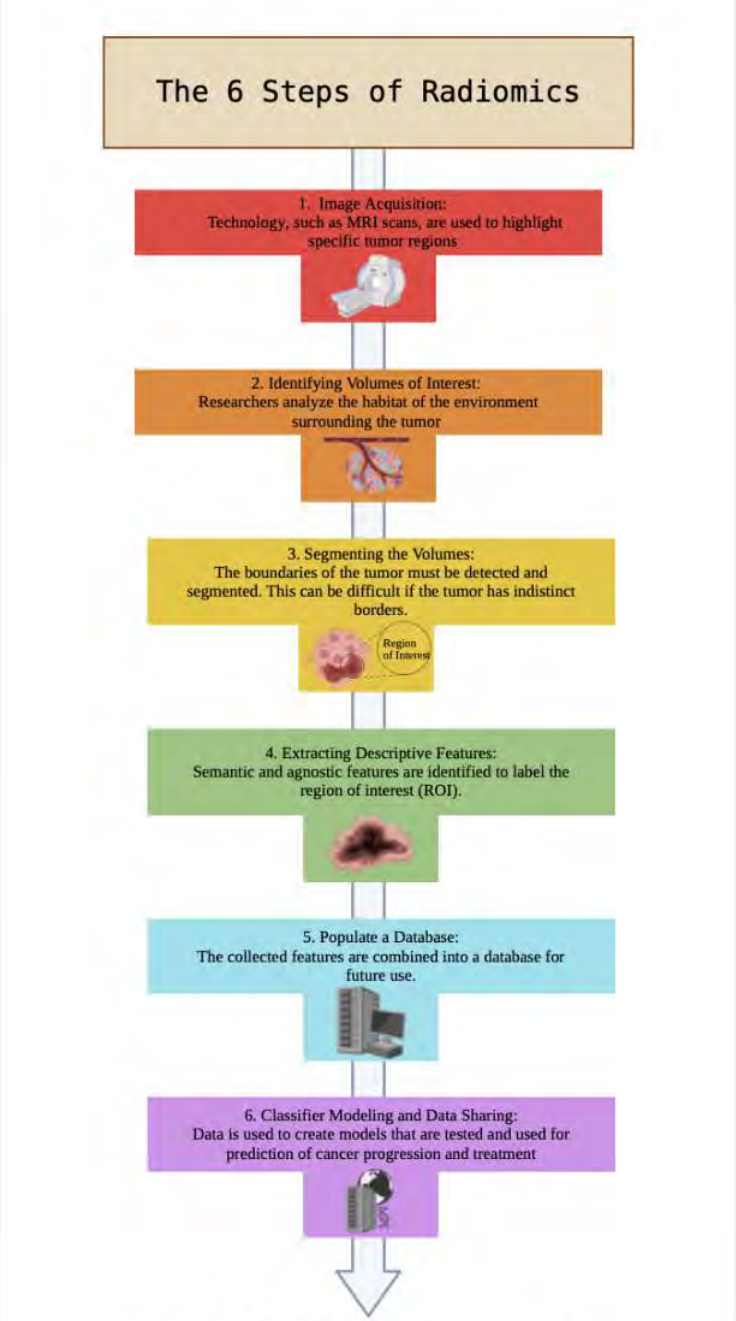


Fig 3: The 6 Steps of Metastasis

Radiomics has already been implemented into imaging. A computer program called PyRadiomics used to detect and study features from images is used by scientists to look at two regions of interests (ROIs). They grouped the features into three types:

1. Shape Features: 14 features showing the size and shape of an ROI
2. First-Order Feature: 18 features depicting contrast

3. Texture Features: 75 features depicting patterns in an ROI

Extra details are studied using a wavelet transform, and in total, 851 features were studied from each area (Nowakowski et al.). Radiomics serves as a promising future for cancer detection,as proven through a study conducted on the subject of MRI radiomics and nutritional-inflammatory biomarkers. In this study (Yan et al.), patients of cervical cancer undergoing concurrent chemoradiotherapy were divided into groups of high-risk and low-risk of progression free survival (PFS), where radiomics was applied with clinical features and inflammatory markers to form a model with an accurate prediction of PFS at 1 year, 3 years, and 5 years. The factors alone were not able to correctly make a prediction, but combined together and applied with radiomics, the model performed more accurately predicting outcomes of PFS (Yan et al.). While technology such as biopsies can only locate diversity in one portion of a chosen tumor, radiomics have the ability to capture heterogeneity across an entire tumor (Mayerhoefer et al.). These diverse features are put to full potential as machine learning techniques process them, increasing the impact of the information. However, since there may be too many radiomic features to handle, researchers reduce the amount of features by removing or combining others together. This process is called "dimensionality reduction" and ensures that features that may not be as useful to show differences between patients, are not used (Mayerhoefer et al.). As a result, radiomics is considered as a highly possible intervention strategy for detecting metastasis.

Radiomics Regarding Metastasis

Currently, radiomics is now a more commonly used tool for cancer metastasis detection. Instead of using a T2-weighted imaging tool (T2WI), which uses an MRI scan for pathology ("T1 vs T2 MRI") apparent diffusion coefficient (ADC) images (*Apparent Diffusion Coefficient*), which uses water molecules as a way of determining cellular density, are being used to provide physicians more information. They show how quickly cells are multiplying and how many cells there are, which can help physicians understand potential metastasis (He et al.). Furthermore, radiomic analyses have been used to aid in patients diagnosed with brain metastases (BM). Radiomics allow for the use of artificial intelligence (AI) and medical images to foreshadow outcomes and improve standard of living for patients. The use of radiomics can improve the treatment plans of cancer metastasis, including stereotactic radiosurgery (SRS). SRS is a form of radiation therapy meant for brain tumors, where a singular high-dose of radiation can be directed to a targeted area. Using radiomics means that the effectiveness of SRS can be increased. Additionally, PET specifically, dynamic amino acid PET, has the potential to be applied with radiomics, helping decipher necrosis from progression. Radiomics analyses are able to discriminate necrosis from disease progression within an area under the curve (AUC) value above 0.8, indicating a good performance for a model (Nowakowski et al.). Since the application of using AI can be conducted on most computers and laptops, they offer the ability to have

medical data being evaluated almost anywhere in the world, with running electricity and wifi, which is attractive for health care facilities in underprivileged regions.

Condition of Developing and Underprivileged Region

The highest incident cancers in developing countries include stomach, lung, liver, breast, and cervical cancer (Jones, "Cancer in the Developing World: A Call to Action").1 in 3 cancer deaths are caused by tobacco, while infections cause 1 in 4 cancer deaths (Jones, "Cancer in the Developing World: A Call to Action"). Infectious agents implicated in >90% of cancers include human papillomavirus causing cervical cancer, *helicobacter pylori* causing stomach cancer, and hepatitis B and C viruses causing liver cancer (Jones, "Cancer in the Developing World: A Call to Action"). Some of the concerns of cancer in developing countries are that many cancers are diagnosed at too late a stage for treatment to prove effective (Jones, "Cancer in the Developing World: A Call to Action"). In fact, 80% of cancer patients are diagnosed with late-stage incurable diseases, diseases that could have been prevented if diagnosed earlier. Additionally, in a study done regarding delays and barriers to cancer care in low-and middle-income countries, it was found that low-income countries (LIC's) had longer presentation intervals compared to lower-middle income or upper-middle income countries, signifying a lack of technology and advancement for cancer research (Brand et al., "Delays and Barriers to Cancer Care in Low‐ and Middle‐Income Countries: A Systematic Review"). In the same review conducted for delays and barriers to cancer care in low-and middle-income countries, 214 studies reviewed nine categories of barriers to early detection. These included health literacy, cancer stigma, access to primary care, inaccurate diagnosis, limited access to diagnostics, poor coordination of care, geographical, financial issues, and sociocultural barriers (Brand et al., "Delays and Barriers to Cancer Care in Low‐ and Middle‐Income Countries: A Systematic Review"). Health literacy refers to the knowledge a community has of signs, risk factors, or being skilled in navigating the health care system. Cancer stigma refers to the fear or embarrassment of cancer diagnoses from a patient to their family. Primary care includes the inability to receive appointments and conflict with transportation. This leads to trouble accessing diagnostics and relates to the geography of a region that can limit treatment. In terms of treatment itself, it is not uncommon for physicians to give false assurance and inaccurate diagnoses that lead to eventual delay of treatment. Poor condition of care refers to patients who did not resume follow-ups after an initial appointment or were not aware of the need for continuous treatment. Financial issues are related to the lack of health insurance or trouble affording treatment, and socio-cultural barriers include aggressive behavior from certain family members or religious barriers that prevent certain individuals from receiving treatment from another individual of the opposite gender (Brand et al., "Delays and Barriers to Cancer Care in Low‐ and Middle‐Income Countries: A Systematic Review").

These barriers are present in most if not all LIC's. In fact, policy implications have stated that focusing on one barrier out of the nine may not reduce delay of detection (Brand et al., "Delays and Barriers to Cancer Care in Low‐ and Middle‐Income Countries: A Systematic Review"). Paying attention to more than one barrier is needed to effectively reduce the delay of

detection. Additionally, the United States National Cancer Institute of Surveillance, Epidemiology, and End Results (SEER) program declared a staging classification for cancer consisting of three stages (Unger-Saldaña, "Challenges to the Early Diagnosis and Treatment of Breast Cancer in Developing Countries"). The first is localized, the second is regional, and the third is distant, where each stage of progressing cancer is continuous due to metastasis. A study done regarding the challenges to early diagnosis and treatment of breast cancer, two key challenges faced in developing countries, reported that most spread stages were in low-and middle income countries (LMICs) while cancer was diagnosed at the local stage in high-income countries (HICs). Furthermore, 70% of breast cancer patients were diagnosed as localized compared to the 20-60% of patients diagnosed at this early stage in LMICs (Unger-Saldaña, "Challenges to the Early Diagnosis and Treatment of Breast Cancer in Developing Countries"). Deaths from cancer in LMICs are also seen increasingly more compared to HICs. In LMCIs such as Gambia or Algeria, their 5-year survival rates for breast cancer was 12% and 38.8%, respectively. However, the 5-year survival rate for HICs such as Japan and Australia are 81.6% and 80.7%, respectively (Unger-Saldaña, "Challenges to the Early Diagnosis and Treatment of Breast Cancer in Developing Countries").

Accessing healthcare is one of the key factors in high mortality rates, late diagnosis, or few early detection cases. There are five components of access: affordability, acceptability, accessibility, accommodation, and availability (Unger-Saldaña, "Challenges to the Early Diagnosis and Treatment of Breast Cancer in Developing Countries"). Availability is the supply of health services or facilities fit for a patient's needs. Accessibility can be defined as the relationship between a patient and their geographical surroundings that can impact their resources, distance of travel, and expenses. Accommodation includes factors such as operation hours or telephone services. Affordability is the ability of a patient to pay for their treatment, and acceptability is a patient's personal perspective of their services, health personnel, or clinic type. Also linked with accessing healthcare is socioeconomic status (SES). SES measures income and education that can directly impact affordability and access, which is a determining factor for morbidity, mortality, and the ability to utilize healthcare. In the study mentioned above, it was reported that decreased SES led to increased breast cancer clinical stage and decreased 5-year survival rates. Additionally, people with low SES may face long waiting intervals between appointments, facing multiple health services before being treated at a cancer hospital, or extended distances of travel to reach a hospital, all forms of accommodation, availability and affordability (Unger-Saldaña, "Challenges to the Early Diagnosis and Treatment of Breast Cancer in Developing Countries"). Because of this, developing strong preventive strategies and focusing research into the cause of cancer is key. More specifically, implementation is top priority if cancer can be halted through interventions.

Cancer Research and Progress in Developing Countries

There have been many programs implemented to solve the issue of increasing cancer cases and health disparities in LLMICs. The National Cancer Institute (NCI) has been working

on addressing disparities and encouraging an increase in health-equity (*Research Areas - Cancer Disparities*). For example, NCI researches the biology behind disparities and studies the factors that lead up to it. They work to find interventions at a community level to reduce disparities and document population based data to view the magnitude of the issue and what areas need improvement. The SEER program operates similar to the latter. They collect quantitative and qualitative data such as age, sex, geography, and cancer incidence to track their progress and once more view the most needed areas of improvement. The Center for Cancer Control Research in Persistent Poverty Areas also works with intervention to prevent other cancers from developing (*Research Areas - Cancer Disparities*). Furthermore, the center for cancer research (CCR), has led multiple studies regarding imaging, biomarkers, and the microenvironment of cancerous tumors. For example, CCR's Laboratory of Human Carcinogenesis (LHC) has been investigating cancer biomarkers with the potential to improve treatment of lung cancer. They have also looked into possible reasons for disparities between African American and European American lung survival, where they found that systematic and systemic racism was a large contributing factor (*Research*). Members of CCR's Molecular Imaging Branch have also tested prostate cancer imagine methods for African American men, lessening disparities. Studies prior noted that the microenvironment difference between African American and European American men, exacerbated using aspirin, was reported to significantly reduce prostate cancer in African Americans (*Research*). Additionally, CCR's Genitourinary Malignancies Branch works with detecting the correlation between biomarkers and increased risk of prostate cancer, as well as locating how treatment can affect an individual's cancer. CCR's Laboratory of Cellular Oncology has also worked to find a promising treatment where an individual's own immune system is activated to fight cancer through a common viral infection (*Research*). As a result, early detection imaging tools such as radiomics, may be incorporated to decrease the cancer mortality rate amongst marginalized groups.

Potential Use of Radiomics in Underprivileged Regions

Scientists have focused on research and finding efficient ways to bring radiomics to developing countries. The two main benefits for introducing radiomics are early detection and cost reduction. In terms of early detection, scientists developed multiple tools to aid with that issue, with one being optical approaches (*Website*). Using optics has the benefit of being inexpensive and available for many. They are used most commonly in classifying skin lesions that are concerning to the point of sending biopsy or surgery. However, the depth of penetration to a tumor is shallow and can be fixed only with optical coherence (OCT) or optoacoustic imaging (OA), technology that is still being tested. Additionally, ultrasounds are among the available tools used for early detection, being able to image tissues of higher depth in real time. However, they are considered to be too operator dependent and techniques called "elastography" that can image tissue stiffness are being used to solve this problem (*Website*).

Scientists have applied radiomics to study early detection in many types of cancer, the most prominent being lung cancer. This is most likely due to the high number of cases of lung

cancer and the aggression of the disease. The National Lung Screening Trial (NLST) conducted multiple studies regarding lung cancer, and in one of them, a "radiomic score" was developed using machine learning by researchers. There are 10 features of small lung nodules in the score and it performed more accurately than using just the volume of the nodule, being as accurate as the standard tool used to predict the risk of lung cancer in 1-2 years, utilizing the Brock model (*Website*). The radiomic score was able to identify outlying cases of cancer in the high or low risk range, being over 93% accurate. The limitations of this model include its large false positives and overdiagnosis of neoplasms that could be left alone and would not cause death or other clinical symptoms. Image-based, genomic, molecular, and cellular biomarkers are being developed to decrease the amount of false positives and overtreatment, done by researchers in the Early Detection Research Network (EDRN) (*Website*).

Radiomics can also solve the problem of cost reduction. The incorporation of radiomics - combined with technology such as CT, MRI, and AI - can lessen the need for biopsies or surgeries that produce the same result but with greater expenses. Once more, early detection serves as a sub-category in cost reduction. If cancer can be detected earlier, then tumors will not have the time to possibly grow. This means that removing the tumor or treating the patient will be easier and require less expensive equipment. In fact, the World Health Organization launched the *Guide to cancer early diagnosis* in 2017, stating that early diagnosis will directly lead to less expenses (World Health Organization: WHO). Asides from early detection, diagnosing errors pose a massive expense to any country's economy. Diagnostic errors include delayed diagnosis, over-diagnosing, misreading imaging tests or misdiagnosis of any form of cancer (Pc). In the United States alone, the cost of healthcare is worth up to 2.7 trillion, with the cost of diagnostic testing taking up more than 10% of healthcare costs and steadily rising (Newman-Toker et al., "How Much Diagnostic Safety Can We Afford, and How Should We Decide? A Health Economics Perspective"). Wasted diagnostic tests are worth more than 25 billion dollars, and these errors can cause death or disability, harming more than a million people in the United States from recent estimates (Newman-Toker et al., "How Much Diagnostic Safety Can We Afford, and How Should We Decide? A Health Economics Perspective"). Out of those harmed, many face the direct consequences of diagnostic errors, including forensic costs, expenses to treat the true condition, or costs for medicine. Defensive medicine, or diagnostic tests used against malpractice lawsuits, can cost up to 45-60 billion US dollars annually, if not more (Newman-Toker et al., "How Much Diagnostic Safety Can We Afford, and How Should We Decide? A Health Economics Perspective"). In terms of the global economy, cancer is estimated to cost up to 25.2 trillion in international dollars from 2020-2050, with China and the US facing the highest expenses, making up for 24.1% and 20.8% of the total cost, respectively (*Website*). Due to lack of data from developing countries, an exact number is not able to define the expense of diagnostic errors for a country. However, the magnitude of diagnostic errors impacting developed countries such as the United States can only be reflected, if not more, for developing countries. In a study done for analyzing the cost-effectiveness of a radiogenomic screening program for the identification of BRCA and 2 (Breast Cancer Genes), the use of radiogenomic

models and clinical/history family data increased BRCA carrier detection rate by 41.8%. The model decreased BRCA-related cancer incidences by 23.7% and increased the cost per woman by only €2.51 per person annually. The study concluded that radiomics have the ability to lower the long-term healthcare costs by enhancing detection rates and accuracy (Di Pilla et al.).

Limitations of Radiomics in Underprivileged Regions

Despite the numerous benefits radiomics can bring to developing countries, there is an equally large expanse of limitations. In order for radiomics to be implemented, there first must be technology paired along. This includes MRI and PET scans, MRI being the most prevalent out of the two. In fact, MRI is used worldwide for oncological, cardiovascular, and neurovascular conditions (Murali et al., "Bringing MRI to Low- and Middle-Income Countries: Directions, Challenges and Potential Solutions"). However, only 34% of the world has access to MRI scanners. In LMICs, the availability of this scanner is significantly lower than in HICs, with 26.53 MRI units per million population (pmp) compared to 1.12 MRI units pmp in LMICs (Murali et al., "Bringing MRI to Low- and Middle-Income Countries: Directions, Challenges and Potential Solutions").

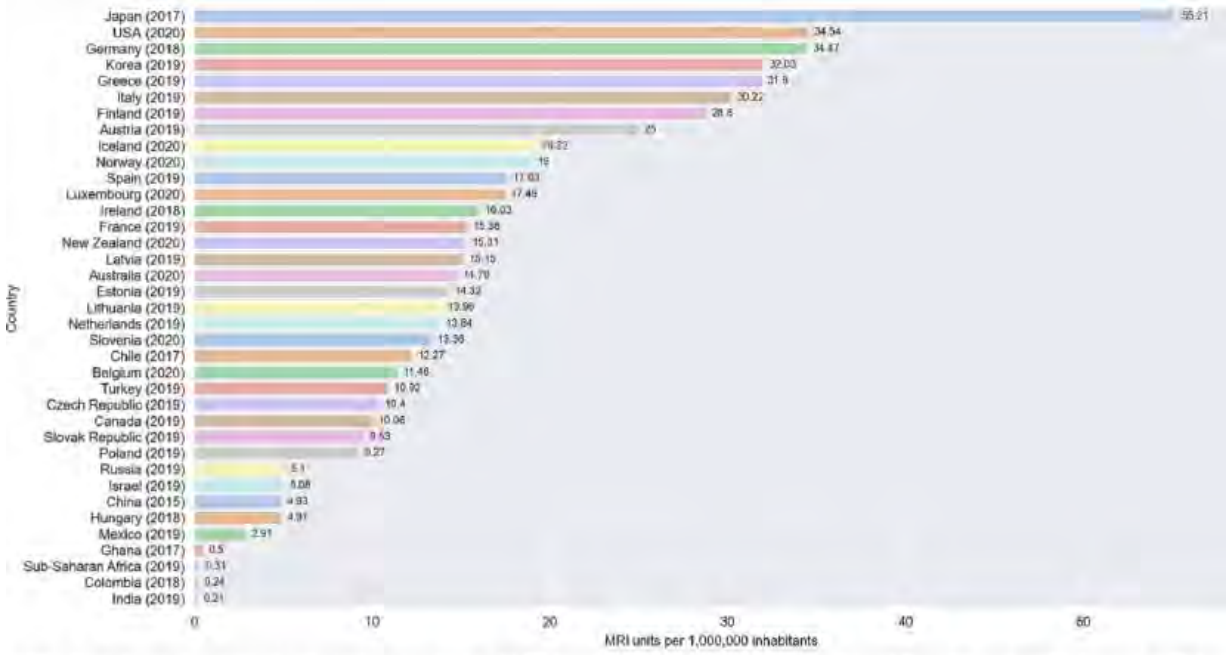


Fig 4: A bar chart displaying the number of MRI units per million population (Murali et al., "Bringing MRI to Low- and Middle-Income Countries: Directions, Challenges and Potential Solutions")

This correlates with an MRI's magnetic field strength: LMICs have fewer high-field (HF) scanners than HICs, and scanners often require more travel distance for communities in LMICs to reach compared to HICs (Murali et al., "Bringing MRI to Low- and Middle-Income Countries: Directions, Challenges and Potential Solutions"). MRIs are also technically difficult to

implement. They extract investment heavy expenses and trained professionals to operate smoothly. Purchasing and installing MRI scanners can also cost up to $1 million, excluding the expense of hiring workers to maintain and manage it. This proves to be a limitation for most LMICs, who only have a small budget for healthcare. Compared to HICs, who spent 12.46% of their GDP on healthcare, LMICs only spent 5.41% of their GDP on healthcare (Murali et al., "Bringing MRI to Low- and Middle-Income Countries: Directions, Challenges and Potential Solutions"). A developing country's poor financial state, compounded with an unreliable energy supply, geographical limitations, and lack of investment by governments make installing MRIs extremely difficult. Along with cost, the technical aspect of maintaining an MRI scanner is challenging as well. Areas with low bandwidth and Internet blackouts must find a way to update MRI software so data is not lost. And even before data can be collected, obtaining data can pose another challenge. Most LMCIs lack regulation of healthcare information and data protection, decreasing the ability of system communication (Murali et al., "Bringing MRI to Low- and Middle-Income Countries: Directions, Challenges and Potential Solutions").

Once a developing country receives the proper technology to implement radiomics, issues regarding the stability of radiomics arise. Even in HICs, radiomics has not been fully used in clinical practice because of challenges such as false positives, lack of standardization, and limited testing that are needed to confirm findings (Unger-Saldaña, "Challenges to the Early Diagnosis and Treatment of Breast Cancer in Developing Countries"). A study addressing these issues in locally advanced cervical cancer (LACC) analyzed MRI scans from 275 patients across 8 groups: 5 for training the model and 3 for testing it. Despite the models built being accurate - reaching 0.94 for specific tumor regions and even above 0.98 when combining different sequences - limitations were proven through this study. The ROIs were marked manually, which can be extremely time consuming, and an automatic system must be created for realistic usage in real life settings. Specific benefits of radiomics, such as early detection, also have their own limitations. Researchers have commonly proposed mammography screening to increase early detection, but that comes along with many barriers. In HICs, mammography has been adopted for early detection in many clinics. This technology has been one of the forms of radiomics LMICs have attempted to implement. However, mammography is not realistic for LMICs, who lack the infrastructure, workers, and financial resources to implement this technology (Unger-Saldaña, "Challenges to the Early Diagnosis and Treatment of Breast Cancer in Developing Countries"). More often than not, countries who use mammography end up with less efficient and more costly screening, adding on to the financial burdens they already face. The ability of mammography has been questioned in recent years as well, with evidence that reduced breast cancer wasn't due to mammography screening but due to adjuvant therapy. In fact, screening technology, including mammography, is the cause of around 30% of overdiagnosis in breast cancers. Out of 2000 women 50 years of age or older, there is only one less breast cancer death but 10 women treated unnecessarily and 200 women who experience false-positives, with half of them receiving a useless biopsy (Unger-Saldaña, "Challenges to the Early Diagnosis and Treatment of Breast Cancer in Developing Countries"). The unreliability of mammography in

HICs concerns researchers, who believe that the negative effects will impact LMICs even more than HICs (Unger-Saldaña, “Challenges to the Early Diagnosis and Treatment of Breast Cancer in Developing Countries”).

Discussion

Currently radiomics has proven to be beneficial in clinical settings to a certain extent, allowing for higher accuracy of models and early detection in cancer, a severe issue in many developing countries. Many studies have been conducted proving the legitimacy of radiomics, with models serving high accuracy results and the benefit of lower cost. LMICs, who have high cancer mortality rates that are not decreasing at the moment, would benefit heavily from radiomics, as cost can be reduced and early detection can be installed. However, there are still barriers to implementing radiomics in developing countries. LMICs face lack of funding, geographical barriers, and issues with accessibility, all limitations to the benefit of radiomics. They face the issue of installing technology needed for radiomics, and evidence has proven an unnecessary increase in cost of healthcare after installing a certain type of radiomics. In the future, scientists and researchers should prioritize making technology such as MRI and CT scans easier to access so radiomics can be implemented, working around geographical barriers, and encouraging more funding for LMICs so clinical data will not be lost and automated systems can make radiomics more efficient.

Works Cited

Apparent Diffusion Coefficient. https://doi.org/10.1016/B978-0-323-48552-4.00018-4. Accessed 24 Jan. 2025.

Brand, Nathan R., et al. "Delays and Barriers to Cancer Care in Low‐ and Middle‐Income Countries: A Systematic Review." The Oncologist, vol. 24, no. 12, Aug. 2019, p. e1371.

c=AU, and o=The State of Queensland. Conditions That Cannot Be Cured. https://www.qld.gov.au/health/support/end-of-life/care/conditions. Accessed 17 Dec. 2024.

Di Pilla, A., et al. "A Cost-Effectiveness Analysis of an Integrated Clinical-Radiogenomic Screening Program for the Identification of BRCA 1/2 Carriers (e-PROBE Study)." Scientific Reports, vol. 14, no. 1, Jan. 2024, pp. 1–8.

Fass, Leonard. "Imaging and Cancer: A Review." Molecular Oncology, vol. 2, no. 2, May 2008, p. 115.

Gillies, Robert J., et al. "Radiomics: Images Are More than Pictures, They Are Data." Radiology, vol. 278, no. 2, Nov. 2015, p. 563.

He, Dong, et al. "MRI-Based Radiomics Models to Assess Prostate Cancer, Extracapsular Extension and Positive Surgical Margins." Cancer Imaging, vol. 21, no. 1, July 2021, pp. 1–9.

Huang, J., et al. "Editorial: Cancer Burden, Prevention and Treatment in Developing Countries." Frontiers in Public Health, vol. 10, Feb. 2023, https://doi.org/10.3389/fpubh.2022.1124473.

Jones, Sinéad B. "Cancer in the Developing World: A Call to Action." BMJ : British Medical Journal, vol. 319, no. 7208, Aug. 1999, p. 505.

Mayerhoefer, Marius E., et al. "Introduction to Radiomics." Journal of Nuclear Medicine, vol. 61, no. 4, Apr. 2020, p. 488.

Murali, Sanjana, et al. "Bringing MRI to Low- and Middle-Income Countries: Directions, Challenges and Potential Solutions." NMR in Biomedicine, vol. 37, no. 7, July 2024, p. e4992.

Newman-Toker, David E., et al. "How Much Diagnostic Safety Can We Afford, and How Should We Decide? A Health Economics Perspective." BMJ Quality & Safety, vol. 22, no. Suppl 2, Oct. 2013, p. ii11.

Nowakowski, Alexander, et al. "Radiomics as an Emerging Tool in the Management of Brain Metastases." Neuro-Oncology Advances, vol. 4, no. 1, Sept. 2022, p. vdac141.

Pc, Simonson Goodman Platzer. "The Devastating Impact of Diagnostic Errors on Cancer Patients." Simonson Goodman Platzer PC, 11 Oct. 2023, https://www.simonsonlegal.com/blog/2023/october/the-devastating-impact-of-diagnostic-errors-on-c/.

Pouliot, Normand, et al. "Investigating Metastasis Using In Vitro Platforms." Madame Curie Bioscience Database [Internet], Landes Bioscience, 2013.

Research. https://ccr.cancer.gov/research/health-disparities/research. Accessed 1 Dec. 2024.

Research Areas - Cancer Disparities. 12 May 2015, https://www.cancer.gov/research/areas/disparities.

"T1 vs T2 MRI." *Mrimaster*, 20 Sept. 2023, https://mrimaster.com/t1-vs-t2-mri/.

Unger-Saldaña, Karla. "Challenges to the Early Diagnosis and Treatment of Breast Cancer in Developing Countries." *World Journal of Clinical Oncology*, vol. 5, no. 3, Aug. 2014, p. 465.

Wang, Tianyuan, et al. "Innovative Imaging Techniques for Advancing Cancer Diagnosis and Treatment." *Cancers*, vol. 16, no. 14, July 2024, https://doi.org/10.3390/cancers16142607.

Website. https://aacrjournals.org/cebp/article/29/12/2556/71998/Radiomics-Improves-Cancer-Screening-and-Early.

World Health Organization: WHO. "Early Cancer Diagnosis Saves Lives, Cuts Treatment Costs." *World Health Organization: WHO*, 3 Feb. 2017, https://www.who.int/news/item/03-02-2017-early-cancer-diagnosis-saves-lives-cuts-treatment-costs.

Yan, Qi, et al. "MRI Radiomics and Nutritional-Inflammatory Biomarkers: A Powerful Combination for Predicting Progression-Free Survival in Cervical Cancer Patients Undergoing Concurrent Chemoradiotherapy." *Cancer Imaging : The Official Publication of the International Cancer Imaging Society*, vol. 24, no. 1, Oct. 2024, p. 144.